

Математическое моделирование
Mathematical modeling

УДК 330.4
<https://doi.org/10.32362/2500-316X-2024-12-1-101-110>



НАУЧНАЯ СТАТЬЯ

Применение беггинга в прогнозировании временных рядов

Я.В. Грамович,
Д.Ю. Мусатов,
Д.А. Петрусевич[@]

МИРЭА – Российский технологический университет, Москва, 119454 Россия
[@] Автор для переписки, e-mail: petrusevich@mirea.ru, petrdenis@mail.ru

Резюме

Цели. Цель работы состоит в построении различных моделей беггинга, сопоставлении точности их прогнозов на тестовый период со стандартными моделями и получении выводов о возможности дальнейшего использования техники беггинга при моделировании временных рядов.

Методы. Исследуется применение беггинга к случайной составляющей временного ряда, формируемой после удаления тренда и сезонной части. Строится серия псевдовыборок, совмещающихся в новую случайную составляющую. На основе полученной компоненты строится новая модель ряда. По мнению многих авторов такой подход позволяет повысить точность модели временного ряда, лучшим образом оценив распределение.

Результаты. В теоретической части приведены характеристики различных моделей беггинга. Разница между ними сводится к оценке смещения, получаемой из-за того, что измерения, которые составляют псевдовыборки, не являются случайными. Представлен вычислительный эксперимент, в котором модели временных рядов строятся по индексу денежных доходов населения макроэкономической статистики Российской Федерации и по курсу акций Сбербанка. Прогнозы на тестовый период, полученные стандартными, нейросетевыми моделями и моделями на основе беггинга для некоторых временных рядов, сравниваются в вычислительном эксперименте. В самой простой реализации беггинг показал результаты, сравнимые со стандартными моделями ARIMA и ETS и несколько уступающие нейросетевым моделям для сезонных рядов; для несезонных рядов лучшие результаты дали стандартные модели ARIMA и ETS, модели беггинга дали близкие результаты. Обе группы моделей существенно превзошли результат нейросетевых моделей.

Выводы. При использовании беггинга лучшие результаты получены при моделировании сезонных временных рядов. Качество прогнозов моделей беггинга несколько уступает качеству прогнозов нейросетевых моделей, но оказывается на том же уровне, что у стандартных моделей ARIMA и ETS. Модели на основе беггинга следует использовать для моделирования временных рядов, различные функции над значениями ряда при построении псевдовыборок должны быть исследованы в дальнейшей работе.

Ключевые слова: динамические ряды, макроэкономическая статистика, ARIMA, псевдовыборка перекрывающихся блоков, псевдовыборка перекрывающихся блоков, стационарный беггинг

• Поступила: 21.06.2023 • Доработана: 19.07.2023 • Принята к опубликованию: 14.12.2023

Для цитирования: Грамович Я.В., Мусатов Д.Ю., Петрусевич Д.А. Применение беггинга в прогнозировании временных рядов. *Russ. Technol. J.* 2024;12(1):101–110. <https://doi.org/10.32362/2500-316X-2024-12-1-101-110>

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

RESEARCH ARTICLE

Implementation of bagging in time series forecasting

Ian V. Gramovich,
Danila Yu. Musatov,
Denis A. Petrushevich @

MIREA – Russian Technological University, Moscow, 119454 Russia

@ Corresponding author, e-mail: petrushevich@mirea.ru, petrdenis@mail.ru

Abstract

Objectives. The purpose of the article is to build different models of bagging, to compare the accuracy of their forecasts for the test period against standard models, and to draw conclusions about the possibility of further use of the bagging technique in time series modeling.

Methods. This study examines the application of bagging to the random component of a time series formed after removing the trend and seasonal part. A bootstrapped series combining into a new random component is constructed. Based on the component thus obtained, a new model of the series is built. According to many authors, this approach allows the accuracy of the time series model to be improved by better estimating the distribution.

Results. The theoretical part summarizes the characteristics of the different bagging models. The difference between them comes down to the bias estimate obtained, since the measurements making up the bootstraps are not random. We present a computational experiment in which time series models are constructed using the index of monetary income of the population, the macroeconomic statistics of the Russian Federation, and the stock price of Sberbank. Forecasts for the test period obtained by standard, neural network and bagging-based models for some time series are compared in the computational experiment. In the simplest implementation, bagging showed results comparable to ARIMA and ETS standard models, while and slightly inferior to neural network models for seasonal series. In the case of non-seasonal series, the ARIMA and ETS standard models gave the best results, while bagging models gave close results. Both groups of models significantly surpassed the result of neural network models.

Conclusions. When using bagging, the best results are obtained when modeling seasonal time series. The quality of forecasts of seigniorage models is somewhat inferior to the quality of forecasts of neural network models, but is at the same level as that of standard ARIMA and ETS models. Bagging-based models should be used for time series modeling. Different functions over the values of the series when constructing bootstraps should be studied in future work.

Keywords: dynamic series, macroeconomic statistics, ARIMA, nonoverlapping block bootstrap (NBB), moving block bootstrap (MBB), stationary bagging (SB)

• Submitted: 21.06.2023 • Revised: 19.07.2023 • Accepted: 14.12.2023

For citation: Gramovich Ia.V., Musatov D.Yu., Petrusевич D.A. Implementation of bagging in time series forecasting. *Russ. Technol. J.* 2024;12(1):101–110. <https://doi.org/10.32362/2500-316X-2024-12-1-101-110>

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

В рамках представленной работы рассматривается применение беггинга [1–5] в моделировании временных рядов. Использование беггинга в моделировании временных рядов можно считать выражением общей идеи построения более точной модели на основе нескольких имеющихся в наличии. Подход к составлению взвешенной комбинации прогнозов нескольких моделей временного ряда и усреднения нескольких прогнозов рассмотрен в работах [6, 7]. Основное отличие беггинга от объединения прогнозов моделей временных рядов состоит в том, что при его применении объединяются только компоненты шума. Основная цель в обоих подходах – повысить качество прогнозов на основе построения комбинации прогнозов нескольких моделей временных рядов.

Рассматриваемый подход является актуальным в связи с целесообразностью повышения точности прогнозирования временных рядов на основе наилучшей оценки распределения случайной составляющей. Статья содержит новые результаты исследований, выраженные в экспериментальной реализации моделей, построенных на основе беггинга временных рядов и сравнении результатов прогнозирования с результатами, полученными при помощи альтернативных моделей ARIMA¹ и нейросетевых моделей. Цель работы состоит в построении различных моделей беггинга, сопоставлении точности их прогнозов на тестовый период со стандартными моделями и в получении выводов о возможности дальнейшего использования техники беггинга при моделировании временных рядов.

Временной ряд представляется в виде комбинации трех частей: сезонной составляющей S_t , тренда T_t и шума R_t в аддитивном (1) или мультипликативном (2) виде:

$$y_t = S_t + T_t + R_t, \quad (1)$$

$$y_t = S_t \times T_t \times R_t, \quad (2)$$

где t – время.

¹ Autoregressive integrated moving average model – модель авторегрессии и скользящего среднего или модель Бокса – Дженкинса.

Беггинг применяется к составляющей шума R_t . Изначально эта стратегия успешно применялась в задаче классификации, где под ней подразумевается построение ансамблевой модели, заключающейся в обучении независимых классификаторов на разных выборках [8]. После этого предсказания, полученные каждой моделью, усредняются для получения окончательного результата (может применяться взвешенное усреднение в зависимости от того, насколько точны предсказания каждой модели, участвующей в ансамбле, на тестовой выборке). Таким образом, повышается точность предсказания.

Кроме идеи объединения моделей, в основе беггинга лежит «бутстреп» (bootstrap). Этот подход состоит в замене неизвестного распределения данных (которое характеризует рассматриваемый временной процесс) на выстраиваемое исследователем эмпирическое распределение. При использовании бутстрепа в задачах классификации данные не имеют временного измерения, поэтому их можно перемешивать случайным образом. Все становится сложнее при применении подобных идей к временным рядам, где различные значения выборки должны следовать друг за другом в соответствии с временными измерениями, даже если их выбирать случайно. Здесь идея преобразуется в построение множества псевдовыборок на основе исходных данных временного ряда. Фактически несколько раз (число моделей задается пользователем) на основе некоторого принципа из данных ряда выбираются значения, которые представляют собой новую временную последовательность. Так как обычно значений временного ряда много, можно выстроить множество «новых временных рядов» на основе исходного, случайно выбирая новые значения для каждой псевдовыборки. Предполагается, что характеристики исследуемого временного ряда, будут близки к параметрам получаемых псевдовыборок.

РАССМАТРИВАЕМЫЕ МЕТОДЫ БЕГГИНГА

Существуют следующие подходы к получению псевдовыборки из значений временного ряда.

1. Построение псевдовыборки из неперекрывающихся блоков (nonoverlapping block bootstrap, block bootstrap, circular bootstrap, NBB) [9, 10]. Данные временного ряда делятся на некоторое

количество непересекающихся блоков. Длина блока – настраиваемый параметр. При построении псевдовыборок каждый блок может с некоторой вероятностью попасть в любую из них. Например, выстроим блоки длиной 3 элемента из ряда с 12 значениями: $X = \{X_1, \dots, X_{12}\}$:

$$(X_1, X_2, X_3), (X_4, X_5, X_6), (X_7, X_8, X_9), (X_{10}, X_{11}, X_{12}).$$

При составлении псевдовыборки из них можно выбирать любые блоки с возвратом. Если длина псевдовыборки равна 12, то можно взять 4 блока, например:

$$(X_4, X_5, X_6), (X_1, X_2, X_3), (X_{10}, X_{11}, X_{12}), (X_4, X_5, X_6).$$

Обратим внимание на то, что блоки могут повторяться, измерения в псевдовыборке не обязаны идти в том же временном порядке, что и в исходных данных, поэтому стационарность исходного временного ряда не обязана сохраняться.

2. Построение псевдовыборки из перекрывающихся блоков (moving block bootstrap, MBV) [11–13]. Блоки, на которые делятся данные временного ряда, могут перекрываться. Длина блока – настраиваемый параметр. При построении псевдовыборок каждый блок может с некоторой вероятностью попасть в любую. В целом, этот случай отличается от первого тем, что блоки могут накладываться друг на друга. Пример из предыдущего пункта можно преобразовать так:

$$(X_1, X_2, X_3), (X_3, X_4, X_5), (X_5, X_6, X_7), (X_7, X_8, X_9), \\ (X_9, X_{10}, X_{11}), (X_{10}, X_{11}, X_{12}).$$

Обратим внимание на то, что начало каждого блока (кроме первого) накладывается на окончание предыдущего блока. Число перекрывающихся элементов, естественно, настраивается. В целом, дальнейшее построение псевдовыборки следует пункту 1, поэтому стационарность исходного ряда, если она есть, не гарантирует стационарность псевдовыборок.

3. Построение стационарной псевдовыборки [14]. Отличается от первых двух случаев тем, что исследователи задались идеей сохранения свойства стационарности для выделяемых псевдовыборок при условии, что исходный временной ряд X стационарен. Длина блоков не фиксируется. Вместо этого задается некоторая вероятность окончания блока p . Первый элемент блока X_i подбирается случайно, а затем каждый последующий элемент либо попадает в блок с вероятностью $(1 - p)$, либо блок завершается и начинается новый. Длины блоков L_1, L_2, \dots подчиняются геометрическому

распределению, поэтому вероятность получить блок длиной l :

$$p(L_j = l) = (1 - p)^{l-1} p.$$

У блока настраивается длина L_j и начальное положение X_i . Получаем набор блоков $B_j(i, L_j) = \{X_1^*, X_2^*, \dots, X_{L_j}^*\}$. Здесь звездочкой обозначен тот факт, что значения, выбираемые из ряда, не обязаны образовывать непрерывный промежуток, но элементы выбираются вслед за начальным элементом X_i псевдовыборки: $X_1^* = X_i$. На рис. 1 схематично представлен процесс подбора элементов временного ряда в псевдовыборку при применении стационарного беггинга: X_i – это последовательность значений временного ряда, X_i^* – подобранная при беггинге псевдовыборка). Каждый последующий элемент должен по времени находиться за ранее выбранным элементом (X_{i+1}^* всегда позже момента, соответствующего элементу ряда X_i^*). При этом, между ними могут быть промежутки.

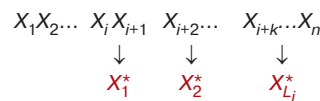


Рис. 1. Пример выбора элементов временного ряда X в псевдовыборку X^* при использовании стационарного беггинга (элемент X_{i+1}^* по времени всегда идет после ранее выбранного X_i^*)

В работе [15] исследуется вопрос выбора оптимальной длины блока и сделан вывод о том, что длина должна быть пропорциональна кубическому корню из длины временного ряда.

В представленной работе также рассматривается и четвертый метод, во многом повторяющий стационарный беггинг. Основное отличие состоит в запрете использовать блоки (значения в следующем блоке могли относиться к более раннему промежутку времени, чем в предыдущем). Вместо этого фактически используется один блок, в котором каждое предыдущее значение относится к более раннему измерению, чем следующее. При необходимости выровнять длину псевдовыборки с длиной ряда используется интерполяция.

В [16] автору удалось построить сравнение методов по смещению математического ожидания (которое появляется в связи с тем, что из временного процесса невозможно выделять независимые величины), в [15] дается более простая оценка смещения для математического ожидания E и дисперсии V :

$$B(\hat{E}(b)) = \frac{A_1}{b} + \bar{\sigma} \left(\frac{1}{b} \right), \\ B(\hat{V}(b)) = \frac{A_2}{b} + \bar{\sigma} \left(\frac{1}{b} \right). \quad (3)$$

Здесь b – длина блока в схеме беггинга, A_1, A_2 – константы, подробности расчета которых приведены в [16]. Так, при рассмотрении оценок первого порядка различные подходы к беггингу теоретически остаются идентичными.

Метод MBV (перекрывающиеся блоки) обладает меньшими моментами второго порядка по сравнению с NBV (неперекрывающиеся блоки) и стационарным беггингом [15, 16]. Оценки для каждого метода приведены в выражениях (4)–(6):

$$V_{NBV}(\hat{E}(b)) = \frac{4\pi^2 g_1(0)}{3n^3} b + \bar{\sigma} \left(\frac{b}{n^3} \right), \quad (4)$$

$$V_{NBV}(\hat{V}(b)) = \frac{4\pi^2 g_2(0)}{3n^3} b + \bar{\sigma} \left(\frac{b}{n^3} \right),$$

$$V_{MBV}(\hat{E}(b)) = \frac{2\pi^2 g_1(0)}{n^3} b + \bar{\sigma} \left(\frac{b}{n^3} \right), \quad (5)$$

$$V_{MBV}(\hat{V}(b)) = \frac{2\pi^2 g_2(0)}{n^3} b + \bar{\sigma} \left(\frac{b}{n^3} \right),$$

$$V_{SB}(\hat{E}(b)) = \frac{4\pi^2 g_1(0) + 2\pi G_1}{n^3} b + \bar{\sigma} \left(\frac{b}{n^3} \right), \quad (6)$$

$$V_{SB}(\hat{V}(b)) = \frac{4\pi^2 g_2(0) + 2\pi G_2}{n^3} b + \bar{\sigma} \left(\frac{b}{n^3} \right),$$

где g_1, g_2, G_1, G_2 – функции, вид и свойства которых разобраны в [15, 16], n – число элементов временного ряда. У метода на основе перекрывающихся блоков MBV моменты второго порядка ниже, чем у NBV. Оценка смещения для стационарного беггинга существенно отличается по виду выражения от двух других случаев, поэтому сравнение затруднительно. Считается, что дисперсия при стационарном беггинге выше. При этом у него есть и преимущества. В [14] исследованы свойства стационарного беггинга и показано, что псевдовыборка представляет собой марковскую цепь, порядок которой зависит от того, сколько совпадающих блоков попадает в псевдовыборку.

В различных статистических пакетах, в основном, реализован алгоритм MBV, как теоретически выигрывающий у других основных стратегий проведения беггинга. Различные модификации беггинга для временных рядов широко применяются для моделирования и прогнозирования временных процессов [2–5, 17].

Алгоритм обработки значений временного ряда для применения одной из стратегий бутстрепа представлен в работах [17, 18]. Его блок-схема отражена на рис. 2.

Итак, в вычислительном эксперименте представлены стандартные модели ARIMA и ETS²,

² Exponential time smoothing – экспоненциальное сглаживание.

нейросетевые модели (LSTM³, GRU⁴, рекуррентная нейронная сеть RNN⁵, полносвязный многослойный перцептрон). Их сравнение представлено в табл. 1.

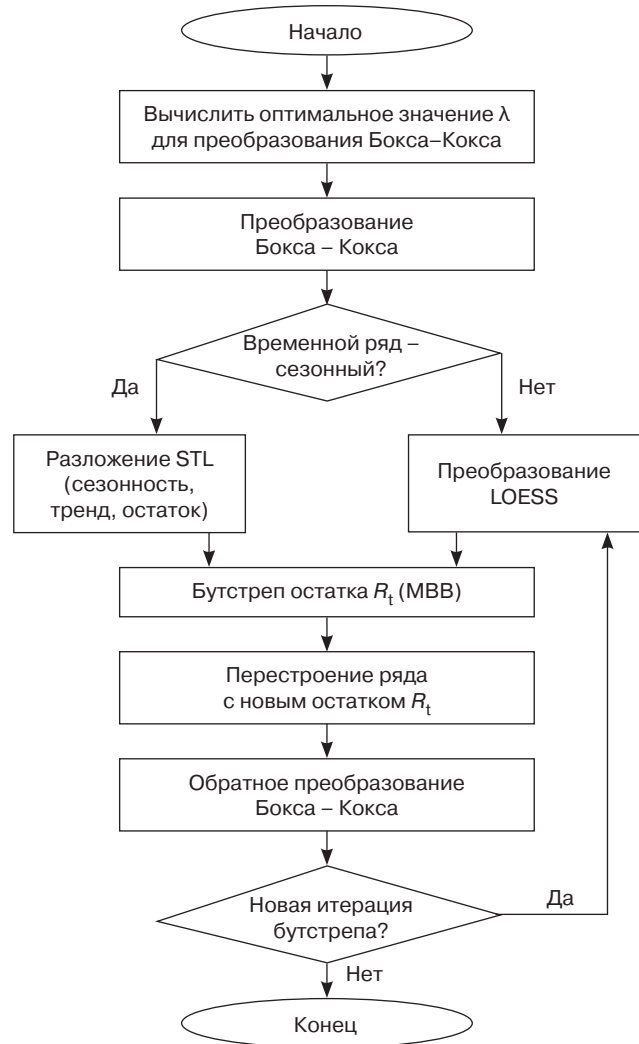


Рис. 2. Иллюстрация примера выбора элементов временного ряда X в псевдовыборку X^* при использовании стационарного беггинга (элемент X_{i+1}^* по времени всегда идет после ранее выбранного X_i^*). λ – параметр преобразования Бокса – Кокса; LOESS (locally estimated scatterplot smoothing) – метод локального сглаживания диаграммы рассеивания; STL (seasonal and trend decomposition using LOESS) – метод разложения временных рядов на тренд, сезонность и остатки

Цель работы состоит в сравнении точности прогноза моделей, построенных с использованием различных подходов к беггингу, друг с другом и с другими моделями, которые часто применяются для моделирования и прогнозирования временных рядов.

³ Long-short term memory – долгосрочная короткая память.

⁴ Gated recurrent unit – управляемый рекуррентный блок.

⁵ Recurrent neural network.

Таблица 1. Сравнение групп моделей, участвующих в вычислительном эксперименте

Группа моделей	Алгоритм обучения (принцип подгонки модели под значения ряда)	Дополнительные показатели сравнения моделей
Стандартные (ARIMA, ETS)	Принцип максимального правдоподобия	Информационные критерии Акаике, Байеса (Шварца)
Нейросетевые (LSTM, GRU, RNN, полносвязные нейронные сети) [19]	Алгоритм обратного распространения ошибки (с добавлением batch normalization, dropout) / градиентный спуск	–
Модели на основе беггинга	После разделения на тренд-сезонность остаток с помощью STL обработка остатка и перестройка модели с новым остатком	–

ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

В рамках вычислительного эксперимента рассматривается несколько моделей временного ряда: денежные доходы населения (НИИ)⁶ и реальный объем сельскохозяйственного производства (AGR)⁷ по данным макроэкономической статистики РФ, а также акции Сбербанка на московской бирже⁸. В работе не рассматриваются экономические вопросы, данные взяты для моделирования и прогнозирования. Все данные, кроме последнего года, используются для обучения. Тестовым периодом, на который делается прогноз, является последний год временного ряда. Следует подчеркнуть, что начало мирового экономического кризиса в 2008 г. и начало кризиса вокруг украинского государственного переворота 2014 г. исключены из рассмотрения, т.к. поведение показателей в это время сильно меняется (меняется математическое ожидание, дисперсия ряда, появляется гетероскедастичность). Данные предыдущего и следующего годов относительно кризисного года склеиваются. График для ряда денежных доходов (отношение среднедушевого денежного дохода в текущем

месяце к этому же показателю за соответствующий месяц прошлого года) и его функций автокорреляции (autocorrelation function, ACF) и частной автокорреляции (partial autocorrelation function, PACF) [1] представлены на рис. 3 и 4. Графики для ряда реального объема сельскохозяйственного производства (относительный показатель изменения объема сельскохозяйственного производства по отношению к некоторому периоду) представлены на рис. 5 и 6.



Рис. 3. Временной ряд индекса денежных доходов населения (в %) по данным макроэкономической статистики РФ за 1993–2019 гг.

В качестве метрик сходства измеряются оценки MAE и RMSE^{9, 10} [1]. Результаты обработки индекса денежных доходов населения представлены в табл. 2 (жирным шрифтом отмечены наилучшие модели по различным критериям, точность – 0.01). Кроме моделей на основе беггинга и стандартных моделей ARIMA и ETS [20] в эксперименте представлены модели на основе нейронных сетей GRU, LSTM, RNN [21–24].

⁹ Mean absolute error – средняя абсолютная ошибка.

¹⁰ Root mean square error – среднеквадратичная ошибка.

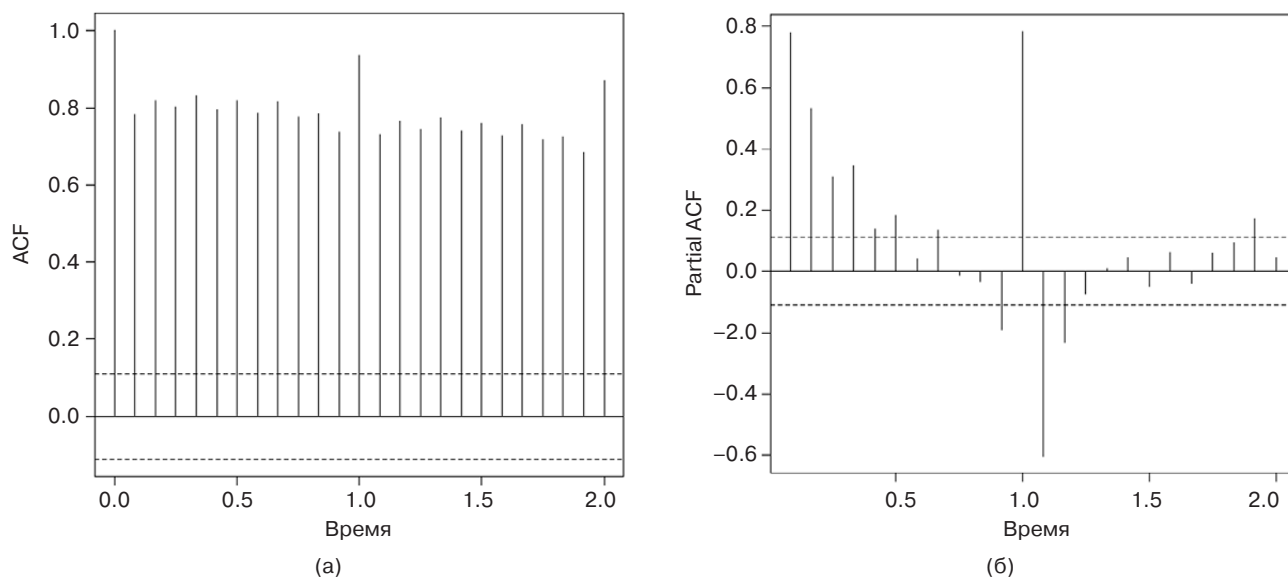


Рис. 4. Графики функций ACF (а) и PACF (б) для временного ряда денежных доходов населения по данным макроэкономической статистики РФ

Таблица 2. Модели индекса денежных доходов по данным макроэкономической статистики РФ и их прогнозы на тестовый период¹¹

Модель временного ряда	MAE	RMSE
NBB	4.67	5.53
MBB	4.78	5.57
Стационарный беггинг	4.10	4.91
Метод LOESS	3.49	4.57
ARIMA	5.86	7.01
ETS	6.57	8.47
RNN	3.88	4.45
Модель LSTM	5.91	6.63
Модель GRU	3.94	4.36

Модели временного ряда на основе беггинга показывают лучшие результаты, чем стандартные модели рядов ARIMA/ETS. Среди них лучший прогноз дала модель на основе стационарного беггинга. При этом качество прогноза модели на основе стационарного беггинга уступает некоторым нейросетевым моделям (RNN и GRU) и методу LOESS (разложение ряда STL).

В эксперименте 2 рассматривается индекс реального объема сельскохозяйственного производства России за период 2000–2020 гг. На рис. 5 и 6 представлены графики ряда и функций ACF и PACF. Все

рассматриваемые модели настраивались на периоде обучения 2000–2020 гг. (из него удалены кризисные годы 2008 и 2014, данные склеены), результаты их прогнозов на тестовый период (2021 г.) сопоставлены в табл. 3.

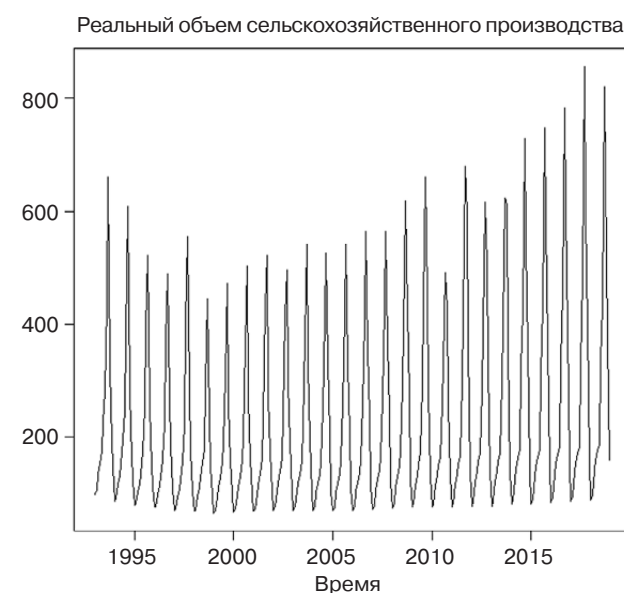


Рис. 5. Временной ряд реального объема сельскохозяйственного производства (в %) по данным макроэкономической статистики РФ¹²

¹¹ Единый архив экономических и социологических данных. Динамические ряды макроэкономической статистики РФ. Индекс денежных доходов населения. http://sophist.hse.ru/hse/1/tables/HNI_M_I.htm. Дата обращения 01.09.2023. [Unified archive of economic and sociological data. Dynamic series of macroeconomic statistics of the Russian Federation. Index of money incomes of the population (in Russ.). http://sophist.hse.ru/hse/1/tables/HNI_M_I.htm. Accessed September 01, 2023.]

¹² Единый архив экономических и социологических данных. Динамические ряды макроэкономической статистики РФ. Индекс реального объема сельскохозяйственного производства. http://sophist.hse.ru/hse/1/tables/AGR_M_I.htm. Дата обращения 01.09.2023. [Unified archive of economic and sociological data. Dynamic series of macroeconomic statistics of the Russian Federation. Index of real volume of agricultural production (in Russ.). http://sophist.hse.ru/hse/1/tables/AGR_M_I.htm. Accessed September 01, 2023.]

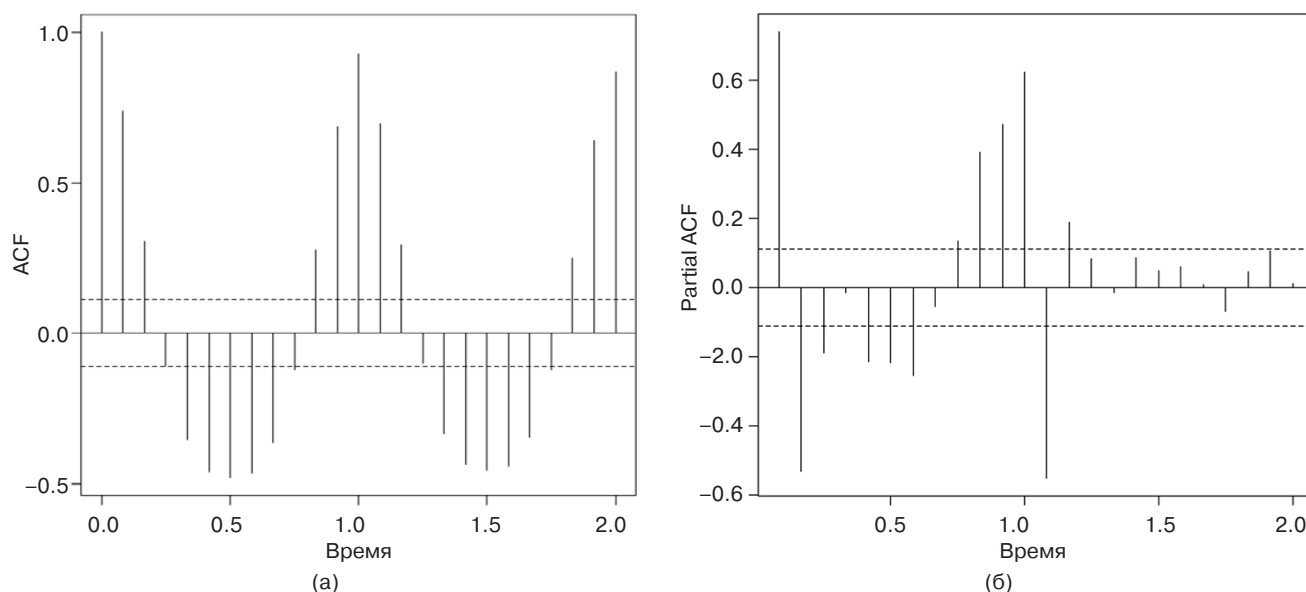


Рис. 6. Графики функций ACF (а) и PACF (б) для временного ряда реального объема сельскохозяйственного производства по данным макроэкономической статистики РФ (безразмерный)

Таблица 3. Модели индекса реального объема сельскохозяйственного производства по данным макроэкономической статистики РФ и их прогнозы на тестовый период

Модель временного ряда	MAE	RMSE
NBB	15.01	22.47
MBB	16.63	25.80
Стационарный беггинг	17.11	25.59
ARIMA	13.24	18.51
ETS	17.22	25,40
Модель LSTM	8.78	15.41
Модель GRU	10.11	16.34
RNN	10.51	16.17

В этом эксперименте среди моделей на основе беггинга лучший результат показал подход NBB (на основе непересекающихся блоков), имея приблизительно равные характеристики по качеству прогноза на тестовый период со стандартными моделями ARIMA/ETS. При этом нейросетевые модели LSTM, GRU и RNN превосходят по прогнозу стандартные модели и модели на основе беггинга (первая – существенно, последние две – незначительно).

Отдельно рассмотрим ряд курса биржевых акций – акций Сбербанка РФ¹³. Этот ряд обладает гетероскедастичностью. В связи с тем, что курс акций несезонный, для каждой нейросетевой системы возможными остаются только два подхода: сделать прогноз сразу на весь тестовый период (интегральный)

¹³ Курс акций ПАО Сбербанк (SBER). <https://www.moex.com/ru/issue.aspx?board=TQBR&code=SBER>. Дата обращения 01.09.2023. [Sberbank (SBER) stock price (in Russ.). <https://www.moex.com/ru/issue.aspx?board=TQBR&code=SBER>. Accessed September 01, 2023.]

или делать пошаговые прогнозы, объявляя каждый новый шаг частью обучающей выборки для перехода к следующему моменту времени. Графики функций ACF и PACF представлены на рис. 7.

Результаты прогнозирования представлены в табл. 4.

Таблица 4. Модели временного ряда курса акций Сбербанка РФ

Модель временного ряда	MAE	RMSE
NBB	23.78	25.53
MBB	24.60	26.39
Стационарный беггинг	20.94	22,61
ARIMA	11.23	42.11
ETS	4.95	20.68
Сеть RNN	80.53	86.39
Модель LSTM	76.95	81.40
Модель GRU	24.66	85.05

Лучшие результаты показывают «классические» методы моделирования рядов – модели ETS и ARIMA. Стационарный беггинг показывает чуть худшие результаты, но существенно превосходит все нейросетевые модели. Следует обратить внимание на то, что стандартные модели ARIMA/ETS лучше описывают временной ряд статистически в случае отсутствия сезонности. Основная идея беггинга – определить свойства шумовой составляющей ряда. Очевидно, это имеет смысл делать для рядов, обладающих сезонными или циклическими закономерностями. Моделирование шума для несезонных рядов не приводит к улучшению прогнозирования (стандартные модели дали лучший прогноз, чем модели на основе применения беггинга).

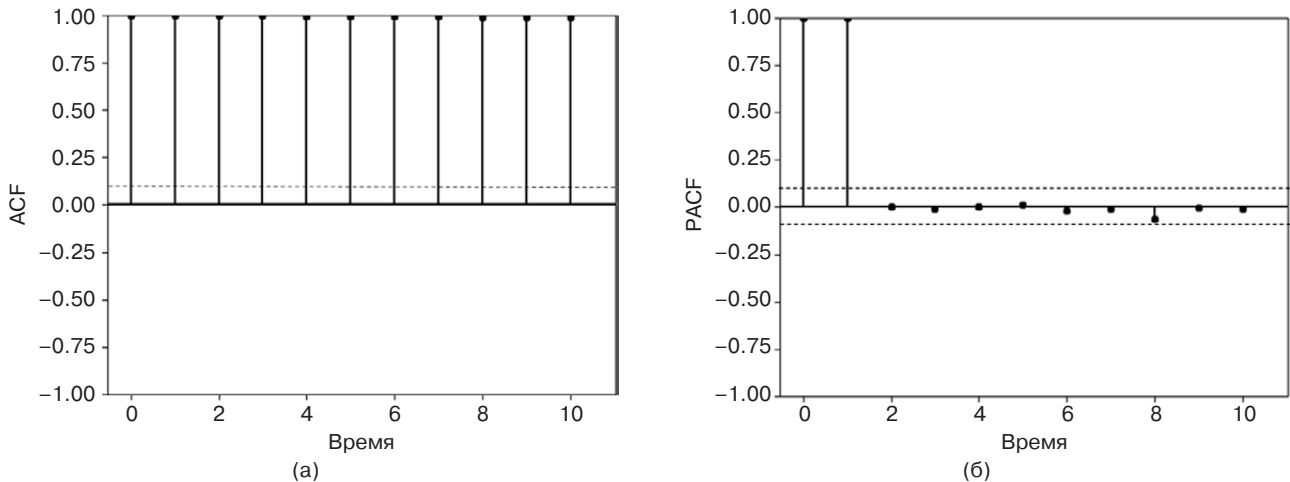


Рис. 7. Графики функции ACF (а) и PACF (б) для временного ряда курса акций Сбербанка РФ (безразмерный)

ЗАКЛЮЧЕНИЕ

В работе представлен анализ различных подходов к беггингу временных рядов и примеры их применения к несезонным и сезонным временным рядам. В вычислительных экспериментах результаты моделей, применяющих беггинг, сравниваются с прогнозами стандартных моделей (ARIMA и ETS) и моделей на основе нейронных сетей (RNN, LSTM, GRU).

При обработке несезонного временного ряда моделирование шумовой составляющей не привело к улучшению моделирования всего ряда и к улучшению его прогноза. В этом эксперименте наилучшие результаты среди всех трех групп моделей дали стандартные модели ARIMA/ETS. Обратим внимание, что часто применяющиеся при моделировании процессов различной природы нейросетевые модели дали прогноз, худший по качеству по сравнению с ARIMA/ETS (табл. 4).

При моделировании сезонного временного ряда лучшие результаты показали нейросетевые модели, которые сейчас активно применяются при моделировании временных рядов, и метод LOESS. Модели

на основе беггинга превосходили стандартные модели ARIMA и ETS. Можно констатировать, что при беггинге удалось лучше смоделировать остаток ряда (который получается при удалении тренда и сезонной составляющей ряда). Таким образом, работу над различными схемами бутстрепа следует продолжать, повышая их точность. Кроме этого, представляется возможным повышать точность моделирования и прогнозирования, работая отдельно над трендом, сезонностью и остатком. При этом не удастся определить, какой именно вид бутстрепа будет лучшим образом моделировать остаток заданного ряда. Каждый из видов оказывается лучшим для своего набора сезонных временных рядов. В представленной работе различные подходы к бутстрепу реализованы в самом простом виде. Из результатов экспериментов следует, что работу следует продолжать, редактируя различные возможности по построению бутстрепов, а также комбинируя различные подходы к моделированию тренда, шума и остатка.

Вклад авторов. Все авторы в равной степени внесли свой вклад в исследовательскую работу.

Authors' contribution. All authors equally contributed to the research work.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Hyndman R.J., Athanasopoulos G. *Forecasting: Principles and Practice*. 2nd ed. Publisher OTexts; 2018. 382 p. ISBN 978-0-9875-0711-2
- Özen K., Yildirim D. Application of bagging in day-ahead electricity price forecasting and factor augmentation. *Energy Econ.* 2021;103(2):105573. <http://doi.org/10.1016/j.eneco.2021.105573>
- Meira E., Oliveira F.L.C., de Menezes L.M. Forecasting natural gas consumption using Bagging and modified regularization techniques. *Energy Econ.* 2022;106(3):105760. <http://doi.org/10.1016/j.eneco.2021.105760>
- Kim D., Baek J-G. Bagging ensemble-based novel data generation method for univariate time series forecasting. *Expert Syst. Appl.* 2022;203(1):117366. <http://doi.org/10.1016/j.eswa.2022.117366>
- Liu X., Liu A., Chen J.L., Li G. Impact of decomposition on time series bagging forecasting performance. *Tourism Manag.* 2023;97(2):104725. <https://doi.org/10.1016/j.tourman.2023.104725>
- Beletskaya N., Petrusевич D. Linear Combinations of Time Series Models with Minimal Forecast Variance. *J. Commun. Technol. Electron.* 2023;67(1):144–158. <http://doi.org/10.1134/S1064226922130022>

7. Musatov D., Petrushevich D. Modeling of forecasts variance reduction at multiple time series prediction averaging with ARMA(1, q) functions. In: *Proceedings of MIP Computing-V 2022: V International Scientific Workshop on Modeling, Information Processing and Computing*. 2022;3091:1–11. <https://doi.org/10.47813/dnit-mip5/2022-3091-1-11>
8. Голов В.А., Петрусевич Д.А. О выборе метода распознавания астрономических объектов на основе анализа исходных данных, полученных по программе Sloan Digital Sky Survey DR14. *Russian Technological Journal*. 2021;9(3):66–77. <https://doi.org/10.32362/2500-316X-2021-9-3-66-77>
[Golov V.A., Petrushevich D.A. Data analysis methods in astronomic objects classification (Sloan Digital Sky Survey DR14). *Russ. Technol. J.* 2021;9(3):66–77 (in Russ.). <https://doi.org/10.32362/2500-316X-2021-9-3-66-77>]
9. Carlstein E. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.* 1986;14(3):1171–1179. <https://doi.org/10.1214/aos/1176350057>
10. Carlstein E., Do K-A., Hall P., Hesterberg T., Künsch H.R. *Matched-Block Bootstrap for Dependent Data*. Research Report 74. Seminar für Statistik. ETH Zentrum. Zürich, Switzerland; 1995. 25 p.
11. Künsch H.R. The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* 1989;17(3):1217–1261. <https://doi.org/10.1214/aos/1176347265>
12. Götze F., Künsch H.R. Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.* 1996;24(5):1914–1933. <http://doi.org/10.1214/aos/1069362303>
13. Bühlman P., Künsch H.P. Block length selection in the bootstrap for time series. *Comput. Stat. Data Anal.* 1999;31(3):295–310. [https://doi.org/10.1016/S0167-9473\(99\)00014-6](https://doi.org/10.1016/S0167-9473(99)00014-6)
14. Politis D.N., Romano J.P. The stationary bootstrap. *J. Am. Stat. Assoc.* 1994;89(428):1303–1313. <https://doi.org/10.1080/01621459.1994.10476870>
15. Politis D.N., White H. Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Rev.* 2004;23(1):53–70. <https://doi.org/10.1081/ETC-120028836>
16. Lahiri S.N. Theoretical comparison of block bootstrap methods. *Ann. Statist.* 1999;27(1):386–404. <https://doi.org/10.1214/aos/1018031117>
17. Bergmeir C., Hyndman R. J., Benitez J.M. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *Int. J. Forecast.* 2016;32(2):303–312. <https://doi.org/10.1016/j.ijforecast.2015.07.002>
18. Petropoulos F., Hyndman R.J., Bergmeir C. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *Eur. J. Oper. Res.* 2018;268(2):545–554. <http://doi.org/10.1016/j.ejor.2018.01.045>
19. Haykin S. *Neural Networks and Learning Machines*. Pearson Education; 2011. 936 p. ISBN 978-0-1330-0255-3
20. Box G., Jenkins G. *Time Series Analysis: Forecast and Management*. John Wiley & Sons; 2015. 712 p. ISBN 978-1-1185-6749-18
21. Rasjid Z.E., Setiawan R., Effendi A. A Comparison: Prediction of Death and Infected COVID-19 Cases in Indonesia Using Time Series Smoothing and LSTM Neural Network. *Procedia Comput. Sci.* 2021;179(5):982–988. <http://doi.org/10.1016/j.procs.2021.01.102>
22. Liu X., Lin Z., Feng Z. Short-term offshore wind speed forecast by seasonal ARIMA – A comparison against GRU and LSTM. *Energy*. 2021;227:120492. <http://doi.org/10.1016/j.energy.2021.120492>
23. Shahid F., Zameer A., Muneeb M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons & Fractals*. 2020;140:110212. <https://doi.org/10.1016/j.chaos.2020.110212>
24. Amalou I., Mouhni N., Abdali A. Multivariate time series prediction by RNN architectures for energy consumption forecasting. *Energy Rep.* 2022;8:1084–1091. <https://doi.org/10.1016/j.egy.2022.07.139>

Об авторах

Грамович Ян Вадимович, студент, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). <https://orcid.org/0009-0000-2908-7611>

Мусатов Данила Юрьевич, студент, ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). Scopus Author ID 57469172700, <https://orcid.org/0000-0003-0673-5393>

Петрусевич Денис Андреевич, к.ф.-м.н., доцент, кафедра высшей математики Института искусственного интеллекта ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: petrushevich@mirea.ru, petrdenis@mail.ru. Scopus Author ID 55900513600, ResearcherID AAA-6661-2020, SPIN-код РИНЦ 7999-6345, <https://orcid.org/0000-0001-5325-6198>

About the authors

Ian V. Gramovich, Student, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). <https://orcid.org/0009-0000-2908-7611>

Danila Yu. Musatov, Student, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). Scopus Author ID 57469172700, <https://orcid.org/0000-0003-0673-5393>

Denis A. Petrushevich, Cand. Sci. (Phys.-Math.), Associate Professor, Higher Mathematics Department, Institute of Artificial Intelligence, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: petrushevich@mirea.ru, petrdenis@mail.ru. Scopus Author ID 55900513600, ResearcherID AAA-6661-2020, RSCI SPIN-code 7999-6345, <https://orcid.org/0000-0001-5325-6198>