

Сравнение подходов к применению беггинга при моделировании временных рядов

Н. В. Белецкая^{*,2}, Д. А. Петрусевич^{*}, М. Е. Савельев^{*}

*** Российский Технологический Университет (МИРЭА), Москва, Россия*

***** Институт проблем передачи информации им. А.А. Харкевича, Москва, Россия*

Поступила в редколлегию 26.06.2025 г. Принята 01.07.2025 г.

Аннотация—В представленной работе произведено сравнение применения беггинга при различных размерах псевдовыборок (бутстреп) на данных временных рядов из набора М3. Стратегия беггинга в классификации интерпретируется как усиление простых классификаторов. При переходе в область, где появляется временная составляющая, основная идея преобразуется к попытке определить свойства распределения данных, к которому принадлежит остаток временного ряда, полученный после удаления тренда и сезонной части. На основе данных остатка временного ряда формируется несколько новых временных рядов (псевдовыборок), которые затем усредняются (или к ним применяется функция, которая даёт новые значения для остатка). Ряд восстанавливается из обновлённого остатка, тренда и сезонности. При выполнении этой операции предполагается, что дисперсия остатка временного ряда снижается, позволяя сделать более точный прогноз. Интерес представляет установление взаимосвязи между точностью прогноза модели, построенной на обновлённом остатке, и параметрами построения псевдовыборок. В представленной работе исследуется взаимосвязь длины псевдовыборки, из которой конструируется обновлённый остаток, и точности прогноза на тестовый период. В вычислительном эксперименте представлено несколько подходов: линейный беггинг, беггинг на основе подвижной выборки (перекрывающихся блоков) MBV, круговой беггинг (CB), построение стационарной выборки (стационарный беггинг). На примере набора данных М3 сравнивается точность подходов к беггингу друг с другом и со стандартными моделями ARIMA/ETS.

КЛЮЧЕВЫЕ СЛОВА: беггинг, псевдовыборка, MBV, NBV, CB, SB, LB, стационарный беггинг, линейный беггинг.

DOI: 10.53921/18195822_2025_25_2_197

1. ВВЕДЕНИЕ

Беггинг, применённый к данным временных рядов, состоит в выделении множества подпоследовательностей измерений из данных остатка временного ряда (построении псевдовыборки, бутстреп), усреднении или применении другой функции (медианы, минимума, максимума и т.д.) к выделенным подпоследовательностям и конструировании новой модели ряда с обновлённым остатком. Такой подход нацелен на улучшение качества описания остатка при разделении данных ряда на части тренд – сезонность – остаток [1]. Исходя из предположения, что сезонность и тренд описаны качественно, эти части удаляются из ряда. Отдельной обработке тренда, а также тренда и сезонности, посвящена работа [2]. Теоретически остаток должен быть случайным, но на практике в нём всегда присутствуют элементы неучтённых параметров, в связи с тем, что сами значения ряда зависят не только от своих предыдущих значений, но и от внешних параметров, которые в модель не входят. Существуют следующие подходы к бутстрепу на основе значений временного ряда.

- Построение последовательности из неперекрывающихся блоков (nonoverlapping block bootstrap, block bootstrap, NBV) [3, 4], подробно рассмотренное в работе [5]. Ряд разделяется на набор неперекрывающихся блоков, размер которых задан заранее. Блоки выбираются в псевдопоследовательность с возвращением и могут перемешиваться без соблюдения последовательности по времени. Этот метод особенно полезен в ситуациях, когда исходный временной ряд имеет сложную структуру или когда данные не соответствуют стандартным предположениям о распределении.
- Построение псевдовыборки из перекрывающихся блоков или подвижной псевдовыборки (moving block bootstrap, MBV) [6–8] отличается от предыдущего тем, что блоки могут накладываться друг на друга. Длины перекрываемых отрезков и длина блока – настраиваемые параметры.
- Круговая псевдовыборка (circular block bootstrap, CBV), в целом, аналогична подвижной выборке из пункта 2, но при нехватке элементов для последнего блока (например, в случае, когда длина выстраиваемой псевдовыборки не делится нацело на число блоков) недостающие элементы берутся в начале ряда [9]. Результаты моделирования с использованием круговой выборки представлены в работе [10]. Подход особенно актуален в контексте анализа временных данных, где наблюдения имеют естественную упорядоченность и могут быть зависимыми от предыдущих значений. Круговая псевдовыборка позволяет учитывать эти особенности, обеспечивая при этом необходимую гибкость в оценке параметров и неопределенности.
- Линейная псевдовыборка (LPV) [11] основана на построении матрицы ковариации, оценивающей взаимосвязи между значениями выборки. Благодаря этому, удаётся снизить взаимосвязи между значениями ряда, избегая приближения к линейной зависимости между значениями и появления гетероскедастичности.
- Построение стационарной псевдовыборки [12, 13]. Выделяемая псевдовыборка сохраняет стационарность 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. Таким образом, выделенные элементы сохраняют свой порядок по времени. Выбирать элементы из ряда можно не подряд, но по временной составляющей они следуют друг за другом по возрастанию. Длины блоков L_1, L_2, \dots подчиняются геометрическому распределению.

В работах [7, 12, 14, 15] исследуется оптимальная длина псевдовыборки. Она оценивается теоретически как $\sqrt[3]{l}$, где l – длина временного ряда. В представленной работе проведены эксперименты, в которых реализованы стационарный, линейный круговой беггинг и MBV на наборе данных [16] с целью проверить это значение и оценить на большом наборе данных практическую оптимальную величину длины псевдовыборки.

2. СМЕЩЕНИЕ МАТ.ОЖИДАНИЯ И ДИСПЕРСИИ

В [14] построена простая оценка смещения $B()$ для математического ожидания E и дисперсии V :

$$B(\hat{E}(b)) = \frac{A_1}{b} + \bar{o}\left(\frac{1}{b}\right),$$

$$B(\hat{V}(b)) = \frac{A_2}{b} + \bar{o}\left(\frac{1}{b}\right).$$

Здесь b – длина блока в схеме беггинга, $A_{1,2}$ – константы, подробности расчета которых приведены в [12]. Так, при рассмотрении оценок первого порядка различные подходы к беггингу теоретически остаются идентичными. В [12] построено сравнение методов по смещению математического ожидания (которое появляется в связи с тем, что из временного процесса невозможно выделять независимые величины). Метод MBV (перекрывающиеся блоки) обладает меньшими моментами второго порядка по сравнению с NBV (неперекрывающиеся блоки) и со стационарным беггингом [12, 14]. Оценки для каждого метода приведены в выражениях 2:

$$\begin{aligned} V_{MBV}(\hat{E}(b)) &= \frac{2\pi^2 g_1(0)}{n^3} b + \bar{o}\left(\frac{b}{n^3}\right), \\ V_{MBV}(\hat{V}(b)) &= \frac{2\pi^2 g_2(0)}{n^3} b + \bar{o}\left(\frac{b}{n^3}\right), \\ V_{NBV}(\hat{E}(b)) &= \frac{4\pi^2 g_1(0)}{3n^3} b + \bar{o}\left(\frac{b}{n^3}\right), \\ V_{NBV}(\hat{V}(b)) &= \frac{4\pi^2 g_2(0)}{3n^3} b + \bar{o}\left(\frac{b}{n^3}\right), \\ V_{SB}(\hat{E}(b)) &= \frac{4\pi^2 g_1(0) + 2\pi G_1}{n^3} b + \bar{o}\left(\frac{b}{n^3}\right), \\ V_{SB}(\hat{V}(b)) &= \frac{4\pi^2 g_2(0) + 2\pi G_2}{n^3} b + \bar{o}\left(\frac{b}{n^3}\right). \end{aligned}$$

где $g_{1,2}$, $G_{1,2}$ – функции, вид и свойства которых разобраны в [12, 14], n – число элементов временного ряда. У метода на основе перекрывающихся блоков MBV моменты второго порядка ниже, чем у NBV. Видно, что с ростом длины блока, который встраивается в псевдовыборку, снижается теоретический предел отклонения мат.ожидания и дисперсии от характеристик исходного ряда. 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 [13] 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.

В рамках представленного ниже вычислительного эксперимента проверяются прогнозы моделей, полученных при различном подходе к способу построения и длине псевдовыборки. В работе [10] представлена полученная экспериментально зависимость ошибки прогноза на тестовый период от длины псевдовыборки для кругового беггинга и метода на основе перекрывающихся блоков MBV.

3. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Модели, реализующие различные подходы к беггингу, были проверены на наборе данных МЗ [16]. На основе каждого ряда построены псевдовыборки различной относительной длины. Для того, чтобы сформировать обновлённый ряд, к ним применены функции: среднее арифметическое, среднее геометрическое, медиана. Графики зависимости точности прогноза (ошибки RMSE на тестовом периоде в 12 последних временных шагов) от длины псевдопоследовательности представлены в приложениях на рисунках 1, 2, 3, 4 в приложениях. Таблица 4 содержит точность лучших моделей в сравнении со стандартными моделями ARIMA/ETS. Длина бутстрепа выражается относительно через l – длину временного ряда. Сезонность рядов

учитывалась автоматически, специально в вычислительном эксперименте не обрабатывалась. Заметим, что характер зависимости совпадает для стационарного беггинга, бутстрепа на основе движущихся блоков и круговой выборки. Наилучшие результаты показали стационарный беггинг, круговой беггинг, беггинг на основе подвижных блоков MBV с длинами блока 0.03*l* и 0.07*l* при применении функций медианы и среднего геометрического к псевдовыборкам, хотя в силу простоты чаще всего применяется вычисление среднего арифметического над псевдовыборками. В случае линейного беггинга лучшие предсказания сильно уступают другим методам, но, при этом, заметен тренд на улучшение точности при увеличении длины псевдовыборки.

Рассмотрим отдельно подмножество сезонных временных рядов. Хотя при беггинге работа идёт с остатками, полученными после удаления сезонности, всё же часть информации о взаимосвязях между различными моментами времени в остатке сохраняется. Здесь почти для всех экспериментов с зафиксированной длиной ряда лучшей функцией над псевдовыборкой является среднее арифметическое (см. рисунки 5, 6). Единственное исключение – ряды с периодом наблюдений в год, где на некоторых длинах псевдовыборок медиана чуть лучше (см. рисунок 7). При этом, почти всегда есть один локальный минимум между длинами 0.14 и 0.20. Что же касается обработки всех рядов в целом, на рисунке 8 видно, что наилучшие значения получены как раз в этом локальном минимуме. В случае линейного беггинга наблюдается тренд снижения точности с ростом длины выборки. Причём, наилучшие результаты были получены именно при применении линейного беггинга. При использовании других подходов зависимость остаётся прямой. В таблице 5 приведены лучшие модели для сезонных рядов набора данных. Для них наиболее удачные результаты дал линейный беггинг. И в этом эксперименте наилучшие показатели у стандартных моделей ARIMA/ETS, которые продолжают демонстрировать хорошие показатели для рядов с ярко выраженными трендом и сезонностью.

В таблицах 1, 2, 3 приведены точности прогнозирования стандартных моделей и моделей на основе беггинга на подмножествах рядов набора M3 [16] с заданным периодом наблюдений: годовым, квартальным и месячным соответственно. Для рядов с периодом наблюдений в год лучшим оказался круговой беггинг, для квартальных лучшие показатели у кругового беггинга и беггинга MBV на основе пересекающихся блоков, для рядов длиной в месяц лучшие показатели дал линейный беггинг. Все эти модели превзошли в среднем стандартные модели временных рядов ARIMA/ETS.

Таблица 1. Модели на основе беггинга, ETS, ARIMA для рядов с периодом наблюдений в год из набора данных M3

Модель бутстрепа	<i>RMSE</i>	<i>MAE</i>
СВ (круговой)	1191.29	1034.92
MBV	1208.18	1049.52
Стационарный беггинг	1280.35	1105.28
Линейный беггинг	1235.22	1074.34
ARIMA	1322.90	1151.96
ETS	1438.16	1257.00

4. ЗАКЛЮЧЕНИЕ

В представленной работе и связанных с ней исследованиях [5, 10] реализовано несколько видов беггинга – стратегии обработки временных рядов, нацеленной на обработку остатков. Теоретически считается, что при построении модели, полностью учитывающей поведение ряда, математические модели тренда и сезонности настроены на поведение ряда так, что оставшаяся часть распределена равномерно вокруг нуля на всём отрезке по времени. Но на практике

Таблица 2. Модели на основе беггинга, *ETS*, *ARIMA* для рядов с периодом наблюдений в квартал из набора данных *M3*

Модель бутстрепа	<i>RMSE</i>	<i>MAE</i>
СВ (круговой)	626.60	536.19
МВВ	624.10	533.26
Стационарный беггинг	839.43	740.48
Линейный беггинг	672.71	577.28
ARIMA	691.81	594.43
ETS	776.71	664.26

Таблица 3. Модели на основе беггинга, *ETS*, *ARIMA* для рядов с периодом наблюдений в месяц из набора данных *M3*

Модель бутстрепа	<i>RMSE</i>	<i>MAE</i>
СВ (круговой)	826.17	690.72
МВВ	825.89	691.03
Стационарный беггинг	853.90	765.67
Линейный беггинг	811.40	673.36
ARIMA	821.59	686.16
ETS	892.27	741.83

обычно такого описания не удаётся достичь, поэтому некоторая информация, обладающая системностью, с остатке присутствует. Основная идея беггинга – уточнение математической модели остатка, что позволяет лучше спрогнозировать его поведение. В представленной работе идёт анализ моделей беггинга самих по себе, но ясно, что эту концепцию перспективно использовать вместе с повышением качества прогнозирования моделей сезонности и тренда. Мы предполагаем в дальнейшем построить гибридные модели, объединяющие беггинг для обработки остатка и вейвлет-анализ для обработки тренда и сезонности.

В вычислительном эксперименте проанализированы подходы к беггингу: стационарный беггинг, построение круговой выборки, выборки на основе движущихся (перекрывающихся) блоков и линейный беггинг. На наборе данных *M3* лучшие показатели прогнозов получены при применении стационарного, кругового подхода и МВВ. Отдельно для сезонных рядов наилучшие результаты получены при применении линейного беггинга. Интересно отметить, что у кругового, стационарного беггинга и метода на основе перекрывающихся блоков, в целом, одна и та же качественная зависимость от длины псевдовыборки. При этом, хотя в [7, 12, 14] теоретически была получена оценка оптимальной длины псевдопоследовательности, анализ этой зависимости для набора *M3* показал наличие нескольких локальных минимумов. В случае линейного беггинга наблюдается повышение точности моделей при увеличении длины псевдовыборки, хотя, в целом, точность линейного беггинга хуже. Возможно, при дальнейшем увеличении длины псевдовыборки можно получить сопоставимые с другими видами беггинга результаты.

Отдельно была выделена подгруппа сезонных рядов. И на ней на рядах всех периодов наблюдения, за исключением периода в год, наилучшие результаты получены при применении линейного беггинга. Хотя при применении беггинга происходит удаление тренда и сезонности часть информации о взаимосвязях между точками временного ряда всё равно сохраняется. Можно предположить, что более сложная обработка данных, применяемая в линейном беггинге, позволяет это учитывать лучше, чем при применении других подходов.

Подчеркнём, что на всех подмножествах данных модели на основе беггинга продемонстрировали более высокую точность, чем стандартные модели временных рядов ARIMA/ETS.

СПИСОК ЛИТЕРАТУРЫ

1. Liu X., Liu A., Chen J.L., Li G. Impact of decomposition on time series bagging forecasting performance. *Tourism Management*. 2023; vol.97, p.104725. doi:10.1016/j.tourman.2023.104725
2. Мохнаткина У.С., Парфенов Д.В., Петрусевич Д.А. Анализ подходов к определению тренда в структуре временного ряда. *Российский Технологический Журнал*. 2024;12(3), 2024, Том 12(3), с.93–103 doi: 10.32362/2500-316X-2024-12-3-93-103
3. Carlstein, E. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *The Annals of Statistics*, 1986, vol.14, p.1171–1179
4. Carlstein E., Do K-A., Hall P., Hesterberg T., Kunsch H. R. Matched-block bootstrap for dependent data. *Research Report 74*, Seminar ftir Statistik, ETH, 1995, Zurich
5. Грамович Я.В., Мусатов Д.Ю., Петрусевич Д.А. Применение беггинга в прогнозировании временных рядов. *Российский Технологический Журнал*. 2024, Том 12(1), с.101-110. doi: 10.32362/2500-316X-2024-12-1-101-110.
6. Kunsch, H. R. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 1989, vol.17, p.1217–1261
7. Buhlman P., Kunsch H.P. Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis*, 1999, vol.31(3), p.295–310. doi: 10.1016/S0167-9473(99)00014-6
8. Gotze F., Kunsch H.R. Second-order correctness of the blockwise bootstrap for stationary observations. *The Annals of Statistics*, 1996, vol.24(5), p.1914–1933. doi: 10.1016/S0167-9473(99)00014-6
9. Bergmeir C., Hyndman R. J., Benitez J.M. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting*, 2016, vol. 21, p. 303–312. doi: 10.1016/j.ijforecast.2015.07.002.
10. Петрусевич Д.А., Дорошенко Г.В., Мусатов Д.Ю., Пашков Б.Б., Савельев М.Е. Применение подвижного и кругового беггинга при моделировании временных рядов. *Сборник трудов III Национальной научно-практической конференции «ФУНДАМЕНТАЛЬНЫЕ, ПОИСКОВЫЕ, ПРИКЛАДНЫЕ ИССЛЕДОВАНИЯ И ИННОВАЦИОННЫЕ ПРОЕКТЫ»*(г. Москва, 27-28 мая 2024 г.) – Москва, 2024 г. — с. 124–128. URL: <https://elibrary.ru/item.asp?id=72645318>
11. McMurry T.L. , Politis D.N. Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 2010, vol. 31(6), p.471–482
12. Lahiri S.N. Theoretical comparison of block bootstrap methods. *The Annals of Statistics*, 1999, vol. 27(1), p.386-404
13. Politis D.N., Romano J.P. The stationary bootstrap. *Journal of the American Statistical Association*. *Journal of the American Statistical Association*, 1994; vol. 89, p. 1303–1313. doi: 10.1080/01621459.1994.10476870
14. Politis D.N., White H. Automatic Block-Length Selection for the Dependent Bootstrap. *Econometric Reviews*, 2004; vol. 23(1), p. 53-70. doi: 10.1081/ETC-120028836
15. Petropoulos F., Hyndman R.J., Bergmeir C. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, 2018, vol. 268(2), p. 545–554. doi: 10.1016/j.ejor.2018.01.045.
16. Makridakis S., Hibon M. The M3-competition: results, conclusions and implications *International Journal of Forecasting*, 2000; vol.16(4), p.451-476. doi:10.1016/S0169-2070(00)00057-1

5. ПРИЛОЖЕНИЯ

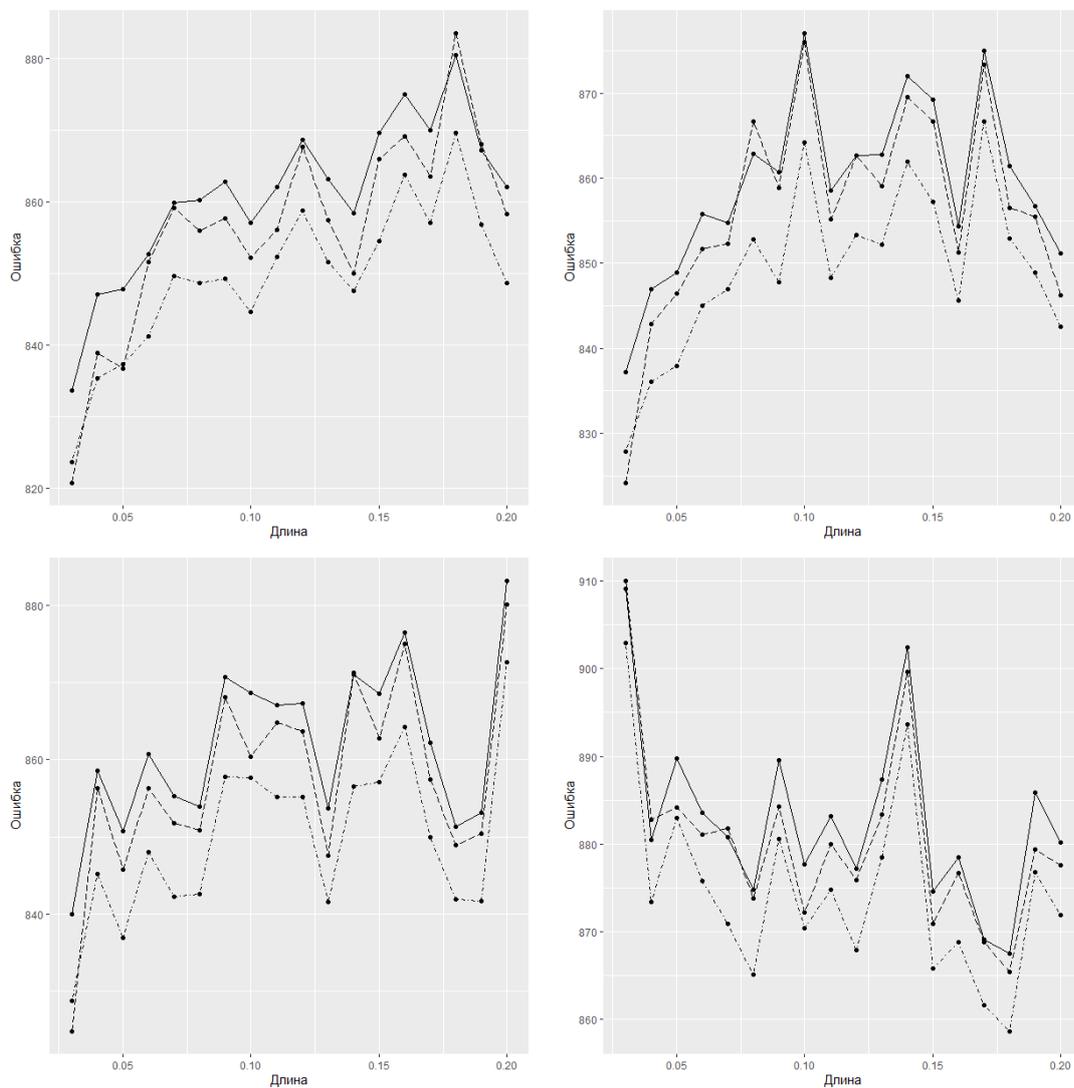


Рис. 1. Графики зависимости оценки прогнозов моделей для временных рядов с периодом наблюдений в месяц по набору данных МЗ (слева сверху круговой беггинг, справа - метод MBV, слева снизу стационарный беггинг, справа - линейный беггинг)

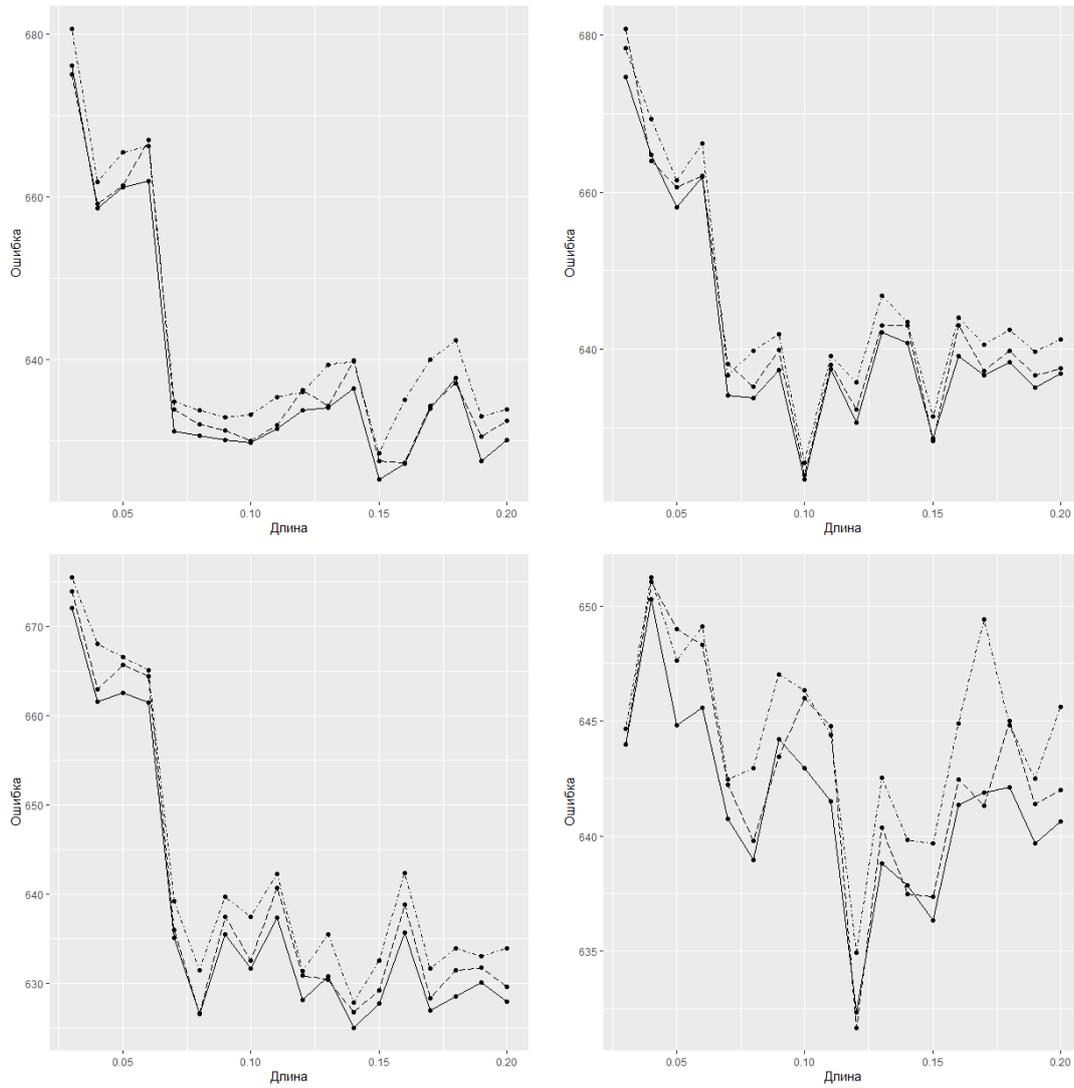


Рис. 2. Графики зависимости оценки прогнозов моделей для временных рядов с периодом наблюдений в квартал по набору данных МЗ (слева сверху круговой беггинг, справа - метод МВВ, слева снизу стационарный беггинг, справа - линейный беггинг)

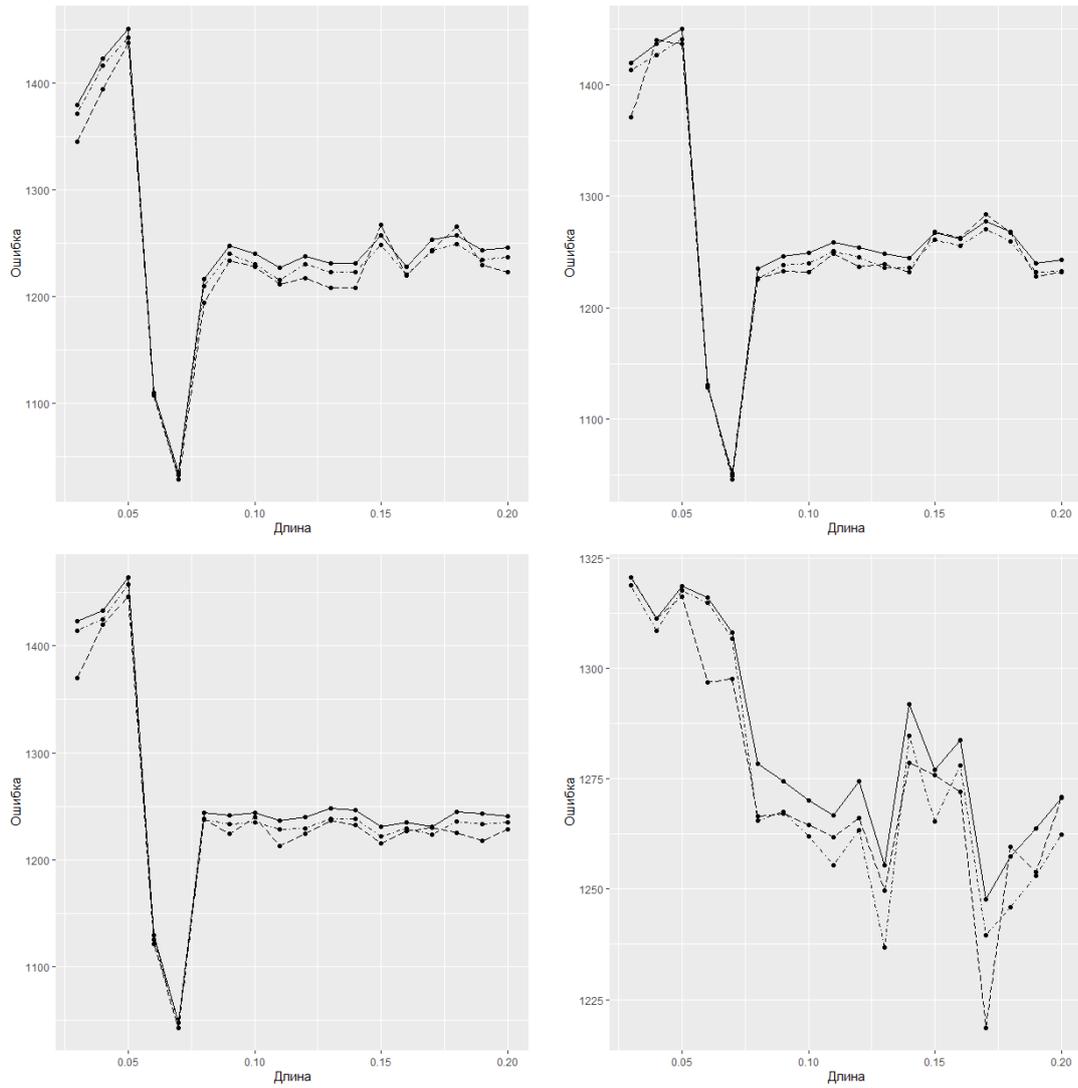


Рис. 3. Графики зависимости оценки прогнозов моделей для временных рядов с периодом наблюдений в год по набору данных МЗ (слева сверху круговой беггинг, справа - метод МВВ, слева снизу стационарный беггинг, справа - линейный беггинг)

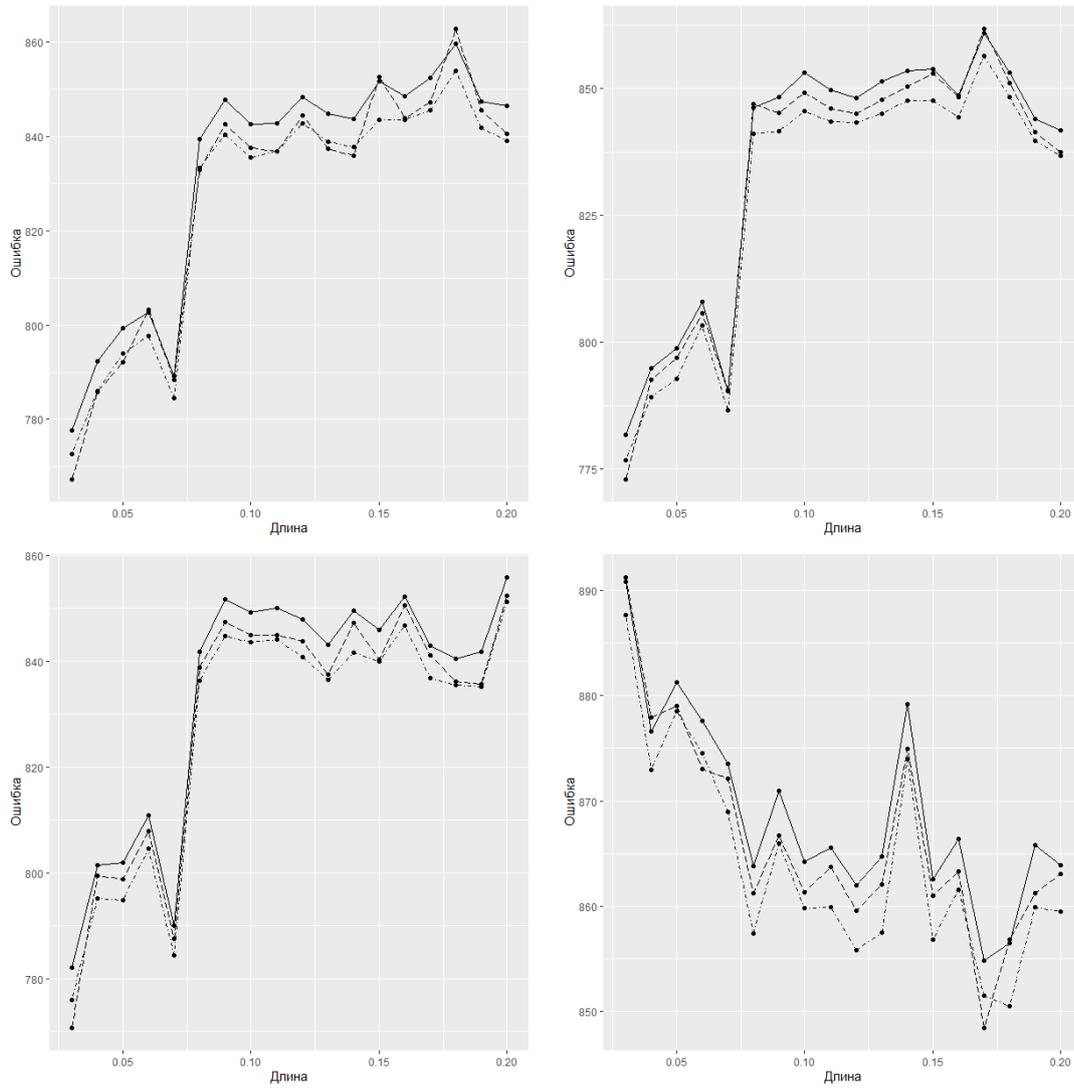


Рис. 4. Графики зависимости оценки прогнозов моделей для временных рядов по набору данных МЗ (слева сверху круговой беггинг, справа - метод МВВ, слева снизу стационарный беггинг, справа - линейный беггинг)

Таблица 4. Модели на основе беггинга, *ETS*, *ARIMA* для набора данных МЗ

Модель, функция бутстрепа	Длина бутстрепа	<i>RMSE</i>	<i>MAE</i>
СВ (круговой), среднее арифметическое	0.03 <i>l</i>	810.01	686.98
СВ (круговой), среднее геометрическое	0.03 <i>l</i>	772.69	653.78
СВ (круговой), медиана	0.03 <i>l</i>	767.26	649.72
СВ (круговой), среднее арифметическое	0.07 <i>l</i>	789.14	671.38
СВ (круговой), медиана	0.07 <i>l</i>	788.40	670.46
СВ (круговой), среднее геометрическое	0.07 <i>l</i>	784.55	666.73
МВВ, среднее арифметическое	0.03 <i>l</i>	781.69	661.38
МВВ, медиана	0.03 <i>l</i>	772.94	654.09
МВВ, среднее геометрическое	0.03 <i>l</i>	776.83	656.39
МВВ, среднее арифметическое	0.07 <i>l</i>	790.42	672.44
МВВ, медиана	0.07 <i>l</i>	790.61	672.42
МВВ, среднее геометрическое	0.07 <i>l</i>	786.58	668.54
Стационарный, среднее арифметическое	0.03 <i>l</i>	782.19	662.72
Стационарный, медиана	0.03 <i>l</i>	770.69	652.23
Стационарный, среднее геометрическое	0.03 <i>l</i>	776.03	656.46
Стационарный, среднее арифметическое	0.07 <i>l</i>	789.14	671.38
Стационарный, медиана	0.07 <i>l</i>	788.40	670.46
Стационарный, среднее геометрическое	0.07 <i>l</i>	784.55	666.73
Линейный, среднее геометрическое	0.17 <i>l</i>	851.5	725.91
<i>ARIMA</i>	–	840.94	716.83
<i>ETS</i>	–	980.98	841.00

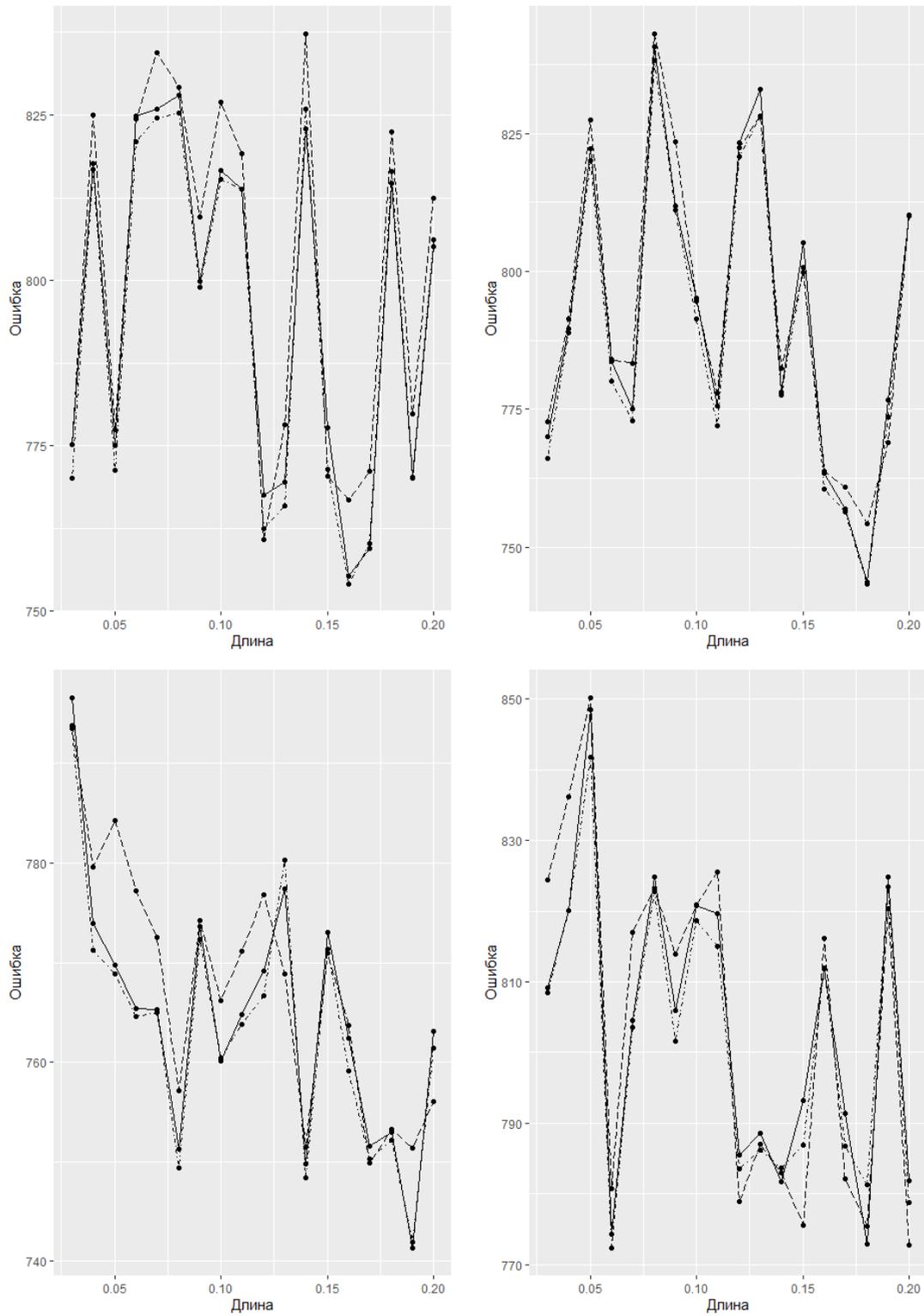


Рис. 5. Графики зависимости оценки прогнозов моделей для сезонных временных рядов с периодом наблюдений в месяц по набору данных МЗ (слева сверху круговой беггинг, справа - метод МВВ, слева снизу стационарный беггинг, справа - линейный беггинг)

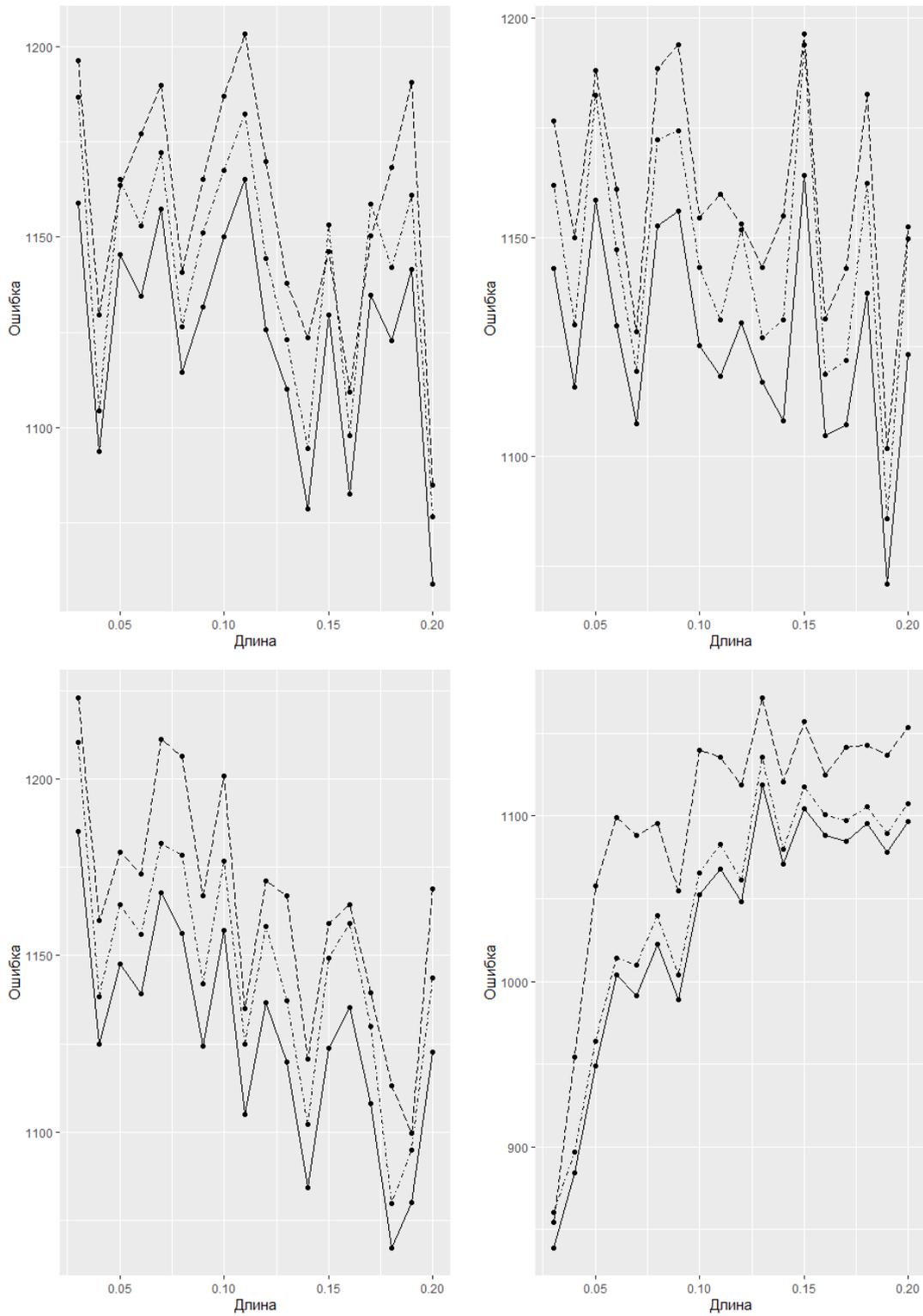


Рис. 6. Графики зависимости оценки прогнозов моделей для сезонных временных рядов с периодом наблюдений в квартал по набору данных МЗ (слева сверху круговой беггинг, справа - метод МВВ, слева снизу стационарный беггинг, справа - линейный беггинг)

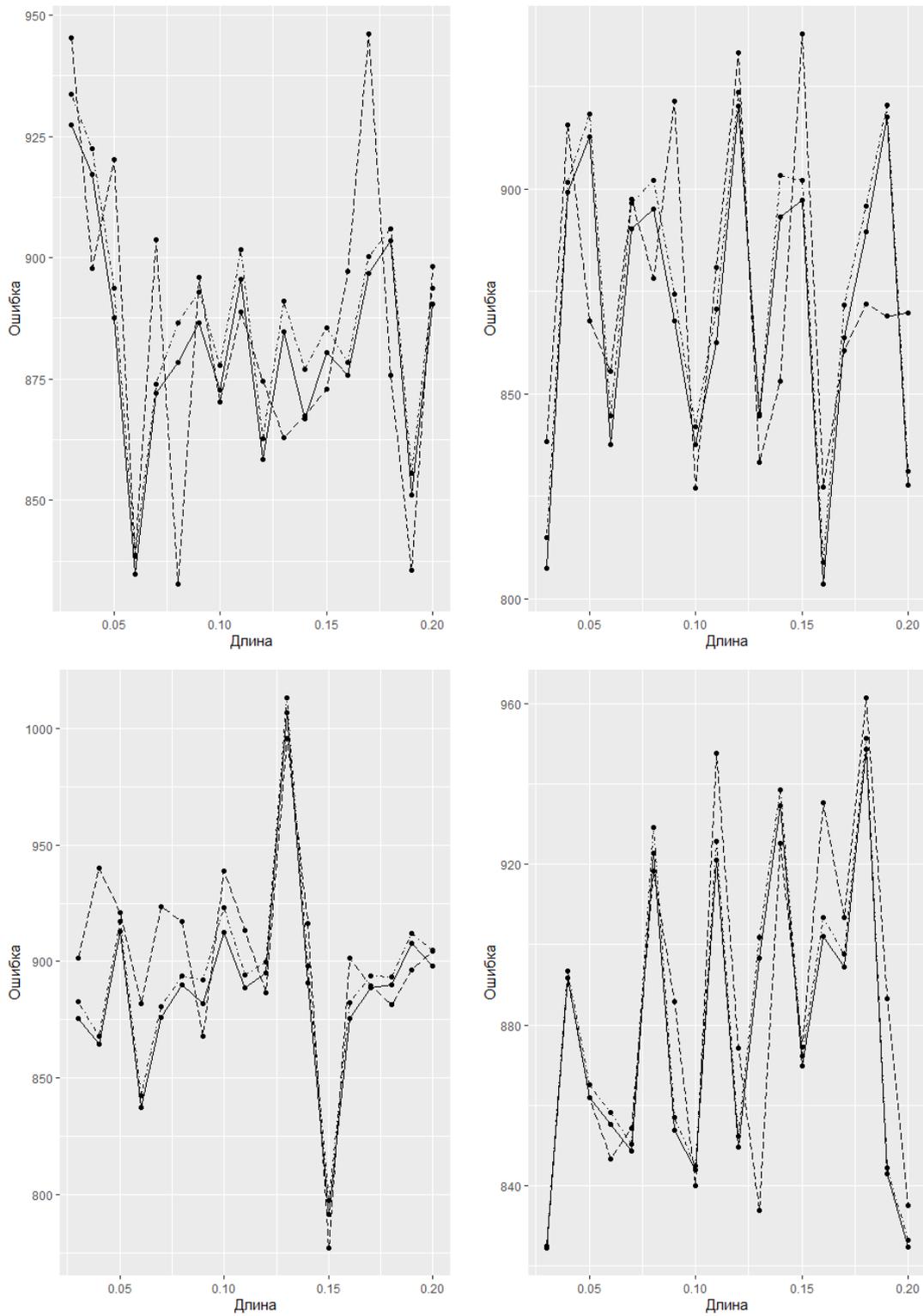


Рис. 7. Графики зависимости оценки прогнозов моделей для сезонных временных рядов с периодом наблюдений в год по набору данных МЗ (слева сверху круговой беггинг, справа - метод МВВ, слева снизу стационарный беггинг, справа - линейный беггинг)

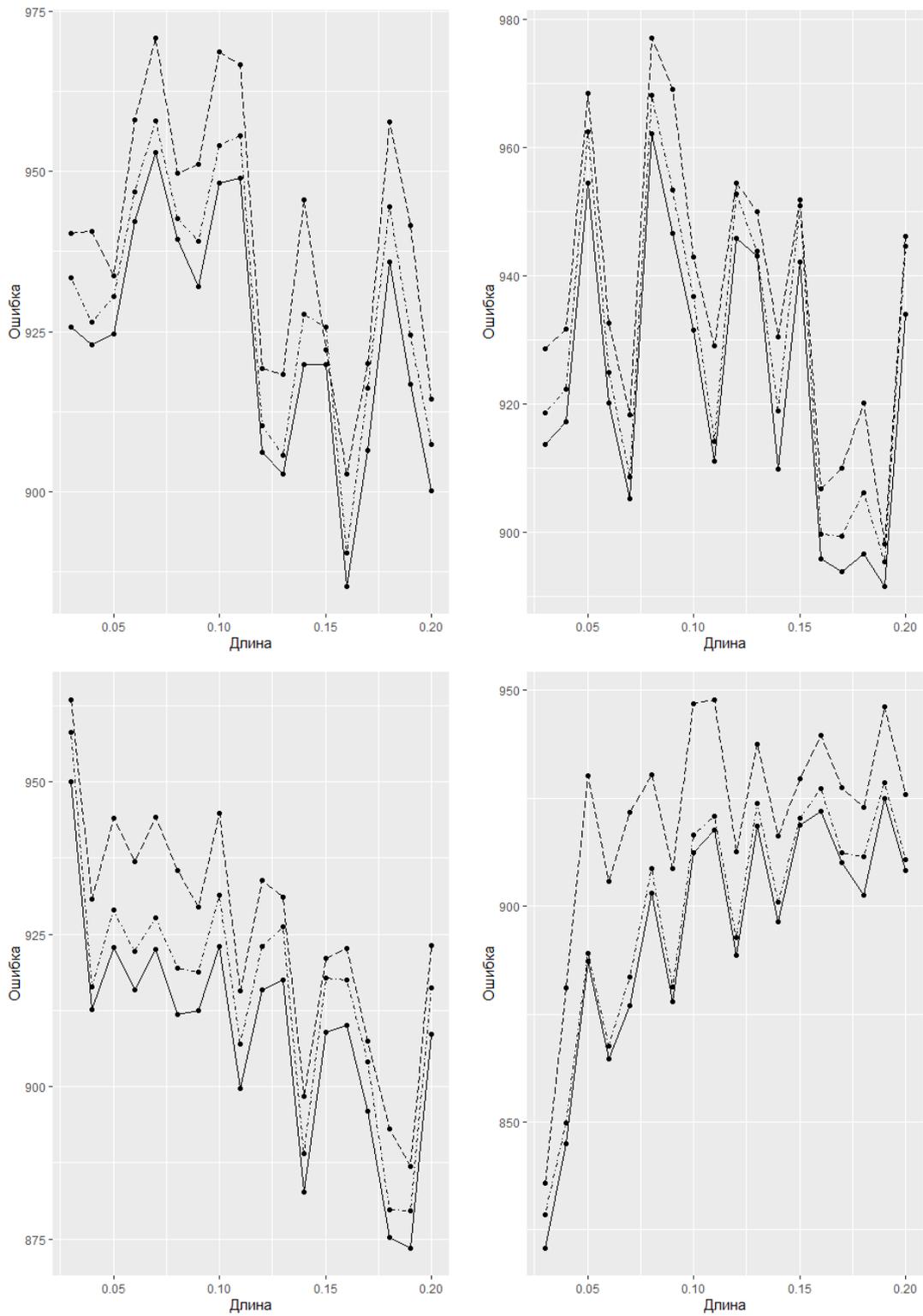


Рис. 8. Графики зависимости оценки прогнозов моделей для сезонных временных рядов по набору данных МЗ (слева сверху круговой беггинг, справа - метод МВВ, слева снизу стационарный беггинг, справа - линейный беггинг)

Таблица 5. Модели на основе беггинга, *ETS*, *ARIMA* для сезонных рядов набора данных M3

Модель, функция бутстрепа	Длина бутстрепа	<i>RMSE</i>	<i>MAE</i>
СВ (круговой), среднее арифметическое	0.16 <i>l</i>	885.20	748.60
СВ (круговой), среднее геометрическое	0.16 <i>l</i>	890.51	750.25
МВВ, среднее арифметическое	0.19 <i>l</i>	891.50	751.77
МВВ, среднее геометрическое	0.19 <i>l</i>	895.38	751.29
Стационарный, среднее арифметическое	0.19 <i>l</i>	873.46	734.27
Стационарный, среднее геометрическое	0.19 <i>l</i>	879.65	735.60
Линейный, среднее арифметическое	0.03 <i>l</i>	820.69	685.56
Линейный, медиана	0.03 <i>l</i>	835.85	697.37
Линейный, среднее геометрическое	0.19 <i>l</i>	828.52	691.40
ARIMA	–	641.97	525.69
ETS	–	767.41	637.71

Comparison of approaches to implementation of bagging in time series modelling

N. V. Beletskaya, D. A. Petrushevich, M. E. Savelev

In the paper comparison of approaches to bagging with use of bootstraps of various size has been made. The bagging strategy in classification is interpreted as strengthening of simple classifiers. In case if the time component appears, the main idea is converted to an attempt to determine the properties of the data distribution to which the remainder of the time series belongs, obtained after deleting the trend and the seasonal part. Based on the data of the remainder of the time series, several new time series (bootstraps) are formed. Then they are averaged (or a function is applied to them). Thus, new value of remainder part is constructed. The series is reconstructed from the updated remainder, trend and seasonality. This operation assumes that the variance of the time series remainder is reduced, allowing more accurate prediction. It is of interest to establish the relationship between the accuracy of the prediction of the model built on the updated remainder and parameters of bootstraps. In the paper the relationship between the length of the bootstrap from which the updated remainder is constructed and the accuracy of the forecast for the test period is investigated. Several approaches are presented in the computational experiment: linear bagging, moving blocks bootstrap MBB, circular bagging (CB), construction of stationary bootstraps (stationary bagging). Using the M3 dataset as an example, the accuracy of the bagging approaches is compared to each other and to the accuracy of the standard ARIMA/ETS models.

KEYWORDS: bagging, bootstrap, MBB, NBB, CB, SB, LB, stationary bagging, linear bagging.