

УДК 519.65

DOI: 10.46960/1816-210X_2023_1_33

ПОСТРОЕНИЕ РОБАСТНЫХ ИНТЕРВАЛЬНЫХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ДИНАМИКИ СТРУКТУРНО-СЛОЖНОЙ СИСТЕМЫ

Е.В. НикульчевORCID: 0000-0003-1254-9132 e-mail: nikulchev@mirea.ru

МИРЭА – Российский технологический университет

*Москва, Россия***А.А. Червяков**ORCID: 0000-0002-5638-8361, e-mail: achervyakov@roskazna.ru

Федеральное казначейство Минфина России

Москва, Россия

Предложен тип робастных интервальных моделей, отличающихся от общепризнанных тем, что интервалы, определяющие нечувствительность модели, относятся не к параметрам динамических уравнения, а к выходным измеряемым рядам. Результатом моделирования является интервал около каждой моделируемой точки, что позволяет прогнозировать систему в форме интервалов, обеспечивая надежность и гарантированную точность прогноза в форме интервала, включающего действительные значения.

Рассмотрена аналитическая система информационной поддержки принятия решений Федерального казначейства Минфина России, управляющая более чем 40 000 временных рядов. Робастные интервальные модели, построенные на основе использования нескольких классических аппроксимирующих моделей временных рядов, позволяют оперировать с динамикой на качественном макроуровне.

Ключевые слова: структурно-сложные системы, информационно-аналитические системы, прогнозирование, временные ряды, робастные модели, робастные интервальные модели, динамические системы.

ДЛЯ ЦИТИРОВАНИЯ: Никульчев, Е.В. Построение робастных интервальных моделей прогнозирования динамики структурно-сложной системы / Е.В. Никульчев, А.А. Червяков // Труды НГТУ им. Р.Е. Алексеева. 2023. № 1. С. 33-41. DOI: 10.46960/1816-210X_2023_1_33

CONSTRUCTION OF ROBUST INTERVAL MODELS PREDICTING DYNAMICS OF A STRUCTURALLY COMPLEX SYSTEM

E.V. NikulchevORCID: 0000-0003-1254-9132 e-mail: nikulchev@mirea.ru

MIREA – Russian Technology University

*Moscow, Russia***A.A. Chervyakov**ORCID: 0000-0002-5638-8361, e-mail: achervyakov@roskazna.ru

Federal Treasury of The Ministry of Finance of the Russian Federation

Moscow, Russia

Abstract. The presented type of robust interval models is distinct from well-recognized ones in that the intervals defining model insensitivity are associated not with dynamic equation parameters, but with measured output series. The result of the simulation is an interval adjacent to every simulated point which enables system prediction in the form of intervals, providing robustness and guaranteed accuracy of the prognosis in the form of interval including actual values. Analytical system controlling over 40.000 time series is considered as informational support in making decisions by Federal Treasury of Ministry of Finance of Russian Federation. Robust interval models constructed using multiple classic approximation models of time series allow to operate dynamics on qualitative micro-level.

Key words: structurally complex systems, informational and analytical systems, prediction, time series, robust models, robust interval models, dynamic systems.

FOR CITATION: Nikulchev E.V., Chervyakov A.A. Construction of robust interval models predicting dynamics of a structurally complex system. Transactions of NNSTU n.a. R.E. Alekseev. 2023. № 1. Pp. 33-41.
DOI: 10.46960/1816-210X_2023_1_33

1. Введение

Под динамической системой подразумевают модель эволюции некоторого процесса, состояние которой определяется на основании начального состояния и уравнений динамики [1]. Задачи прогнозирования на основе исторических данных представляют собой процесс определения динамической системы, поведение которой соответствовало бы наблюдаемому временному ряду. Современные методы построения прогнозирующих моделей [2, 3], включая методы искусственного интеллекта [4, 5], заключаются в поиске наиболее подходящих систем, с заданной степенью достоверности и адекватности описывающих эволюцию динамической системы. Для динамической системы понятия грубых и робастных моделей противоположно чувствительности [6], при этом грубость моделей часто подразумевает малые возмущения вблизи невозмущенных состояний, т.е. робастность представляет собой более общее понятие [7]. Как правило, робастная модель определяется как нечувствительная в динамическом поведении, динамических и частотных характеристиках при изменении параметров модели (*параметрическая робастность*). В противовес интервальным заданиям параметров можно ввести динамическую робастность по выходу. Решение, определяемое динамической системой с фиксированными параметрами, имеет интервальный вид. Назовем такую модель *интервальной робастной моделью*. В задаче построения модели по временному ряду, т.е. некорректной (обратной) задаче динамики под робастными интервальными моделями можно понимать динамические системы, решение которой задано значениями с доверительными интервалами. При этом интервал достоверно и адекватно включает исходный временной ряд и дает достоверный прогноз. В этих условиях несколько совершенно различных моделей могут определять исходный интервал около наблюдаемых состояний системы. Таким образом, для временного ряда $X(t)$, ($t = 0, 1, 2, \dots$) определены функции $f_i(x(t)) = y(t+1) \pm \varepsilon_i$, где f_i – различные аналитические функции, или функциональные разложения; t – дискретное время (номер точки); $x(t)$ – значение состояния системы в момент t ; y – вычисленное значение; ε_i – отклонение; i – номер функции; тогда интервальная робастная модель представляет собой объединение интервалов $[\min_i \varepsilon_i, \max_i \varepsilon_i]$ в каждой точке моделируемого или прогнозируемого временного ряда. В качестве робастной модели предлагается использовать объединение доверительных интервалов всех используемых моделей [8].

Справедливость и смысл предложенной операции и определения имеют место в случае стационарных процессов, а также для диссипативных систем, фазовые портреты которых, устойчивые или неустойчивые, имеют предельные циклы и не имеют точек бифуркаций [1]. К циклическим диссипативным системам относятся многие финансовые ряды, особенно агрегированные портфельные финансовые инструменты, включающие десятки тысяч рядов – как макропоказателей, так и быстро меняющихся микрорядов. Декомпозиция портфеля представляет собой большой набор связанных между собой систем (в терминах системного анализа – структурно-сложных динамических объектов) [9, 10]. Вместе с тем, структурная сложность и большое количество параметров уменьшают требования к качеству прогнозирования динамики.

В статье рассматривается система поддержки принятия решений Федерального казначейства Минфина России (ФК) [11], включающая более 20 000 кассовых планов клиентов ФК. В итоге система содержит плановые остатки на счетах, а также фактические остатки на их счетах (итого – более 40 000 временных рядов). Все входящие ряды по своей сути представляют собой временные ряды с сезонными и циклическими компонентами и плохо прогно-

зируемыми выбросами и структурными сдвигами. В [11] дается описание логической структуры системы принятия решений, оперирующую робастными моделями [12-14]. В системе происходит разделение всех поступающих данных на значимые и незначимые, прогнозируемые и непрогнозируемые с последующей обработкой их по разным трекам. Незначимые прогнозируются моделями машинного обучения без участия эксперта, значимые также прогнозируются моделями машинного обучения, но им на вход дополнительно подается экспертное мнение о структурных сдвигах и непрогнозируемых доходах/расходах по счету. Такой подход позволяет оперативно и с минимальными затратами со стороны экспертов формировать аналитические материалы непосредственно для обеспечения поддержки принятия решений.

2. Построение робастной интервальной модели

Определение 1. Пусть задан временной ряд $X = x(0), x(1), \dots, x_n, \dots$, и существует эволюционная функция $y(n) = f(x(n))$, (n – дискретное время). При этом $x(n+1)$ принадлежит интервалу $y \pm \varepsilon: x(n+1) \in [y - \varepsilon; y + \varepsilon]$. Тогда при $n \rightarrow \infty$, если для $\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0$, такое что для всякого состояния $\forall x(n)$ из $\|x(0) - y(0)\| < \delta$ следует $\|x(n) - y(n)\| < \varepsilon$.

Такую модель будем называть *робастной интервальной моделью*.

Определение 2. Для интервальных моделей f_1, f_2 задана операция объединения моделей, при этом исходный интервал в каждой точке $x(n)$ определяется как $[\min_{i=1,2} \varepsilon_i, \max_{i=1,2} \varepsilon_i]$.

Лемма 1. Пусть $f_1(x)$ – не удовлетворяет условию определения 1. Тогда может существовать функция $f_2(x)$, являющаяся робастной интервальной моделью исходной системы. Если рассматривать операцию объединения моделей, то результирующая модель тоже будет являться робастной. Доказательство очевидно, поскольку объединение интервалов по определению 2 включает робастный интервал.

Следствие. В случае расходящихся моделей интервал увеличивается, но свойство робастности не исчезает для диссипативных стационарных систем.

Рассмотрим задачу таргетирования остатков ФК. Данные, поступающие в систему, делятся на два типа: 1) плановые остатки – временные ряды с периодом в один (текущий) год и помесечной детализацией; 2) фактические остатки – временные ряды с периодом от начала текущего года до текущего дня и подневной детализацией. Выходной итоговый сводный прогноз должен быть временным рядом с периодом в один (текущий) год и подневной детализацией. Соответственно, для его построения актуально решение таких задач: 1) разбиение месячных итоговых значений на дневные для рядов плановых остатков и 2) прогнозирование с заданным горизонтом (до конца года) для рядов фактических остатков и преобразованных в подневные рядов плановых остатков.

3. Построение моделей на исторических данных плановых остатков

Для моделирования были взяты архивные данные за период 2015-2020 гг. и отобраны значимые ряды (например, доходы/расходы фондов ФСС, ПФР, ФФОМС) и агрегированные временные ряды нескольких незначимых (например, поступления муниципального уровня). Для решения задачи интерполяции было применено шесть следующих моделей.

1. Среднее взвешенное значение.
2. Тройное экспоненциальное сглаживание (Хольта – Уинтерса).
3. ARIMA (Бокса – Дженкинса).
4. Метод линейной регрессии.
5. Метод опорных векторов.
6. Метод градиентного бустинга над решающими деревьями.

Произведено исследование подходов обработки данных для улучшения точности результатов интерполирования:

- 1) исходные / предобработанные данные;
- 2) все исторические данные / данные месяц-к-месяцу;
- 3) относительные (доля/вес от итогового значения за месяц) / абсолютные значения.

На рис. 1-3 приведены графики, отражающие результаты моделирования, и сравнение с истинными значениями.

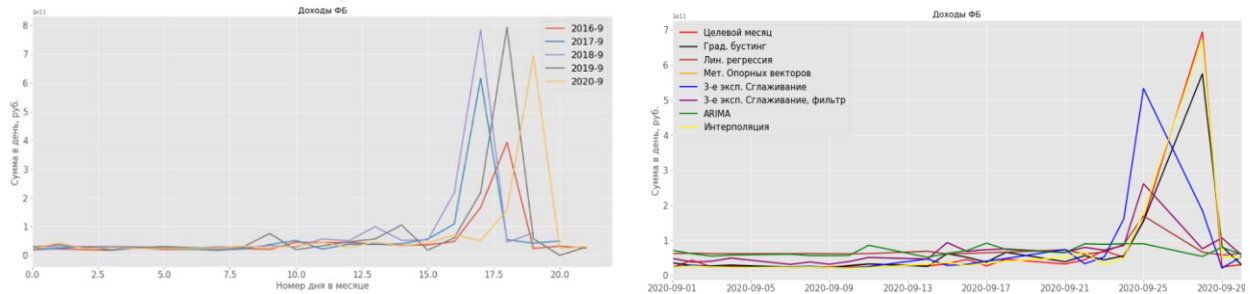


Рис. 1. Доходы, федеральный уровень:

a) плановые значения с сентября 2016 по сентябрь 2020 гг.; б) графики прогнозных значений моделей

Fig. 1. Revenue, federal level:

a) target values since September 2016 till September 2020; b) predictive value curves for models

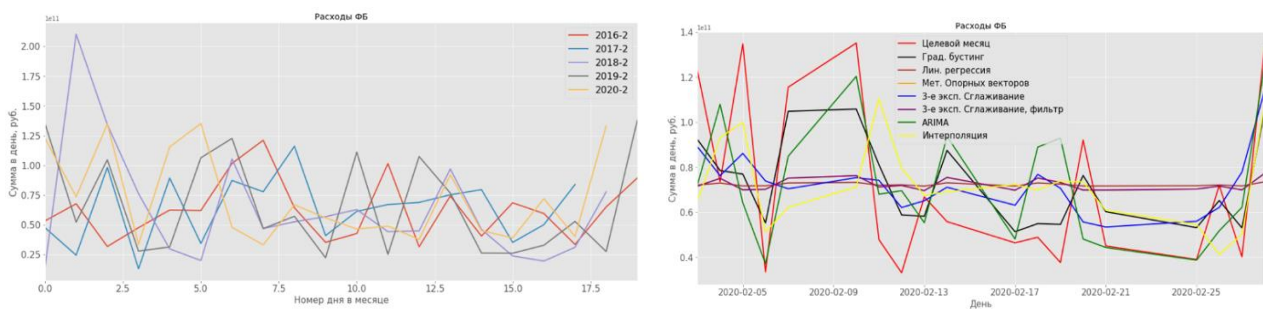


Рис. 2. Расходы, федеральный уровень:

a) плановые значения с февраля 2016 по февраль 2020; б) графики прогнозных значений моделей

Fig. 2. Expenses, federal level:

a) target values since February 2016 till February 2020; b) predictive value curves of models

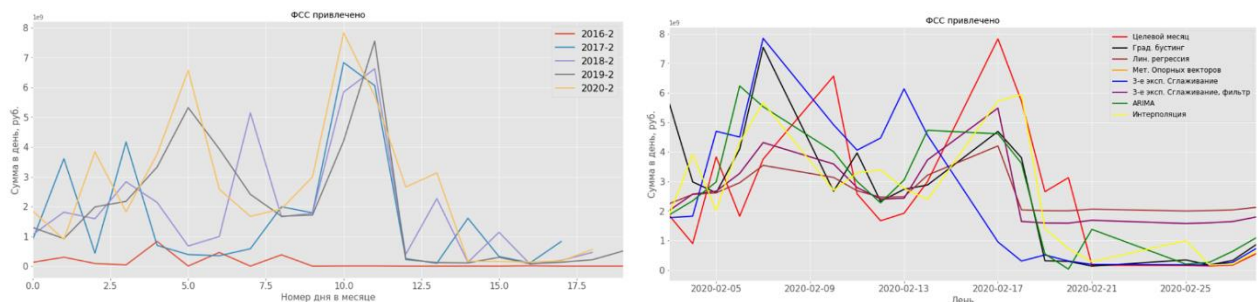


Рис. 3. ФСС привлечено:

a) плановые значения за февраль 2016-2020; б) графики прогнозных значений моделей

Fig. 3. Raised by Social Insurance Fund:

a) target values for February 2016-2020; b) predictive value curves of models

Как видно из графиков, от месяца к месяцу наблюдаются хорошо выраженная цикличность как результат регулярно повторяющихся процессов «выплат» и «сборов». Однако внутри конкретного месяца наблюдаются сдвиги пиковых значений на 1-2 дня в разные го-

ды. Это связано либо с изменениями производственного календаря, что можно относительно просто учесть в моделях, либо с решением клиентов «о назначении днем выплат» именно этого конкретного дня месяца в этом году, что учесть без участия эксперта невозможно.

Для оценки точности моделей использовалась метрика MAE – средняя абсолютная ошибка (т.е. на сколько в среднем ошибается прогнозная модель). При этом ошибка измеряется в единицах, являющихся долей/весом значения суток от общей суммы месяца, выраженная в процентах (доля/вес, умноженная на 100 %). Результаты приведены в табл. 1, из которой видно, что выделить приоритетный метод сложно. Одни ряды лучше обрабатываются с помощью ARIMA, другие – с помощью градиентного бустинга. В совокупности они все дают в среднем одну величину интервала на множестве рядов.

Таблица 1.
Сводная таблица мера ошибки MAE, %

Table 1.
Summary table, measure of error (MAE, %)

Показатель	Сред. взвешенное	Эксп. сглаживание	ARIMA	Линейная регрессия	Градиент. бустинг	Опорные вектора
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
Доходы, федеральный уровень	2,11	3,60	1,99	4,16	1,91	3,19
Нефтегазовые доходы, федеральный уровень	2,27	4,01	5,69	4,80	2,16	3,83
Расходы, федеральный уровень	2,49	2,42	1,87	2,64	2,28	2,64
Поступления по источникам	7,59	6,48	8,32	5,43	4,01	4,07
Выплаты по источникам	6,37	5,69	6,01	6,57	6,53	6,64
Вр. распоряжение	2,66	2,37	1,40	2,27	2,20	2,31
Вр. распоряжение	1,94	1,90	1,01	1,92	1,76	1,92
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
ФСС, привлечено	3,78	3,55	2,92	3,91	3,17	3,78
ФСС, возвращено	3,32	2,48	1,68	2,70	2,47	2,41
ФФОМС, привлечено	4,12	3,63	3,72	4,61	3,32	3,72
ФФОМС, возвращено	4,41	2,48	3,90	4,97	4,32	4,81
ПФР, привлечено	5,31	4,35	5,16	5,59	4,88	5,28
ПФР, возвращено	5,08	4,41	4,22	4,79	4,77	4,77
Поступления, муниципальный уровень	0,77	0,84	0,84	1,02	0,75	1,07
Выбытия, муниципальный уровень	0,68	0,83	0,71	0,76	0,56	0,90
Средняя итоговая MAE	3,53	3,27	3,30	3,74	3,01	3,42

4. Построение прогнозной модели изменения остатков

Были взяты архивные данные за период 2015-2020 гг. Для анализа были отобраны значимые ряды (например, доходы/расходы фондов ФСС, ПФР, ФФОМС) и агрегированные временные ряды нескольких незначимых (например, поступления муниципального уровня).

Для решения задачи экстраполяции с горизонтом прогнозирования на год вперед с подневной дискретизацией было изучено и применено 4 модели:

- 1) среднее взвешенное значение;
- 2) тройное экспоненциальное сглаживание (Хольта-Уинтерса);
- 3) ARIMA (Бокса-Дженкинса);
- 4) линейная регрессия.

Для оценки точности моделей использовалась метрика MAPE – средняя абсолютная процентная ошибка (т.е. на сколько процентов в среднем ошибается прогнозная модель). Данные по фактическим остаткам на счетах имеют своей особенностью существенные структурные сдвиги (рис. 4), они обусловлены многими причинами, например, решениями клиента о разделении остатков между несколькими своими счетами или передачей в «резерв» какой-то величины остатка. В СППР реализована [11] возможность учета мнения экспертов по выбору основной модели.



Рис. 4. Структурные сдвиги остатков ФСС

Fig. 4. Structural changes of Social Insurance Fund balance

Сводные результаты приведены в табл. 2. Видно, что сложно выделить приоритетную и наиболее достоверную модель. Хотя некоторые ряды лучше обрабатываются с помощью ARIMA, а некоторые – с помощью экспоненциального сглаживания, однако все модели в совокупности дают в среднем одну величину ошибки на множестве рядов.

5. Построение робастной интервальной модели

Построим робастную интервальную модель, исходя из введенных в п. 2 понятий. На вход в структурно-сложную систему подадим максимально агрегированные временные ряды, в качестве моделей машинного обучения возьмем простые скользящие средние, показавшие наихудшие результаты в исследованиях отдельных рядов, прогнозировать будем ежемесячные данные. Имея данные за 2015-2020 гг., эмулируем поведение системы в 2020 г. Последовательно добавляя в систему данные очередного месяца, будем экстраполировать увеличивающиеся ряды с помощью скользящих средних и агрегировать в итоговый прогноз с интервалом в 99 %. На выходе (рис. 5) получим соответственно 11 пар верхних и нижних интервалов в виде набора отдельных рядов с размерностью от 11 до 1 точки.

Таблица 2.
Сводная таблица ошибки MAPE, %

Table 2.
Summary table, mean absolute percentage error (MAPE, %)

Показатель	Средневзв.	Эксп. сглаживание	ARIMA	Лин. регр.
<i>1</i>	2	3	4	5
Вр. распоряжение БУ АУ юр. лица, привлечено	26,3	16,1	25	23
Вр. распоряжение БУ АУ юр. лица, возвращено	20,1	18	14,6	10
ФСС, привлечено	41,4	20	19,8	29,8
ФСС, возвращено	38,9	30,5	27,6	30,9
ФФОМС, привлечено	22,7	18	18,5	16
ФФОМС, возвращено	26,9	20,1	15,7	23,3
ПФР, привлечено	19,7	6,8	4,19	14,5
ПФР, возвращено	17,5	9	11,2	15,1
Поступления, муниципальный уровень,	23,7	16,4	15,2	10,4
Выбытия, муниципальный уровень	13,7	5,1	5,3	15,8
Средняя итоговая MAPE	25,09	16	15,709	18,88

Объединяя соседние первые точки верхних и нижних рядов доверительных интервалов, получим набор интервалов (коридор), в котором должны находиться фактические значения итогового прогноза (рис. 6). Несмотря на всю грубость модели, фактические данные остатка на ЕКС в ходе эксперимента вышли бы за нижнюю границу интервала лишь однажды в августе 2020 г.



Рис. 5. Построение робастной интервальной модели на 2020 год

Fig. 5. Construction of a robust interval model for 2020

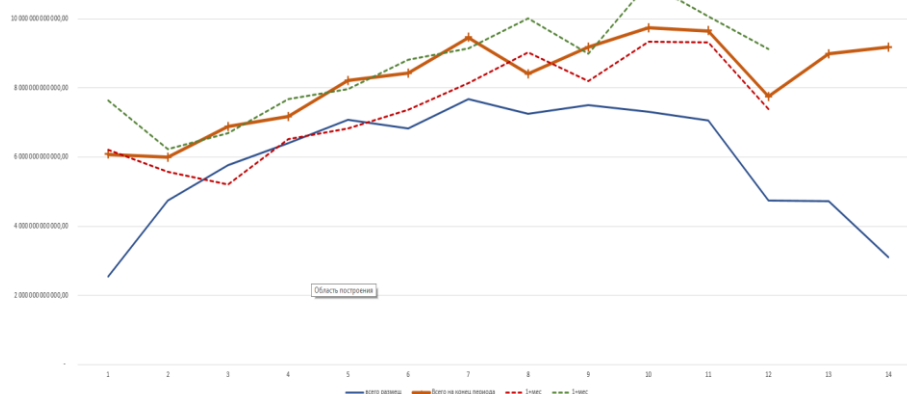


Рис. 6. Сравнение робастной интервальной модели с фактическими данными

Fig. 6. Comparing the robust interval model with actual data

6. Обсуждение результатов и выводы

Устойчивость итогового прогноза обеспечивается большим количеством входящих временных рядов. Важным параметром является количество исходных рядов. Проведем построение итогового прогноза на основе ограниченного количества рядов, при этом абсолютные значения их суммы масштабируем до величины итогового интервального прогноза, чтобы получить сопоставимые по абсолютной величине кривые (рис. 7).

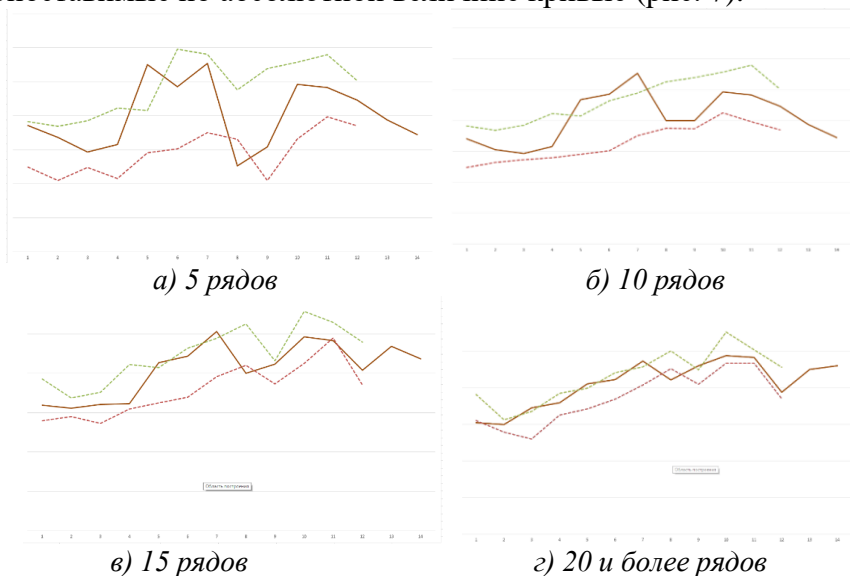


Рис. 7. Зависимость итогового интервального прогноза от количества исследуемых временных рядов

Fig. 7. Summary interval prognosis dependence from number of studied time series

Как видно на рис. 7, при росте количества входящих временных рядов, на совокупности которых строится итоговый прогноз, уменьшается волатильность как самих итоговых данных, так и интервала прогнозных значений. После определенного порога – в модельном примере это 20 рядов – получаем точные прогнозы, даже несмотря на все допущения, на основе использования простых скользящих средних. Из введенных определений и построенных моделей видно, что предложенный подход эффективно применим для диссипативных систем (с выраженным устойчивым или неустойчивым) трендом. Важным условием приме-

нимости является отсутствие точек бифуркации. В контексте рассматриваемой системы такие переходы могут быть вызваны кризисами, однако проведенные ранее исследования [15] показали, что в финансовых инструментах кризис переносит притягивающую область (аттрактор) в иную область пространства состояний, сохраняя топологическую эквивалентность фазовых портретов.

Таким образом, применение робастных интервальных моделей для структурно-сложных систем с большим количеством входных данных позволяет адекватно моделировать исследуемые процессы на макроуровне, обеспечивая достоверные интервальные оценки прогнозирования временных рядов.

Библиографический список

1. **Никульчев, Е.В.** Геометрический подход к моделированию нелинейных систем по экспериментальным данным [Текст] / Е.В. Никульчев. – М. : МГУП, 2007. – 162 с.
2. **Zeng, Z.** Bayesian median autoregression for robust time series forecasting / Z. Zeng, M. Li // *International Journal of Forecasting*. 2021. V. 37. № 2. P. 1000-1010.
3. **Jeon, Y.** Robust recurrent network model for intermittent time-series forecasting / Y. Jeon, S. Seong // *International Journal of Forecasting*. 2022. V. 38. № 4. P. 1415-1425.
4. **Liu, Z.** A robust time series prediction method based on empirical mode decomposition and high-order fuzzy cognitive maps / Z. Liu, J. Liu // *Knowledge-Based Systems*. 2020. V. 203. P. 106105.
5. **Hewamalage, H.** Recurrent neural networks for time series forecasting: Current status and future directions / H. Hewamalage, C. Bergmeir, K. Bandara // *International Journal of Forecasting*. 2021. V. 37. № 1. P. 388-427.
6. **Оморов, Р.О.** Алгебраический метод исследования робастности интервальных динамических систем [Текст] // *Научно-технический вестник информационных технологий, механики и оптики*. 2020. Т. 20. № 3. С. 364-370.
7. **Горяшко, А.П.** Робастная оптимизация, как новый подход к решению задач с неопределенностью [Текст] / А.П. Горяшко, Е.В. Никульчев // *Известия высших учебных заведений. Проблемы полиграфии и издательского дела*. 2011. № 1. С. 077-086.
8. **Петрушин, В.Н.** Эмпирическая оценка интервальных решений [Текст] / В.Н. Петрушин, Е.В. Никульчев, И.А. Чертихина // *Известия высших учебных заведений. Проблемы полиграфии и издательского дела*. 2011. № 6. С. 051-063.
9. **Музыкин, С.Н.** Системный анализ [Текст] / С.Н. Музыкин, Ю.М. Родионова. – М.: МГУПИ, 2003. – 199 с.
10. **Касты, Дж.** Большие системы: сложность, связность, катастрофы [Текст] / Дж. Касты. – М.: МИР, 1982. – 216 с.
11. **Албычев, А.С.** Структура системы поддержки принятия решений при таргетировании остатка денежных средств на едином казначейском счете Федерального казначейства [Текст] / А.С. Албычев, А.А. Червяков, Е.В. Никульчев // *International Journal of Open Information Technologies*. 2022. Т. 10. № 12. С. 116-122
12. **Costa, A.S.** A robust hierarchical nominal multicriteria classification method based on similarity and dissimilarity [Текст] / A. S. Costa, S. Corrente, S. Greco, J. R. Figueira, J. Borbinha // *European Journal of Operational Research*. 2020. V. 286. №3. P. 986-1001.
13. **Kandakoglu, M.** A robust multicriteria clustering methodology for portfolio decision analysis / M. Kandakoglu, G. Walther, S. B. Amor // *Computers & Industrial Engineering*. 2022. V. 174. P. 108803.
14. **Lourenco, J.C.** PROBE – a multicriteria decision support system for portfolio robustness evaluation / J. C Lourenco, A. Morton, C. A. Costa // *Decision support systems*. 2012. V. 54. № 1. P. 534-550.
15. **Никульчев, Е.В.** Модели хаоса для процессов изменения курса акций [Текст] / Е.В. Никульчев, М.В. Волович // *Exponenta Pro: Математика в приложениях*. 2003. № 1. С. 49-52.

*Дата поступления
в редакцию: 20.12.2022*

*Дата принятия
к публикации: 02.02.2023*