

**ПОВЫШЕНИЯ КАЧЕСТВА ПРОГНОЗА
ПОСРЕДСТВОМ МЕТОДОВ ОБЪЕДИНЕНИЯ ПРОГНОЗОВ**

Русанов Михаил Александрович

*руководитель центра информационно-аналитических систем,
Югорский научно-исследовательский институт информационных технологий
Ханты-Мансийск, Россия
E-mail: rusanovma@uriit.ru*

Шергин Сергей Николаевич

*кандидат физико-математических наук, доцент
ФГБОУ ВО «Югорский государственный университет»
Ханты-Мансийск, Россия
E-mail: ssn@ugrasu.ru*

Татьянкин Виталий Михайлович

*кандидат технических наук, доцент
БУ ВО «Сургутский государственный университет»
Ханты-Мансийск, Россия
E-mail: bambar@bk.ru*

Предмет исследования: статья посвящена сравнению методов прогнозирования и методов объединения прогнозов при прогнозировании среднемесячной заработной платы для некоторых регионов Российской Федерации за период с января 2013 года по июль 2022 года.

Цель исследования: проверить гипотезу о повышении качества прогноза посредством методов объединения прогнозов.

Методы и объекты исследования: для сравнения методов прогнозирования и методов объединения прогнозов используются временные ряды среднемесячной заработной платы в Ханты-Мансийском автономном округе – Югре, Ямало-Ненецком автономном округе и Свердловской области. В работе рассматриваются методы прогнозирования: TBATS, ARI-MA, экспоненциальное сглаживание, ETS, Theta, STL, полиномиальная регрессия и подход в объединении прогнозов методом Грейнджера-Раманатхана.

Основные результаты исследования: в работе представлены результаты сравнения методов прогнозирования и подхода в объединении прогнозов методом Грейнджера-Раманатхана. Временные ряды среднемесячной заработной платы были взяты из статистических сборников Росстата. Горизонт прогноза задавался на 12, 18 и 24 точки. Метод Грейнджера-Раманатхана показал, что в большинстве случаев получается повысить качество прогноза методом объединения частных прогнозов.

Ключевые слова: прогнозирование, объединение прогнозов, метод Грейнджера-Раманатхана.

**IMPROVING THE QUALITY OF THE FORECAST
THROUGH METHODS OF COMBINING FORECASTS**

Mikhail A. Rusanov

*Head of the center for information and analytical systems,
Yugorsky research institute of information technologies
Khanty-Mansiysk, Russia
E-mail: rusanovma@uriit.ru*

Sergey N. Shergin

*Candidate of Physics and Mathematics,
Associate Professor of the Department of Information Technology,
Yugra State University
Khanty-Mansiysk, Russia
E-mail: ssn@ugrasu.ru*

Vitaliy M. Tatjankin

*Candidate of Technical Sciences,
Associate Professor of the Department of Information Technology,
Surgut State University
Khanty-Mansiysk, Russia
E-mail: bambar@bk.ru*

Subject of research: The article is devoted to the comparison of forecasting methods and methods of combining forecasts when forecasting average monthly wages for some regions of the Russian Federation for the period from January 2013 to July 2022.

Purpose of the study: To test the hypothesis of improving forecast quality through forecast pooling methods.

Methods and objects of research: Time series of average monthly wages in the Khanty-Mansi Autonomous Okrug – Yugra, Yamal-Nenets Autonomous Okrug and Sverdlovsk region are used to compare forecasting methods and methods of combining forecasts. The paper considers forecasting methods: TBATS, ARIMA, exponential smoothing, ETS, Theta, STL, polynomial regression and the approach in combining forecasts by the Granger-Ramanathan method.

Main results of research: The paper presents the results of comparison of forecasting methods and approach in combining forecasts by Granger-Ramanathan method. Time series of average monthly wages were taken from statistical collections of Rosstat. The forecast horizon was set at 12, 18 and 24 points. The Granger-Ramanathan method showed that in most cases it is possible to improve the quality of the forecast by combining private forecasts.

Keywords: forecasting, forecast pooling, Granger-Ramanathan method.

Введение

При проведении исследования был подготовлен обзор литературы по вопросам улучшения качества прогнозов социально-экономических показателей за счет использования методов объединения прогнозов, которые, как показывает практика, демонстрируют лучшие результаты по сравнению с частными моделями прогнозирования.

Вначале были проанализированы статьи, содержащие подходы поиска косвенных показателей, использующихся при прогнозировании социально-экономических показателей. Например, в статье [1] говорится об использовании интеллектуального анализа текста для анализа настроений в новостных онлайн-порталах для предсказания тенденции цен на акции. В статье [2] используется подход применения графа знаний для выявления наиболее значимых связанных показателей с исследуемым показателем. Подобные исследования [1, 2] требуют проведения дополнительных работ по сбору большого объема данных, предобработке и разметке данных, что в разы увеличивает объем исследований.

В статье [3] представлена ретроспектива темы объединенных прогнозов за 50 лет и охватывает все основные направления исследования в этой области.

В данной работе представлен обзор литературы по объединению прогнозов, а также ссылки на доступные реализации программного обеспечения с открытым исходным кодом. Обсуждаются вопросы потенциала и ограничений различных методов объединения прогно-

зов. Также во введении статьи отмечается рост статей WoS до 13,8%, относящихся к теме объединения прогнозов среди всех статей, относящихся к прогнозированию.

Отдельно авторы выделяют вклад соревнования M-Competition [4, 5] в развитие темы прогнозирования временных рядов.

В статье [6] приводится обзор методов прогнозирования, который рассматривает качество различных методов для прогнозирования экономических показателей, а также разные оценки для валидации качества прогноза.

Среди российских исследователей можно выделить группу исследователей Института экономики РАН [7]. В основе работ лежит подход в объединении прогнозов методами Грейнджера-Раманатхана, метод матрицы парных предпочтений и метод линейной комбинации частных прогнозов с различными весами (веса определяются по матрице коэффициентов парной корреляции).

Для построения частных прогнозов использовались методы: гармонических весов (MGV), адаптивного экспоненциального сглаживания с использованием трэкинг-сигнала (MAEKS), экспоненциального сглаживания (MEKS) и модель Бокса-Дженкинса (ARIMA).

Оценка и сравнение методов проводилось по двум оценкам: средняя абсолютная ошибка и средняя относительная ошибка. Лучший результат показали методы Грейнджера-Раманатхана. Авторы статьи отмечают, что методы хорошо описывают не только плавные изменения, но и кризисные годы с резкими перепадами. Также у авторов статьи [7] можно отметить еще ряд статей посвященных сравнению методов прогнозирования [8, 9, 10].

Результаты и обсуждение

Постановка задачи

Целью статьи является проверка гипотезы о повышении качества прогноза посредством методов объединения прогнозов.

Для проверки гипотезы были решены задачи:

- 1) Подготовлены временные ряды для показателя «Среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по субъектам Российской Федерации с 2013 года (по месяцам), рублей» для ХМАО-Югры, ЯНАО и Свердловской области.
- 2) Рассчитаны прогнозы частными моделями прогнозирования и методом объединения прогнозов Грейнджера-Раманатхана.
- 3) Проведен сравнительный анализ качества прогноза частными моделями прогнозирования и методом объединения прогнозов Грейнджера-Раманатхана.

Решение

Для проверки гипотезы были выбраны частные модели прогнозирования: TBATS, ARIMA, exponential smoothing (экспоненциальное сглаживание), ETS, Theta, STL, полиномиальная регрессия (Polinom) и метод объединения прогнозов Грейнджера-Раманатхана.

Для проведения эксперимента использовались программные реализации методов прогнозирования из библиотеки skitime языка программирования python. Метод объединения прогнозов был написан на языке программирования python по описанию, представленному в статье [7].

Модель для временных рядов с множественной сезонностью TBATS рассматривалась в работах [11, 12]. TBATS это аббревиатура из:

- Trigonometric seasonality;
- Box-Cox transformation;
- ARMA errors;
- Trend;
- Seasonal components.

Модель была разработана для прогнозирования временных рядов с несколькими сезонными периодами. Например, ежедневные данные могут иметь как недельный, так и годовой компонент сезонности. Также и почасовые данные могут иметь несколько периодов: дневной, недельный или годовой. В рассматриваемом методе преобразование по методу Вох-Сох применяется к исходному временному ряду, а затем он моделируется как линейная комбинация экспоненциально сглаженного тренда, сезонного компонента и компонента ARMA. Сезонные компоненты моделируются тригонометрическими функциями через ряды Фурье.

Модель ARIMA представляет собой обобщение модели авторегрессионного скользящего среднего и адаптируется к данным временных рядов для прогнозирования будущих точек. Модели ARIMA могут быть особенно эффективными в тех случаях, когда данные свидетельствуют о нестационарности.

Модель экспоненциального сглаживания (модель Хольта-Винтерса) является одним из методов прогнозирования с использованием экспоненциального сглаживания. Рассматриваемая модель применялась в работе [13]. Сглаживание состоит в создании взвешенного скользящего среднего, вес которого определяется по схеме: чем старше информация об изучаемом явлении, тем меньше значение для текущего прогноза.

Модель экспоненциального сглаживания ETS (M, A, N) реализована так, что в модели учитывается мультипликативная ошибка («M» на первой позиции), аддитивный тренд («A» на второй позиции) и сезонность («N» на третьей позиции) [13].

Theta (theta-метод) основан на концепции преобразования временного ряда с помощью некоторого коэффициента [14, 15]. Полученный ряд сохраняет среднее значение и тенденцию исходных данных, а не их отклонения. Такие ряды называются theta-рядами. Их основной качественной характеристикой является усовершенствование адекватности долгосрочной тенденции данных или увеличение краткосрочных характеристик, в зависимости от значения коэффициента. При применении данной модели исходный временной ряд представляется в виде двух и более theta-рядов. Прогноз делается отдельно для каждого ряда, а итоговый прогноз – это усредненное значение полученных значений.

Модель декомпозиции с сезонностью и трендом (STL) раскладывает временной ряд на сезонный компонент, объединенный тренд, циклический компонент и компонент «погрешности» [16].

Полиномиальная регрессия – это частный случай линейной регрессии, когда подбирается полиномиальное уравнение для данных с криволинейной зависимостью между целевой переменной и независимыми переменными [17].

Для объединения прогнозов использовался метод Грейнджера-Раманатхана с 3, 4 и 5 моделями для объединения. Для каждого объединения прогнозов в сравнении результатов прогнозирования указано свое название: Grandger-3, Grandger-4 и Grandger-5.

Для проведения эксперимента использовались временные ряды «Среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по субъектам Российской Федерации с 2013 года (по месяцам), рублей» для регионов РФ: ХМАО-Югра, ЯНАО, Свердловская область.

Результаты оценки качества прогноза частными методами прогнозирования и методом объединения прогнозов приведены в таблицах 1-3.

Таблица 1

Результат прогнозирования показателя по данным для ЯНАО

| Горизонт прогноза | 12 месяцев | | | 18 месяцев | | | 24 месяца | | |
|-------------------|------------|-------|-------|------------|------|------|-----------|------|------|
| Модель | Оценки | | | Оценки | | | Оценки | | |
| | SMAPE | MAE | MdAE | SMAPE | MAE | MdAE | SMAPE | MAE | MdAE |
| ExpSmooth | 0.041 | 5459 | 3087 | 0.036 | 4516 | 2957 | 0.031 | 3973 | 2856 |
| ETS | 0.064 | 8202 | 6553 | 0.042 | 5304 | 3776 | 0.043 | 5485 | 4520 |
| Theta | 0.083 | 10929 | 10894 | 0.070 | 9014 | 7476 | 0.068 | 8775 | 8888 |
| ARIMA | 0.086 | 10987 | 9240 | 0.045 | 5729 | 4512 | 0.050 | 6373 | 5329 |

| | | | | | | | | | |
|---------------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|
| Polinom | 0.082 | 10808 | 8922 | 0.087 | 11211 | 9000 | 0.095 | 11438 | 9873 |
| STL | 0.057 | 7388 | 5783 | 0.041 | 5146 | 4175 | 0.038 | 4832 | 4093 |
| TBATS | 0.049 | 6542 | 7410 | 0.062 | 7461 | 4008 | 0.049 | 5930 | 3662 |
| Grandger – 3 | 0.028 | 3601 | 2072 | 0.038 | 4873 | 3439 | 0.033 | 4140 | 3211 |
| Grandger – 4 | 0.032 | 4093 | 3749 | 0.036 | 4491 | 4090 | 0.031 | 4004 | 2761 |
| Grandger – 5 | 0.030 | 3943 | 2733 | 0.035 | 4394 | 2454 | 0.030 | 3861 | 2360 |

Таблица 2

Результат прогнозирования показателя по данным для ХМАО-Югры

| Горизонт прогноза | 12 месяцев | | | 18 месяцев | | | 24 месяца | | |
|---------------------|--------------|-------------|-------------|--------------|-------------|-------------|--------------|-------------|-------------|
| Модель | Оценки | | | Оценки | | | Оценки | | |
| | SMAPE | MAE | MdAE | SMAPE | MAE | MdAE | SMAPE | MAE | MdAE |
| ExpSmooth | 0,049 | 4939 | 4380 | 0,061 | 6030 | 3564 | 0,056 | 5418 | 3095 |
| ETS | 0,059 | 5701 | 3196 | 0,068 | 6631 | 4195 | 0,061 | 5261 | 5254 |
| Theta | 0,073 | 7194 | 5988 | 0,078 | 7578 | 6066 | 0,082 | 7805 | 4994 |
| ARIMA | 0,048 | 4711 | 2313 | 0,077 | 7426 | 4304 | 0,068 | 6588 | 3947 |
| Polinom | 0,109 | 10747 | 9137 | 0,095 | 9271 | 7674 | 0,098 | 8935 | 8122 |
| STL | 0,053 | 5262 | 4136 | 0,068 | 6643 | 4429 | 0,066 | 6465 | 3647 |
| TBATS | 0,042 | 4115 | 3500 | 0,051 | 4999 | 2865 | 0,057 | 5569 | 2631 |
| Grandger – 3 | 0,038 | 3738 | 2450 | 0,049 | 4934 | 3231 | 0,053 | 5134 | 2986 |
| Grandger – 4 | 0,031 | 3022 | 2031 | 0,045 | 4435 | 2771 | 0,043 | 4069 | 3015 |
| Grandger – 5 | 0,038 | 3809 | 2323 | 0,054 | 5320 | 3360 | 0,047 | 4295 | 3993 |

Таблица 3

Результат прогнозирования показателя по данным для Свердловской области

| Горизонт прогноза | 12 месяцев | | | 18 месяцев | | | 24 месяца | | |
|---------------------|--------------|------------|------------|--------------|------------|------------|--------------|-------------|-------------|
| Модель | Оценки | | | Оценки | | | Оценки | | |
| | SMAPE | MAE | MdAE | SMAPE | MAE | MdAE | SMAPE | MAE | MdAE |
| ExpSmooth | 0.017 | 941 | 602 | 0.052 | 2774 | 2435 | 0.054 | 2821 | 2812 |
| ETS | 0.046 | 2505 | 2288 | 0.073 | 3853 | 3545 | 0.073 | 3843 | 3569 |
| Theta | 0.051 | 2816 | 2731 | 0.080 | 4261 | 3958 | 0.091 | 4772 | 4295 |
| ARIMA | 0.021 | 1136 | 875 | 0.065 | 3465 | 3546 | 0.068 | 3595 | 3107 |
| Polinom | 0.047 | 2642 | 1835 | 0.048 | 2647 | 1762 | 0.053 | 2815 | 1972 |
| STL | 0.046 | 2520 | 2703 | 0.091 | 4808 | 4317 | 0.080 | 4229 | 4148 |
| TBATS | 0.022 | 1222 | 770 | 0.055 | 2941 | 2315 | 0.072 | 3743 | 3658 |
| Grandger – 3 | 0.016 | 924 | 661 | 0.023 | 1173 | 912 | 0.035 | 1802 | 1550 |
| Grandger – 4 | 0.013 | 707 | 556 | 0.016 | 878 | 367 | 0.036 | 1813 | 1397 |
| Grandger – 5 | 0.020 | 1118 | 896 | 0.032 | 1776 | 932 | 0.025 | 1258 | 1086 |

Заключение и выводы

В статье приводится сравнение методов прогнозирования и метода объединения прогнозов Грейнджера-Раманатхана на примере прогнозирования показателя «Среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций по субъектам Российской Федерации с 2013 года (по месяцам), рублей» для регионов РФ: ХМАО-Югра, ЯНАО, Свердловская область. Результаты сравнения показывают, что метод объединения прогнозов имеет более высокое качество прогнозирования, чем частные модели прогнозирования.

В статье применяется метод Грейнджера-Раманатхана без ограничения параметров, что не позволяет описать влияние каждой модели, входящей в объединение, на итоговый результат. Но в дальнейшем планируется рассмотреть метод Грейнджера-Раманатхана с ограничениями и линейную комбинацию объединения прогнозов.

Литература

1. Xie, Y. Stock Market Forecasting Based on Text Mining Technology: A Support Vector Machine Method / Y. Xie, H. Jiang // *Journal of Computers*. – 2019. – V. 12, № 1. – P. 500–510. – DOI: 10.17706/jcp.12.6.500-510.
2. Tilly, S. Macroeconomic forecasting with statistically validated knowledge graphs / S. Tilly, G. Livan. – URL: <https://arxiv.org/pdf/2104.10457.pdf> (date of application: 08.12.2022).
3. Wang, X. Forecast combinations: an over 50-year review / X. Wang, B.J. Hyndman, F. Li, Y. Kang. – URL: <https://arxiv.org/pdf/2205.04216.pdf> (date of application: 08.12.2022).
4. Makridakis, S. The M4 Competition: 100,000 time series and 61 forecasting methods / S. Makridakis, E. Spiliotis, V. Assimakopoulos // *International Journal of Forecasting*. – 2020. – V. 36, № 1. – P. 54–74. – DOI: 10.1016/j.ijforecast.2019.04.014.
5. Макридакис Соревнования – Makridakis Competitions // Интернет-ресурс wiki. – URL: https://wikisru.ru/wiki/Makridakis_Competitions (дата обращения 08.12.2022). – Текст : электронный.
6. Makridakis, S. Statistical and Machine Learning forecasting methods: Concerns and ways forward / S. Makridakis, E. Spiliotis, V. Assimakopoulos // *PLOS ONE*. – 2018. – V. 27. – DOI: 10.1371/journal.pone.0194889.
7. Френкель, А. А. Сравнительный анализ методов построения объединенного прогноза / А. А. Френкель, Н. Н. Волкова, А. А. Сурков, Э. И. Романюк. – Текст : электронный // *Вопросы статистики*. – 2017. – URL : <https://voprstat.elpub.ru/jour/article/view/535/488> (дата обращения 08.12.2022).
8. Френкель, А. А. Использование методов гребневой регрессии при объединении прогнозов / А. А. Френкель, Н. Н. Волкова, А. А. Сурков, Э. И. Романюк. – Текст : непосредственный // *Финансы: теория и практика*. – 2018. – DOI: 10.26794/2587-5671-2018-22-4-6-17.
9. Сурков, А. А. Применение метода попарных сравнений при объединении экономических прогнозов / А. А. Сурков. – URL: <https://cyberleninka.ru/article/n/primenenie-metoda-poparnyh-sravneniy-pri-obedinenii-ekonomicheskikh-prognozov/viewer> (дата обращения 08.12.2022). – Текст : электронный.
10. Сурков, А. А. Объединение экономических прогнозов с использованием экспертной информации / А. А. Сурков. – URL: <https://cyberleninka.ru/article/n/obedinenie-ekonomicheskikh-prognozov-s-ispolzovaniem-ekspertnoy-informatsii/viewer> (дата обращения 08.12.2022). – Текст : электронный.
11. De Livera, A. M. Forecasting time series with complex seasonal patterns using exponential smoothing / A. M. De Livera, R. J. Hyndman, R. D. Snyder // *Journal of the American Statistical Association*. – 2012. – V. 106, № 496. – P. 1513–1527. – DOI: 10.1198/jasa.2011.tm09771
12. Hyndman, R. J. Forecasting: principles and practice / R.J. Hyndman, G. Athanasopoulos. – URL: <https://otexts.com/fpp3/> (date of application: 19.04.2020).
13. Иванов, С. А. Использование модели Хольта для прогнозирования изменения температурного режима в закрытом грунте / С. А. Иванов, И. Ю. Квятковская. – Текст : непосредственный // *Вестник СГТУ* – 2016. – Т. 82, № 1. – 4 с.
14. Assimakopoulos, V. The theta model: a decomposition approach to forecasting / V. Assimakopoulos, K. Nikolopoulos // *International Journal of Forecasting*. – 2000. – V. 16, №4. – P. 521-530. – DOI: 10.1016/S0169-2070(00)00066-2
15. Hyndman, R. J. Unmasking the Theta method / R. J. Hyndman, B. Billah // *International Journal of Forecasting*. – 2003. – V. 19, №2. – P. 287-290. – DOI: 10.1016/S0169-2070(01)00143-1
16. Cleveland, R. B. STL: A Seasonal-Trend Decomposition Procedure Based on LOESS / R. B. Cleveland, W. S. Cleveland, J.E. McRae, I. Terpenning // *Journal of Official Statistics*. – 1990. – V. 6. – P. 3-73.
17. Proietti, T. Real time estimation in local polynomial regression, with application to trend-cycle analysis / T. Proietti, A. Luati // *The Annals of Applied Statistics*. – 2009. – V. 2, №4. – P. 1523-1553. – DOI: 10.1214/08-AOAS195.