МОДЕЛИ СЦЕНАРНОГО ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ КРИЗИСОВ НА ОСНОВЕ ГИБРИДНОГО ПОДХОДА¹

В. М. Савинова, С. А. Ярушев

Российский экономический университет имени Г. В. Плеханова,
Москва, Россия

В рамках данной статьи рассматривается предложенный авторами полупараметрический метод макроэкономического прогнозирования в периоды резких изменений в экономике. Подбор блоков данных производится на основе методов кластеризации, наиболее близких к текущим экономическим условиям. Основным методом кластеризации является метод ближайшего соседа. Временные ряды делятся на блоки, а затем к самому последнему блоку наблюдений осуществляется подбор наиболее близкого блока, что отражает идею согласования направленных движений рядов. В качестве базовой модели прогнозирования авторами используется модель ARIMA. Показаны преимущества подхода к прогнозированию во время великой рецессии – экономического спада 2008 г. для таких переменных, как инфляция, безработица и реальные личные доходы. Предложенный метод превосходит параметрические линейные, нелинейные, одномерные и многомерные альтернативные методы за период 2007–2019 гг. В статье приведены расчеты, полученные в результате компьютерного эксперимента с использованием языка Python для данных инфляции и цен на нефть за указанный период. Представленный подход в перспективе может использоваться в интеллектуальных методах машинного обучения, таких как нейронные сети.

Ключевые слова: гибридные модели, прогнозирование, временной ряд, ARIMA, кластерный анализ, метод ближайшего соседа.

MODELS OF SCENARIO FORECASTING OF ECONOMIC CRISES ON THE BASIS OF HYBRID APPROACH

Victoria M. Savinova, Sergei A. Yarushev

Plekhanov Russian University of Economics,
Moscow, Russia

The present article studies a semi-parametrical method of macro-economic forecast in periods of sharp changes in economy put forward by the authors. Data blocks are chosen on the basis of clusterization methods, which are as close to the current economic conditions as possible. The key method of clusterization is the method of the closest neighbor. Time series is split into blocks and then the closest block is chosen for the last block of observation, which demonstrates the idea of coordination of directive series movements. As a basic model of forecasting the authors use ARIMA model. The authors show advantages of this approach for forecasting during the great recession – the economic slump of 2008 for such variables as inflation rate, unemployment and real private income. This method demonstrates its superiority in comparison with parametrical linear, non-linear, single-dimension and multi-dimension alternative methods for the period 2007–2019. The article provides calculations obtained as a result of computer experiment using Python language for data on inflation rate and oil prices for the mentioned period. This approach in future can be used in intellectual methods of machine teaching, such as neuron networks.

Keywords: hybrid models, forecast, time series, ARIMA, cluster analysis, method of the closest neighbor.

¹ Статья подготовлена по результатам исследования, проведенного в рамках внутреннего гранта «Разработка методики прогнозирования цен на финансовые инструменты на базе нейронных сетей» ФГБОУ ВО «РЭУ им. Г. В. Плеханова».

Введение

Великая рецессия (экономический спад) стала отрезвляющим опытом во многих отношениях. Одним из таких уроков были трудности, с которыми сталкиваются эконометрические модели при прогнозировании показателей в условиях кризисных явлений.

Данная статья тесно связана с недавними исследованиями, которые подчеркивают растущие трудности в прогнозировании макроэкономических переменных и выступают за использование изменяющихся во времени параметров или стохастической волатильности, чтобы отразить динамику инфляции.

Великая рецессия усилила проблемы, с которыми сталкиваются профессионалы при прогнозировании изменения цен. В результате проведенного исследования открывается новый взгляд на загадку прогноза инфляции. Важно отметить, что информация с финансовых рынков (в частности, цены на жилье) имеет тенденцию улучшать прогнозы во время великой рецессии [1].

Предлагаемый подход к прогнозированию объединяет гибкость непараметрического метода ближайшего соседа (NN) с моделью авторегрессионного интегрированного скользящего среднего (ARIMA) [6]. Первый этап представленной модели – это деление интересующих временных рядов на блоки. Далее осуществляется поиск ближайшего блока к самому последнему блоку наблюдений. В рамках подхода возможно использование двух алгоритмов для кластеризации данных в блоки. Первый метод (сопоставление по уровням) позволяет сравнивать непосредственно два блока данных; второй метод (соответствие по отклонениям) показывает отклонения от локального среднего двух последовательностей ряда.

В рамках исследования была построена модель прогнозирования на основе скорректированного прогноза ARMA/ARIMA. Коррекция погрешности, рассчитанной с учетом соответствия блока данных (воз-

можно, неправильно заданной параметрической начальной оценки), близка к параметрически управляемой непараметрической регрессии. Результирующая схема показала высокую точность прогнозов во время резких изменений данных.

Оценка эффективности прогнозирования с использованием метода ближайшего соседа во время великой рецессии дала положительный прогностический результат, в частности, по таким переменным, как занятость по данным платежных ведомостей, промышленное производство и реальный личный доход.

Прогноз, построенный статистически (основываясь на тесте [3]), оказался более точным по сравнению с результатами, полученными из линейных моделей для 60% рядов в нашей выборке.

Сравнение проводилось между квартальными агрегатами наших ежемесячных прогнозов и оценками профессиональных прогнозистов и прогнозов GreenBook в условиях великой рецессии (при условии, что наши прогнозы являются высококонкурентными). Оценивались также долгосрочные прогнозы на срок от 3 до 12 месяцев вперед.

Алгоритм сопоставления метода ближайшего соседа

Кратко рассмотрим эконометрическую структуру метода ближайшего соседа.

Пусть $Y_T = \{y_1, y_2, \dots, y_t\}$ обозначает набор наблюдений. Мы предполагаем общую нелинейную авторегрессионную структуру $y_t = g(y_{t-1}, \dots, y_{t-k}) + \epsilon_t$, учитывая, что $E(\epsilon_t | Y_{t-1}) = 0$, где g – авторегрессионная функция.

Цель алгоритма ближайшего соседа – оценить функцию условного ожидания.

Для этой цели сформируем следующую непараметрическую оценку функции $g(y_{t-1}, \dots, y_{t-k})$:

$$\hat{g} = \frac{\sum I(\text{dist}(y_{j-k:j-1}, y_{t-k:t-1}) \leq \text{dist}_{m(T)} y_j)}{\sum I(\text{dist}(y_{j-k:j-1}, y_{t-k:t-1}) \leq \text{dist}_{m(T)})},$$

где $\hat{g}$ – функция условного ожидания;

I – индикаторная функция;

$dist$ – функция расстояния;

$y_{j-k:j-1}$ – диапазон наблюдений от элемента $j-k$ до элемента $j-1$;

$y_{t-k:t-1} = \{y_{t-k}, \dots, y_{t-1}\}$;

$dist_{m(T)}$ – пороговый уровень, при котором используются только совпадения расстояний меньшей размерности;

y_j – j -е наблюдение;

t – время;

k – размерность шага (блока наблюдения).

Возможность медленного роста числа ближайших соседей m по мере увеличения числа наблюдений T является решающим условием для последовательности оценки ближайшего соседа.

Суммирование производится по количеству совпадений. Это число в свою очередь зависит от размера выборки T , особенно при $\frac{m}{T} \rightarrow 0$, так как $m, T \rightarrow \infty$.

С операционной точки зрения наши цели состоят в следующем:

1) объединить данные в блоки C_j из k элементов, при этом $C_j = \{[y_{j1}, y_{j1+1}, \dots, y_{j1+k-1}], [y_{j2}, y_{j2+1}, \dots, y_{j2+k-1}], \dots, [y_{jm}, y_{jm+1}, \dots, y_{jm+k-1}]\}$; этот кластер содержит tm множеств/блоков наблюдений (размерностью k);

2) соотнести самые последние k наблюдений $[y_{T-k+1}, \dots, y_T]$ с ближайшим кластером;

3) составить прогноз наблюдения y_{T+1} .

Рассмотрим два основных класса алгоритмов сопоставления (функции расстояния): сопоставление по уровням и сопоставление по отклонению от локального среднего.

Сопоставление с уровнями имеет смысл, когда уровень ряда данных обладает важным экономическим или прогнозным значением либо когда ряд является явно стационарным. Сопоставление с отклонениями от локального среднего значения имеет больший смысл, когда исследователь хочет сопоставить направленное движение ряда в отличие от общего уровня.

Модель прогнозирования

Относительно базового алгоритма ближайшего соседа произведем корректиров-

ку текущего прогноза из базовой модели ARMA/ARIMA с учетом ошибок прогноза модели из предыдущих аналогичных периодов времени, где сходство определяется функцией расстояния.

Этот подход близок к параметрически управляемой непараметрической регрессионной модели прогнозирования [2], которая показала потенциальные преимущества для конечной выборки использования параметрической пробной оценки с целью уменьшения смещения относительно непараметрической регрессии без штрафов за отклонения.

Без ограничения общности предположим, что первая последовательность – это та, которая соответствует текущей последовательности, т. е. имеющая наибольшее сходство или наименьшее расстояние [4]. Для составления прогноза использовалась следующая формула:

$$\hat{y}_{t+1} = (y_{k+1} - \hat{y}_{k+1, ARMA}) + \hat{y}_{t+1, ARMA},$$

где $\hat{y}_{t+1, ARMA}$ – прогноз на один шаг от модели ARMA/ARIMA.

Модель ARMA/ARIMA была выбрана в качестве вспомогательной [10], поскольку данные являются ежемесячными с высокой изменчивостью.

Мы должны усреднить прогнозы по верхним соответствиям m , что и делается в эмпирическом разделе.

Метод ближайшего соседа имеет явно локальный характер, как подчеркивалось во многих публикациях [5].

Заметим, что информация, содержащаяся в данных, относящихся к верхнему m , полностью соответствует текущей последовательности и имеет полный вес на этапе корректировки, тогда как ряды, далекие от этой последовательности, не имеют веса. Эта локальная природа оценки верна для непараметрической оценки в целом.

Сопоставим локальное поведение оценки ближайшего соседа с глобальными методами, такими как линейная модель ARMA/ARIMA, которая использует все исторические данные с одинаковым весом. Указанный подход может быть неуместным в периоды кризисов, когда экономи-

ческие отношения, преобладающие в обычное время, имеют тенденцию разрушаться. Кроме того, корректировка предварительного прогноза по модели ARMA/ARIMA является важным первым шагом в процедуре прогнозирования.

Ежемесячные макроэкономические данные [7] обычно хорошо описываются динамикой модели ARMA/ARIMA, являющейся популярной моделью (в сокращенной форме) роста потребления и безработицы среди многих других макроэкономических рядов [11].

Важный компонент скользящего среднего, часто встречающийся при моделировании месячных данных, предполагает необходимость сначала удалить эту динамику. Общее решение заключается в выборе параметрического ориентира, кото-

рый близок к истинной непараметрической функции.

В качестве данных для прогнозирования были взяты временные ряды, характеризующие инфляцию в Российской Федерации и цены на нефть за период 2007–2019 гг. ежемесячно. В качестве инструмента моделирования использовался язык Python 3.7.

Результаты

В рамках исследования были проведены расчеты с использованием описанного метода. На первом этапе осуществлен кластерный анализ с использованием алгоритма ближайшего соседа. Затем была построена модель ARIMA [8].

Таким образом, был выполнен прогноз и проведено сопоставление (рис. 1).

Рис. 1. Прогноз стоимости нефти марки Brent

Отклонение прогноза составило 5% от фактического значения. Был построен также прогноз показателя инфляции в

Российской Федерации с использованием того же метода (рис. 2).

Рис. 2. Прогноз показателя инфляции в Российской Федерации

Выводы

В результате использования метода ARMA/ARIMA точность и качество существующих прогнозов не превосходят построенную в ходе экспериментов модель.

Гибридная модель прогнозирования объединила в себе алгоритмы кластеризации и ARIMA [9]. Полученный алгоритм доказал свою эффективность при резких изменениях временных рядов. Данная модель позволяет прогнозировать кризисные ситуации, т. е. сильные движения ряда – резкие рост и спад.

Предложенный нами подход более точен, чем линейные альтернативы из-за комбинации условий. Метод хорошо справляется с резкими изменениями данных, которые не имели место в предыдущих исследованиях, основанных на более стабильных выборках (конца 1980-х – начала 1990-х гг.). Еще одним преимуществом подхода является простота вычислений и гибкость.

Несмотря на то что в качестве базовых рассматриваются только канонические ARIMA-модели, теоретически может ис-

пользоваться любое количество линейных или нелинейных моделей. В базовой модели должны быть оценены только два представляющих интерес параметра: длина совпадения и параметр, управляющий количеством верхних совпадений, по которым выполняется усреднение. Эти параметры выбираются на основе предыдущих показателей прогнозирования вне выборки, используя критерий среднеквадратичной ошибки.

Использование финансовой информации не ухудшает предсказательную силу нашего подхода вне кризиса.

Предложенная методология может рассматриваться в качестве дополнительного инструмента для сбора доступных методов в целях прогнозирования макроэкономических переменных.

В перспективе модели на основе кластеризации могут применяться при нелинейных методах прогнозирования, таких как искусственные нейронные сети, деревья решений и пр.

Список литературы

1. Дубовик М. В., Зайцева Е. В., Литвишко О. В. Ценностный подход как методология оценки инновационного развития // Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика. – 2018. – Т. 45. – № 3. – С. 422–431.
2. Литвишко О. В. Роль рынка первичных размещений в модернизации российской экономики // Вестник Российского экономического университета имени Г. В. Плеханова. – 2011. – № 5 (41). – С. 52–57.
3. Титов В. А., Вейнберг Р. Р., Савинова В. М. Реализация мультиагентной модели «Стартап-инвестор – корпорация и государство» // Фундаментальные исследования. – 2017. – № 12-1. – С. 217–222.
4. Borodin A., Veynberg R., Litoishko O. Methods of Text Processing when Creating Chatbots // Humanitarian Balkan Research. – 2019. – Vol. 3 (5). – P. 108–111.
5. Del Negro M., Giannoni M., Schorfheide F. Inflation in the Great Recession and New Keynesian Models : Staff Reports 618. – New York : Federal Reserve Bank of New York, 2014.
6. Diebold F., Mariano R. Comparing Predictive Accuracy // Journal of Business and Economic Statistics. – 1995. – N 13. – P. 253–263.

7. Farmer J. D., Sidorowich J. J. Predicting Chaotic Time Series // *Physical Review Letters*. – 1987. – Vol. 59 (8). – P. 845–848.
8. Farmer R. The Stock Market Crash of 2008 Caused the Great Recession: Theory and Evidence // *Journal of Economic Dynamics and Control*. – 2012. – Vol. 36 (5). – P. 693–707.
9. Ferrara L., Guegan D., Rakotomalahy P. GDP Nowcasting with Ragged-Edge Data: A Semi-Parametric Modeling // *Journal of Forecasting*. – 2010. – N 29. – P. 186–199.
10. Guerrón-Quintana P., Zhong M. Macroeconomic Forecasting in Times of Crises // *Finance and Economics Discussion Series 2017–2018*. – Washington : Board of Governors of the Federal Reserve System, 2017. – URL: <https://doi.org/10.17016/FEDS.2017.018>
11. Kavousi-Fard A., Samet H., Marzbani F. A New Hybrid Modified Firefly Algorithm and Support Vector Regression Model for Accurate Short Term Load Forecasting // *Expert Systems with Applications*. – 2014. – N 41. – P. 6047–6056.

References

1. Dubovik M. V., Zaytseva E. V., Litvishko O. V. Tsennostnyy podkhod kak metodologiya otsenki innovatsionnogo razvitiya [Value Approach as Methodology of Estimating the Innovation Development]. *Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta. Seriya: Ekonomika. Informatika* [Academic Bulletin of the Belgorod State University. Series: Economics. Information Science], 2018, Vol. 45, No. 3, pp. 422–431. (In Russ.).
2. Litvishko O. V. Rol rynka pervichnykh razmeshcheniy v modernizatsii rossiyskoy ekonomiki [The Role of IPO in Modernizing Russian Economy]. *Vestnik Rossiyskogo ekonomicheskogo universiteta imeni G. V. Plekhanova* [Vestnik of the Plekhanov Russian University of Economics], 2011, No. 5 (41), pp. 52–57. (In Russ.).
3. Titov V. A., Veynberg R. R., Savinova V. M. Realizatsiya multiagentnoy modeli «Startup-investor – korporatsiya i gosudarstvo» [Realizing the Multi-Agent Model 'Start-Up-Investor-Corporation and State']. *Fundamentalnye issledovaniya* [Fundamental Research], 2017, No. 12-1, pp. 217–222. (In Russ.).
4. Borodin A., Veynberg R., Litvishko O. Methods of Text Processing when Creating Chatbots. *Humanitarian Balkan Research*, 2019, Vol. 3 (5), pp. 108–111.
5. Del Negro M., Giannoni M., Schorfheide F. Inflation in the Great Recession and New Keynesian Models : Staff Reports 618. New York, Federal Reserve Bank of New York, 2014.
6. Diebold F., Mariano R. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 1995, No. 13, pp. 253–263.
7. Farmer J. D., Sidorowich J. J. Predicting Chaotic Time Series. *Physical Review Letters*, 1987, Vol. 59 (8), pp. 845–848.
8. Farmer R. The Stock Market Crash of 2008 Caused the Great Recession: Theory and Evidence. *Journal of Economic Dynamics and Control*, 2012, Vol. 36 (5), pp. 693–707.
9. Ferrara L., Guegan D., Rakotomalahy P. GDP Nowcasting with Ragged-Edge Data: A Semi-Parametric Modeling. *Journal of Forecasting*, 2010, No. 29, pp. 186–199.
10. Guerrón-Quintana P., Zhong M. Macroeconomic Forecasting in Times of Crises. *Finance and Economics Discussion Series 2017–2018*. Washington, Board of Governors of the Federal Reserve System, 2017. Available at: <https://doi.org/10.17016/FEDS.2017.018>

11. Kavousi-Fard A., Samet H., Marzbani F. A New Hybrid Modified Firefly Algorithm and Support Vector Regression Model for Accurate Short Term Load Forecasting. *Expert Systems with Applications*, 2014, No. 41, pp. 6047–6056.

Сведения об авторах

Виктория Михайловна Савинова

старший преподаватель
кафедры информатики
РЭУ им. Г. В. Плеханова.
Адрес: ФГБОУ ВО «Российский
экономический университет
имени Г. В. Плеханова»,
117997, Москва, Стремянный пер., д. 36.
E-mail: Savinova.VM@rea.ru

Сергей Александрович Ярушев

старший преподаватель
кафедры информатики
РЭУ им. Г. В. Плеханова.
Адрес: ФГБОУ ВО «Российский
экономический университет
имени Г. В. Плеханова»,
117997, Москва, Стремянный пер., д. 36.
E-mail: YArushev.SA@rea.ru

Information about the authors

Victoria M. Savinova

Senior Lecturer
of the Department for Informatics
of the PRUE.
Address: Plekhanov Russian
University of Economics,
36 Stremyanny Lane, Moscow, 117997,
Russian Federation.
E-mail: Savinova.VM@rea.ru

Sergei A. Yarushev

Senior Lecturer
of the Department for Informatics
of the PRUE.
Address: Plekhanov Russian
University of Economics,
36 Stremyanny Lane, Moscow, 117997,
Russian Federation.
E-mail: YArushev.SA@rea.ru