УДК 519.86
DOI: http://dx.doi.org/10.21686/2500-3925-2017-3-4-9

Д.В. Домащенко, Э.Е. Никулин
Российский экономический университет имени Г.В. Плеханова, Москва, Россия

Прогнозирование рядов динамики рыночных индикаторов на основе нелинейной авторегрессионной нейронной сети

Современная практика экономических исследований активно полагается на математические модели, позволяющие выявлять в статистических данных скрытые закономерности и строить на их основании прогнозы. Линейные модели прогнозирования рядов динамики, основанные на векторной авторегрессии (VAR) являются наиболее распространенными. Однако связи между рядами динамики в экономике часто имеют сложно идентифицируемый характер, поэтому нелинейные авторегрессионные (NAR) модели показывают более достоверные результаты. Для их реализации обычно используются нейронные сети, которые не предоставляют возможности оценки прогноза в виде математического ожидания и стандартного отклонения. Поэтому предлагаемая в статье модель сочетает в себе два блока: VAR и NAR. NAR используется для построения прогноза на заданное количество точек, а VAR для оценки прогноза в виде математического ожидания и стандартного отклонения. Оценка достоверности модели проводилась на дневных данных валютного курса USD/RUB и цен на нефть марки «Брент» с 1.01.2016 по 1.03.2017. Средняя точность прогнозирования тренда для курса доллара США к рублю составила 54,9%, для цены нефти – 54,0%. При этом относительная ошибка прогнозирования курса доллара составила от 1,09% (для первой точки) до 2,01% (для десятой точки), относительная ошибка прогнозирования цен на нефть составила от 1,28% (для первой точки) до 4,58% (для десятой точки). Таким образом, модель представляет достаточно точные для принятия инвестиционных решений прогнозы, при этом производится оценка прогнозов на основании тестирования NAR блока на исторических данных и на основании прогноза VAR блока в форме математического ожидания и стандартного отклонения.

Ключевые слова: *искусственная нейронная сеть, векторная авторегрессия, прогнозирование, курс доллара, цены на нефть.*

Denis V. Domashchenko, Edvard E. Nikulin
Plekhanov Russian University of Economics, Moscow, Russia

Forecasting time series of the market indicators based on a nonlinear autoregressive neural network

The modern practice of economic research relies heavily on mathematical models that make it possible to reveal hidden regularities in statistical data and make forecasts on their basis. Linear models based on vector autoregression (VAR) are the most common. However, the relationship between time series in the economy is often difficult to identify, so non-linear autoregressive (NAR) models show more reliable results. Artificial neural networks (ANNs) are usually used for implementation of these models, but ANNs do not provide the possibility of estimating the forecast in the form of mathematical expectation and a standard deviation. Therefore, the model proposed in the article combines two blocks: VAR and NAR. NAR is used to construct a prediction for a given number of points, and VAR for estimating the forecast in the form of a mathematical expectation and a standard deviation. The evaluation of the model was carried out on the daily data: exchange rate USD / RUB and "Brent" oil from 1.01.2016 to 1.03.2017. The average accuracy of forecasting the trend for the dollar was 54.9%, for the oil prices it was 54.0%. In this case, the relative error in predicting the dollar rate was from 1.09% (for the first point) to 2.01% (for the tenth point); the relative error in forecasting oil prices was from 1.28% (for the first point) to 4.58 % (for the tenth point). Thus, the model showed accurate results when predicting dynamic series and can be used to solve other forecasting problems. In particular, it is expedient to use the model as one of the factors when making investment decisions. In addition, the evaluation of forecasts is done on the basis of testing the NAR block of historical data and on the basis of VAR block forecast in the form of mathematical expectation and standard deviation.

Keywords: *artificial neural network, vector autoregression, forecasting, exchange rate USD/RUB, oil price.*

Введение

Современная практика экономических исследований активно полагается на математические модели, позволяющие выявлять в статистических данных скрытые закономерности и строить на их основании прогнозы. На данный момент наиболее распространенными являются модели, основывающиеся на векторной авторегресии (VAR). Первая модель векторной авторегрессии была предложена К. Симсом в 1980 году [1]. Значительными преимуществами предложенной модели в сравнении с моделями, основанными на дифференциальных уравнениях, были простота идентификации, реализации и решение проблемы переоценки (overfitting) модели. С тех пор методология построения авторегрессионных моделей сильно продвинулась, в основном благодаря простоте и универсальности методов. Были разработаны модели VEC [2], ARIMA [3] и GARCH [4], снимающие ограничения на использование только стационарных временных рядов. Данные модели являются наиболее

распространенными решениями для прогнозирования рядов динамики, т.к. при простоте реализации позволяют получать достаточно точные результаты [5,6,7].

Все перечисленные модели являются линейными, что является значительным недостатком с учетом нелинейного характера связей между экономическими рядами динамики. Для решения данной проблемы используются модели, основанные на нелинейной авторегресии (NAR), предложенные в [8,9]. Для реализации NAR моделей применяются динамические искусственные нейронные сети, которые показывают лучшие результаты при работе с зашумленными нестационарными временными рядами [10,11]. Так, в работе [12] проведено сравнение модели нелинейной авторегрессионной нейронной сети с экзогенными входами (NARX) и ARIMA-модели. По результатам моделирования среднеквадратичная ошибка NARX-модели на порядок меньше ошибки ARIMA-модели (NARX – 0.0004, ARIMA – 0.0061). В статье [13], показано превосходство NARX модели над обобщенными моделями авторегрессионной условной гетероскедастичности (GARCH, EGARCH).

Однако главным недостатком NAR-модели является невозможность оценки прогноза классическими метриками математической статистики: математическим ожиданием и стандартным отклонением. Поэтому для решения задачи прогнозирования было принято решение использовать модель, включающую два блока:

– нелинейный авторегрессионный блок, основанный на нелинейной авторегрессионной нейронной сети (NAR);

– линейный авторегрессионный блок, основанный на векторной авторегресии (VAR).

NAR-блок используется для получения прогноза рядов динамики на заданное количество точек, благодаря нелинейному характеру данного блока. Совместное использование VAR и NAR блоков позволяет устранить недостатки обоих типов моделей. VAR служит для построения прогнозов в виде математического ожидания и стандартного отклонения, что выступает в качестве оценки прогноза, получаемого с помощью NAR блока, представляющего прогноз в виде точечных значений.

Таким образом, целью настоящей работы является разработка модели прогнозирования рядов динамики рыночных индикаторов и тестирование модели на исторических данных (back-test). Оценка результатов прогнозирования производится по средней абсолютной ошибке и по ошибке определения направления тренда.

1. Модель прогнозирования временных рядов

Блок векторной авторегрессии (VAR)

VAR описывается следующим уравнением:

$$y_t = a_0 + A_1 y_{t-1} + \\ + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \varepsilon_t = \\ = a_0 + \sum_{n=1}^{p} A_n y_{t-n} + \varepsilon_t$$

где p – количество лагов;
a_0 – вектор констант;
$A_1 \ldots A_p$ – матрицы параметров модели (коэффициенты авторегрессии);
y_t – вектор значений временных рядов;
y_{t-p} – вектор предыдущих значений временных рядов;
ε_t – вектор случайных ошибок.

VAR модель работает со стационарными рядами динамики, поэтому перед построением прогноза необходимо их приведение к стационарным. Для этого с помощью расширенного теста Дики-Фуллера определяется порядок интегрированности временного ряда и каждый ряд приводится к стационарному путем вычисления разностей.

С целью определения оптимального количества лагов модели используются информационные критерии Акайаке (AIC) [14] и Байеса (BIC) [15], при получении различных результатов по критериям выбирается минимальный лаг.

Для построения прогноза используется метод Монте-Карло, который генерирует 1000 точек для каждой точки прогноза. По полученным 1000 значениям рассчитывается математическое ожидание и стандартное отклонение, которые и являются прогнозом.

В качестве метрики для оценки ошибки модели используется среднеквадратичная ошибка (MSE).

Блок нелинейной авторегрессионной нейронной сети (NAR)

$$y_t = F(y_{t-1}, y_{t-2}, \ldots, y_{t-p}) + \varepsilon_t$$

где $F(\cdot)$ – нелинейная функция, аппроксимируемая нейронной сетью;
y_t – вектор значений временных рядов;
y_{t-p} – вектор предыдущих значений временных рядов;
ε_t – вектор случайных ошибок.

Архитектура NAR нейронной сети изображена на рис. 1.

Для обучения нейронной сети используются выборки данных, которые делятся в следующем соотношении:

Обучающая выборка – 70%,
Тестовая выборка – 15%,
Валидационная выборка – 15%.

Обучение нейронной сети производится с помощью алгоритма Левенберга-Маквардта, при этом для оценки производительности нейронной сети используется среднеквадратичная ошибка (MSE).

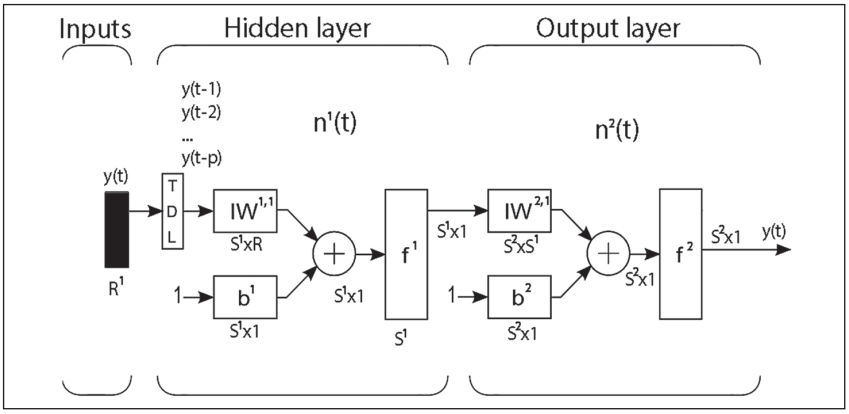

Рис. 1. Архитектура NAR

Тестирование NAR блока на исторических данных

Тестирование модели на исторических данных (back-test) проводится по следующему алгоритму:

1) Модель обучается на окне из *n* точек;

2) Строится прогноз на *k* точек;

3) Рассчитываются абсолютная ошибка и ошибка определения тренда для каждой из *k* точек полученного прогноза;

4) Окно для обучения модели сдвигается на одну точку вперед;

5) Операции 2–4 повторяются до тех пор, пока модель не достигнет последних точек временных рядов;

6) Рассчитываются средняя абсолютная ошибка и ошибка определения тренда.

Ошибка определения тренда рассчитывается следующим образом:

Рассчитывается тренд для каждой *k* точки прогноза и соответствующий ей фактический тренд, по формуле:

$$tr_k = sign(p_k - p_0) \qquad (1)$$

где tr_k – тренд;
p_k – значение *k* точки прогноза;
p_0 – значение последней из *n* точек, входящих в обучающее окно (или последнее фактическое значение).

Если значения спрогнозированного и фактического тренда не совпадают, тогда переменной $tr_err_k^i$ присваивается значение 1; если совпадают, присваивается 0.

Ошибка определения тренда рассчитывается как количество неверно определенных трендов *k* точки ко всем прогнозам *k* точки.

$$t_error_k = \frac{\sum tr_err_k^i}{n} \qquad (2)$$

где tr_error_k – ошибка определения тренда для *k* точки;
$tr_err_k^i$ – *i* неверный прогноз тренда для *k* точки;
n – количество прогнозов.

Точность определения тренда выражается следующим образом:

$$t_pr_k = 1 - tr_error_k \qquad (3)$$

2. Оценка модели и прогноз

Для оценки качества прогнозирования модели использовались дневные данные по USD/RUB и цен на нефть марки «Брент» с 1.01.2016 по 1.03.2017.

Для оценки NAR блока на исторических данных (back-test) модели задавались следующие параметры:

– количество точек прогноза: 10;

– количество лагов: 2;

– количество нейронов: 5;

– окно для обучения, точек: 200.

Временные ряды содержат 415 значений, с учетом 200 точек, которые необходимы для обучения сети. Таким образом, работа модели тестировалась на 215 исторических точках.

Результаты оценки прогнозирования для доллара и цен на нефть представлены в таблице 1.

Из данных таблицы 1 можно оценить среднюю относительную ошибку и среднюю точность определения тренда.

Относительная ошибка:

– при прогнозировании курса доллара: 1,09 % для 1-й точки, 2,01 % для 10-й точки

– при прогнозировании цен на нефть: 1,28 % для 1-й точки, 4,58 % для 10-й точки

Средняя точность определения тренда:

– при прогнозировании курса доллара: 54,9 %

– при прогнозировании цен на нефть: 54,0 %

На рис. 2 и 3 изображены графики фактических и спрогнозированных (представлен прогноз на первую точку) значений курса дол-

Таблица 1

Оценка модели прогнозирования

		Точки прогноза									
		1	2	3	4	5	6	7	8	9	10
USD/RUB	Средняя абсолютная ошибка	0,45	0,61	0,71	0,83	0,92	1,0	1,08	1,14	1,17	1,24
	Точность определения тренда	0,54	0,55	0,59	0,56	0,56	0,55	0,54	0,57	0,53	0,50
Цены на нефть «Brent»	Средняя абсолютная ошибка	0,68	1,00	1,25	1,54	1,71	1,85	2,04	2,18	2,27	2,43
	Точность определения тренда	0,50	0,52	0,53	0,55	0,53	0,58	0,53	0,53	0,57	0,56

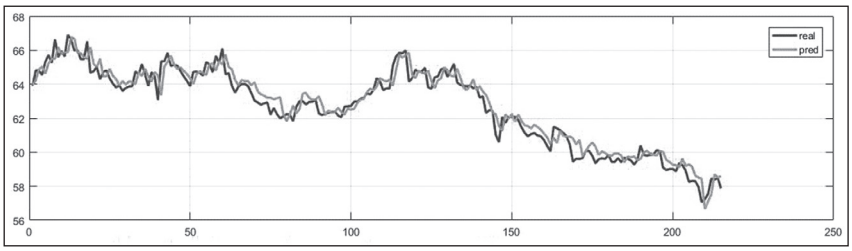

Рис. 2. Динамика фактического (real) и спрогнозированного (pred) значений временного ряда USD/RUB

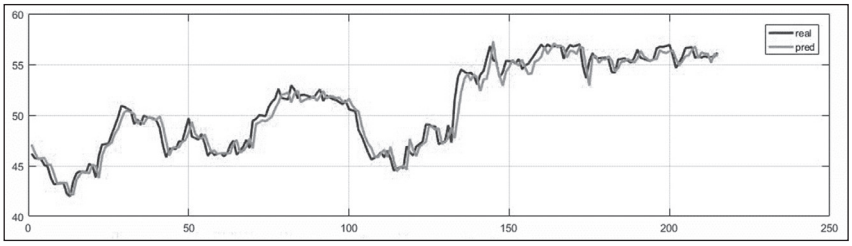

Рис. 3. Динамика фактического (real) и спрогнозированного (pred) значений временного ряда нефть «Брент»

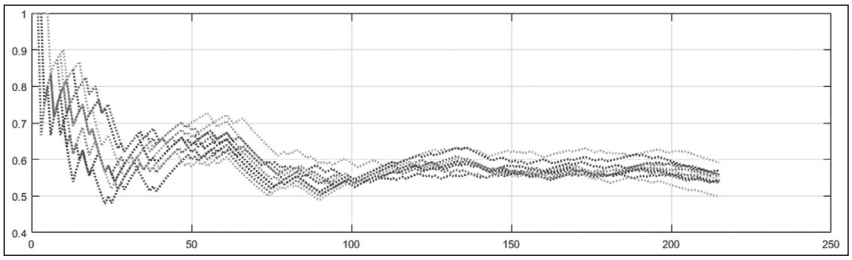

Рис. 4. Динамика точности определения тренда временного ряда USD/RUB

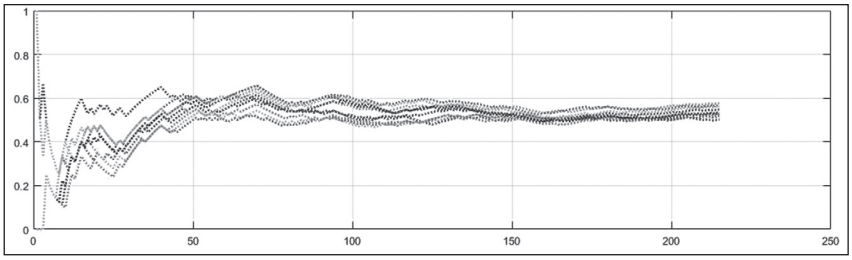

Рис. 5. Динамика точности определения тренда временного ряда нефть «Брент»

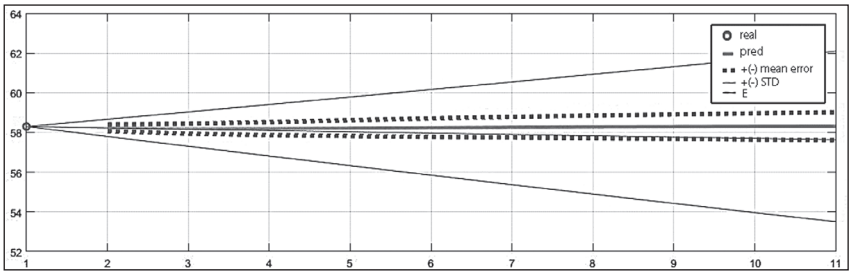

Рис. 6. Прогноз USD/RUB

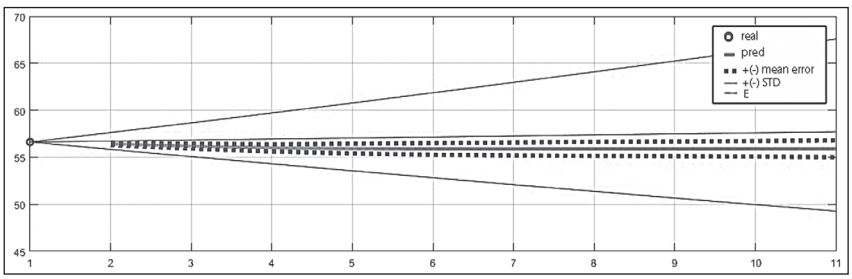

Рис. 7. Прогноз цен на нефть «Брент»

лара (USD/RUB) и цены нефти «Брент» по результатам тестирования модели на 215 исходных точках.

Рис. 4 и 5 отображают зависимость вероятности правильного определения тренда от количества тестовых точек для значений курса доллара (USD/RUB) и цены нефти «Брент». Расчет значений для каждой точки осуществляется по формулам (1)–(3) для каждого момента времени, принадлежащего тестовой выборке. По данным графикам можно судить о стабильности работы модели. Так, если, начиная с некоторого момента времени на графике не выделяется определенный тренд – значения колеблются около установившегося значения, тогда можно сделать вывод о стабильной работе модели.

Как видно, из рис. 4 и рис. 5 значения точности определения тренда не имеют значительных отклонений от конечного значения, поэтому можно говорить о надежности модели.

На рис. 6 и 7 представлены прогнозы на USD/RUB и цену Нефти “Брент” на 10 прогнозных точек в виде следующих графиков: последнее фактическое значение временного ряда (real), NAR прогноз (pred), средняя абсолютная ошибка NAR прогноза (+(–) mean error), стандартное отклонение VAR (+(–) STD), математическое ожидание VAR (E).

На рис. 8 и 9 представлены увеличенные графики прогнозов.

Прогнозы, представленные на рис. 6–9, служат для иллюстрации графической формы прогнозов модели. Временной горизонт прогноза составляет 10 дней.

К тестированию VAR блока предъявляются значительно меньшие требования, чем к NAR блоку, т.к. основной задачей VAR блока является оценка прогноза NAR блока. Поэтому для оценки качества

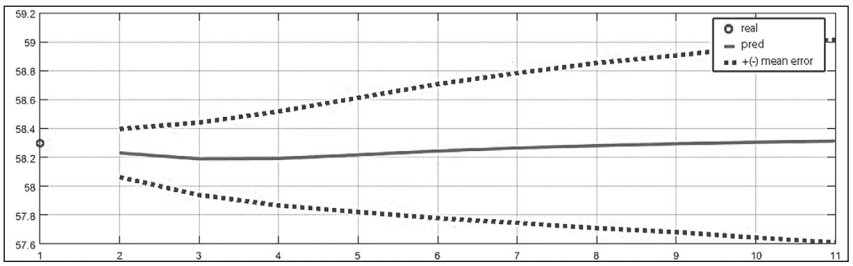

Рис. 8. Прогноз USD/RUB (увел.)

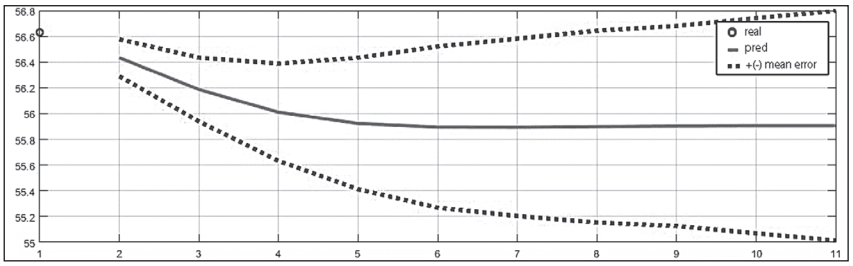

Рис. 9. Прогноз цен на нефть «Брент» (увел.)

модели достаточно результатов средней квадратичной ошибки при методе тестирования аналогичном представленному для NAR блока.

Для VAR блока по критериям AIC и BIC было определено оптимальное количество лагов равное 1. Значения средней квадратичной ошибки VAR блока для курса USD/RUB находятся в диапазоне от 0.07% до 0.11 %, для цен на нефть – от 0.05% до 0.09 %.

Заключение

В рамках настоящей работы предложена модель прогнозирования рядов динамики, включающая два блока: NAR и VAR. NAR блок необходим для построения точечного прогноза, а VAR для оценки прогноза в виде классических метрик статистики: математического ожидания и стандартного отклонения. Также проведена оценка модели на дневных данных USD/RUB, нефть «Брент» с 1.01.2016 по 1.03.2017 и построен прогноз на 10 точек, с учетом рассчитанных при тестировании значений средней абсолютной ошибки модели. Средняя точность прогнозирования тренда для курса доллара составила 54,9%, для цены нефти – 54,0%. Относительная ошибка прогнозирования курса доллара находится в пределах от 1,09 % (для первой точки) до 2,01 % (для десятой точки), относительная ошибка прогнозирования цен на нефть – от 1,28 % (для первой точки) до 4,58 % (для десятой точки). В итоге модель показала достаточно точные результаты при прогнозировании динамических рядов и может быть использована для решения прочих задач прогнозирования. В частности, модель целесообразно использовать в качестве одного из факторов при принятии инвестиционных решений.

Литература

1. *Sims C. A.* Macroeconomics and reality // Econometrica: Journal of the Econometric Society. 1980. P. 1–48.

2. *Johansen S.* Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models // Econometrica: Journal of the Econometric Society. 1991. P. 1551–1580.

3. *Box G. E. P., Jenkins G. M.* Time Series Models for Forecasting and Control // San Francisco. 1970.

4. *Engle R. F.* Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation // Econometrica: Journal of the Econometric Society. 1982. P. 987–1007.

5. *Zhang X.* and *Frey R.* Improving ARMA-GARCH forecasts for high frequency data with regime-switching ARMA-GARCH // Journal of Computational Analysis & Applications. 2015. 18(1).

6. *Kambouroudis D.S., McMillan D.G.* and *Tsakou K.* Forecasting Stock Return Volatility: A Comparison of GARCH, Implied Volatility, and Realized Volatility Models // Journal of Futures Markets. 2016. 36(12). P. 1127–1163.

References

1. *Sims C. A.* Macroeconomics and reality // Econometrica: Journal of the Econometric Society. 1980. P. 1–48.

2. *Johansen S.* Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models // Econometrica: Journal of the Econometric Society. 1991. P. 1551–1580.

3. *Box G. E. P., Jenkins G. M.* Time Series Models for Forecasting and Control // San Francisco. 1970.

4. *Engle R. F.* Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation // Econometrica: Journal of the Econometric Society. 1982. P. 987–1007.

5. *Zhang X.* and *Frey R.* Improving ARMA-GARCH forecasts for high frequency data with regime-switching ARMA-GARCH // Journal of Computational Analysis & Applications. 2015. 18(1).

6. *Kambouroudis D.S., McMillan D.G.* and *Tsakou K.* Forecasting Stock Return Volatility: A Comparison of GARCH, Implied Volatility, and Realized Volatility Models // Journal of Futures Markets. 2016. 36(12). P. 1127–1163.

7. *Corrêa J.M., Neto A.C., Júnior L.T., Franco E.M.C.* and *Faria A.E.* Time series forecasting with the WARIMAX-GARCH method. Neurocomputing. 2016. P. 805–815.

8. *Leontaritis I. J., Stephen A.* Billings. Input-output parametric models for non-linear systems part I: deterministic non-linear systems. // International journal of control. 1985. 41.2. P. 303–328.

9. *Chen S., Billings S. A.* Representations of non-linear systems: the NARMAX model // International Journal of Control. 1989. T. 49. No. 3. P. 1013–1032.

10. *Darrat A. F., Zhong M.* On testing the random-walk Hypothesis: A model-comparison approach // Financial Review. 2000. T. 35. No. 3. P. 105–124.

11. *Jiang C.* and *Song F.* Sunspot Forecasting by Using Chaotic Time-series Analysis and NARX Network // *JCP*. 2011. *6*(7). P. 1424–1429.

12. *Diaconescu E.* The use of NARX neural networks to predict chaotic time series // *Wseas Transactions on computer research*. 2008. *3*(3). P.182–191.

13. *Chaudhuri T.D.* and *Ghosh I.* Artificial Neural Network and Time Series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework // *arXiv preprint*. 2016. *arXiv:1607.02093*

14. *Akaike H.* Maximum likelihood identification of Gaussian autoregressive moving average models // Biometrika. 1973. P. 255–265.

15. *Adkinson M. D.* et al. Alternative models of climatic effects on sockeye salmon, Oncorhynchus nerka, productivity in Bristol Bay, Alaska, and the Fraser River, British Columbia // Fisheries Oceanography. 1996. T. 5. No. 3–4. P. 137–152.

7. *Corrêa J.M., Neto A.C., Júnior L.T., Franco E.M.C.* and *Faria A.E.* Time series forecasting with the WARIMAX-GARCH method. Neurocomputing. 2016. P. 805–815.

8. *Leontaritis I. J., Stephen A.* Billings. Input-output parametric models for non-linear systems part I: deterministic non-linear systems. // International journal of control. 1985. 41.2. P. 303–328.

9. *Chen S., Billings S. A.* Representations of non-linear systems: the NARMAX model // International Journal of Control. 1989. T. 49. No. 3. P. 1013–1032.

10. *Darrat A. F., Zhong M.* On testing the random-walk Hypothesis: A model-comparison approach // Financial Review. 2000. T. 35. No. 3. P. 105–124.

11. *Jiang C.* and *Song F.* Sunspot Forecasting by Using Chaotic Time-series Analysis and NARX Network // *JCP*. 2011. *6*(7). P. 1424–1429.

12. *Diaconescu E.* The use of NARX neural networks to predict chaotic time series // *Wseas Transactions on computer research*. 2008. *3*(3). P.182–191.

13. *Chaudhuri T.D.* and *Ghosh I.* Artificial Neural Network and Time Series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework // *arXiv preprint*. 2016. *arXiv:1607.02093*

14. *Akaike H.* Maximum likelihood identification of Gaussian autoregressive moving average models // Biometrika. 1973. P. 255–265.

15. *Adkinson M. D.* et al. Alternative models of climatic effects on sockeye salmon, Oncorhynchus nerka, productivity in Bristol Bay, Alaska, and the Fraser River, British Columbia // Fisheries Oceanography. 1996. T. 5. No. 3–4. P. 137–152.

Сведения об авторах

Денис Викторович Домащенко
Кандидат экономических наук, Доцент
Российский Экономический Университет имени Г.В. Плеханова, Москва, Россия
Эл. почта: dendv@rambler.ru

Эдвард Евгеньевич Никулин
Российский Экономический Университет имени Г.В. Плеханова, Москва, Россия
Эл. почта: edvardnikulin@gmail.com

Information about the authors

Denis V. Domashchenko
Cand. Sci. (Economics)
Plekhanov Russian University of Economics, Moscow, Russia
E-mail: dendv@rambler.ru

Edvard E. Nikulin
Plekhanov Russian University of Economics, Moscow, Russia
E-mail: nikulinedvard@gmail.ru