

МАТЕМАТИЧЕСКИЕ, СТАТИСТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ В ЭКОНОМИКЕ

DOI: 10.36683/2306-1758/2023-2-44/13-19

УДК 519.237.5:004

JEL: B41, C31, C35

Барбашова Е. В.

РЕГРЕССИОННЫЕ МОДЕЛИ С ПЕРЕМЕННОЙ СТРУКТУРОЙ В ЦИФРОВОЙ ЭКОНОМИКЕ

Барбашова Екатерина Вадимовна

кандидат экономических наук, доцент
Среднерусский институт управления – филиал РАНХиГС
г. Орел, Россия
e-mail: work.67@mail.ru
ORCID: 0000-0002-7842-2890

Ekaterina V. Barbashova

Candidate of Economic Sciences, Associated Professor
Central Russian Institute of Management – Branch of RANEPA
Orel, Russia
e-mail: work.67@mail.ru
ORCID: 0000-0002-7842-2890

В статье выявлена эффективность применения фиктивных переменных в регрессионных моделях, адекватно описывающих динамику показателей социально-экономических процессов сложной структуры. Доказано их преимущество перед моделями, построенными по участкам монотонного изменения показателей: для оценивания параметров модели, в которых качественные характеристики заданы в виде структурных переменных не приходится выходить за рамки классического метода наименьших квадратов, что существенно упрощает процедуру линейного регрессионного анализа. Показано, что коэффициенты моделей с фиктивными переменными легко интерпретировать, они наглядно представляют структуру динамического процесса при относительно небольших затратах времени на аналитическую обработку данных. Это позволяет использовать регрессионные модели с переменной структурой для обработки результатов массовых социологических опросов.

The article discusses the effectiveness of dummy variable use in regression models that adequately describe indicators dynamics of socio-economic processes of complex structure. Their advantage over the models built on areas of monotonous indicators change has been proved. To estimate the parameters of the model in which qualitative characteristics are given as structural variables, one does not have to go beyond the classical least squares method, which greatly simplifies the procedure of linear regression analysis. It is shown that coefficients of dummy variables models are easy to interpret, they clearly represent the dynamic process structure with relatively little time for analytical data processing. This makes it possible to use regression models with variable structure to process the results of mass sociological surveys.

Ключевые слова: фиктивные переменные, регрессионные модели, динамический процесс, индексы, моделирование.

Keywords: dummy variables, regression models, index process, indices, modeling.

Барбашова Е. В. Регрессионные модели с переменной структурой в цифровой экономике // Экономическая среда. – 2023. – № 2 (44). – С. 13-19. – <http://dx.doi.org/10.36683/2306-1758/2023-2-44/13-19>.

Barbashova E.V. Regression Models with Variable Structure in Digital Economy. *Economic environment*. 2023; 2 (44): 13-19. (In Russ.). – <http://dx.doi.org/10.36683/2306-1758/2023-2-44/13-19>.

В наших предыдущих публикациях [4; 5; 6] на эмпирической базе «ВЦИОМ-СПУТНИК» по динамике индексов социальных оценок в 2016-2022 гг. были раскрыты дополнительные возможности, предоставляемые методами параметрической и непараметрической статистики при обращении к полной информации по массовым опросам россиян. Настоящая статья посвящена актуальному вопросу теории и практики построения регрессионных моделей с переменной структурой [3; 22], отражающих изменения в уровне и скорости протекания социально-экономических процессов [10; 14; 16], обусловленных влиянием тех или иных ключевых факторов [1; 13; 19; 20]. В математическом отношении это модели линейной регрессии, в которые, помимо основного предиктора (временной переменной в случае моделей динамики), введены структурные переменные, отражающие ту или иную качественную характеристику. Эти переменные в эконометрике носят название «фиктивные переменные» [8], как правило, они дихотомичны, т. е. принимают два значения: 0 – признака нет и 1 – признак есть. Чаще всего такие модели встречаются в социологии, в экономике они встречаются реже.

Исследователи отмечают ряд положительных свойств моделей с фиктивными переменными [2; 3; 24]. Основное из них – то, что фиктивные переменные позволяют объединить в одной модели выборки, имеющие отличия; в принципе, их можно рассматривать отдельно, но объединение может дать более качественную модель. Второе преимущество – интервалы между наблюдениями не обязательно должны быть одинаковыми, причём в выборке могут быть пропущенные наблюдения. Немаловажным также является то, что коэффициенты при фиктивных переменных наглядно представляют структуру динамического процесса, их легко интерпретировать.

Различают два основных случая введения фиктивных переменных в модели регрессии. В первом случае они предназначены для описания скачков в поведении результирующей переменной. Так, в нашей работе [4] приведён пример скачка показателя «неуверенность респондентов в суждениях», наблюдаемого на рубеже 2017-2018 гг. С помощью процедуры обобщённой линейной модели (ОЛМ) в [4] показано, что неуверенность респондентов в оценке ситуации в стране и в личной жизни скачкообразно возрастает в процессе обсуждения предстоящей пенсионной реформы.

Другой крайний случай – скачкообразное изменение скорости динамики результирующей переменной, когда на графике временного ряда наблюдается «излом». В работе [4] показано, что подобные изломы, характерные для индексов позитивной и негативной ситуации в стране и личной жизни, происходят также на рубеже 2017-2018 гг.

При более внимательном рассмотрении динамики индексов социальных оценок можно заметить, что оба крайних случая изменения структуры моделей (скачок показателя и излом) встречаются одновременно, но либо преобладают скачки уровня показателя, и тогда нужно применять модели *ANOVA* (модели дисперсионного анализа), либо преобладают изломы рядов динамики, и тогда следует применять модели *ANCOVA* (модели ковариационного анализа) [7; 9; 18].

Покажем, что введение в модели линейной регрессии фиктивных переменных решает обе задачи, позволяя оценить и скачки, и изломы в динамике результирующих переменных. В качестве эмпирической базы используем данные агентства «ВЦИОМ-СПУТНИК» по динамике индекса оптимизма россиян в 2017-2023 гг. Этот индекс имеет «социологическую» природу, но задаваемый респондентам вопрос носит экономический характер: «Есть разные точки зрения по поводу экономического кризиса в нашей стране. Как Вы думаете, мы переживаем сейчас самые тяжелые времена, или они позади, или ещё впереди?» Индекс оптимизма, иначе именуемый «индексом социальных ожиданий», показывает, насколько оптимистично россияне оценивают экономическое будущее страны. Показатель рассчитывается как разность между ответом «они уже позади» и суммой ответов «переживаем их сейчас» и «они еще впереди». Индекс может принимать значение от -100 до 100 пунктов. Чем выше значение индекса, тем оптимистичнее выглядят перспективы России. Нулевое значение фиксирует баланс оптимистичных и пессимистичных прогнозов¹.

На рисунке 1 представлены графики динамики вероятности различных вариантов ответов: вероятность экономического пессимизма как сумма ответов «переживаем их сейчас» и «они еще впереди», вероятность экономического оптимизма (вариант «они уже позади») и вероятность варианта «затрудняюсь ответить», в период от января 2017 по февраль 2023 год.

Рассматриваемый временной период включает по меньшей мере три знаменательных события, во многом определившие экономический оптимизм/пессимизм россиян [12; 15; 17; 21; 23]: 1) обсуждение пенсионной реформы с января 2018 года; 2) всплеск пандемии КОВИД-19 с января 2020 года; 3) проведение специальной военной операции на Украине с марта 2022 года (на рисунке 1 *а* начала этих событий отмечены вертикальными точечными линиями).

В ракурсе поставленной в статье задачи наибольший интерес представляют два временных диапазона. Первый из них, представленный в более крупном масштабе на рисунке 1 *б* с января 2018 по январь 2020 года, отвечает наиболее интенсивному обсуждению россиянами

¹ ВЦИОМ : официальный сайт. – URL: <https://wciom.ru/>.

пенсионной реформы. Этот временной диапазон содержит два цикла (рисунок): цикл спада экономического оптимизма (январь 2018 – январь 2019 года) и цикл подъёма экономического оптимизма (февраль 2019 – январь 2020 года) и является эмпирической базой для отработки техники конструирования моделей с переменной структурой. Второй временной диапазон – с мая 2020 по март 2022 года – отвечает «стационарной» динамике показателей; он выбран для проведения факторного анализа с целью выявления взаимосвязи компонентов индекса оптимизма.

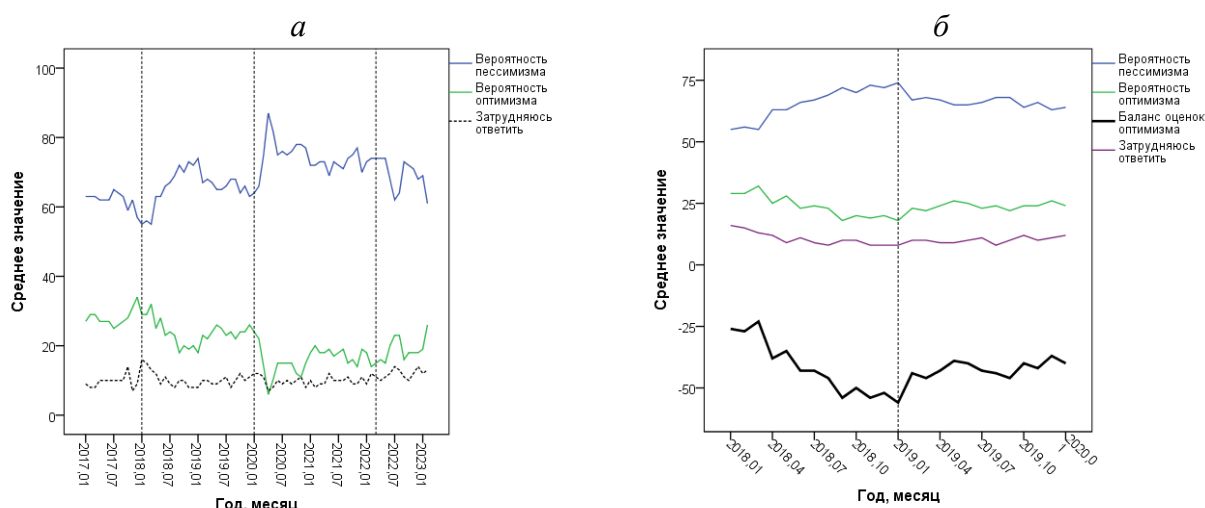


Рисунок 1 – Динамика вероятности вариантов ответов на вопрос: «Как Вы думаете, мы переживаем сейчас самые тяжелые времена, или они позади, или еще впереди?»:

а – в период январь 2017 года – февраль 2023 года;

б – в период январь 2018 года – январь 2020 года

В рамках традиционного подхода к моделированию динамики компонентов индекса оптимизма на первом временном диапазоне (с января 2018 по январь 2020 года) необходимо рассматривать этапы спада и подъёма экономического оптимизма в отдельности. В результате применения процедуры «подбор кривой» пакета анализа данных SPSS-22 оказалось, что если все четыре линейные модели адекватно отображают этап спада экономического оптимизма, то на этапе подъёма экономического оптимизма из четырёх моделей адекватна только одна – для вероятности пессимизма (таблица 1).

Таблица 1 – Сводка характеристик качества линейных моделей динамики составляющих индекса экономического оптимизма на этапах спада и подъёма

Составляющая индекса экономического оптимизма	Этап спада экономического оптимизма		Этап подъёма экономического оптимизма	
	Коэффициент детерминации	Критерий Фишера / значимость	Коэффициент детерминации	Критерий Фишера / значимость
Вероятность пессимизма, %	0,914	116,828 / 0,000	0,331	4,958 / 0,050
Вероятность оптимизма, %	0,826	52,234 / 0,000	0,078	0,843 / 0,380
Баланс оценок, %	0,894	93,147 / 0,000	0,332	3,016 / 0,113
Неуверенность респондентов, %	0,745	32,150 / 0,000	0,279	3,869 / 0,078

Поясним присутствие в таблице 1 результирующей переменной «неуверенность респондентов»). Как было показано в нашей предыдущей работе [4] на примере индексов социальной ситуации, динамика ответов «затрудняюсь ответить» отражает динамику неуверенности респондентов в суждениях. Аналогично факторный анализ данных второго временного диапазона (с мая 2020 по март 2022 года) показал, что совокупность четырёх составляющих индекса экономического оптимизма россиян имеет двумерную структуру, в которой первый главный фактор, объясняющий 71,9 % общей дисперсии (после вращения по критерию «варимакс»), нагружен на положительном направлении оси составляющими «Вероятность оптимизма» и «Баланс оценок», и на отрицательном составляющей «Вероятность пессимизма», а второй главный фактор, объясняющий 28,1 % общей дисперсии, нагружен составляющей «Затрудняюсь ответить». Эти два фактора объясняют почти 100 % общей дисперсии, и двухфакторную структуру составляющих можно считать адекватной моделью их взаимосвязи.

Перейдём к моделированию динамики составляющих индекса экономического оптимизма на временном диапазоне с января 2018 по январь 2020 года «в целом». Будем искать МНК-оценки параметры моделей вида:

$$Y = a + bt + g_1D + g_2Dt + e, \quad (1)$$

где t – временная переменная (номер месяца),

D – структурная переменная, принимающая значение 0 до точки излома и значение 1 после этой точки,

g_1 и g_2 – дифференциальный свободный член и дифференциальный угловой коэффициент, соответственно, с помощью процедуры линейной регрессии пакета анализа данных SPSS-22.

Не останавливаясь на деталях, приведём таблицу 2, содержащую характеристики качества моделей динамики составляющих индекса экономического оптимизма на всем временном интервале с января 2018 по январь 2020 года.

Таблица 2 – Сводка характеристик качества моделей динамики составляющих индекса экономического оптимизма на временном интервале с января 2018 по январь 2020 год

Составляющая индекса экономического оптимизма	Коэффициент детерминации	Критерий Фишера / значимость
Вероятность пессимизма, %	0,882	52,417 / 0,000
Вероятность оптимизма, %	0,776	24,206 / 0,000
Баланс оценок, %	0,859	42,742 / 0,000
Неуверенность респондентов, %	0,678	14,746 / 0,000

Из сравнения таблиц 1 и 2 следует, что в результате введения в модели структурной переменной все четыре модели становятся адекватными.

В качестве примера приведём явный вид модели динамики вероятности пессимизма со структурной переменной:

$$Y_{\text{расч}} = 53,923 + 1,692t + 17,380D - 1,969Dt. \quad (2)$$

Графики динамики эмпирических и расчётных значений показателя приведены на рисунке 2, здесь же показаны верхняя и нижняя девяностопятипроцентные доверительные границы индивидуальных значений прогноза.

На графике просматривается не только изменение наклона аппроксимирующей прямой, но и «скачок» предсказанных значений показателя, отвечающий временной метке излома «февраль 2019 г.».

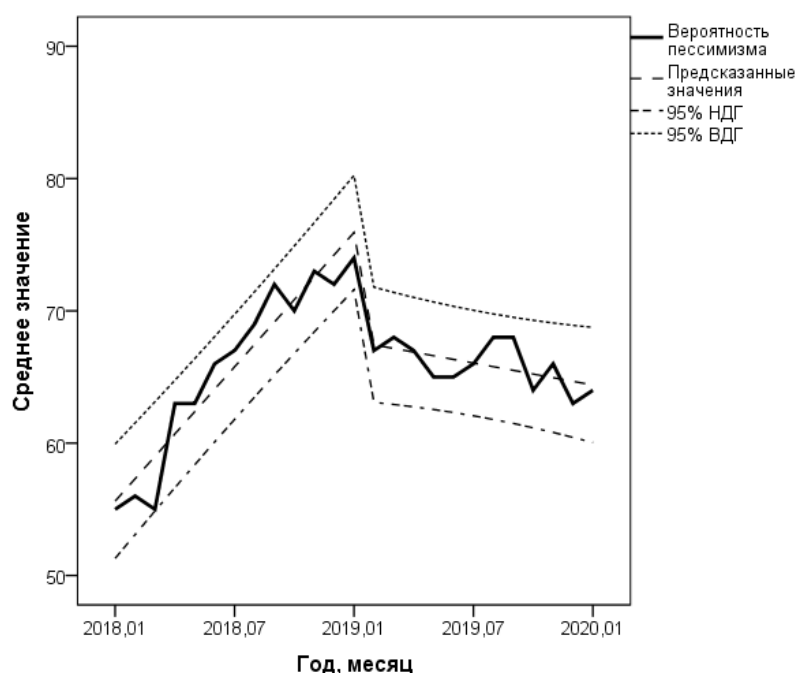


Рисунок 2 – Аппроксимация динамики вероятности экономического пессимизма моделью с фиктивной переменной

Для интерпретации параметров модели (1) рассчитаем предсказанные значения показателя, отвечающие временным меткам «январь 2019 г.» и «февраль 2019 г.». Подставив в уравнение регрессии (2) значения временной переменной $t=13$ и $t=14$, а также значения фиктивной переменной $D=0$ и $D=1$, соответственно, получаем:

$$Y_{\text{расч}}(t=13) = 53,923 + 1,692 \cdot 13 = 75,919\%; \quad (3)$$

$$Y_{\text{расч}}(t=14) = 53,923 + 1,692 \cdot 14 + 17,380 \cdot 1 - 1,969 \cdot 1 \cdot 14 = 67,424\%; \quad (4)$$

«Скачок» показателя составит величину:

$$\Delta Y_{\text{расч}} = Y_{\text{расч}}(t=14) - Y_{\text{расч}}(t=13) = 67,424 - 75,919 = -8,495\%. \quad (5)$$

Сравнивая выражения (3)-(5), получаем формулу для расчёта величины «скачка» показателя:

$$\Delta Y_{\text{расч}} = b + g_1 + g_2 \cdot t_0, \quad (6)$$

где t_0 – временная координата сдвига показателя.

Легко видеть, что интерпретация параметров модели a , b и g_2 такова: a – начальный уровень показателя, b – коэффициент наклона линии регрессии до точки сдвига показателя, $b+g_2$ – коэффициент наклона линии регрессии после сдвига.

Сведём параметры моделей динамики составляющих индекса экономического оптимизма в таблицу 3.

Поскольку все параметры приведенных в таблице 3 моделей статистически значимы по t -критерию, имеется возможность сравнить их величины для различных составляющих индекса экономического оптимизма. Так, наклон составляющей «баланс оценок» до и после сдвига равен разности наклонов составляющих оптимизма и пессимизма, и для этой же составляющей наблюдается максимум сдвига показателя. В наименьшей мере сдвиг и излом показателя сказывается на составляющей «Неуверенность респондентов», тем не менее оба эффекта статистически значимы и требуют социологической оценки.

Таблица 3 – Параметры кусочно-линейных моделей динамики составляющих индекса экономического оптимизма на временном интервале с января 2018 по январь 2020 года

Составляющая индекса экономического оптимизма	a	b	g_1	g_2	Наклон до сдвига	Наклон после сдвига	Сдвиг показателя при $t_0=14$
Вероятность пессимизма, %	53,923	1,692	17,380	-1,969	1,692	-0,277	-8,495
Вероятность оптимизма, %	31,308	-1,088	-9,368	1,189	-1,088	0,101	6,190
Баланс оценок, %	-22,615	-2,780	-26,748	3,158	-2,780	0,378	14,684
Неуверенность респондентов, %	14,769	-0,604	-8,012	0,779	-0,604	0,175	2,290

Подведём итоги. Приведённые примеры убедительно свидетельствуют об эффективности применения фиктивных переменных в регрессионных моделях, адекватно описывающих динамику показателей социально-экономических процессов сложной структуры. Их преимущество перед моделями, построенными по участкам монотонного изменения показателей, в том, что для оценивания параметров модели «в целом» не приходится выходить за рамки классического метода наименьших квадратов, при этом достаточно ограничиться «стандартной» процедурой линейного регрессионного анализа. Коэффициенты моделей с фиктивными переменными легко интерпретировать, они наглядно представляют структуру динамического процесса.

Список источников:

1. Авров, А. П. О применении индексного факторного анализа в мультипликативных моделях обобщающих статистических показателей / А. П. Авров // Развитие территорий. – 2022. – № 1(27). – С. 78-82. – DOI 10.32324/2412-8945-2022-1-78-82. – EDN XXVIJC.
2. Артемова, Н. В. Методические подходы прогнозирования валового сбора продукции растениеводства / Н. В. Артемова, С. А. Ильминская // Вестник ОрелГИЭТ. – 2021. – № 1(55). – С. 12-21. – DOI 10.36683/2076-5347-2021-1-55-12-21. – EDN YTRXKJ.
3. Базилевский, М. П. Метод построения неэлементарных линейных регрессий на основе аппарата математического программирования / М. П. Базилевский // Проблемы управления. – 2022. – № 4. – С. 3-14. – DOI 10.25728/pu.2022.4.1. – EDN PNHVHL.
4. Барбашова, Е. В. Информационное обеспечение мониторинга социальных процессов: проблемы социолого-математического моделирования / Е. В. Барбашова, О. В. Ляковская // Образование и наука без границ: фундаментальные и прикладные исследования. – 2022. – № 16. – С. 106-113. – DOI 10.36683/2500-249X/2022-16/106-113. – EDN VNQRZL.
5. Барбашова, Е. В. О моделировании динамики социально-политических процессов (по материалам мониторинга ВЦИОМ) / Е. В. Барбашова, О. В. Ляковская // Современные проблемы физико-математических наук: материалы VIII Всероссийской научно-практической конференции с международным участием (25 - 26 ноября 2022 г., г. Орёл) / под общей редакцией кандидата физико-математических наук, доцента Т. Н. Можаровой. – Орёл: ОГУ имени И. С. Тургенева, 2022. – С. 155-168.
6. Барбашова, Е. В. Применение функции Харрингтона для оценки инвестиционной привлекательности регионов РФ / Е. В. Барбашова, И. В. Гайдакина, А. А. Музалевская // Вестник ОрелГИЭТ. – 2022. – № 1(59). – С. 13-19. – DOI 10.36683/2076-5347-2022-1-59-13-19. – EDN MQGYYY.
7. Бурганов, Р. Т. Модернизированная производственная функция Кобба - Дугласа как инструмент исследования влияния цифровой трансформации на экономику региона / Р. Т. Бурганов // Среднерусский вестник общественных наук. – 2022. – Т. 17, № 3. – С. 161-183. – DOI 10.22394/2071-2367-2022-17-3-161-183. – EDN FFBVKR.
8. Доугерти К. Введение в эконометрику. – Москва: ИНФРА-М, 2009. – 465 с.
9. Замышляева, Е. Л. Информационная прозрачность и анализ больших данных о системе предпринимательства на основе регуляторов должной финансовой отчетности / Е. Л. Замышляева, Н. В. Парушина // Среднерусский вестник общественных наук. – 2020. – Т. 15, № 1. – С. 162-185. – DOI 10.22394/2071-2367-2020-15-1-162-185. – EDN OKQZXC.

10. Илюхина, И. Б. Оценка результатов государственной политики по формированию региональной инновационной среды Орловской области / И. Б. Илюхина, С. А. Ильминская, Ю. Г. Голоктионова // Вестник ОрелГИЭТ. – 2022. – № 1(59). – С. 209-213. – DOI 10.36683/2076-5347-2022-1-59-209-213. – EDN НКТХСР.
11. Малкина, М. Ю. Институциональные факторы политического доверия в современной России / М. Ю. Малкина, В. Н. Овчинников, К. А. Холодилин // Журнал институциональных исследований. – 2020. – Т. 12, № 4. – С. 77-93. – DOI 10.17835/2076-6297.2020.12.4.077-093. – EDN QCXRWV.
12. Матвеев, В. В. Роль государства в развитии цифровой экономики: российский опыт / В. В. Матвеев, А. В. Агеев // Вестник ОрелГИЭТ. – 2020. – № 2(52). – С. 245-249. – DOI 10.36683/2076-5347-2020-2-52-245-249. – EDN VVEVUQ.
13. Мицек, Е. Б. Анализ факторов динамики основных макроэкономических переменных Российской Федерации / Е. Б. Мицек, С. А. Мицек // Вопросы управления. – 2020. – № 1(62). – С. 47-62. – DOI 10.22394/2304-3369-2020-1-47-62. – EDN КРҮСНГ.
14. Павленок, Д. В. Современные методы оценки эффективности инновационного развития регионов в рамках ускоренной цифровизации / Д. В. Павленок // Научные Записки ОрелГИЭТ. – 2020. – № 3(35). – С. 29-33. – EDN JVNINI.
15. Полянин, А. В. Инвестиционное поведение населения в период глобальной турбулентности / А. В. Полянин, О. В. Попова, К. А. Суровнева // Среднерусский вестник общественных наук. – 2021. – Т. 16, № 3. – С. 166-182. – DOI 10.22394/2071-2367-2021-16-3-166-182. – EDN МКРАQT.
16. Савина, А. Г. Построение и реализация динамической модели компетенций в контексте цифрового университета / А. Г. Савина, Л. И. Малявкина // Образование и наука без границ: фундаментальные и прикладные исследования. – 2021. – № 14. – С. 88-92. – DOI 10.36683/2500-249X/2021-14/88-92. – EDN WNNFVJ.
17. Самородова, Е. М. Цифровизация, экономическая безопасность и конкурентоспособность национальных экономик: международные измерения / Е. М. Самородова // Образование и наука без границ: фундаментальные и прикладные исследования. – 2022. – № 15. – С. 40-44. – DOI 10.36683/2500-249X/2022-15/40-44. – EDN NYEFWK.
18. Скроботов, А. А. Структурные сдвиги в моделях коинтеграции / А. А. Скроботов // Прикладная эконометрика. – 2021. – № 3(63). – С. 117-141. – DOI 10.22394/1993-7601-2021-63-117-141. – EDN ZWMYCY.
19. Стрельцов, А. В. Использование методов стратегического анализа при формировании конкурентоспособности предприятий / А. В. Стрельцов, Г. И. Яковлев // Экономическая среда. – 2022. – № 1(39). – С. 5-10. – DOI 10.36683/2306-1758/2022-1-39/5-10. – EDN ANYGUL.
20. Чекулина, Т. А. Оценка качества жизни населения на региональном уровне с учетом дифференциации денежных доходов / Т. А. Чекулина, О. В. Лясковская, Е. В. Барбашова // Вестник ОрелГИЭТ. – 2019. – № 2(48). – С. 208-219. – EDN HPGQTV.
21. Шманев, С. В. Современное состояние цифровой экономики в России и перспективы ее развития / С. В. Шманев // Известия Юго-Западного государственного университета. Серия: Экономика. Социология. Менеджмент. – 2021. – Т. 11, № 3. – С. 44-52. – EDN DXRMZW.
22. Linets, G. I. Evaluation of the efficiency of reference machine learning models for buffer memory prediction when transforming a self-similar input stream of packets into a stream having exponential distribution under the condition of equality of mathematical expectations and median flows / G. I. Linets, R. A. Voronkin, S. V. Govorova // Modern Science and Innovations. – 2022. – No. 3(39). – P. 20-32. – DOI 10.37493/2307-910X.2022.3.2. – EDN LOKEDF.
23. Novikova, N. V. Regional aspects of studying the digital economy in the system of economic growth drivers / N. V. Novikova, E. V. Strogonova // Journal of New Economy. – 2020. – Vol. 21, No. 2. – P. 76-93. – DOI 10.29141/2658-5081-2020-21-2-5. – EDN FYTRUQ.
24. Ray, S. Does personality and demographic variances of individual investors challenge the assumption of rationality? A two staged regression modeling-artificial neural network approach / S. Ray // Economic Environment. – 2022. – No. 3(41). – P. 34-47. – DOI 10.36683/2306-1758/2022-3-41/34-47. – EDN STGAOX.