

5.2.3. РЕГИОНАЛЬНАЯ И ОТРАСЛЕВАЯ ЭКОНОМИКА
(ЭКОНОМИЧЕСКИЕ НАУКИ)

Научная статья

УДК 311, 519.237.5, 631.11, 004.8

DOI: 10.53914/issn2071-2243_2023_3_251

EDN: ITNCVJ

**Моделирование уровня производства сельскохозяйственной
продукции в крестьянских (фермерских) хозяйствах
методами эконометрики и машинного обучения**

**Анна Владимировна Уколова¹, Баярма Шагдаровна Дашиева^{2✉},
Денис Витальевич Быков³, Александр Евгеньевич Ульянов⁴,
Дмитрий Эдуардович Храмов⁵**

^{1, 2, 3, 4, 5}Российский государственный аграрный университет –

МСХА имени К.А. Тимирязева, Москва, Россия

²dashieva.b.sh@rgau-msha.ru[✉]

Аннотация. В рамках Государственной программы развития сельского хозяйства ... государство предоставляет ряд мер поддержки К(Ф)Х в целях стимулирования их развития. Несмотря на это, К(Ф)Х сталкиваются с рядом проблем, поэтому проведение статистического анализа результатов их деятельности является важным инструментом при разработке мер повышения эффективности хозяйствования. Представлены результаты эконометрического моделирования уровня производства продукции сельского хозяйства в крестьянских (фермерских) хозяйствах Липецкой области. Построены три модели парной регрессии, характеризующие зависимость уровня доходов от реализации продукции сельского хозяйства в расчете на единицу площади от обеспеченности работниками и сельскохозяйственной техникой, а также урожайности зерновых и зернобобовых культур от затрат на минеральные удобрения. Оценка параметров моделей производилась классическим методом наименьших квадратов (МНК), взвешенным методом наименьших квадратов (ВМНК), а также методами машинного обучения. Параметры моделей статистически значимы, и они могут быть использованы для прогнозирования и определения резервов роста. Первая модель характеризуется более тесной связью между переменными, при этом у половины К(Ф)Х имеются резервы роста доходов от реализации сельскохозяйственной продукции при фиксированном уровне обеспеченности работниками. По третьей модели получен положительный коэффициент полной регрессии, показывающий прирост урожайности на 0,6 ц/га при увеличении затрат на 1 тыс. руб. в расчете на 1 га посевной площади. С точки зрения качества аппроксимации, МНК сравнивался с многослойным перцептроном (MLP), методом k ближайших соседей (KNN), по которым не был получен существенно отличающийся коэффициент детерминации, и методом градиентного бустинга (GB), позволившего получить наибольший R^2 . Моделью градиентного бустинга с фактором затрат на минеральные удобрения и средства защиты растений в расчете на единицу посевной площади объясняется 66,9% вариации урожайности зерновых и зернобобовых культур.

Ключевые слова: сельское хозяйство, крестьянское (фермерское) хозяйство, уровень производства продукции, корреляционно-регрессионный анализ, метод наименьших квадратов, взвешенный метод наименьших квадратов, методы машинного обучения

Для цитирования: Уколова А.В., Дашиева Б.Ш., Быков Д.В., Ульянов А.Е., Храмов Д.Э. Моделирование уровня производства продукции сельского хозяйства в крестьянских (фермерских) хозяйствах методами эконометрики и машинного обучения // Вестник Воронежского государственного аграрного университета. 2023. Т. 16, № 3(78). С. 251–262. https://doi.org/10.53914/issn2071-2243_2023_3_251-262.

5.2.3. REGIONAL AND SECTORAL ECONOMICS
(ECONOMIC SCIENCES)

Original article

**Modeling the level of agricultural production in peasant (farm)
households using econometrics and Machine Learning Methods**

**Anna V. Ukolova¹, Bayarma Sh. Dashieva^{2✉},
Aleksandr E. Ulianckin³, Denis V. Bykov⁴,
Dmitriy E. Khramov⁵**

^{1, 2, 3, 4, 5}Russian Timiryazev State Agrarian University, Moscow, Russia

²dashieva.b.sh@rgau-msha.ru[✉]

Abstract. Within the framework of the State Program for the Development of Agriculture the State provides a number of support measures for peasant (farm) households (PFH) in order to stimulate their development. Despite this, PFHs face a number of problems. Therefore, statistical analysis of results of their activities is an important tool in the development of measures for improving the efficiency of management. The authors present the results of econometric modeling of the level of agricultural production in peasant (farm) households of Lipetsk Oblast. Three paired regression models were built characterizing the dependence of the level of income obtained by selling agricultural products per unit area from the availability of human resources and agricultural machinery, as well as the dependence of yield of grain and leguminous crops from the cost of mineral fertilizers. The parameters of models were estimated by the classical ordinary least squares method (OLS), weighted ordinary least squares method (WLS), and Machine Learning Methods. The parameters of models are statistically significant and can be used to predict and determine the growth reserves. The first model is characterized by a closer relationship between the variables, and half of the assessed PFHs have reserves for the growth of income from selling agricultural products at a fixed level of human resource availability. The third model gave a positive coefficient of full regression showing an increase in yield by 0.6 c/ha with an increase in costs by 1 thousand rubles per 1 hectare of sown area. In terms of quality of approximation, the OLS was compared with a multilayer perceptron (MLP) using the method of k -nearest neighbors (KNN) that did not provide a significantly different coefficient of determination, and the gradient boosting (GB) method, which allowed obtaining the largest R^2 . The gradient boosting model with the factor of cost of mineral fertilizers and plant protection products per unit of sown area explains the 66.9% of variation in the yield of grain and leguminous crops.

Key words: agriculture, peasant (farm) households, level of agricultural production, correlation-regression analysis, ordinary least squares method, weighted least squares method, Machine Learning Methods

For citation: Ukolova A.V., Dashieva B.Sh., Bykov D.V., Ulianckin A.E., Khramov D.E. Modeling the level of agricultural production in peasant (farm) households using econometrics and Machine Learning Methods. *Vestnik Voronezhskogo gosudarstvennogo agrarnogo universiteta = Vestnik of Voronezh State Agrarian University*. 2023;16(3):251-262. (In Russ.). https://doi.org/10.53914/issn2071-2243_2023_3_251-262.

Аграрно-экономические реформы, проводимые в России в конце XX в., заложили основы многоукладной экономики в современном ее проявлении, позволили частично осуществить ряд земельных преобразований, принципиально изменить организационно-правовые формы большей части сельхозпроизводителей. К настоящему времени сформировалась аграрная структура, в которую вошли как представители крупного бизнеса, так и малые, и средние предпринимательские агроструктуры.

Официальная статистика выделяет в аграрном секторе российской экономики три категории хозяйств:

- 1) сельскохозяйственные предприятия;
- 2) крестьянские (фермерские) хозяйства;
- 3) хозяйства населения.

Крестьянское (фермерское) хозяйство как форма хозяйствования представляет собой объединение граждан, связанных родством и (или) свойством, имеющих в общей собственности имущество и лично участвующих в сельскохозяйственном производстве и деятельности, связанной с ним [2]. Удельный вес К(Ф)Х в структуре продукции сельского хозяйства Российской Федерации увеличился с 1% в 1992 г. до 16% в 2022 г., что свидетельствует об увеличении их значимости как представителей малого бизнеса в сельском хозяйстве в обеспечении продовольственной безопасности, создании новых рабочих мест, сохранении сельского образа жизни и развитии сельских территорий.

В рамках Государственной программы развития сельского хозяйства и регулирования рынков сельскохозяйственной продукции, сырья и продовольствия [4] государство предоставляет ряд мер поддержки К(Ф)Х в целях стимулирования их развития [7]. Несмотря на это, К(Ф)Х сталкиваются с рядом проблем, таких как нехватка финансовых ресурсов, недостаточное знание современных технологий и методов управления, низкая эффективность производства и др. Поэтому проведение статистического анализа результатов деятельности К(Ф)Х является важным инструментом при разработке мер повышения эффективности их хозяйствования. С учетом имеющихся данных эффективность деятельности К(Ф)Х предлагается оценивать показателями уровня производства продукции сельского хозяйства в расчете на единицу площади и уровня продуктивности сельскохозяйственных культур и животных.

Целью представленного исследования являлось построение эконометрических моделей уровня производства продукции сельского хозяйства в крестьянских (фермерских) хозяйствах.

Для достижения поставленной цели в круг задач были включены следующие:

- анализ исходной совокупности К(Ф)Х, исключение нулевых показателей и наблюдений, устранение выбросов;
- расчет относительных показателей, характеризующих уровень производства продукции сельского хозяйства К(Ф)Х;
- эконометрическое нейросетевое моделирование уровня производства продукции сельского хозяйства К(Ф)Х;
- анализ и интерпретация результатов по построенным моделям.

Информационной базой исследования служили данные формы отчетности «Информация о производственной деятельности глав крестьянских (фермерских) хозяйств – индивидуальных предпринимателей за 2021 год» (далее – форма № 1-КФХ), которая представляется в Министерство сельского хозяйства Российской Федерации крестьянскими (фермерскими) хозяйствами, получающими субсидии из бюджета [3].

Для исследования была выбрана Липецкая область, занимающая 7-е место из 18 регионов Центрального федерального округа по числу К(Ф)Х по данным микропереписи 2021 г. и 8-е место в 2021 г. по доле валовой добавленной стоимости (ВДС) сельского, лесного хозяйства, охоты, рыболовства и рыбоводства в валовом региональном продукте (ВРП), по последнему показателю регион находится на 32-м месте среди всех субъектов Российской Федерации (рис. 1) [1].

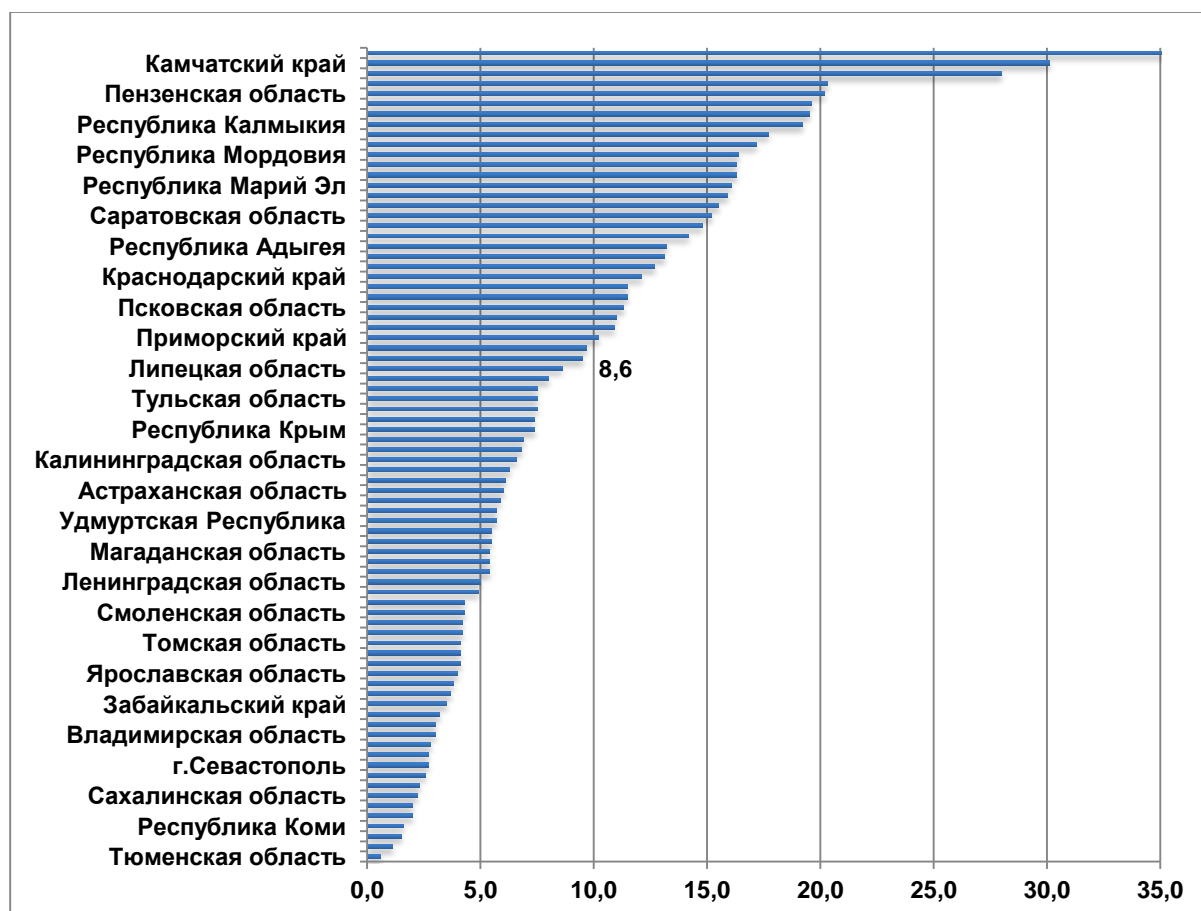


Рис. 1. Удельный вес ВДС сельского, лесного хозяйства, охоты, рыболовства и рыбоводства в ВРП субъектов Российской Федерации в 2021 г., %

По данным формы № 1-КФХ были определены результативные признаки уровня производства продукции сельского хозяйства и продуктивности сельскохозяйственных культур и их факторы (показатели интенсивности):

y_1 – доходы от реализации сельскохозяйственной продукции собственного производства, продуктов ее первичной и промышленной переработки в расчете на 100 га площади земельных угодий, тыс. руб.;

y_2 – урожайность зерновых и зернобобовых культур, ц/га;

x_1 – среднегодовая численность наемных работников и членов К(Ф)Х в расчете на 100 га площади земельных угодий, чел.;

x_2 – обеспеченность сельскохозяйственной техникой в расчете на 100 га площади земельных угодий, шт.;

x_3 – затраты на минеральные удобрения и средства защиты растений в расчете на 1 га убранной площади зерновых и зернобобовых, тыс. руб.

Поскольку более половины всех К(Ф)Х Липецкой области специализируется на производстве продукции растениеводства, в первую очередь зерновых и зернобобовых, одной из результативных переменных была взята урожайность этой группы культур.

В исследовании использовали данные 92,7% К(Ф)Х Липецкой области, имевших доходы от реализации продукции сельского хозяйства в расчете на 100 га площади земельных участков в размере до 20 млн руб. и сдавших форму отчетности № 1-КФХ в 2021 г. (общее количество таких субъектов – 451). Авторами были исключены К(Ф)Х с нулевой земельной площадью и доходами, а также с резко выделяющимися доходами – от 20 млн руб. и выше.

Для изучения тесноты связи между изучаемыми признаками были рассчитаны парные коэффициенты корреляции Пирсона.

Сила связи между y_1 (доходы в расчете на 100 га площади земельных участков) и первым факторным признаком x_1 (численность работников К(Ф)Х на 100 га площади земельных угодий) – умеренная ($r_{y_1x_1} = 0,483$), а со вторым фактором x_2 (число сельскохозяйственной техники в расчете на 100 га площади земельных участков) – слабая ($r_{y_1x_2} = 0,342$).

Между двумя факторными признаками x_1 и x_2 наблюдается тесная взаимосвязь – $r_{x_1x_2} = 0,714$, следовательно, включать их одновременно в уравнение регрессии нецелесообразно, поэтому были построены только парные модели. Между фактором y_2 (урожайность зерновых и зернобобовых культур, ц/га) и фактором x_3 (затраты на минеральные удобрения и средства защиты растений в расчете на 1 га убранной площади зерновых и зернобобовых) также наблюдается слабая связь ($r_{y_2x_3} = 0,391$).

По диаграммам рассеяния (рис. 2) можно предположить наличие прямой линейной связи средней (ниже средней) тесноты между y_1 и x_1 , y_1 и x_2 в условиях гетероскедастичности остатков.

Гистограммы по каждому признаку показывают, что распределения хозяйств по изучаемым признакам имеют правостороннюю скошенность, особенно сильную по двум факторам – x_1 и x_2 .

Таким образом, могут быть построены модели:

$$\hat{y}_1 = a + bx_1 ; \quad (1)$$

$$\hat{y}_1 = a + bx_2 . \quad (2)$$

Для урожайности зерновых и зернобобовых культур был подобран фактор x_3 (затраты, связанные с приобретением минеральных удобрений, а также средств защиты растений) и следующая модель регрессии:

$$\hat{y}_2 = a + bx_3 . \quad (3)$$

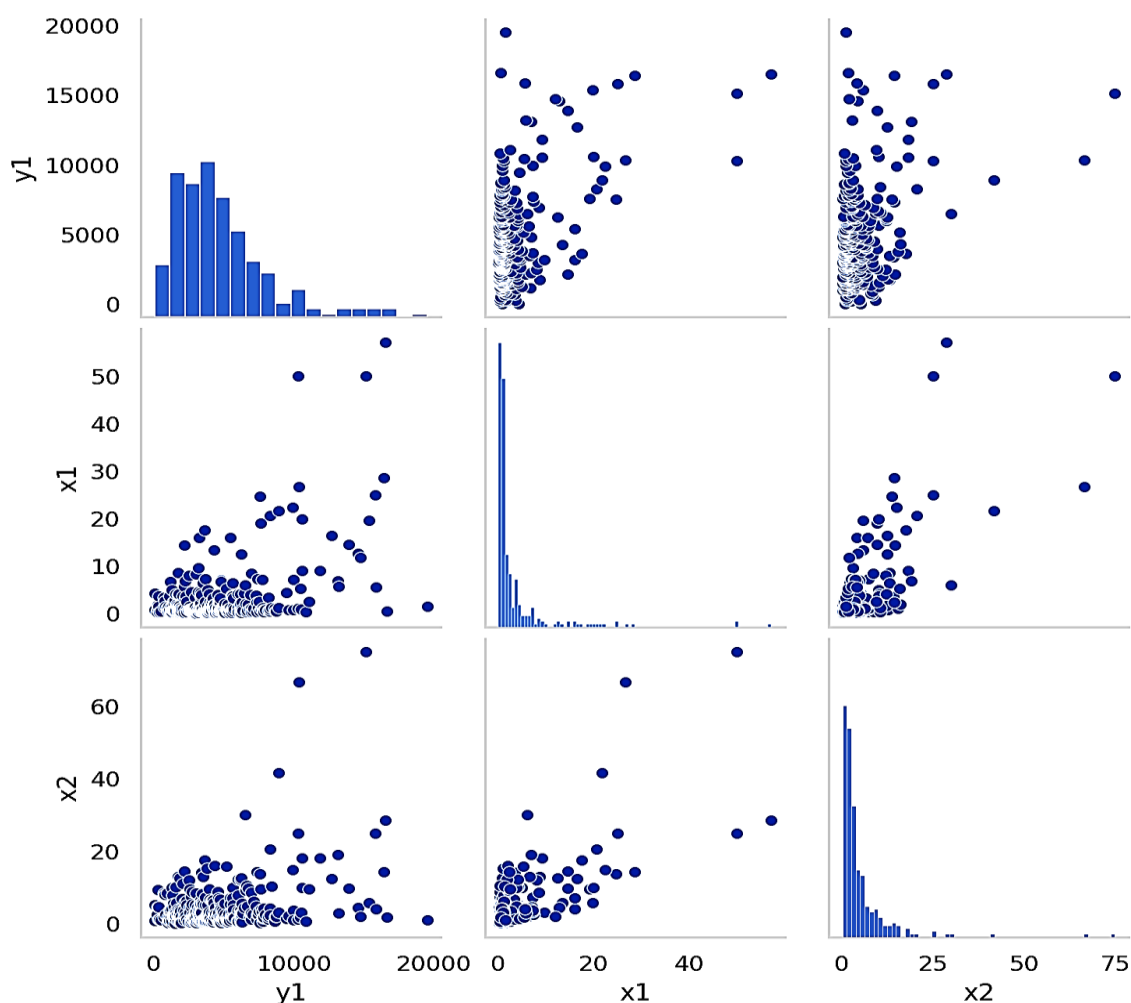


Рис. 2. Диаграммы рассеяния и гистограммы по изучаемым признакам

Источник: составлено авторами на основе данных отчетов [3, 6].

Оценка параметров линейной функции может быть проведена с помощью метода наименьших квадратов (МНК), для реализации которого необходимо выполнение предпосылок, в частности отсутствие гетероскедастичности в остатках [10]. Обычно данная предпосылка МНК нарушается в пространственных выборках, тогда как нарушение другой предпосылки МНК – об отсутствии автокорреляции в остатках – присуще в основном моделям, построенным по данным динамических рядов. Поскольку по графикам было предположено наличие гетероскедастичности, необходимо статистически проверить эту гипотезу.

Существует множество тестов на гетероскедастичность остатков, в представленной работе применен тест Уайта [8], основанный на зависимости, описываемой следующей функцией:

$$\hat{e}^2 = a_0 + b_1x + b_2x^2. \quad (4)$$

Зависимость (4) может быть использована в дальнейшем для устранения гетероскедастичности остатков.

Вначале выдвигаются нулевая и альтернативная гипотезы о наличии или отсутствии гетероскедастичности в остатках:

- H_0 – гетероскедастичность отсутствует;
- H_a – гетероскедастичность присутствует.

В работе при проверке статистических гипотез использовался критический уровень значимости 5%.

Определяются оценки параметров модели (4) с помощью МНК, проводится дисперсионный анализ с использованием F-критерия Фишера. Если будет принята альтернативная гипотеза, то зависимость дисперсии остатков (e^2) от значений фактора x может быть описана квадратичной функцией.

Если применить МНК к модели с гетероскедастичными остатками, то будут получены неэффективные оценки генеральных параметров, что может стать причиной недостоверности параметров, невозможности их интерпретации и использования модели в целях прогнозирования. Поэтому в случае гетероскедастичности остатков следует применять взвешенный МНК (ВМНК), «скорректировав» расчеты параметров уравнения регрессии с учетом значений ковариационной матрицы остатков [12].

Для основных расчетов в работе был использован язык программирования Python. С целью изучения тесноты связи между изучаемыми признаками построена матрица парных коэффициентов корреляции с использованием библиотек pandas, seaborn. Для оценки параметров уравнений регрессии с помощью МНК обычного метода наименьших квадратов применена функция OLS из библиотеки statsmodels, которой на вход подаются значения зависимой и независимой переменных.

Полученные модели были проверены на наличие гетероскедастичности остатков по тесту Уайта: на вход функции het_white библиотеки statsmodels переданы векторы остатков по МНК и значений независимой переменной. Для устранения гетероскедастичности использовался ВМНК, для этого на вход функции WLS библиотеки statsmodels были переданы векторы зависимой и независимой переменных, а также веса, которые были рассчитаны как обратные величины выравненных квадратов остатков, полученных по функции, использующейся в тесте Уайта.

Для решения задачи регрессии могут использоваться различные методы машинного обучения:

- нейронные сети и, в частности, модель многослойного персептрона (Multilayer Perceptron, MLP);
- метод k ближайших соседей (K-nearest Neighbors, KNN);
- градиентный бустинг (Gradient Boosting, GB) и др.

Реализация данных методов была осуществлена при помощи специализированной библиотеки scikit-learn для языка Python.

Оценки параметров уравнения регрессии (1) получены с помощью МНК:

$$\hat{y}_1 = 3941,9 + 241,1x_1.$$

Уравнение регрессии в целом и каждый из его параметров статистически значимы на уровне 5%, о чем свидетельствуют следующие значения:

- p -value критериев F-Фишера (p -value (F) = 0,000);
- t -Стьюдента (p -value (t_a) = 0,000) (см. табл.).

Что касается теста Уайта, то он показал наличие гетероскедастичности в остатках (p -value (F) = 0,008).

Линейные модели регрессии на основе методов оценивания OLS и WLS

Метод оценивания	Модель	Параметры уравнения		Тест Уайта на гетероскедастичность остатков		Оценка достоверности				R^2
						уравнения регрессии		условного начала		
		a	b	F	p -value (F)	F	p -value (F)	t_a	p -value (t_a)	
OLS	1	3941,9	241,1	4,97	0,008	95,9	0,000	21,4	0,000	0,233
WLS	1	3958,4	223,0	0,08	0,923	91,3	0,000	23,6	0,000	0,225
OLS	2	4015,4	155,3	2,49	0,084	41,8	0,000	19,1	0,000	0,117
OLS	3	26,5	0,6	0,19	0,830	38,9	0,000	24,3	0,000	0,153

Источник: составлено авторами на основе данных отчетов [3, 6].

В результате был реализован ВМНК: $\hat{y}_1 = 3958,4 + 223,0x_1$, его оценки являются предпочтительными, поскольку тест Уайта не показал наличие гетероскедастичности остатков (p -value (F) = 0,923). Уравнение в целом и его параметры достоверны при уровне критической значимости 5%.

По первому уравнению регрессии можно сделать вывод о том, что увеличение численности работников в К(Ф)Х на 1 чел. в расчете на 100 га земельных угодий приведет к росту доходов от реализации продукции сельского хозяйства в среднем на 223 тыс. руб. в расчете на 100 га земельных участков.

С помощью полученного уравнения регрессии можно выявить резервы роста для хозяйств, у которых фактические значения доходов от реализации продукции не превысили их выравненные значения по уравнению (рис. 3, а). Ниже линии регрессии расположены К(Ф)Х, у которых при имеющихся трудовых ресурсах есть возможность увеличить доходы от реализации продукции. Было определено 180 таких хозяйств, средний резерв роста доходов которых составлял 1892,6 тыс. руб. в расчете на 100 га земельных угодий.

Уравнение регрессии при той же результативной переменной и числе сельскохозяйственной техники в расчете на 100 га площади земли в качестве фактора (2) имеет следующий вид: $\hat{y}_1 = 4015,4 + 155,3x_2$, гетероскедастичность в данной модели не была обнаружена. Коэффициент полной регрессии свидетельствует о том, что при увеличении числа сельскохозяйственной техники на единицу доходы возрастут в среднем на 155,3 тыс. руб. (в расчете на 100 га земельных угодий по каждой переменной).

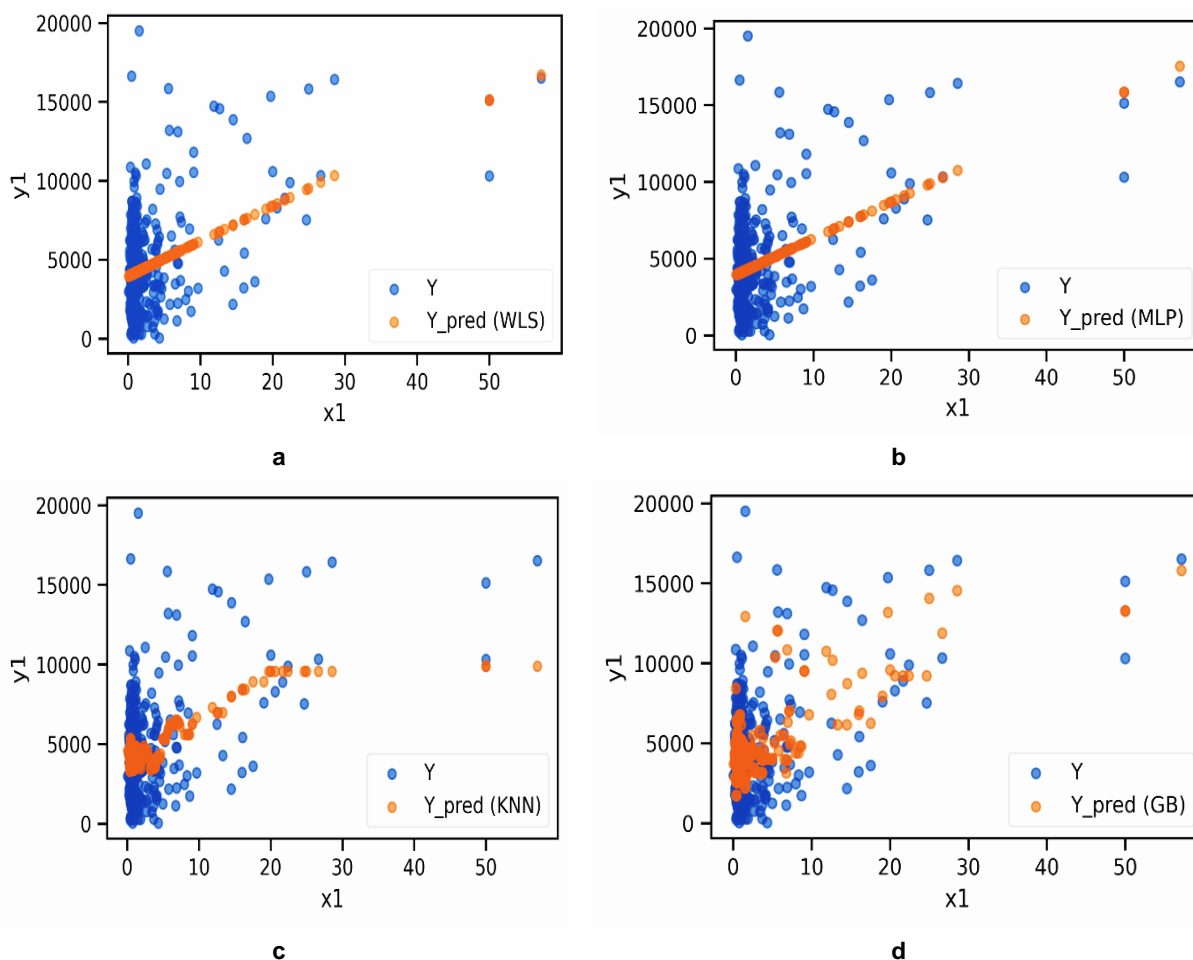


Рис. 3. Фактические (Y) и выравненные значения (Y_{pred}) доходов К(Ф)Х (y_1) в зависимости от уровня обеспеченности рабочей силой (x_1), полученные разными методами: а – WLS; б – MLP; в – KNN; д – GB

Источник: составлено авторами на основе данных отчетов [3, 6].

При моделировании урожайности зерновых и зернобобовых культур (3) получено уравнение регрессии также с помощью МНК: $\hat{y}_2 = 26,5 + 0,6x_3$. Уравнение регрессии в целом и все параметры статистически значимы. Коэффициент полной регрессии показывает, что с увеличением затрат на минеральные удобрения на 1 тыс. руб. в расчете на 1 га убранной площади урожайность зерновых и зернобобовых культур увеличится в среднем на 0,6 ц/га.

В таблице мы можем наблюдать значения коэффициента детерминации R^2 как метрики качества для каждого метода определения оценок уравнения линейной регрессии. Стоит отметить, что R^2 для WLS завышается в силу особенности метода, поэтому его нельзя использовать как показатель качества модели.

Для сравнения качества моделирования значений результативных переменных была построена и использована модель MLP со следующими гиперпараметрами:

- 4 скрытых слоя с числом нейронов соответственно 150, 100, 100, 100;
- число эпох обучения – 1000;
- функция активации – rectified linear unit function (ReLU);
- метод корректировки весов – stochastic gradient-based optimizer (Adam);
- начальный темп обучения – 0,01.

Значения гиперпараметров были подобраны эмпирическим путем.

Для первой модели метод MLP позволил получить наибольший коэффициент детерминации ($R^2 = 0,242$) (рис. 3, *b*).

Для второй модели данный метод не показал лучшего результата, коэффициент детерминации остался на уровне 0,117.

Для третьей модели методом MLP удалось повысить R^2 незначительно ($R^2 = 0,155$).

С использованием методов KNN и GB для всех трех моделей удалось получить более высокие значения R^2 , что говорит об их более высокой предсказательной силе в сравнении с методом OLS (результаты моделирования приведены на рисунке 3, *c, d*). Так, по методу KNN коэффициент детерминации для трех моделей соответственно равен 0,237, 0,161, 0,189, по методу GB – соответственно 0,627, 0,551, 0,669.

Таким образом, для первой модели регрессии рекомендуется использовать метод WLS ввиду гетероскедастичности остатков либо методы GB, MLP. Для моделей 2 и 3 наиболее предпочтителен метод OLS или метод GB.

Полученные модели можно использовать в целях прогнозирования и для классификации К(Ф)Х с целью принятия качественных управленческих решений на государственном уровне по развитию малого предпринимательства.

Тема статистического анализа результатов деятельности крестьянских (фермерских) хозяйств в России не так широко распространена в среде научно-исследовательских организаций и ученых, в том числе в силу ограниченного доступа к информации.

Аналізу состояния и динамики развития К(Ф)Х в России, оценке основных факторов эффективности посвящены работы сотрудников Института сельского хозяйства РАН (Ушачев И.Г., Бондаренко Л.В., Чекалин В.С. [9]) и Национального исследовательского университета «Высшая школа экономики» (руководитель авторского коллектива Гохберг Л.М. [5]).

Исследования, связанные с эконометрическим моделированием доходов домохозяйств, проводят сотрудники различных научных учреждений за рубежом [17]. Так, ученые из Малайзии (Yee G.P., Rusiman M.S., Ismail S., Suparman S., Hamzah F.M. & Shafi M.A.) изучают зависимость валового совокупного дохода домохозяйств от географических и демографических характеристик с использованием модели множественной линейной регрессии с комбинацией метода К-средних для предварительной кластеризации домохозяйств на однородные группы по размеру доходов. Полученные результаты позволяют правительству страны проводить гибкую политику, направленную на повышение уровня жизни населения, проживающего в сельской местности, и принимать взвешенные решения для оптимизации экономической ситуации в стране [17].

Работы многих исследователей посвящены проблеме гетероскедастичности. Так, коллектив ученых из Мельбурнского университета и Австралийского медицинского исследовательского института (WEHI) разрабатывает моделирование групповой гетероскедастичности в псевдообъемных данных одноклеточной РНК-секвенции (You Y., Dong X., Wee Y.K., Maxwell M.J., Alhamdoosh M., Smyth G.K., Hickey P.F., Ritchie M.E. & Law C.W.) [19]. Проведение тестов на гетероскедастичность в непараметрической регрессии на основе анализа остатков, включая методы Детте и Мунка, было осуществлено сотрудниками научной школы Сианьского университета (Китай) (Zhang L. и Mei C.) [20].

Эффективность модели многослойного персептрона для решения задачи регрессии была доказана в научных исследованиях, связанных с прогнозированием краткосрочных тенденций фондового рынка (Namdari A., Durrani T.S.) [15], оценкой содержания металлических элементов в почве (Li F., Zhang X., Lu A., Xu L., Ren D. & You T.) [14], изучением эрозии почвы и подверженности оползням (Egbueri J.C.) [11]. Исследованию методов k -ближайших соседей и градиентного бустинга посвящены научные публикации, связанные с прогнозированием производительности водно-болотных угодий (Lee B.-H., Scholz M.) [13], роста реального ВВП (Yoon J.) [18] и др.

Проведенное авторами исследование крестьянских (фермерских) хозяйств по данным формы № 1-КФХ может быть продолжено в направлении построения классификации и описания типов К(Ф)Х с помощью системы показателей специализации [16], интенсификации, размеров и др., а также в направлении построения типологических регрессий.

Выводы

На уровень производства продукции сельского хозяйства субъектов малого предпринимательства, в том числе К(Ф)Х, оказывает влияние множество факторов, одними из приоритетных являются уровень обеспеченности рабочей силой и сельскохозяйственной техникой, между которыми была установлена тесная связь, препятствующая их одновременному включению в модель регрессии.

По изучаемой совокупности К(Ф)Х выявлена правосторонняя асимметрия по всем изучаемым признакам, показывающая, что большинство К(Ф)Х характеризуется невысокими уровнями доходов от реализации сельскохозяйственной продукции собственного производства, продуктов первичной и промышленной переработки, обеспеченности рабочей силой и сельскохозяйственной техникой в расчете на единицу земельной площади. С помощью парных коэффициентов корреляции Пирсона выявлена умеренная сила связи уровня доходов от реализации сельскохозяйственной продукции с обеспеченностью рабочей силой ($R = 0,483$) и слабая – с обеспеченностью сельскохозяйственной техникой ($R = 0,342$); между урожайностью зерновых и зернобобовых культур и затратами на минеральные удобрения и средства защиты растений рассчитан коэффициент корреляции, равный 0,391.

Для оценки параметров моделей регрессии использовали метод наименьших квадратов (МНК) – в случае выполнения условий применения метода и взвешенный метод наименьших квадратов (ВМНК) – при нарушении предпосылки о гомоскедастичности остатков. Из трех анализируемых регрессионных моделей при помощи теста Уайта выявлена гетероскедастичность остатков в модели зависимости доходов от реализации сельскохозяйственной продукции от среднегодовой численности наемных работников, которая была устранена в результате применения ВМНК. В модели зависимости доходов от реализации сельскохозяйственной продукции от обеспеченности сельскохозяйственной техникой и в модели зависимости урожайности зерновых и зернобобовых от затрат на минеральные удобрения и средства защиты растений остатки гомоскедастичны.

Построенные эконометрические модели позволяют оценить изменения эндогенных переменных в результате влияния факторов. Так, увеличение численности работников в $K(\Phi)X$ на 1 чел. в расчете на 100 га земельных угодий приведет к росту доходов от реализации продукции сельского хозяйства в среднем на 223 тыс. руб. в расчете на ту же единицу площади, при этом около половины хозяйств имеют средний резерв роста доходов от увеличения трудовых ресурсов в размере 1892,6 тыс.руб. в расчете на 100 га земельных угодий.

Установлено также, что при увеличении числа сельскохозяйственной техники на единицу доходы возрастут в среднем на 155,3 тыс. руб. (в расчете на 100 га земельных угодий по каждой переменной), а с увеличением затрат на минеральные удобрения на 1 тыс. руб. в расчете 1 га убранной площади урожайность зерновых и зернобобовых культур увеличится в среднем на 0,6 ц/га.

С точки зрения качества аппроксимации, метод наименьших квадратов сравнивался с многослойным перцептроном (MLP), методом k ближайших соседей (KNN), которые не показали существенно лучший результат, и методом градиентного бустинга (GB), позволившего получить наибольший коэффициент детерминации.

Таким образом, в рамках изучаемой совокупности данных, для увеличения качества аппроксимации функции уровня производства продукции сельского хозяйства $K(\Phi)X$ в зависимости от уровня обеспеченности рабочей силой и сельскохозяйственной техникой, наиболее эффективным методом является градиентный бустинг. Однако МНК позволяет получить коэффициенты регрессии и измерить на их основе меру влияния факторов на зависимую переменную, что делает данный метод высоко востребованным в эконометрических исследованиях.

Список источников

1. Национальные счета [Электронный ресурс] // Федеральная служба государственной статистики (Росстат). Официальный сайт. URL: <https://rosstat.gov.ru/statistics/accounts> (дата обращения: 01.03.2023).
2. О крестьянском (фермерском) хозяйстве: Федеральный закон от 11 июня 2003 г. № 74-ФЗ [Электронный ресурс]. URL: <http://www.kremlin.ru/acts/bank/19605> (дата обращения: 01.08.2023).
3. Об утверждении формы отчета о финансово-экономическом состоянии товаропроизводителей агропромышленного комплекса за 2021 год и срока его представления: Приказ Министерства сельского хозяйства Российской Федерации от 13 июля 2021 г. № 465 [Электронный ресурс]. URL: <http://publication.pravo.gov.ru/Document/View/0001202108300007> (дата обращения: 01.03.2023).
4. О Государственной программе развития сельского хозяйства и регулирования рынков сельскохозяйственной продукции, сырья и продовольствия на 2008–2012 годы: Постановление Правительства Российской Федерации от 14 июля 2007 г. № 446 (с последующими изменениями и дополнениями) [Электронный ресурс]. URL: <https://base.garant.ru/2162858/> (дата обращения: 18.03.2023).
5. Прогноз научно-технологического развития агропромышленного комплекса Российской Федерации на период до 2030 года: утвержден приказом Минсельхоза России № 3 от 12 января 2017 г. Министерство сельского хозяйства Российской Федерации, Национальный исследовательский университет «Высшая школа экономики»; Л.М. Гохберг (руководитель авторского коллектива и др.]. Москва: НИУ ВШЭ, 2017. 139 с.
6. Формы отчетов $K(\Phi)X$ и СПок [Электронный ресурс] // Управление сельского хозяйства Липецкой области. Официальный сайт. URL: https://ush48.ru/industries/small_forms_of_management/formy-otchetov-kfkh-i-spok/ (дата обращения: 01.03.2023).
7. Уколова А.В., Быков Д.В. Типизация личных подсобных хозяйств методом нейросетевого кластерного анализа // Экономика сельского хозяйства России. 2023. № 6. С. 97–108. DOI: 10.32651/236-97.
8. Уколова А.В., Ульянов А.Е. Статистический анализ результатов сельскохозяйственной переписи 2017 года по типам ферм США // Экономика сельского хозяйства России. 2023. № 1. С. 109–118. DOI: 10.32651/231-109.
9. Ушачев И.Г., Бондаренко Л.В., Чекалин В.С. Основные направления комплексного развития сельских территорий России // Вестник Российской академии наук. 2021. № 91(4). С. 316–325. DOI: 10.31857/S0869587321040113.
10. Drichoutis A.C., Lazaridis P., Nayga R.M. Heteroskedasticity, the single crossing property and ordered response models // Economics Bulletin. 2006. Vol. 3(31). Pp. 1–6.
11. Egbueri J.C. Soil erosion and landslide susceptibility insights based on hierarchical clustering and multilayer perceptron networks: a Nigerian case study // International Journal of Environmental Science and Technology. 2023. Vol. 20. Pp. 10763–10786. DOI: 10.1007/s13762-022-04714-7.

12. Griffiths W.E., Carter Hill R., Lim G.C. Using EViews for Principles of Econometrics. New York: John Wiley & Sons, Inc, 2012. 466 p.
13. Lee B.-H., Scholz M. A comparative study: Prediction of constructed treatment wetland performance with k-nearest neighbors and neural networks // *Water, Air, and Soil Pollution*. 2006. Vol. 174(1-4). Pp. 279–301. DOI: 10.1007/s11270-006-9113-2.
14. Li F., Zhang X., Lu A. et al. Estimation of metal elements content in soil using x-ray fluorescence based on multilayer perceptron // *Environmental Monitoring and Assessment*. 2022. Vol. 194. Article no. 95. DOI: 10.1007/s10661-022-09750-x.
15. Namdari A., Durrani T.S. A multilayer feedforward perceptron model in neural networks for predicting stock market short-term trends // *SN Operations Research Forum*. 2021. Vol. 2. Article no. 38. DOI: 10.1007/s43069-021-00071-2.
16. Ukolova A.V., Dashieva B.S. Study of the labor resources of peasant (farm) households by production type // In: *Environmental Footprints and Eco-Design of Products and Processes*. Singapore: Springer, 2022. Pp. 229–241.
17. Yee G.P., Rusiman M.S., Ismail S. et al. K-means clustering analysis and multiple linear regression model on household income in Malaysia // *IAES International Journal of Artificial Intelligence (IJ-AI)*. 2023. Vol. 12(2). Pp. 731–738. DOI: 10.11591/ijai.v12.i2.pp731-738.
18. Yoon J. Forecasting of real GDP growth using machine learning models: Gradient Boosting and Random Forest Approach // *Computational Economics*. 2021. Vol. 57. Pp. 247–265. DOI: 10.1007/s10614-020-10054-w.
19. You Y., Dong X., Wee Y.K. et al. Modeling group heteroscedasticity in single-cell RNA-seq pseudo-bulk data // *Genome Biology*. 2023. Vol. 24. Article no. 107. DOI: 10.1186/s13059-023-02949-2.
20. Zhang L., Mei C. Testing heteroscedasticity in nonparametric regression models based on residual analysis // *Applied Mathematics – A Journal of Chinese Universities*. 2008. Vol. 23(3). Pp. 265–272. DOI: 10.1007/s11766-008-1648-0.

References

1. Natsional'nye scheta. Federal'naya sluzhba gosudarstvennoj statistiki (Rosstat). Ofitsial'nyj sajt [National accounts. Federal State Statistics Service (Rosstat). Official website]. URL: <https://rosstat.gov.ru/statistics/accounts>. (In Russ.).
2. O krest'yanskom (fermerskom) khozyajstve: Federal'nyj zakon ot 11 iyunya 2003 g. № 74-FZ [Concerning Peasant Private Farm Holdings: Federal Law No. 74-FZ of June 11, 2003]. URL: <http://www.kremlin.ru/acts/bank/19605>. (In Russ.).
3. Ob utverzhdenii formy otcheta o finansovo-ekonomicheskom sostoyanii tovaroproizvoditelej agropromyshlennogo kompleksa za 2021 god i sroka ego predstavleniya: Prikaz Ministerstva sel'skogo khozyajstva Rossijskoj Federatsii ot 13 iyulya 2021 g. № 465 [On approval of the format of the Report on the financial and economic status of agricultural producers for 2021 and Report provision form: Order of the Ministry of Agriculture of the Russian Federation No. 465 of July 13, 2021]. URL: <http://publication.pravo.gov.ru/Document/View/0001202108300007>. (In Russ.).
4. O Gosudarstvennoj programme razvitiya sel'skogo khozyajstva i regulirovaniya rynkov sel'skokhozyajstvennoj produkcii, syr'ya i prodovol'stviya na 2008-2012 gody: Postanovlenie Pravitel'stva Rossijskoj Federatsii ot 14 iyulya 2007 g. № 446 (s posleduyuschimi izmeneniyami i dopolneniyami) [On the State Program for the Development of Agriculture and Regulation of Agricultural Products, Raw Materials and Food Markets for 2008-2012: Decree of the Government of the Russian Federation No. 446 of July 14, 2007 (as subsequently amended)]. URL: <https://base.garant.ru/2162858/>. (In Russ.).
5. Prognoz nauchno-tekhnologicheskogo razvitiya agropromyshlennogo kompleksa Rossijskoj Federatsii na period do 2030 goda: utverzhen prikazom Minzel'khoza Rossii № 3 ot 12 yanvarya 2017 g. Ministerstvo sel'skogo khozyajstva Rossijskoj Federatsii, Natsional'nyj issledovatel'skij universitet "Vysshaya Shkola Ekonomiki"; L.M. Gohberg (rukovoditel' avtorskogo kolektiva et al.) [Forecast of scientific and technological development of Agro-Industrial Complex of the Russian Federation for the period up to 2030: approved by Order of the Ministry of Agriculture of the Russian Federation No. 3 of January 12, 2017. Ministry of Agriculture of the Russian Federation, National Research University "Higher School of Economics"; L.M. Gohberg (Head of the authoring team et al.)]. Moscow: Higher School of Economics Press; 2017. 139 p. (In Russ.).
6. Formy otchetov K(F)Kh i SPoK. Upravlenie sel'skogo khozyajstva Lipetskoj oblasti. Ofitsial'nyj sajt [Formats of Reports of Peasant Private Farm Holdings and Agricultural Consumer Co-operatives. Department of Agriculture of Lipetsk Oblast. Official website]. URL: https://ush48.ru/industries/small_forms_of_management/formy-otchetov-kfkh-i-spok/. (In Russ.).
7. Ukolova A.V., Bykov D.V. Tipizatsiya lichnykh podsobnykh khozyajstv metodom nejrosetevogo klaster'nogo analiza [Typing of household plots by the method of neural network cluster analysis]. *Ekonomika sel'skogo khozyajstva Rossii = Economics of Agriculture of Russia*. 2023;6:97-108. DOI: 10.32651/236-97. (In Russ.).
8. Ukolova A.V., Ulyanckin A.E. Statisticheskij analiz rezul'tatov sel'skokhozyajstvennoj perepisi 2017 goda po tipam ferm SShA [Statistical analysis of the results of the 2017 Agricultural Census by types of farms in the USA]. *Ekonomika sel'skogo khozyajstva Rossii = Economics of Agriculture of Russia*. 2023;1:109-118. DOI: 10.32651/231-109. (In Russ.).
9. Ushachev I.G., Bondarenko L.V., Chekalin V.S. Osnovnye napravleniya kompleksnogo razvitiya sel'skikh territorij Rossii [Main directions of integrated development of rural areas of Russia]. *Vestnik Rossijskoj akademii nauk = Bulletin of the Russian Academy of Sciences*. 2021;91(4):316-325. DOI: 10.31857/S0869587321040113. (In Russ.).

10. Drichoutis A.C., Lazaridis P., Nayga R.M. Heteroskedasticity, the single crossing property and ordered response models. *Economics Bulletin*. 2006;3(31):1-6.
11. Egbueri J.C. Soil erosion and landslide susceptibility insights based on hierarchical clustering and multilayer perceptron networks: a Nigerian case study. *International Journal of Environmental Science and Technology*. 2023;20:10763-10786. DOI: 10.1007/s13762-022-04714-7.
12. Griffiths W.E., Carter Hill R., Lim G.C. Using EViews for Principles of Econometrics. New York: John Wiley & Sons, Inc; 2012. 466 p.
13. Lee B.-H., Scholz M. A comparative study: Prediction of constructed treatment wetland performance with k-nearest neighbors and neural networks. *Water, Air, and Soil Pollution*. 2006;174(1-4):279-301. DOI: 10.1007/s11270-006-9113-2.
14. Li F., Zhang X., Lu A. et al. Estimation of metal elements content in soil using x-ray fluorescence based on multilayer perceptron. *Environmental Monitoring and Assessment*. 2022;194:95. DOI: 10.1007/s10661-022-09750-x.
15. Namdari A., Durrani T.S. A multilayer feedforward perceptron model in neural networks for predicting stock market short-term trends. *SN Operations Research Forum*. 2021;2:38. DOI: 10.1007/s43069-021-00071-2.
16. Ukolova A.V., Dashieva B.S. Study of the labor resources of peasant (farm) households by production type // In: Environmental Footprints and Eco-Design of Products and Processes. Singapore: Springer; 2022:229-241.
17. Yee G.P., Rusiman M.S., Ismail S. et al. K-means clustering analysis and multiple linear regression model on household income in Malaysia. *IAES International Journal of Artificial Intelligence (IJ-AI)*. 2023;12(2):731-738. DOI: 10.11591/ijai.v12.i2.pp731-738.
18. Yoon J. Forecasting of real GDP growth using machine learning models: Gradient Boosting and Random Forest Approach. *Computational Economics*. 2021;57:247-265. DOI: 10.1007/s10614-020-10054-w.
19. You Y., Dong X., Wee Y.K. et al. Modeling group heteroscedasticity in single-cell RNA-seq pseudo-bulk data. *Genome Biology*. 2023;24:107. DOI: 10.1186/s13059-023-02949-2.
20. Zhang L., Mei C. Testing heteroscedasticity in nonparametric regression models based on residual analysis. *Applied Mathematics – A Journal of Chinese Universities*. 2008;23(3):265-272. DOI: 10.1007/s11766-008-1648-0.

Информация об авторах

А.В. Уколова – кандидат экономических наук, доцент, и.о. зав. кафедрой статистики и кибернетики ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А.Тимирязева», <https://orcid.org/0000-0002-2806-6365>, statmsha@rgau-msha.ru.

Б.Ш. Дашиева – кандидат экономических наук, доцент кафедры статистики и кибернетики ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А.Тимирязева», <https://orcid.org/0000-0002-3155-6517>, dashieva.b.sh@rgau-msha.ru.

Д.В. Быков – ассистент кафедры статистики и кибернетики ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А.Тимирязева», <https://orcid.org/0000-0001-6287-3462>, bykovdv@rgau-msha.ru.

А.Е. Ульянкин – ассистент кафедры статистики и кибернетики ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А.Тимирязева», <https://orcid.org/0000-0003-2978-5473>, aeulianckin@rgau-msha.ru.

Д.Э. Храмов – ассистент кафедры статистики и кибернетики ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А.Тимирязева», <https://orcid.org/0009-0009-7873-7030>, khramovde@rgau-msha.ru.

Information about the authors

A.V. Ukolova, Candidate of Economic Sciences, Docent, Acting Head of the Dept. of Statistics and Cybernetics, Russian Timiryazev State Agrarian University, <https://orcid.org/0000-0002-2806-6365>, statmsha@rgau-msha.ru.

B.Sh. Dashieva, Candidate of Economic Sciences, Docent, the Dept. of Statistics and Cybernetics, Russian Timiryazev State Agrarian University, <https://orcid.org/0000-0002-3155-6517>, dashieva.b.sh@rgau-msha.ru.

D.V. Bykov, Assistant, the Dept. of Statistics and Cybernetics, Russian Timiryazev State Agrarian University, <https://orcid.org/0000-0001-6287-3462>, bykovdv@rgau-msha.ru.

A.E. Uliankin, Assistant, the Dept. of Statistics and Cybernetics, Russian Timiryazev State Agrarian University, <https://orcid.org/0000-0003-2978-5473>, aeulianckin@rgau-msha.ru.

D.E. Khramov, Assistant, the Dept. of Statistics and Cybernetics, Russian Timiryazev State Agrarian University, <https://orcid.org/0009-0009-7873-7030>, khramovde@rgau-msha.ru.

Статья поступила в редакцию 20.06.2023; одобрена после рецензирования 28.07.2023; принята к публикации 01.08.2023.

The article was submitted 20.06.2023; approved after reviewing 28.07.2023; accepted for publication 01.08.2023.

© Уколова А.В., Дашиева Б.Ш., Быков Д.В., Ульянкин А.Е., Храмов Д.Э., 2023