

Современные направления прогнозирования урожайности сельскохозяйственных культур на основе использования эконометрических моделей

Марина Юрьевна Архипова,

Артем Игоревич Смирнов

Национальный исследовательский университет «Высшая школа экономики», г. Москва, Россия

Сельское хозяйство является одной из важнейших отраслей экономики и основным поставщиком продуктов питания и сырья для многих отраслей промышленности. Сельскохозяйственный сектор России в последнее время переживает обновление и подъем благодаря интенсификации и применению современных инновационных технологий, контролю за состоянием полей с помощью космических фотоснимков на основе систем компьютерного зрения. Вместе с тем, остается еще широкий пласт задач, требующий оперативного решения. Одной из таких задач является разработка новых моделей и методов, позволяющих прогнозировать основные результирующие показатели развития сельского хозяйства и обладающих преимуществом по сравнению с существующими моделями. Для повышения точности прогнозных моделей необходимо опираться на широкий спектр доступных статистических показателей и новый современный эконометрический инструментарий.

В статье представлен комплекс методических разработок построения моделей урожайности сельскохозяйственных культур на основе использования новых эконометрических моделей, работающих по урезанной выборке (не включающей область возможных отрицательных значений), статистических оценок применяемых показателей, в составе которых акцентируется внимание на экологической компоненте, а также структурных и общеэкономических индикаторах. Предлагаемые модели позволяют получать более точные прогнозы по сравнению с традиционными популярными моделями, основанными на методе наименьших квадратов. Работа опирается на данные Росстата по 100 сельскохозяйственным полям, расположенным в муниципальных образованиях 43 регионов России, выбранных пропорционально объему продукции растениеводства данного региона.

Результаты исследования представляют интерес для международных и российских организаций различного уровня, деятельность которых связана как с вопросами принятия управленческих решений, направленных на обеспечение продовольственной безопасности страны, повышение уровня и качества жизни населения, так и организаций, призванных на местах обеспечивать современные условия ведения сельского хозяйства.

Ключевые слова: прогнозирование, эконометрическое моделирование, урожайность сельскохозяйственных культур, экологические факторы, сельскохозяйственный анализ.

JEL: C50, O13.

doi: <https://doi.org/10.34023/2313-6383-2020-27-5-65-75>.

Для цитирования: Архипова М.Ю., Смирнов А.И. Современные направления прогнозирования урожайности сельскохозяйственных культур на основе использования эконометрических моделей. Вопросы статистики. 2020;27(5):65-75.

Current Trends in Crop Yield Forecasting Based on the Use of Econometric Models

Marina Yu. Arkhipova,

Artem I. Smirnov

National Research University Higher School of Economics, Moscow, Russia

Agriculture is one of the most important branches of the national economy and the main supplier of food and raw materials for many industries. Agricultural sector in Russia has recently been undergoing renewal and growth due to the intensification and application of modern innovative technologies for monitoring the state of fields using satellite images based on computer vision systems. At the same time, there is still a number of problems and challenges that require prompt solutions. One of them is developing new forecasting models and methods for key resulting indicators of agricultural development and have an advantage over existing models. To improve the accuracy of forecasting models, it is necessary to rely on a broad range of available statistical indicators and new modern econometric tools.

The paper presents a set of methodological developments for modeling and forecasting crop yields based on the use of new econometric models that allow working with a truncated regression by limiting the range of possible negative values, statistical estimations of the introduced indicators that focus on the ecological component, as well as structural and general economic indicators. The suggested models allow obtaining

more accurate forecasts compared to traditional popular models based on the least squares method. The work relies on Rosstat data for 100 agricultural fields located in municipalities of 43 regions of Russia, selected in proportion to the volume of crop production in this region.

The results of this study are of interest to international and Russian organizations of various levels, whose activities are related to the issues of making managerial decisions aimed at ensuring food security of the country, improving the level and quality of life of the population, as well as organizations designed to provide modern conditions for farming on the ground.

Keywords: forecasting, econometric modeling, crop yield, environmental factors, agricultural analysis.

JEL: C50, O13.

doi: <https://doi.org/10.34023/2313-6383-2020-27-5-65-75>.

For citation: Arkhipova M.Yu., Smirnov A.I. Current Trends in Crop Yield Forecasting Based on the Use of Econometric Models. *Voprosy Statistiki*. 2020;27(5):65-75. (In Russ.)

Введение

Сельское хозяйство является важной частью российской экономики, составляя по данным компании Deloitte в 2018 г. порядка 4,1% валового внутреннего продукта (ВВП)¹, и, согласно данным ООН за 2017 г., сектором занятости около 6,7% российской рабочей силы.

Однако на сегодняшний день не все проекты развития сельского хозяйства реализованы в России и странах мира. Нет единых баз данных, включая спутниковые изображения, позволяющие получать полную информацию о состоянии полей; не выработано единого мнения о факторах, оказывающих статистически значимое влияние на урожайность сельскохозяйственных культур, нет понимания влияния экологической ситуации и благосостояния жителей муниципальных образований на основные характеристики результативности сельскохозяйственного производства. Указанные причины не позволяют проводить комплексные исследования уровня развития и состояния сельского хозяйства, необходимые для принятия управленческих решений на различных уровнях иерархии, включая инвесторов в области сельского хозяйства и небольших фермеров.

Цель исследования заключалась в выделении факторов, оказывающих статистически значимое влияние на урожайность сельскохозяйственных культур, и их использовании для моделирования и прогнозирования урожайности полей в муниципальных образованиях (МО) регионов России.

В работе тестировались следующие гипотезы.

H1: Оказывает ли положительное влияние на урожайность сельскохозяйственных культур благосостояние населения муниципального образования (МО)?

H2: Действительно ли ухудшение экологических факторов (загрязнение воздуха, почвы и воды) в равной степени отрицательно влияет на урожайность сельскохозяйственных культур?

H3: Влияет ли увеличение численности населения МО на урожайность сельскохозяйственных культур?

H4: Повышение оснащённости МО тракторами, внесение минеральных удобрений на гектар засеянной территории положительно сказывается на урожайности сельскохозяйственных культур и характеризует эффективность использования материальных и трудовых вложений в плодородную почву и систему посадки и уборки культур.

Для реализации задач исследования и тестирования выдвинутых гипотез на первом этапе был проведен обзор современных разработок с целью, во-первых, выявления комплексной системы показателей, которые могут быть использованы для моделирования и прогнозирования урожайности сельскохозяйственных культур; во-вторых, выбора класса моделей, которые могут быть положены в основу алгоритма прогнозирования урожайности сельскохозяйственных культур.

Обзор современных исследований

Первые исследования в области анализа и моделирования урожайности сельскохозяйственных культур на основе использования многомерного статистического инструментария появились в середине прошлого века, хотя использование новых неагрономических факторов в анализе урожайности сельскохозяйственных культур началось в 21-м веке в связи с значительным развитием компьютерных технологий.

Среди первых работ такого класса можно выделить работу [1], проведенную по данным южной

¹ URL: <https://ru.investinrussia.com/data/file/obzor-rynka-selskogo-hozyajstva-2019.pdf>.

Италии. При разработке модели урожайности сельскохозяйственных культур ученые предложили учитывать такие факторы, как качество почвы, растительность, климат и менеджмент районов. Также были сделаны первые попытки использования космических фотоснимков полей для численной усредненной оценки растительного покрова и плотности почвы.

Лука Салвати в работе [2] одним из первых предложил использовать для прогнозирования урожайности сельскохозяйственных культур (в евро на гектар) экономические и социальные данные о муниципалитетах. В его работе в качестве факторов были использованы такие показатели, как доля сельского населения в общей численности жителей, доля растениеводства в сельскохозяйственном продукте и продуктивность садоводства.

Исследования, проведенные авторами в работах [3 и 4], дополнили набор значимых переменных при моделировании урожайности сельскохозяйственных культур за счет введения факторов, отвечающих за благосостояние населения, тип почвы, температуру воздуха и осадки, развитие биомассы растений во время вегетации [нормализованный относительный индекс растительности (NDVI)].

Использование статистического и эконометрического инструментария для анализа урожайности сельскохозяйственных культур можно найти в работах [5-16]. Так, например, в работе [5] авторы демонстрируют предсказательную силу нелинейной модели в течение нескольких лет. Исследователи отмечают, что в анализе прогнозирования урожайности сельскохозяйственных культур может существовать значительная нелинейная связь между некоторыми аграрными факторами. Такие модели, как нелинейные наименьшие квадраты (NLS), помогают выявлять эти зависимости.

Использование модели по урезанной выборке (truncated regression) и ее эффективность для прогнозирования случайных величин, ограниченных сверху или снизу, описано авторами в работе [1].

Таким образом, проведенный анализ современных исследований позволил определиться как с набором переменных, так и классом моделей, которые необходимо использовать для моделирования урожайности сельскохозяйственных культур. Следует отметить, что, не смотря на значимость выделенной проблемы для развития сельского хозяйства и обеспечения продовольственной безопасности страны, в России ощу-

щается недостаток исследований, посвященных выделенной проблематике.

Описание показателей и методика исследования

Для решения поставленных задач и проверки выдвинутых гипотез в работе использовались данные по 100 сельскохозяйственным полям, расположенным в муниципальных образованиях 43 регионов России. Отметим, что регионы России, так же как и количество наблюдений от каждого региона, выбирались пропорционально объему продукции растениеводства данного региона в 2018 г. (удельному весу продукции растениеводства в продукции сельского хозяйства по категориям хозяйств по субъектам РФ). Это обеспечило выбор регионов, в которых продукция растениеводства является важной частью сельского хозяйства и экономики региона в целом. Исследование было проведено по данным Росстата.

Для построения модели урожайности сельскохозяйственных культур, расположенных в муниципальных образованиях 43 регионов России, первоначально было отобрано 29 показателей, описывающих социально-экономические и природные характеристики района, в котором находится поле (см. Приложение). Все денежные показатели на этапе предварительного анализа были прологарифмированы [17]. Однако, в ходе предварительного анализа была установлена тесная связь и дублирование информации в ряде переменных. В результате было принято решение оставить для анализа одну зависимую (эндогенную) и 12 независимых (экзогенных) переменных:

Y_1 - независимая (эндогенная) переменная - урожайность в рублях на гектар засеянных территорий муниципального округа. Данную переменную исследователи часто включают в анализ в качестве зависимой переменной, так как она отражает продуктивность поля, как в физическом объеме, так и в денежном [2].

В качестве экзогенных переменных рассматривалась возможность включения в модель следующих показателей:

X_1 - доля растениеводства в сельском хозяйстве МО (%). Предполагается, что чем выше данный показатель, тем выгоднее растениеводство в регионе, больше инвестиций в агрокомплекс, следовательно, выше урожайность.

X_2 - доля посевной площади в общей площади МО. По аналогии с X_1 считается, что чем больше территорий района засеяно, тем выше урожайность.

$\log X_3$ - среднемесячная заработная плата в сельскохозяйственном секторе (ОКВЭД, раздел А). Предполагается, что чем выше данный показатель, тем более мотивированы работники в эффективном ведении хозяйства и, соответственно, выше урожайность полей.

X_4 - количество тракторов на гектар засеянной территории (единиц). Считается, что данный фактор важен и в современной России, где в некоторых регионах наблюдается недостаток оснащения агрокультурного комплекса современной техникой.

X_5 - внесение минеральных удобрений (в тоннах на гектар полей). Данный показатель говорит об уровне материальных и трудовых затрат в создании питательной среды для сельскохозяйственных культур.

$\log X_6$ - численность сельского населения на 1 января 2018 г. Предполагается, что при большей населенности района шире возможности специализации на рынке труда, следовательно, район с большим количеством населения имеет больше возможностей для оптимизации и выше урожайность полей. Также, вероятно, существует эффект большого местного рынка сбыта, что только увеличивает стимулы аграриев производить больше продукции.

X_7 - наличие или отсутствие загрязнения воздуха по критерию превышение предельно допустимой концентрации вредных веществ, рассчитанное Росгидрометом² (фиктивная переменная: 0 - отсутствие предельно допустимой концентрации вредных веществ, 1 - наличие). Согласно российскому законодательству³ установлено ПДК для целого ряда частиц, которые могут вредить как здоровью людей, быть опасными для экологии, так и серьезно мешать эффективному земледелию.

X_8 - количество серы на снежном покрытии (в граммах на квадратный километр). Считается, что недостаток, как и переизбыток серы в почве, может сильно повлиять на злаковые, одно из ключевых семейств культурных растений в сельском хозяйстве России [7].

X_9 - количество азота на снежном покрытии (в килограммах на квадратный километр). Исследователи-химики выявляют, что избыточное содержание азота приводит к увеличению периода вегетации, снижению урожая и его качества, так же, как и недостаток данного элемента. Как и в случае содержания серы - X_8 - недостаток элемента можно отнести к более низкому уровню контроля за почвой, а переизбыток, помимо того, к химическому почвенному загрязнению⁴.

X_{10} - уровень кислотности (рН) снежного покрытия. Вероятно, большинство культур хорошо растут в нейтральных или немного щелочных почвах, к умеренно-щелочным относят, к примеру, чернозем [9]. Существуют также растения, для выращивания которых предпочтительнее использовать подкисленную среду. Однако как излишняя кислотность, так и избыток щелочи вредит почве и всем культурным растениям, соответственно, как X_8 и X_9 может рассматриваться как фактор почвенного загрязнения⁵.

X_{11} - вывоз твердых коммунальных отходов за год (тысяч кубометров на душу населения). Данный фактор отражает эффективность сбора и вывоза бытового мусора, исходя из предпосылки, что в среднем россияне в сельской местности имеют одинаковый уровень выброса мусора, разница лишь в том, в какой мере местные власти справляются с его утилизацией. Считается, что твердые бытовые отходы малоприменимы в сельском хозяйстве⁶ и, мало того, почвы и растительность вблизи мусорных полигонов испытывают всевозможные негативные воздействия⁷.

X_{12} - качество воды водоемов суммарно по классам загрязнения 4 и 5, согласно методике подсчета

² Обзор состояния и загрязнения окружающей среды в Российской Федерации за 2018 год, Федеральная служба по гидрометеорологии и мониторингу окружающей среды (РОСГИДРОМЕТ), Москва, 2019.

³ Постановление Главного государственного санитарного врача Российской Федерации от 22 декабря 2017 года № 165, ГН 2.1.6.3492-17 «Предельно допустимые концентрации (ПДК) загрязняющих веществ в атмосферном воздухе городских и сельских поселений».

⁴ Курс лекций по дисциплине Б.1 В.ОД.1Агрохимия, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кубанский государственный аграрный университет», Краснодар, 2015.

⁵ Продовольственная и сельскохозяйственная организация ООН (FAO), серия учебных и практических пособий Yunga, E-ISBN 978-92-5-408434-9 (PDF).

⁶ Информационное телеграфное агентство России (ИТАР-ТАСС). URL: <https://tass.ru/obschestvo/3995514>.

⁷ Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО». URL: <http://u3a.itmo.ru/otxodyi-xozyajstvennoj.html>.

Гидрохимического Института. К 4 и 5 классам загрязнения поверхностных вод относятся разряды: грязная, очень грязная, экстремально грязная вода [14 и 15]. Классификация производится по величинам комбинаторного индекса загрязненности воды с учетом ряда дополнительных факторов. В работе использованы готовые индексы, предоставленные Гидрохимическим институтом⁸.

Для моделирования и прогнозирования урожайности сельскохозяйственных культур были рассмотрены следующие классы моделей:

1. Традиционные модели, основанные на методе наименьших квадратов (*OLS*). Выбор этой формы модели был определен исходя из ряда статей, посвященных прогнозированию урожайности сельскохозяйственных (например, [8, 13, 15]). В условиях мультиколлинеарности признакового пространства были рассмотрены также ридж-регрессия, ласо-регрессия (*LASSO, Least Absolute Shrinkage*), регрессия с наименьшими углами [*least-angle regression (LARS)*], метод *Elastic net* и др. Описание данного вида моделей и методов можно найти, например, в [18].

2. Нелинейные регрессионные модели. Описание данного вида моделей можно найти, например, в [18].

3. Регрессионная модель по урезанной выборке (*truncated regression*). Преимуществом данной модели является возможность избежать ошибки прогнозирования отрицательных значений, которые могут быть получены с использованием традиционных регрессионных моделей (*OLS*). Выбор данной модели опирался на результаты, полученные в исследовании [1]. Модель имеет нелинейную связь переменных тех наблюдений, которые близки к нулю. Числовые характеристики зависимой переменной в регрессии, оцениваемой по урезанной выборке, выглядят следующим образом:

$$\begin{aligned} Y_i &= x_i\beta + \varepsilon_i, \\ \varepsilon_i &\sim N(0, \sigma^2), Y_i > a, \\ Y_i &= x_i\beta + \xi_i, \\ E[Y_i | Y_i > a] &= x_i\beta + \sigma\lambda_i, \\ V[Y_i | Y_i > a] &= \sigma^2(1 - \gamma_i). \end{aligned}$$

Отличительной особенностью модели по урезанной выборке является невозможность

прямой интерпретации (объяснения) влияния переменных на результирующий признак, как это происходит в моделях, основанных на методе наименьших квадратов. Для интерпретации коэффициентов модели необходимо изучить средние маргинальные эффекты независимых переменных.

$$\widehat{Y}_i = x_i\widehat{\beta} + \widehat{\sigma}\widehat{\lambda}_i, \quad \widehat{ME}_{ij} = \widehat{\beta}_j(1 - \widehat{\gamma}_i).$$

Качество моделей проверялось за счет использования кросс проверки. Отобранные 100 наблюдений были разбиты на 2 части. В первую вошло 75 наблюдений (или три четвертых выборки), во вторую - 25 наблюдений (или одна четвертая выборки). Модели настраивались по первой части, в то время как вторая часть использовалась для проверки их качества и сопоставления предсказательной способности по сравнению другими моделями.

Модели и полученные результаты

Для построения модели для прогнозирования урожайности сельскохозяйственных культур были рассмотрены два класса моделей: традиционная регрессионная модель, основанная на методе наименьших квадратов (*OLS*) и регрессионная модель по урезанной выборке (*truncated regression*), которые были протестированы на данных Росстата за 2018 г. по 100 сельскохозяйственным полям, расположенным в муниципальных образованиях 43 регионов России.

Использование традиционных линейных регрессионных моделей (OLS).

Оценка параметров *OLS*-модели осуществлялась с использованием метода наименьших квадратов и пошаговых алгоритмов регрессионного анализа. В результате была получена модель (1), включающая 9 переменных.

$$\begin{aligned} \widehat{Y}_1 &= -515,75 + 9187,76X_{11} + 25,52X_1 + \\ &+ 38,09 \log X_3 + 4577,73X_4 + 13,32X_5 - 15,43X_2 + \\ &+ 10,92 \log X_6 - 0,19X_8 + 4,2X_9 \end{aligned} \quad (1)$$

График наблюдаемых и предсказанных моделью значений урожайности полей муниципальных образований представлен на рисунке, характеристики моделей - в таблице 1.

⁸ Качество поверхностных вод Российской Федерации, о наиболее загрязненных водных объектах Российской Федерации (приложение к ежегоднику), Федеральная служба по гидрометеорологии и мониторингу окружающей среды, Федеральное государственное бюджетное учреждение «Гидрохимический институт» 2018, Ростов-на-Дону, 2019.

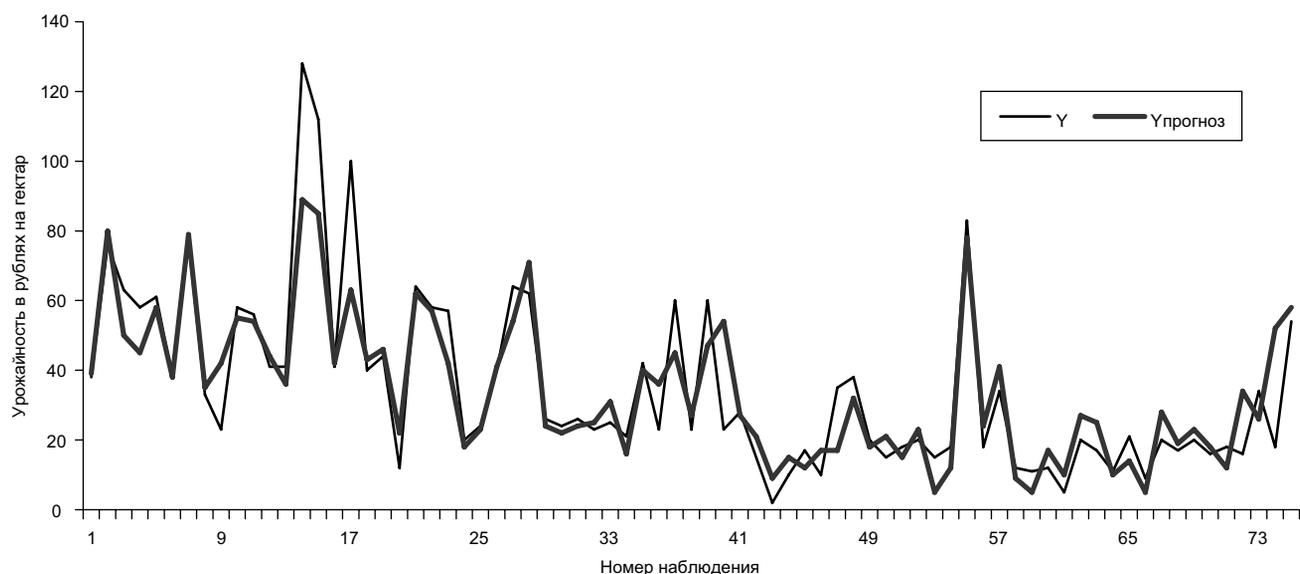


Рис. Наблюдаемые и предсказанные значения урожайности полей муниципальных образований по обучающей выборке

Источник: расчеты авторов по данным Росстата.

Таблица 1

Сравнение эконометрических моделей по критериям информативности

Параметры модели \ Модель	Ridge	Lasso	Model LARS	Elastic net	OLS	Модель по урезанной выборке
Число переменных	27	11	7	12	9	7
AIC	107,45	108,39	130,17	98,67	103,11	102,1
BIC	170,02	133,89	146,40	149,65	126,29	122,96

Полученная модель и все коэффициенты значимы на уровне $\alpha = 0,01$. Исключение составляет признак X_2 , для которого $\alpha = 0,1$, R^2_{adj} составил 0,69, что выше, чем для остальных моделей (с другим набором переменных). Тест VIF показал отсутствие мультиколлинеарности в модели. Проверка остатков на нормальное распределение и гомоскедастичность была проведена с помощью теста Жарка-Бера ($p_{value} = 0,63 > 0,01$) и теста Уайта ($p_{value} = 0,0487 > 0,01$). Изучение всех представленных характеристик модели позволило сделать вывод о ее приемлемой точности и адекватности.

Однако важно отметить, что модель (1) предсказала два отрицательных значения для урожайности, что нельзя признать корректным. Данный факт свидетельствует о нецелесообразности использования данного вида модели для решения задач подобного класса.

Использование нелинейных регрессионных моделей.

На данных выборки была построена и протестирована нелинейная регрессионная модель. Предполагалось, что влияние минимальной и

максимальной температуры и осадков нелинейно, однако включение этих регрессоров в модель как любой формы нелинейной взаимосвязи ухудшало качество модели по критериям информативности. Кроме того, во всех комбинациях нелинейной взаимосвязи показатели температур и осадков оказывались незначимыми. Все это не позволило сделать выбор в пользу нелинейной модели регрессии.

Использование регрессионной модели по урезанной выборке.

Оценка параметров регрессионной модели по урезанной выборке на отобранных данных позволила получить модель (2).

$$\hat{Y}_1 = -9,13 + 1,19X_1 - 0,52X_2 + 0,72\log X_3 + 126,6X_4 + 0,71X_5 + 0,38\log X_6 + 277,19X_{11} + 0,42\lambda_r \quad (2)$$

Модель удовлетворяет требованию нормальности остатков, о чем свидетельствует тест Жарка-Бера ($p_{value} = 0,2 > 0,01$) и является статистически значимой ($p_{value} = 0,000 < 0,01$).

Модель (2) с представленным набором переменных превосходит модели с другими наборами

переменных по критериям информативности Акаике и Шварца и достаточно точно предсказывает изучаемый процесс.

Сопоставление традиционной модели OLS (1) с моделью по урезанной выборке (2) по критериям информативности также позволило сделать выбор в пользу регрессии по урезанной выборке (таблица 1). Модель (2) позволяет получать более точные прогнозы по сравнению с традиционными моделями, основанными на методе наименьших квадратов. Так, если критерии информативности Акаике и Шварца для модели по урезанной выборке составили 102,1 и 122,96 соответственно, для OLS модели - 103,11 и 126,29 соответственно.

Интерпретация результатов модели на основании *средних маргинальных эффектов*, представленных в таблице 2, позволяет говорить, что рост фактора вывезенных за год твердых коммунальных отходов на душу населения, то есть высокая эффективность уборки мусора, положительно влияет на урожайность полей соответствующей территории. Это объясняется недостаточно эффективной с экологической точки зрения обработкой бытового мусора в некоторых регионах России. Также некоторые территории сталкиваются с образованием незаконных свалок, что в совокупности негативно влияет на урожайность сельскохозяйственных культур. В регионах, где вывоз мусора налажен, урожайность сельскохозяйственных полей выше.

Таблица 2

Значения средних маргинальных эффектов для модели по урезанной выборке

Независимая переменная	Значение среднего маргинального эффекта
X_1	1,17
X_2	-0,54
$\log X_3$	1,13
X_4	91,53
X_5	0,58
$\log X_6$	0,29
X_{11}	258,07

Источник: расчеты авторов по данным Росстата.

Другим фактором, оказывающим положительное влияние на результативный признак, является доля растениеводства в сельском хозяйстве МО, что можно объяснить эффектом отдачи от масштаба - в специализирующихся на растениевод-

стве районах выше его продуктивность. В такие районы больше вкладываются агрокультурные холдинги, имеются большие возможности специализации и, соответственно, оптимизации при выращивании зерновых культур, которые требуют существенных трудозатрат и финансовых вложений в виде удобрений, системы контроля и др. На урожайность полей положительно влияют количество тракторов на гектар засеянной территории, что характеризует уровень технической оснащенности агрокультурного бизнеса, а также внесение минеральных удобрений на гектар поля.

Положительная связь наблюдается между урожайностью и уровнем благосостояния жителей МО, мотивирующих их к высокопродуктивной работе.

Еще один вывод модели - положительная связь численности населения муниципального образования и урожайности полей. Происходит это из-за больших возможностей для специализации людей на разных этапах производства продукции и обслуживания ферм и холдингов. Ожидаемое положительное влияние на урожайность полей оказывают инвестиции бизнеса в расчете на душу населения. Данный факт еще раз указывает на то, что, во-первых, сельскохозяйственные поля требуют инвестиций для поддержки уровня урожайности, во-вторых, существует централизация экономической активности в крупных районах. Агрокультурный бизнес заинтересован в наличии большого рынка сбыта продукции, который бы находился недалеко от поля, чтобы сократить транспортные издержки и стоимость хранения продукции.

Негативное влияние на урожайность полей оказывает фактор роста посевной территории относительно всей территории района. Объяснить данный факт, скорее всего, можно тем, что для эффективного земледелия и большой урожайности важны не только размеры полей, но и сопутствующая инфраструктура.

Дальнейшие направления исследования

Дальнейшие направления исследования повышения точности прогнозных моделей урожайности полей видятся в использовании методов интеллектуального анализа данных (нейросетей) [19] и информации космических фотоснимков, которые могут быть использованы наряду с другими независимыми переменными. Согласно ряду

работ, использование космических фотоснимков позволяет повысить качество прогнозных моделей за счет использования новых типов данных.

Заключение

В результате проведения исследования были получены ответы на поставленные задачи и выдвинутые гипотезы. В рамках работы представлена методика прогнозирования урожайности сельскохозяйственных культур на основе использования современного эконометрического инструментария. Значительное внимание уделено повышению точности моделей прогнозирования урожайности сельскохозяйственных культур муниципальных образований регионов России по сравнению с традиционными моделями за счет использования различных типов данных и регрессионных моделей по урезанной выборке.

В ходе проведения анализа было выявлено, что экологические факторы оказывают различное влияние на урожайность. Так если влияние одних факторов оказалось статически незначимым, то важность вывоза твердых коммунальных отходов статистически подтвердилось. Неэффективная система вывоза и уборки отходов домохозяйств, незаконные свалки мусора поблизости полей, негативно влияют на урожайность сельскохозяйственных культур.

Положительным фактором является то, что ряд переменных, отвечающих за экологическую ситуацию в регионах и МО, не вошли в итоговую модель. Это свидетельствует о достаточно благополучной экологической обстановке в ряде регионов России и отсутствии значимого загрязнения воздуха, количества серы и азота, а также допустимом уровне кислотности на снежном покрытии.

Было выявлено также, что на урожайность сельскохозяйственных культур оказывают влияние социально-экономические и технологические факторы, среди которых: уровень благосостояния населения соответствующего МО, уровень оснащения агрокультурного комплекса современной техникой, создание питательной среды для сельскохозяйственных культур за счет внесения удобрений, численность населения в регионе проживания и др.

Негативное воздействие на урожайность оказывает фактор чрезмерно высокой доли засеянных территорий МО, что объясняется недо-

статочно развитой инфраструктурой, нехваткой дорог сельских и городских территорий.

Литература

1. **Basso F.** et al. Evaluating environmental sensitivity at the basin scale through the use of geographic information systems and remotely sensed data: an example covering the Agri basin (Southern Italy) // *Catena*. 2000. Vol. 40. No. 1. P. 19-35.
2. **Salvati L.** et al. Exploring the relationship between agricultural productivity and land degradation in a dry region of Southern Europe // *New Medit.* 2010. Vol. 9. No. 1. P. 35-40.
3. **Pantazi X.E.** et al. Wheat yield prediction using machine learning and advanced sensing techniques // *Computers and Electronics in Agriculture*. 2016. Vol. 121. P. 57-65.
4. **Anders U., Korn O.** Model selection in neural networks // *Neural networks*. 1999. Vol. 12. No. 2. P. 309-323.
5. **De la Casa A.** et al. Soybean crop coverage estimation from NDVI images with different spatial resolution to evaluate yield variability in a plot // *I SPRS journal of photogrammetry and remote sensing*. 2018. Vol. 146. P. 531-547.
6. **Bajracharya D.** Econometric Modeling Vs Artificial Neural Networks: A Sales Forecasting Comparison. 2011.
7. **Demuth H.B.** et al. Neural network design. Martin Hagan. 2014.
8. **Dharmadhikari N.L.** Economic Modeling of Agricultural Production in North Dakota Using Transportation Analysis and Forecasting: дис. - North Dakota State University. 2018.
9. **Haghverdi A., Washington-Allen R.A., Leib B.G.** Prediction of cotton lint yield from phenology of crop indices using artificial neural networks // *Computers and Electronics in Agriculture*. 2018. Vol. 152. P. 186-197.
10. **Jordanova N.** Soil magnetism: Applications in pedology, environmental science and agriculture. Academic Press. 2016.
11. **Molnar C.** Interpretable Machine Learning—A Guide for Making Black Box Models Explainable. Leanpub, np. 2018.
12. **Moshiri S., Cameron N.** Neural network versus econometric models in forecasting inflation // *Journal of forecasting*. 2000. Vol. 19. No. 3. P. 201-217.
13. **Pxldaru R., Roots J., Viira A.H.** Estimating econometric model of average total milk cost: A support vector machine regression approach // *Economics and rural development*. 2005. Vol. 1. No. 1. P. 23-31.
14. **Ranjan R.** et al. Irrigated pinto bean crop stress and yield assessment using ground based low altitude remote sensing technology // *Information Processing in Agriculture*. 2019. Vol. 6. No. 4. P. 502-514.
15. **Zhang C.** et al. Machine-learned prediction of annual crop planting in the US Corn Belt based on historical crop planting maps // *Computers and Electronics in Agriculture*. 2019. Vol. 166. P. 104989.

16. Zhang L., Lei L., Yan D. Comparison of two regression models for predicting crop yield // 2010 IEEE International Geoscience and Remote Sensing Symposium. Ieee, 2010. P. 1521-1524.

17. Архипова М.Ю., Александрова Е.А. Исследования характера связи инновационной и экспортной активности российских предприятий // Прикладная эконометрика. 2014. № 38 (4). С. 88-101

18. Мхитарян В.С. и др. Анализ данных: учебник для академического бакалавриата. Сер. 58 Бакалавр. Академический курс (1-е изд.). М.: Изд-во Юрайт, 2017. 490 с.

19. Ширяев В.И. Финансовые рынки. Нейронные сети, хаос и нелинейная динамика: Учебное пособие / В.И. Ширяев. М.: Либроком, 2013. 232 с.

Приложение

Сводка переменных и источников данных

Символ	Название	Единицы	Источник
Y_1	Урожайность засеянных территорий МО	руб / га	https://www.gks.ru
X_3	Чернозем	1 - да, 0 - нет	https://www.etomesto.ru
X_4	NDVI на начало лета	%	https://eos.com
X_5	NDRE на начало лета	%	https://eos.com
X_6	NDVI на конец лета	%	https://eos.com
X_7	NDRE на конец лета	%	https://eos.com
X_8	Минимальная температура	t °C	https://eos.com
X_9	Максимальная температура	t °C	https://eos.com
X_{10}	Максимальные осадки	мм	https://eos.com
X_{11}	Вывоз твердых коммунальных отходов в год	тысяча кубометров на душу населения	https://www.gks.ru
X_{12}	Доля растениеводства в сельском хозяйстве	%	https://www.gks.ru
X_{13}	Доля сельского населения	%	https://www.gks.ru
X_{14}	Наличие города	1 – да, 0 – нет	https://www.gks.ru
$\log X_{15}$	Среднемесячная зарплата во всех секторах	руб.	https://www.gks.ru
$\log X_{16}$	Среднемесячная зарплата в сельскохозяйственном секторе (Раздел ОКВЭД А)	руб.	https://www.gks.ru
X_{17}	Количество тракторов	шт / га	https://www.gks.ru
X_{18}	Минеральные удобрения	шт / га	https://www.gks.ru
X_{19}	Доля посевной площади от площади МО	%	https://www.gks.ru
$\log X_{20}$	Сельское население на 1 января 2018 г.	человек	https://www.gks.ru
$\log X_{21}$	Объем инвестиций в основной капитал (без учета бюджетных средств) на душу населения	руб.	https://www.gks.ru
X_{22}	Специализированные продуктовые магазины на душу населения	единицы на душу населения	https://www.gks.ru
X_{23}	Загрязнение воздуха (превышение индекса загрязнения атмосферы больше 7)	1 - да, 0 - нет	https://www.igce.ru
X_{24}	Загрязнение воздуха (превышение предельно допустимой концентрации)	1 - да, 0 - нет	https://www.igce.ru
X_{25}	S на снежном покрытии	г / км ² ·мес	https://www.meteorf.ru
X_{26}	N на снежном покрытии	кг / км ² ·мес	https://www.meteorf.ru
X_{27}	pH на снежном покрытии	pH	https://www.meteorf.ru
X_{28}	Качество воды водоемов, процент по классам загрязнения 4 и 5	процент водных ресурсов	https://www.gidrohim.com
X_{29}	Инсоляция	кВт/м ²	https://solargis.com

Информация об авторах

Архипова Марина Юрьевна - д-р экон. наук, профессор департамента статистики и анализа данных, ведущий научный сотрудник научно-учебной лаборатории измерения благосостояния, Национальный исследовательский университет «Высшая школа экономики». 101000, г. Москва, ул. Мясницкая, д. 20. E-mail: archipova@yandex.ru. ORCID <https://orcid.org/0000-0002-9022-7385>.

Смирнов Артем Игоревич - студент 4-го курса бакалавриата, факультет экономических наук, ОП «Экономика и статистика», стажер-исследователь в международной лаборатории институционального анализа экономических реформ Института институциональных исследований, Национальный исследовательский университет «Высшая школа экономики». 101000, г. Москва, ул. Мясницкая, д. 20. E-mail: art.smirnoff.hse@yandex.ru.

Финансирование

Исследование выполнено при поддержке гранта РФФИ №18-010-00564 «Современные тенденции и социально-экономические последствия развития цифровых технологий в России».

References

1. **Basso F.** et al. Evaluating Environmental Sensitivity at the Basin Scale Through the Use of Geographic Information Systems and Remotely Sensed Data: An Example Covering the Agri Basin (Southern Italy). *CATENA*. 2000;40(1): 19-35.
2. **Salvati L.** Exploring the Relationship Between Agricultural Productivity and Land Degradation in a Dry Region of Southern Europe. *New Medit.* 2010;9(1):35-40.
3. **Pantazi X.E.** et al. Wheat Yield Prediction Using Machine Learning and Advanced Sensing Techniques. *Computers and Electronics in Agriculture*. 2016;121:57-65.
4. **Anders U., Korn O.** Model Selection in Neural Networks. *Neural Networks*. 1999;12(2):309-323.
5. **De la Casa A.** et al. Soybean Crop Coverage Estimation from NDVI Images with Different Spatial Resolution to Evaluate Yield Variability in a Plot. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2018;146:531-547.
6. **Bajracharya D.** *Econometric Modeling Vs Artificial Neural Networks: A Sales Forecasting Comparison*. Master's thesis. University of Bores; 2011.
7. **Demuth H.B.** et al. *Neural Network Design*. Martin Hagan; 2014.
8. **Dharmadhikari N.L.** *Economic Modeling of Agricultural Production in North Dakota Using Transportation Analysis and Forecasting: PhD Thesis*. Fargo, North Dakota: North Dakota State University; 2018.
9. **Haghverdi A., Washington-Allen R.A., Leib B.G.** Prediction of Cotton Lint Yield from Phenology of Crop Indices Using Artificial Neural Networks. *Computers and Electronics in Agriculture*. 2018;152:186-197.
10. **Jordanova N.** *Soil Magnetism: Applications in Pedology, Environmental Science and Agriculture*. Academic Press; 2016.
11. **Molnar C.** *Interpretable Machine Learning - A Guide for Making Black Box Models Explainable*. Leanpub, np.; 2018.
12. **Moshiri S., Cameron N.** Neural Network Versus Econometric Models in Forecasting Inflation. *Journal of Forecasting*. 2000;19(3):201-217.
13. **Pxldaru R., Roots J., Viira A.H.** Estimating Econometric Model of Average Total Milk Cost: A Support Vector Machine Regression Approach. *Economics and Rural Development*. 2005;1(1):23-31.
14. **Ranjan R.** et al. Irrigated Pinto Bean Crop Stress and Yield Assessment Using Ground Based Low Altitude Remote Sensing Technology. *Information Processing in Agriculture*. 2019;6(4):502-514.
15. **Zhang C.** et al. Machine-Learned Prediction of Annual Crop Planting in the US Corn Belt Based on Historical Crop Planting Maps. *Computers and Electronics in Agriculture*. 2019;166:104989.
16. **Zhang L., Lei L., Yan D.** Comparison of Two Regression Models for Predicting Crop Yield. In: *Proc. of the 2010 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2010, July 25-30, 2010, Honolulu, Hawaii, USA*. New York, USA: Ieee; 2010. P. 1521-1524.
17. **Arkhipova M., Aleksandrova E.** Study of the Relationship Between Innovation and Export Activity of Russian Firms. *Applied Econometrics*. 2014;36(4):88-101. (In Russ.)
18. **Mkhitarian V.S.** (ed.). et al. *Data Analysis: Textbook for Academic Bachelor Degree*. Ser. 58 Bachelor. Academic course (1st ed.). Moscow: Urait Publishing House; 2016. 490 p. (In Russ.)
19. **Shiryaev V.I.** *Financial Markets. Neural Network, Chaos and Nonlinear Dynamics: Textbook*. Moscow: URSS Publ.; 2013. 232 p. (In Russ.)

About the authors

Marina Yu. Arkhipova - Dr. Sci. (Econ.), Professor, Department of Statistics and Data Analysis, Faculty of Economic Sciences; Leading Research Fellow, Laboratory for Wealth Measurement, National Research University Higher School of Economics (HSE University). 20, Myasnitskaya Ulitsa, Moscow, 101000, Russia. E-mail: archipova@yandex.ru. ORCID <https://orcid.org/0000-0002-9022-7385>.

Artem I. Smirnov - Fourth-Year Student, Bachelor's Programme in Economics and Statistics, Faculty of Economic Sciences; Research Assistant, International Laboratory for Institutional Analysis of Economic Reforms, Center for Institutional Studies, National Research University Higher School of Economics (HSE University). 20, Myasnitskaya Ulitsa, Moscow, 101000, Russia. E-mail: art.smirnoff.hse@yandex.ru.

Funding

This research was funded by the RFBR grant No. 18-010-00564 «Current trends and socio-economic consequences of the development of digital technologies in Russia».