

ИСПОЛЬЗОВАНИЕ ИЕРАРХИЧЕСКОЙ МНОГОФАКТОРНОЙ РЕГРЕССИОННОЙ МОДЕЛИ С ЦЕЛЬЮ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ПОКАЗАТЕЛЕЙ

Овчинников П. В.¹, Ткачев А. Н.², Мирославская М. Д.³
(ФГБОУ ВО «ЮРГПУ (НПИ) имени М.И. Платова»,
Новочеркасск)

Представлена методика определения латентных параметров многоуровневой многофакторной регрессионной модели и способы использования модели на примере прогнозирования динамики показателей, отражающих функционирование сельскохозяйственного производства: урожайности зерновых культур и потребления кормовой пшеницы. Актуальность модели обусловлена возможностью применения в целях прогнозирования главного выходного показателя (первый уровень модели) и его промежуточных составляющих (последующие уровни модели). Выполнена обработка статистических данных, применен метод регрессионного анализа информации, выполнено построение графиков результатов моделирования с использованием MS Excel и среды разработки языка программирования Python. Модель строится на основании иерархической зависимости первого и последующих уровней, при которой входные данные выходного показателя группируются как индивидуальные или общие параметры линейных зависимостей промежуточных переменных. В результате апробации модели были получены данные урожайности зерновых культур в зависимости от типа подкормки удобрениями и соответствующие им доли площадей по данным Российской Федерации в целом и Ростовской области в частности; произведен прогноз урожайности на три года, приведено сравнение с результатами применения факторного анализа. В вопросе определения потребления кормовой пшеницы были получены данные объема чистого потребления пшеницы и объема пшеницы, перерабатываемой на комбикорм.

Ключевые слова: регрессионная модель, иерархическая система, урожайность, потребление, зерновые культуры.

1. Введение

Проблема прогнозирования динамики показателей производства и потребления зерновых культур приобретает особую

¹ Петр Вячеславович Овчинников, к.э.н., доцент (pvo78@yandex.ru).

² Александр Николаевич Ткачев, д.т.н., профессор (tkachev.an@mail.ru).

³ Мария Дмитриевна Мирославская, аспирант (miroslavsky.maryimir@gmail.com).

актуальность в связи с тем, что именно зерновые стали основой отечественного агропромышленного комплекса. При этом представляет интерес получение прогнозных характеристик как с точки зрения использования зерна (в частности, оценка ожидаемых объемов потребления по видам продукции), так и с точки зрения определения ожидаемого уровня валового сбора, урожайности и т.п. В настоящее время методология и инструментарий для решения задач прогнозирования достаточно хорошо разработаны и разнообразны. Наиболее широко применяются различные регрессионные модели, в практической плоскости применение нашли, прежде всего, одноуровневые модели линейной регрессии. В таких моделях предполагается, что выходная переменная является линейной функцией базисных факторов [23, 28–29]. Линейные одноуровневые модели используются и при оценке функциональных зависимостей и прогнозировании сельскохозяйственных показателей, что представлено, в том числе, в работах Зинченко В.Е., Шубнова М.Г., Бисчокова Р.М. и др. [2, 3, 10, 16, 18, 20, 24]. Значительная часть данных моделей являются локальными и позволяют проводить анализ в рамках определенной предметной области. Представляет интерес задача построения регрессионных моделей для иных частных случаев и, в перспективе, – построения общей модели оценки и прогнозирования показателей сельскохозяйственной деятельности.

Предлагаемая в настоящем исследовании модель разрабатывалась на основе принципов метода моделирования структурных уравнений (SEM) и множественной регрессии (MR) [12, 17, 19, 21, 26–27].

Метод моделирования структурных уравнений является элементом эконометрики, используемым на данный момент для исследования социальных и поведенческих структур [25]. В то же время растет популярность применения модели множественной регрессии, которая используется не только в технических, но и в экономических и сельскохозяйственных исследованиях [1, 6]. Недостатком такой модели является слабый прогнозный потенциал на долгосрочный период, так как колебания любого

входного показателя за пределы существующих в тестовом наборе приводит к выбросам и низкой точности прогноза.

Первым этапом моделирования SEM является выделение латентных (скрытых) переменных, которые не могут быть измерены напрямую или представлены в виде откликов независимых переменных с весовыми коэффициентами. Для определения таких факторов наиболее распространенным является факторный анализ, позволяющий применить целый ряд методов: метод главных компонент, метод наименьших квадратов, метод максимального правдоподобия. Для формирования модели использовалась идея разделения переменных на группы общих и индивидуальных факторов с использованием метода наименьших квадратов.

Использование множественной регрессии обусловлено значимым количеством факторов, которые влияют на выходной параметр модели. При этом выявление скрытых факторов позволяет не только снизить размерность набора входных данных на каждом этапе построения модели, но и определить значения этих факторов как самостоятельных переменных, несущих логический информационный смысл при проведении анализа и построения прогностических моделей.

Предполагается, что предлагаемый тип модели может применяться в случае, когда общая результирующая переменная может быть представлена в виде функциональной зависимости от базисных факторов, каждый из которых, в свою очередь, получает представление как функция от факторов более низкого уровня. В результате модель имеет вид иерархической структуры с множеством входных факторов, см. рис. 1.

Количество уровней зависит от моделируемой системы, при этом высшие уровни подчиняются правилу

$$(1) \quad y = f(y_1, y_2, \dots, y_m) = \sum_{k=1}^m \lambda_k y_k.$$

Возможны следующие ситуации:

1. В случае исследования относительных величин (например урожайности) выполняется условие

$$(2) \quad \sum_{k=1}^m \lambda_k = 1; \quad 0 \leq \lambda_k \leq 1.$$

2. В случае исследования абсолютных показателей (например потребления кормовой пшеницы)

$$(3) \quad \sum_{k=1}^m \lambda_k > 1.$$

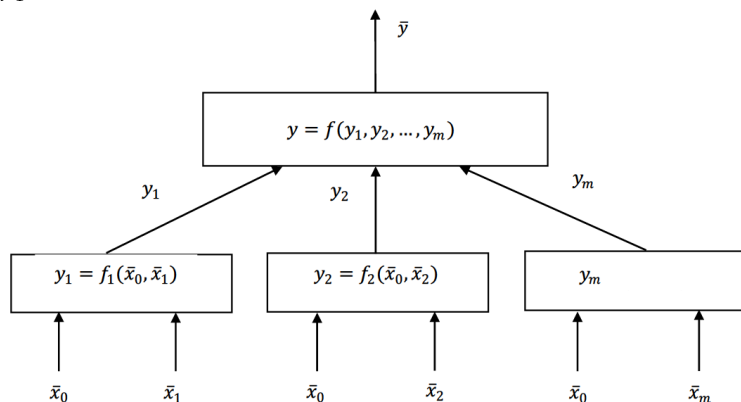


Рис. 1. Структура многоуровневой модели [10]

Математический аппарат, обоснование, валидация представленной модели и идентификация ее коэффициентов была проведена авторами в более ранней публикации [8].

2. Определение параметров модели на различных уровнях

Первый уровень модели представляет из себя множественную регрессию вида:

$$(4) \quad y = a_0 + \sum_{i=1}^n a_i x_i + \varepsilon,$$

где: x_i – входные переменные; a_0 , a_i – коэффициенты, подлежащие определению; ε – ошибка оценки регрессии [8].

Промежуточные переменные y_k , $k = 1, 2, \dots, m$, определяются в виде линейной регрессии вида

$$(5) \quad y_k = f_k(\overline{x_o}, \overline{x_k}) = \sum_{i=1}^{n_0} a_{ki} x_{oi} + \sum_{j=1}^{n_k} a_{kj} x_{kj} + \varepsilon_k,$$

где a_{ki} , a_{kj} – подлежащие определению коэффициенты; ε_k – ошибка оценки регрессии [8].

Общая и индивидуальные части промежуточных регрессионных моделей (моделей 2 уровня) определяются по формулам

$$(6) \quad \begin{cases} u_k = \sum a_i x_{ki}, \\ u_o = \sum a_i x_{oi} + a_o; \end{cases}$$

$$(7) \quad \begin{cases} u_k = \sum a_i x_{ki} + \lambda_k a_k, \\ u_o = \sum a_i x_{oi} + \lambda_o a_o; \end{cases}$$

где a_i , a_o – коэффициенты регрессии (4).

Допущение о вариативности определения общей и индивидуальных частей обусловлено зависимостями промежуточных параметров y_k , где возможны явления:

$$(8) \quad y = \sum_{k=1}^m \lambda_k y_k,$$

$$(9) \quad y = \sum_{k=1}^{m-1} \lambda_k y_k + \lambda_o y_o,$$

где y_o – промежуточный параметр, не имеющей индивидуальной части.

Коэффициенты λ_k также определяются в зависимости от видов промежуточных переменных, для (8) и (9) соответственно:

$$(10) \quad \lambda_k = \frac{\sigma_k^2}{\sigma_y^2 + \sum \sigma_k^2},$$

$$(11) \quad \lambda_k = \frac{\sigma_y^2}{\sigma_k^2},$$

где σ_k^2 , σ_y^2 – показатель дисперсии для регрессий y_k , y .

Для каждой модели 2 уровня определяются коэффициенты вхождения в 1 уровень системы:

$$(12) \alpha_k = \frac{w_k}{\sum_{k=1}^m w_k},$$

$$(13) w_k = u_k + \lambda_o u_o,$$

тогда параметры y_k находятся как

$$(14) y_k = \alpha_k y,$$

$$(15) y_k = \frac{\alpha_k y}{\lambda_k},$$

где y – набор выходных переменных.

При наличии более двух уровней в системе расчет выполняется идентично для всех последующих промежуточных факторов.

3. Апробация модели на примере показателя урожайности зерновых культур

3.1. ПОДГОТОВКА ИСХОДНЫХ ПАРАМЕТРОВ ДЛЯ МОДЕЛИРОВАНИЯ

Для реализации и оценки модели использовались сведения об урожайности зерновых культур на территории РФ, содержащиеся в отчете Росстата за период с 2000 по 2021 гг. В качестве выходного (результатирующего) параметра принят y – общая средняя урожайность в стране (ц/га). Факторами множественной регрессии были выбраны: x_1, x_2 – объем вносимых минеральных и органических удобрений (млн т); x_3 – энергообеспеченность сельскохозяйственных организаций (лс/га); x_4, x_5, x_6 – средняя температура по месяцам в вегетативный период (апрель, май, июнь) в градусах; x_7, x_8, x_9 – средняя сумма осадков по месяцам в вегетативный период (в мм).

Задача оценки урожайности предполагает создание трех регрессионных моделей, взаимосвязанных между собой и соответствующих трем случаям категориям обрабатываемых территорий:

– когда в почву вносились минеральные и органические удобрения; обозначим через y_1 среднюю урожайность удобряе-

мых пахотных площадей; y_{11} – урожайность при внесении минеральных удобрений; y_{12} – урожайность при внесении минеральных и органических удобрений;

– когда поля не удобрялись; урожайность обозначена через y_2 [8].

Представленная задача определения урожайности удобряемых и не удобряемых пахотных площадей соответствует модели (1) с подчинением λ_k условию (2) и определением по формуле (10).

Особенностью задачи является влияние факторов $x_4 - x_9$ на выходной показатель. Урожайность любой сельскохозяйственной культуры зависит от климатических условий. Температурный режим и влажность должны находиться в определенном промежутке, так как любые отклонения за пределы оптимальных значений негативно влияют на урожайность [4–5, 16].

Преобразование факторов осуществлялось следующим образом. Выделим из исходных данных те, которые соответствуют наиболее урожайным годам и для них найдем средние значения показателей. Пусть найденные таким образом значения равны $\tilde{x}_i, i = \overline{1, 9}$. Тогда положим

$$(16) \quad e_i(x_i) = x_i - \tilde{x}_i, i = \overline{1, 2}; e_i(x_i) = -|x_i - \tilde{x}_i| + e_{oi}, i = \overline{3, 9}.$$

Урожайность будем искать в виде

$$(17) \quad y_k = \sum_{i=1}^9 a_{ki} e_i(x_i), \quad k = \overline{1, 2},$$

где $a_{21} = a_{22} = 0$, что соответствует случаю, когда поля не удобрялись, $e_i(x_i)$ – базисные факторы.

3.2. ПОСТРОЕНИЕ ИЕРАРХИЧЕСКОЙ МОДЕЛИ

Среднюю урожайность РФ с учетом различий в обработке полей удобрениями можно представить в виде иерархической структуры, см. рис. 2.

Построим данную модель по данным за период 2000–2019 гг.

Построение математической модели происходит в несколько этапов:

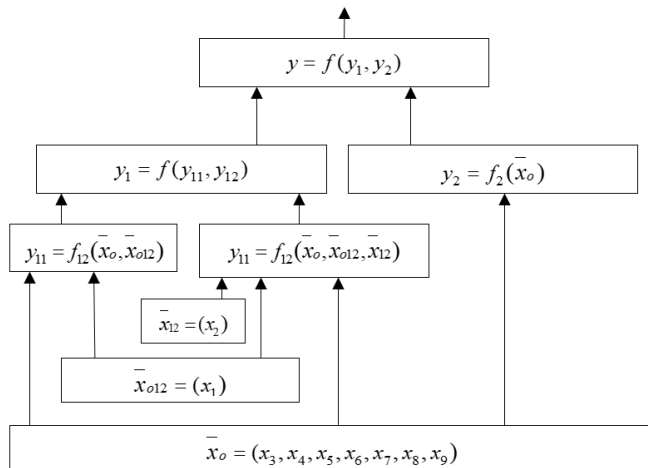


Рис. 2. Иерархическая структура модели урожайности

1 этап. Определение коэффициентов a_0 , a_i регрессии 1 уровня (4) с помощью языка программирования Python с использованием библиотеки `scikit-learn`.

2 этап. Расчет коэффициентов λ_k по формулам (10) или (11) с учетом вариаций (8), (9).

В результате решения задачи получены следующие результаты. Значения λ_k оказались равными: $\lambda_1 = 0,87$, $\lambda_2 = 0,13$. Заметим, что эти результаты хорошо согласуются со статистическими данными относительно доли удобряемых и неудобряемых площадей. Так, в данных за 2021 год приводятся следующие значения показателей: $\lambda_1 = 0,81$, $\lambda_2 = 0,19$. При этом имеющиеся расхождения в величине λ_2 можно объяснить погрешностью оценки, так как это значение мало по сравнению с λ_1 и использованием модельных показателей λ_1 , λ_2 как среднего значения за тестовый период [8].

3 этап. Определение набора выходных переменных 2 уровня по формулам (12)–(14) и коэффициентов модели. Оценка показателей качества модели: σ – скорректированный коэффициент детерминации модели; MSE – среднеквадратическое отклонение модели.

Для модели y_1 были получены следующие значения: $\sigma = 0,89$, $MSE = 0,98$. Для модели y_2 были получены следующие значения: $\sigma = 0,71$, $MSE = 0,77$.

4 этап. Повторение этапа 2 и 3 для следующих уровней (при их наличии).

Для модели y_{11} были получены следующие значения: $\sigma = 0,89$, $MSE = 0,86$. Для модели y_{12} были получены следующие значения: $\sigma = 0,91$, $MSE = 0,99$. Показатели оценки точности регрессий отражают сильную связь входных и выходных переменных. При этом $\lambda_{11} = 0,36$, $\lambda_{12} = 0,5$.

Построенная модель позволила оценить урожайность зерновых и зернобобовых культур при различных системах внесения удобрений и определить доли площадей для каждой из систем, см. таблицу 1.

Таблица 1. Урожайность зерновых и зернобобовых культур [8]

Год	Урожайность			
	y_{11}	y_{12}	y_2	y
2000	14,15	19,51	13,25	15,60
2001	17,50	20,94	17,86	19,40
2002	16,94	20,84	15,66	19,60
2003	15,78	19,40	16,28	17,80
2004	19,01	20,68	18,52	18,80
2005	16,20	16,77	15,89	18,50
2006	19,29	19,34	17,94	18,90
2007	20,18	20,40	17,73	19,80
2008	23,55	24,60	19,40	23,80
2009	22,78	24,65	19,15	22,70
2010	18,78	20,36	15,15	18,30
2011	21,86	23,38	17,25	22,40
2012	19,79	21,69	15,91	18,30
2013	20,31	22,62	17,37	22,00
2014	21,74	25,91	17,71	24,10
2015	22,64	27,48	18,61	23,70
2016	23,36	28,57	17,41	26,20
2017	26,35	31,94	19,20	29,20
2018	24,69	30,94	17,58	25,40
2019	23,66	30,52	15,26	26,70

Множественная модель 1 уровня имеет высокий прогнозный потенциал при оценки будущих значений средней урожайности, который был проверен на среднесрочный период – 3 года.

С использованием модели был построен прогноз урожайности на 2020–2022 год, прогнозное значение при этом составило: 2020 г. – 28,35 ц/га при фактическом объеме 28,6 ц/га, в 2021 г. – 30,35 ц/га при фактическом 26,7 ц/га, в 2022 г. – 34,87 ц/га, при фактическом объеме в 33,6 ц/га. Отклонение факта от прогноза составляет 0,25 ц/га, 3,65 ц/га и 1,14 ц/га в абсолютных значениях и в относительных – 1%, 14% и 4% соответственно.

При использовании в целях прогнозирования первого уровня модели также можно получить прогнозные значения высокой точности, но при влиянии на фактическое значение урожайности неопределенных внешних факторов данная модель имеет больший процент отклонения прогнозного значения от фактического. Так, например, фактическая урожайности 2021 г. была оценена в 26,7 ц/га, но ее модельное значение представлено в виде 31,8 ц/га, разница составляет 19,1%. В то же время иерархическая модель снизила процент отклонения до 13,7%.

На основании полученного прогнозного значения у можно произвести оценку ее составляющих, а именно, значения показателя урожайности при внесении только минеральных, минеральных и органических удобрений и отсутствия удобрений. Для 2020 г.: $y_{11} - 26,16$ ц/га, $y_{12} - 32,78$ ц/га, $y_2 - 17,27$ ц/га; для 2021 г.: $y_{11} - 28,46$ ц/га, $y_{12} - 34,97$ ц/га, $y_2 - 17,67$ ц/га; для 2022 г.: $y_{11} - 33,0$ ц/га, $y_{12} - 39,78$ ц/га, $y_2 - 21,08$ ц/га.

Данная модель также может эффективно применяться для оценки и прогнозирования урожайности для отдельных регионов с учетом возможности вариации вегетационного периода. При усредненных данных вегетационный период определяется с учетом самой значимой территориально культуры. Так, на территории РФ и, в частности, на территории Ростовской области преобладает пшеница, в результате климатические показатели представлены на основании ее вегетационного периода.

Рассматривая данные для Ростовской области в целях построения модели, было определено отсутствие показателя энер-

гообеспеченности в отчетности Ростовской области. Исключение данного базисного фактора из модели не повлекло существенного изменения качества модели и позволило провести анализ, оценку и прогнозирования показателя урожайности зерновых культур. Таким образом, добавление или изменение факторов в модели возможно, но требует проверки значимости в целях предотвращения перегруженности многофакторной модели.

Иерархическая модель урожайности для Ростовской области идентична построения для РФ, с учетом отсутствия фактора хз, и охватывает период с 2000 по 2019 г. в целях настройки модели и период с 2020 по 2022 гг. – в целях проверки качества прогнозирования.

Особенностью данных РО по сравнению с усредненными по РФ являются более динамичные изменения как в сторону увеличения, так и в сторону уменьшения показателя урожайности на период с 2000 г. по 2012 г. и отчетливые тренды роста в периоды 2012 – 2017 гг, 2018 – 2022 гг., что отражает менее явную линейную зависимость по сравнению с данными РФ.

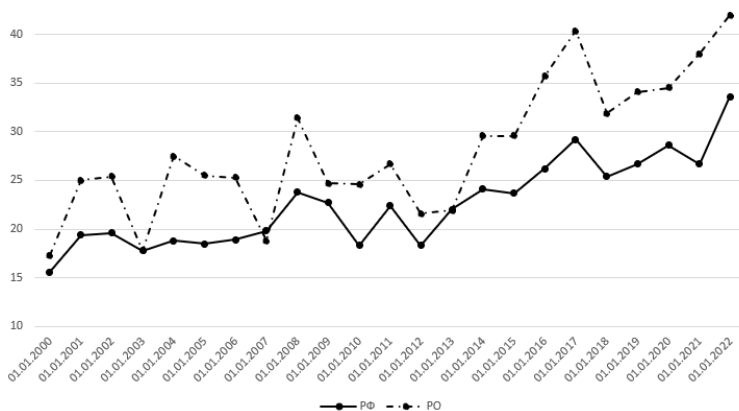


Рис. 3. Изменение показателя урожайности РФ и РО в 2000–2022 гг., ц/га

Характеристики первого уровня иерархической модели РФ: $\sigma = 0,78$, $MSE = 2,18$. При разделении набора на втором уровне

коэффициенты $\lambda_1 = 0,91$, $\lambda_2 = 0,09$ соответствуют удобряемой и неудобряемой территории, средние фактические показатели процента удобряемых территорий РО за 2015–2019 гг.: $\lambda_1 = 0,87$, $\lambda_2 = 0,13$.

Второй уровень модели представлен показателями Y_1 и Y_2 с оценочными параметрами $\sigma = 0,8$, $MSE = 2,27$ и $\sigma = 0,81$, $MSE = 1,35$ для удобряемых и неудобряемых территорий соответственно. Y_1 включает в себя модели y_{11} и y_{12} , характеризующиеся влиянием только минеральных удобрений (y_{11}) и влиянием всех вносимых удобрений (y_{12}). Оценочные характеристики для y_{11} : $\sigma = 0,83$, $MSE = 2,17$; для y_{11} : $\sigma = 0,81$, $MSE = 2,24$; с вкладом в Y_1 $\lambda_1 = 0,45$, $\lambda_2 = 0,46$ соответственно.

Среднее относительное отклонение по всему набору данных 9,02%, при этом с учетом неравномерно изменяющихся показателей урожайности максимальное отклонение составляет 28,7%. Данный показатель отражает более слабую прогнозную точность на среднесрочный период и влияние внешних неучтенных факторов в период 2020–2022 гг., так как среднее относительное отклонение по 20 предыдущим точкам составляет 7,6%.

3.3. СОПОСТАВЛЕНИЕ ИЕРАРХИЧЕСКОЙ МОДЕЛИ И ФАКТОРНОГО АНАЛИЗА

Методика иерархической модели заключается в разделении входных переменных на общую и индивидуальные части. Схожими принципами обладает факторный анализ, который заключается в оценке и группировке входных признаков на скрытые компоненты.

На языке программирования Python определить скрытые компоненты можно с помощью библиотеки `factor_analyzer`, задавая самостоятельно или определяя с помощью «метода локтя» (график каменистой осыпи) оптимальное количество скрытых компонент.

Для набора данных РФ в период 2000–2019 гг. был проведен факторный анализ для выявления скрытых компонент и проверки возможности адекватного разделения переменных

на поднаборы для определения урожайности на удобряемых и неудобряемых площадях.

Исходя из графика (рис. 4) оптимальным количеством компонент является 4, тогда произведем расчет значений факторов. Факторы сгруппировали переменные следующим образом, см. таблицу 2.

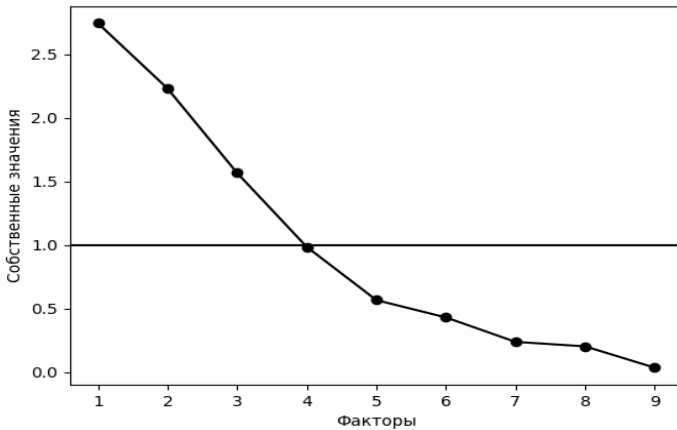


Рис. 4. График каменной осыпи для определения необходимого количества факторов

Таблица 2. Группировка переменных X по 4 факторам

Фактор 1	Фактор 2	Фактор 3	Фактор 4
X_3	X_1	X_5	X_4
X_6	X_2	X_7	
X_8			
X_9			

Очевидно, что фактор 2 описывает влияние удобрений на урожайность, однако в факторе 4 присутствует только одна переменная, что является спорным моментов в оценке правильности выбора указанного количества компонент.

Для проверки выберем в качестве количества формируемых факторов 3, так как именно 3 точки лежат выше уровня собственного значения в размере 1 (таблица 3).

Таблица 3. Группировка переменных X по 3 факторам

Фактор 1	Фактор 2	Фактор 3
x_1	x_5	x_2
x_3	x_7	x_4
x_6	x_9	
x_8		

Фактор 2 полностью представлен показателями осадков, но минеральные и органические удобрения разделились на разные группы факторов. С учетом распределения переменных, построение регрессионной модели урожайности на неудобряемой территории может быть представлено только через фактор 2, то есть с учетом влияния только объема осадков, что противоречит установленному факту зависимости урожайности от колебаний температур [15].

Характеристики регрессионной модели первого уровня на основании 4 факторов: $\sigma = 0,7$, $MSE = 1,75$, что отражает более худшее качество модели по сравнению с общей моделью урожайности по входным переменным X . При этом, используя методику определения доли площадей удобряемых и неудобряемых территорий, результатом являются показатели $\lambda_1 = 0,71$, $\lambda_2 = 0,29$, которые несопоставимы с фактическими представленными ранее показателями.

Определение скрытых факторов для данных урожайности Ростовской области показала схожий результат: «метод локтя» идентичен по числу оптимальных факторов, фактор 4 сформирован одной переменной. Отличие состоит в разделении переменных удобрений по разным компонентам.

Таблица 4. Группировка переменных X по 4 факторам

Фактор 1	Фактор 2	Фактор 3	Фактор 4
x_4	x_3	x_1	x_2
x_5	x_6	x_7	
		x_8	

При использовании трех компонент показатели минеральных и органических удобрений также определяются в разные группы.

Таким образом, в контексте изучения урожайности сельскохозяйственных культур факторный анализ не может полностью удовлетворить запрос на разделение общей урожайности на подтипы в зависимости от существующей подкормки, группировка некоторых переменных к одному фактору не может быть полностью логически обоснована.

4. Потребление кормовой пшеницы

Анализ и оценка объемов потребления зерновых культур являются важным аспектом при производстве зерна. В связи с Распоряжением Правительства РФ «Об утверждении Долгосрочной стратегии развития зернового комплекса Российской Федерации до 2035 года» в период до 2035 г. ожидается увеличение спроса и предложения на продукцию зернового комплекса. На этом основании остро встает вопрос распределения объемов зерна на потребительские и кормовые нужды. Анализ потребления зерновых культур с помощью многофакторных регрессионных моделей представлен авторами в ранней публикации [7]. Проведенный анализ соотношения типов потребления определил, что абсолютное значение объемов продовольственного потребления пшеницы за период 2001–2022 гг. варьируется в пределах 21–24 млн т. В то же время абсолютные и относительные показатели объема кормового потребления растут: так, в 2001 г. доля кормовой пшеницы составляла 35% от общего ее потребления, а к 2022 г. этот показатель вырос до 47%, в абсолютных показателях значения оцениваются в 13 и 21,5 млн т соответственно.

При сравнении объемов производства комбикормов (и доли пшеницы в виде сырья в общей массе) и общего объема кормового потребления было выявлено, что на производство корма для животных уходит только часть пшеницы, так как во многих сельскохозяйственных организациях все еще распространено кормление животных зерновыми культурами без добавления

дополнительных питательных веществ [11]. В связи с этим была поставлена задача оценки объемов пшеницы, направляемых на производство комбикормов и объемов чистого потребления.

В качестве исходных данных выступают показатели, взятые из российских статистических ежегодников Росстата и данных отчета АБ-Центра, а именно: y – потребление кормовой пшеницы (млн т); x_1 – поголовье КРС без учета молочных коров (млн); x_2 – поголовье молочных коров (млн); x_3 – поголовье свиней (млн); x_4 – поголовье овец и коз (млн); x_5 – поголовье птицы (млн).

На основании рекомендаций по питанию сельскохозяйственных животных и птиц была составлена иерархическая схема потребления кормовой пшеницы, см. рис. 5.

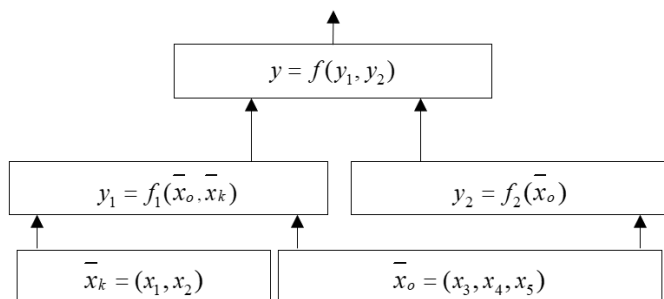


Рис. 5. Иерархическая структура модели потребления кормовой пшеницы

В результате построения множественной регрессионной модели 1 уровня были определены коэффициенты модели, при этом $\sigma = 0,7$, $MSE = 1,33$.

Показатели λ_k определяются по формуле (11) с учетом (9): $\lambda_1 = 0,4$, $\lambda_2 = 0,6$.

Для формирования общей и индивидуальной части использовались формулы (7).

Построенные регрессионные модели 2 уровня определяют показателями оценки: $\sigma = 0,74$, $MSE = 0,84$ для модели y_1

и $\sigma = 0,9$, $MSE = 0,5$ для модели y_2 , что подтверждает значимость результатов моделирования.

Подтверждением достоверности результатов также является среднее значение объема пшеницы в составе произведенных комбикормов по данным с 2015 по 2022 гг. (более ранние значения показателей являются неполными и не могут участвовать в оценке), см. таблицу 5.

Таблица 5. Урожайность зерновых и зернобобовых культур

Период	Объем производства комбикорма, млн т	Объем пшеницы в массе комбикормов, млн т	Модельные значения объема пшеницы на комбикорм, млн т
2015	24,81	9,79	9,42
2016	26,07	10,20	11,72
2017	28,01	10,96	13,64
2018	29,17	11,29	11,69
2019	30,35	11,59	10,54
2020	31,22	11,82	11,78
2021	32,18	12,18	10,38
2022	34,21	12,76	12,56

На основании представленных результатов моделирования можно сделать вывод о применимости модели для анализа и оценки сельскохозяйственных показателей.

5. Вывод

В результате исследования применения иерархической многофакторной регрессионной модели приведена процедура построения регрессионной многоуровневой системы для оценки и прогнозирования урожайности зерновых культур по Российской Федерации в целом и по Ростовской области в частности. Представлена иллюстрация численной процедуры разделения многофакторной модели на уровни и сравнение результатов процедуры с результатами применения факторного анализа для определения скрытых компонент. Применение иерархической модели позволяет не только сгруппировать переменные в фак-

торы, но и логически обосновать эти факторы, представляя их как действительно существующие параметры, а также использовать эти факторы для обособленного анализа.

Для относительных величин (модель урожайности зерновых и зернобобовых культур) построена трехуровневая иерархическая система, проведена оценка качества модели с использованием стандартных методов и специализированного программного обеспечения, построен и подтвержден краткосрочный прогноз на три года вперед.

Для абсолютных показателей (модель потребления кормовой пшеницы) произведено определение состава и объемов входящих в них значений, оценена точность модели.

Оценка предложенной иерархической многофакторной модели на различных показателях агропромышленного комплекса позволяет высоко оценить применимость модели для сельскохозяйственных и агропромышленных нужд.

Литература

1. БЕРГ Н.А., ДЕГТЯРЕВА Н.А. *Принятие управленческих решений в сельском хозяйстве на основе модели множественной регрессии* // Вестник ЧелГУ. – 2023. – №11(481). – С. 167–175.
2. БИСЧОКОВ Р.М. *Анализ, моделирование и прогноз урожайности сельскохозяйственных культур средствами искусственных нейронных сетей* // Вестник Российского университета дружбы народов. Серия: Агрономия и животноводство. – 2022. – Т. 17, №2. – С. 146–157.
3. ЗИНЧЕНКО В.Е. *Прогнозирование урожайности озимых культур и реакции ярового ячменя и озимой пшеницы на различные приемы обработки солонцов в условиях ростовской области*: Автореф. дис. канд. сель.-хоз. наук., Персиановский, 2005. – 22 с.
4. КЛОЧКОВ А.В., СОЛОМКО О.Б., КЛОЧКОВА О.С. *Влияние погодных условий на урожайность сельскохозяйственных культур* // Вестник белорусской государственной сельскохозяйственной академии. – 2019. – №2. – С. 101–105.

5. ЛАМАЖАП Р.Р., ЛИПШИН А.Г. *Влияние климатических условий на урожайность ярового ячменя в республике Тыва* // Вестник КрасГАУ. – 2016. – №12(123). – С. 13–19.
6. МАКСИМОВ Р.А. *Метод определения параметров адаптивной способности с использованием множественного регрессионного анализа взаимосвязи урожайности и ее элементов структуры* // Достижения науки и техники АПК. – 2021. – №6. – С. 4–10.
7. ОВЧИННИКОВ П.В., КОМИССАРОВА М.А., МИРОСЛАВСКАЯ М.Д. *Прогнозирование потребления зерновых культур при помощи многофакторных регрессионных моделей* // XIV Всероссийское совещание по проблемам управления (ВСПУ-2024): сборник научных трудов, 17-20 июня 2024 г., Москва. – С. 2161–2164.
8. ОВЧИННИКОВ П.В., ТКАЧЕВ А.Н., МИРОСЛАВСКАЯ М.Д. и др. *Идентификация смешанных аддитивных регрессионных моделей многоуровневых систем* // Изв. ВУЗов. Сев.-Кавк. регион. Техн. науки. – 2024. – №2. – С. 28–39.
9. ПРОКИНА Л.Н., ПУГАЕВ С.В. *Урожайность озимой пшеницы в зависимости от предшественников, удобрений и известкования* // Аграрная наука Евро-Северо-Востока. – 2022. – №3. – С. 318–326.
10. РОГАЧЕВ А. Ф., МЕЛИХОВА Е. В. *Прогнозирование продуктивности агрокультур на основе ретроспективных данных методом наименьших модулей* // Известия наук. – 2022. – №1(65). – С. 361–369.
11. САВОСТИН Д.С., САВОСТИН С.Д., МАГОМЕДОВ М.Д. и др. *Научное обоснование направлений увеличения объемов производства комбикормов и животноводческой продукции в Российской Федерации* // Экономические системы. – 2022. – №1. – С. 99–109.
12. СЕДЫХ И.А., ИСТОМИН В.А. *Применение иерархических динамических нейро-окрестностных моделей* // XIV Всероссийское совещание по проблемам управления (ВСПУ-2024): Сборник научных трудов, 17-20 июня 2024 г., Москва. – С. 3000–3004

13. СЕМИНЧЕНКО Е.В., СОЛОНКИН А.В. *Влияние климатических факторов на урожай озимой пшеницы и ярового ячменя в условиях сухой степи Нижнего Поволжья* // Siberian Journal of Life Sciences and Agriculture. – 2022. – №3. – С. 58–74.
14. СЕРАЯ Т.М., БОГАТЫРЕВА Е.Н., БИРЮКОВА О.М. и др. *Агроэкономическая эффективность органических удобрений при возделывании озимой пшеницы на дерново-подзолистой легкосуглинистой почве* // Почвоведение и агрохимия. – 2012. – №2(49) – С. 82–96.
15. СОЗИНОВ А.А., ЖЕМЕЛА Г.П. *Улучшение качества зерна озимой пшеницы и кукурузы.* – М.: Колос, 1983. – 270 с.
16. ШУБНОВ М.Г. *Алгоритмы и инструментальные средства нейросетевых технологий моделирования урожайности на основе автокорреляционных функций временных рядов:* Автореф. дис. канд. экон. наук. – Кисловодск, 2013. – 23 с.
17. AFTHANORHAN A., AWANG Z., AIMRAN N. *Five common mistakes for using partial least squares path modeling (pls-pm) in management research* // Contemporary Management Research. – 2020. – Vol. 4, No. 16. – P. 255–278.
18. BHARADIYA J.P., TZENIOS N., REDDY M. *Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches* // Journal of Engineering Research and Reports. – 2023. – Vol. 24, No. 12. – P. 29–44.
19. CHENG E.W. *SEM being more effective than multiple regression in parsimonious model testing for management development research* // Journal of Management Development. – 2001. – Vol. 7, No. 20. – P. 650–667.
20. DISSANAYAKE D.M.P.W., RATHNAYAKE R.M.K.T., CHATHURANGA G. *Crop yield forecasting using machine learning techniques - a systematic literature review* // KDU Journal of Multidisciplinary Studies (KJMS). – 2023. – Vol. 5, No. 1. – P. 54–65.
21. ELANGO VAN N., RAJENDRAN R. *Structural equation modeling-a second-generation multivariate analysis* // National Conference on Indian Business Management «Emerging Man-

- agement Paradigm in Indian Business» at: Sri Ramakrishna Institute of Technology, Coimbatore Volume. – 2015. – P. 33–54.
22. HAYFIELD T., RACINE J.S. *Nonparametric econometrics: the NP package* // Journal of statistical software. – 2008. – Vol. 5, No. 27. – P. 1–32.
 23. ITIDAL A.A., MOHAMMAD A.A., WALEED B.A. *Application of stepwise multiple regression to supersaturated designs data of water pollution in Saudi Arabia* // JP Journal of Biostatistics. – 2024. – Vol. 3, No. 24. – P. 487–515.
 24. MERONI M., WALDNER F., SEGUINI L. et al. *Yield forecasting with machine learning and small data: what gains for grains?* // Agricultural and Forest Meteorolog. – 2021. – Vol. 308–309. – P. 1–13.
 25. NUNKOO R., RAMKISSOON H. *Structural equation modeling and regression analysis in tourism research* // Current Issues in Tourism. – 2011. – No. 15(8) – P. 1–26.
 26. PUTRA W.B.T.S. *Modul Pembelajaran SEM-PLS: Permasalahan, Kepercayaan Umum, Tahapan Spesifikasi dan Evaluasi Model PLS-SEM, HCMs, serta PLS-MGA* // Indonesian School of Research. – 2024. – 85 p.
 27. SADIKAJ G., WRIGHT A.G.C., DUNKLEY D.M. et al. *Multi-level structural equation modeling for intensive longitudinal data: a practical guide for personality researchers* // The Handbook of Personality Dynamics and Processes. – 2021. – P. 855–885.
 28. WANG W., et al. *Forecasting elections with non-representative polls* // Int. Journal of Forecasting. – 2015. – P. 980–981.
 29. ZOU H., HASTIE T. *Regularization and variable selection via the elastic NET* // Journal of the Royal Statistical Society: Series B (Statistical Methodology). – 2005. – Vol. 67, No. 2. – P. 301–320.

USING A HIERARCHICAL MULTI-VARIABLE REGRESSION MODEL FOR THE PURPOSE OF ANALYSIS AND FORECASTING AGRICULTURAL INDICATORS

Petr Ovchinnikov, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, Cand.Sc., Associate Professor (pvo78@yandex.ru).

Alexander Tkachev, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, Doctor of Science, Professor (tkachev.an@mail.ru).

Maria Miroslovskaya, Platov South-Russian State Polytechnic University (NPI), Novocherkassk, Graduate Student (miroslovsky.marymir@gmail.com).

Abstract: The article presents a methodology for determining the latent parameters of a multi-level multifactor regression model and ways of using the model using the example of forecasting the dynamics of indicators reflecting the functioning of agricultural production: grain crop yield and feed wheat consumption. The relevance of the model is due to the possibility of using it for forecasting the main output indicator (the first level of the model) and its intermediate components (subsequent levels of the model). The article processes statistical data, applies the method of regression analysis of information, and plots the results of modeling using MS Excel and the Python programming language development environment. The model is built on the basis of a hierarchical dependence of the first and subsequent levels, in which the input data of the output indicator are grouped as individual or common parameters of linear dependencies of intermediate variables. As a result of testing the model, data on grain crop yields were obtained depending on the type of fertilizer application and the corresponding shares of areas according to the Russian Federation as a whole and the Rostov Region, in particular, a yield forecast for three years was made, and a comparison with the results of applying factor analysis was given. In determining the consumption of feed wheat, data on the volume of net wheat consumption and the volume of wheat processed into compound feed were obtained.

Keywords: regression model, hierarchical system, yield, consumption, grain crops.

УДК 519.6 +330.4

ББК 22.19

*Статья представлена к публикации
членом редакционной коллегии Р.М. Нижегородцевым.*

*Поступила в редакцию 31.10.2024.
Опубликована 31.05.2025.*