

Литература

Аксень Э.М. 2019. Стохастическая динамическая балансовая модель для выпусков отраслей с использованием коэффициентов пропорций инвестирования. *Экономика, моделирование, прогнозирование*. Сборник научных трудов. Минск: НИЭИ Министерства экономики Республики Беларусь. Выпуск 13. С. 125–131.

Аксень Э.М., Читая Г.О., Шишко О.В. 2021. Модель структурной динамики экономики в условиях цифровизации. *Бизнес. Инновации. Экономика*. Сборник научных статей. Минск: Институт бизнеса БГУ. Выпуск 5. С. 199–205.

Бийчук А.Н. 2017. Цифровая трансформация бизнеса в современной экономике. *Экономическая среда*. №2 (20). С. 12–22.

Головенчик Г.Г., Ковалев, М.М. 2019. Цифровая экономика. Минск: БГУ. 395 с.

Марченко В.М., Поддубная О.Н. 2006а. Линейные стационарные ГДР системы: I. Представление решений. *Известия РАН. Теория и системы управления*. №5. С. 24–38.

Марченко В.М., Поддубная О.Н. 2006б. Представление решений гибридных дифференциально-разностных систем. *Дифференциальные уравнения*. Т. 42. №6. С. 741–755.

Поддубная О.Н. 2020. Теоретические и концептуальные подходы к моделированию экономической динамики в контексте цифровой трансформации. *Белорусский экономический журнал*. №4. С. 79–90.

Поддубная О.Н. 2021. Процессно-системный подход в исследовании социально-экономических систем: от концепций к моделям. *Белорусский экономический журнал*. №3. С. 70–82.

Статья поступила 10.02.2022 г.



**СТАТИСТИЧЕСКИЙ АНАЛИЗ УРОЖАЙНОСТИ ЗЕРНОВЫХ
В РЕСПУБЛИКЕ БЕЛАРУСЬ МЕТОДОМ ЧАСТНЫХ
НАИМЕНЬШИХ КВАДРАТОВ**

Белько И.В.,

доктор физико-математических наук, профессор,

Криштапович Е.А.,

Сапун О.Л.,

кандидат педагогических наук, доцент,

Белорусский государственный аграрный технический университет

Повышение урожайности сельскохозяйственных культур в растениеводстве и продуктивности скота в животноводстве являются важными задачами агропромышленного комплекса Республики Беларусь. Итоги статистического анализа показателей хозяйственной деятельности сельхозпредприятий используются во многих ее отраслях. Для проведения такого анализа используются эконометрические методы и пакеты компьютерных программ.

В нашей предыдущей статье (Белько, Криштапович, Сапун, 2021. С. 270) анализ урожайности зерновых проводился методом главных компонент, которые строятся рекуррентным способом по максимуму объяснённой дисперсии (Белько, Криштапович, Лапко, 2013. С. 134). Полученный в статье прогноз на 2020 г. оказался близким к реальному показателю.

Аналогичный метод, который мы будем использовать – это PLS-метод частных наименьших квадратов, его подробное описание приводится в работе (Henseler, Ringle, Sinkovics, 2009. С. 277). Примеры его использования представлены в пособии (Esposito Vinzi, Chin, Henseler, Wang, 2010). Этот метод отличается возможностью применения при минимальных требованиях к исходным статистическим данным.

Наше исследование основано на данных о показателях хозяйственной деятельности за 2013–2020 гг., приведенных в статистическом ежегоднике «Сельское хозяйство Республики Беларусь» 2020 г., где Y – урожайность зерновых, ц/га, X_1 – инвестиции, % к предыдущему году, X_2 – экспорт, тыс. т, X_3 – средние цены (2013–2015 гг. – тыс. руб./т, 2016–2020 гг. – руб./т), X_4 – производительность труда, % к предыдущему году; X_5 – рентабельность, %; X_6 – сроки операций, % соблюдения; X_7 – техника, тыс. шт.; X_8 – погода (дни хорошей погоды); X_9 – удобрения, кг/га (табл. 1).

Таблица 1

Исходные данные

Год		Финансы			Общие факторы			Уход за посевами		
		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
2013	29,7	94,8	15,5	197,3	100,7	13	57	41	72	174
2014	36	64,9	15,5	199,3	107,5	15,9	47	44	51	180
2015	36,5	88,6	16,7	199	100	8	55	42	68	200
2016	31,5	76	18	236	106	8,9	48	44	52	169,8
2017	33,2	116,3	17	276	109	16,5	43	41	45	169,3
2018	26,8	98,9	15,6	328	99	14,3	50	40	61	169,7
2019	30,4	110,3	16,8	345	108	19,7	47	40	55	174
2020	29,31	102,7	18	350	106	24,4	50	39	60	191

Источник: составлено по данным: Сельское хозяйство Республики Беларусь. 2020. Минск: Национальный статистический комитет Республики Беларусь.

В PLS-моделях для построения регрессии на результирующий показатель по блокам наблюдаемых (MV)-переменных строятся латентные (LV)-переменные Y_f, Y_g, Y_h . Их построение основано на использовании множественной регрессии:

$$Y = X \cdot \omega + E,$$

где Y, ω, E – соответственно матрицы-столбцы значений эндогенной переменной, параметров регрессии и случайных остатков, X – матрица значений экзогенных переменных.

Затем коэффициенты регрессии ω оцениваются по методу наименьших квадратов (Елисева, 2009. С. 45): $\hat{\omega} = (X^T X)^{-1} X^T Y$, где X^T – транспонированная матрица, $(X^T X)^{-1}$ – обратная матрица.

Для дальнейшего использования приводим стандартизацию данных по формулам (Елисева, 2009. С. 41): $t_y = (y - y_{сред}) / \sigma_y, t_x = (x - x_{сред}) / \sigma_x$, где t_y, t_x – стандартизованные значения, $y_{сред}, x_{сред}$ – математические ожидания, σ_y, σ_x – стандартные отклонения (табл. 2).

Таблица 2

Стандартизованные данные

Год	Y	Финансы			Общие факторы			Уход за посевами		
		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
2013	-0,59	0,04	-1,10	-1,02	-0,96	-0,39	1,63	-0,20	1,55	-0,39
2014	1,29	-1,70	-1,10	-0,99	0,75	0,15	-0,58	1,42	-0,77	0,13
2015	1,43	-0,32	0,06	-1,00	-1,14	-1,31	1,19	0,34	1,11	1,90
2016	-0,05	-1,05	1,31	-0,45	0,37	-1,15	-0,36	1,42	-0,66	-0,77
2017	0,45	1,30	0,35	0,14	1,12	0,26	-1,46	-0,20	-1,44	-0,81
2018	-1,45	0,28	-1,00	0,92	-1,39	-0,15	0,08	-0,74	0,33	-0,77
2019	-0,38	0,95	0,16	1,17	0,87	0,85	-0,58	-0,74	-0,33	-0,39
2020	-0,70	0,50	1,31	1,24	0,37	1,72	0,08	-1,29	0,22	1,11

Источник: авторская разработка.

Стандартизованные переменные имеют нулевые средние и среднеквадратическое отклонение, равно единице. При проведении регрессии по таким переменным свободный член уравнения равен нулю. При этом коэффициенты корреляции (табл. 3) между исходными переменными и их стандартизациями сохраняют свои значения.

Таблица 3

Таблица корреляций

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Y	1									
x1	-0,45	1								
x2	-0,02	0,20	1							
x3	-0,67	0,66	0,36	1						
x4	0,26	0,11	0,44	0,23	1					
x5	-0,37	0,47	0,19	0,73	0,49	1				
x6	-0,09	-0,19	-0,30	-0,40	-0,82	-0,37	1			
x7	0,65	-0,84	-0,08	-0,77	0,15	-0,65	-0,10	1		
x8	-0,21	-0,03	-0,32	-0,23	-0,84	-0,25	0,98	-0,28	1	
x9	0,47	-0,16	0,18	-0,17	-0,25	0,00	0,46	-0,08	0,45	1

Источник: авторская разработка.

В табл. 1, 2, 3 приводятся исходные и стандартизированные данные за 2013–2020 гг., а также коэффициенты корреляции между ними, которые, как отмечено выше, будут одинаковыми для данных табл. 1 и 2. Стандартизированные данные носят вербальный характер, их единицы измерения не определяются.

Изменение урожайности за 2013–2020 гг. и ее прогноз на 2021 г. отражают данные рис. 1.

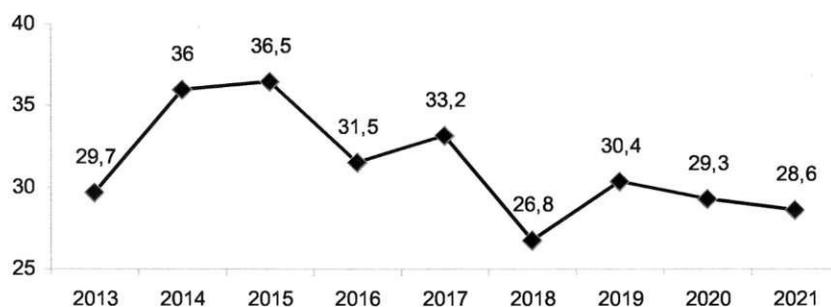


Рис. 1. Урожайность зерновых с прогнозом, ц/га

Источник: авторская разработка.

При проведении регрессии по всем переменным Y , $X_1 - X_9$ проявляется значительный эффект мультиколлинеарности, и уравнение не значимо. В связи с этим нужно искать другие подходы и методы.

Сначала используем PLS-метод разбиения экзогенных переменных на блоки и построения латентных переменных (Esposito Vinzi, Chin, Henseler, Wang, 2010).

В табл. 1 содержатся все наблюдаемые переменные (MV), среди которых на основании корреляций и логического анализа (табл. 3) выделены три блока по три переменных для построения латентных переменных. Показатели блоков характеризуются наличием общих свойств по влиянию на эндогенную переменную Y .

Каждый блок определяет латентную переменную:

Y_f – финансы, с переменными (инвестиции, экспорт, средние цены);

Y_g – общие факторы, с переменными (производительность труда, рентабельность, сроки выполнения операций);

Y_h – уход (наличие техники, погода, удобрение).

На рис. 2 отображена схема зависимостей показателей блоков на эндогенную переменную Y , полученная по уравнениям регрессий (1).

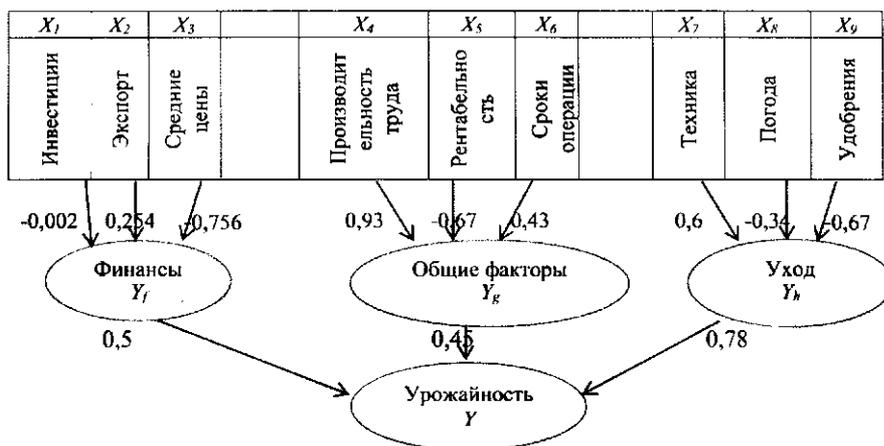


Рис. 2. Схема влияния показателей блоков на эндогенную переменную Y

Источник: авторская разработка.

На этой схеме приведены также коэффициенты уравнений регрессий и их коэффициенты детерминации. В результате становится очевидной возможность получения стандартизованных показателей переменных Y_f , Y_g , Y_h .

После проведения регрессий отдельно по каждому из блоков получаем уравнения:

$$\begin{aligned}
 Y_f &= -0,002X_1 + 0,254X_2 - 0,756X_3 & (R^2 = 0,5); \\
 Y_g &= 0,93X_4 - 0,67X_5 + 0,43X_6 & (R^2 = 0,45); \\
 Y_h &= 0,6X_7 - 0,34X_8 + 0,67X_9 & (R^2 = 0,78).
 \end{aligned}
 \tag{1}$$

Влияние латентных переменных на эндогенную переменную Y можно интерпретировать схематически (рис. 2) как свидетельство достаточно тесной связи между переменными.

Итогом нашего исследования является набор значений стандартизованных латентных переменных Y_f , Y_g , Y_h , их прогнозов на 2021 г., приведенных в табл. 4, и уравнений (1) регрессий Y на эти переменные.

Влияние латентных переменных на итоговую переменную Y можно интерпретировать графически (рис. 3). Прогнозное значение Y на 2021 г. равно 28,6 ц/га.

Уравнение регрессии латентных переменных на эндогенную переменную имеет вид:

$$Y = 0,138 Y_f + 0,214 Y_g + 0,823 Y_h \quad (R^2 = 0,81).$$

Таблица 4

Значения латентных переменных и их прогнозов

	Y	Y _f	Y _g	Y _h
2013	-0,59	0,50	0,05	-0,92
2014	1,29	0,48	0,35	1,22
2015	1,43	0,77	0,32	1,11
2016	-0,05	0,68	0,96	0,57
2017	0,45	-0,02	0,25	-0,18
2018	-1,45	-0,95	-1,16	-1,09
2019	-0,38	-0,84	0,00	-0,60
2020	-0,70	-0,61	-0,78	-0,11
Прогноз на 2021	-0,93	1,08	-0,68	-0,58

Источник: авторская разработка.

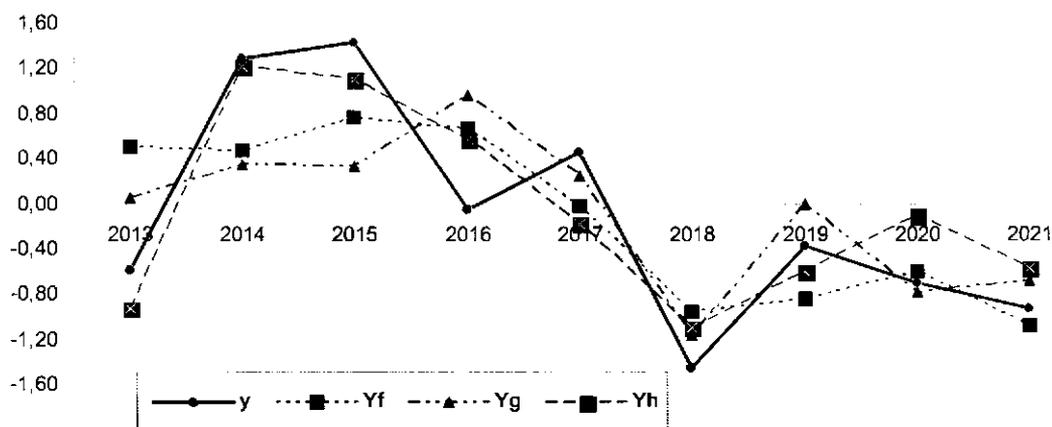


Рис. 3. Влияние латентных переменных на итоговую переменную Y

Источник: авторская разработка.

Связь между переменными, согласно данным рис. 3, достаточно тесная. Тем не менее при равенстве $R^2 = 0,81$ по F -статистике уравнение незначимо ($F = 5,62$, значимость $F = 0,064$). По t -статистике коэффициенты при латентных переменных также незначимы.

Однако, если учесть прогнозные значения переменных Y , Y_f , Y_g , Y_h на 2021 г., то уравнение регрессии примет вид:

$Y = -0,01 + 0,186 Y_f + 0,205 Y_g + 0,81 Y_h$, ($R^2 = 0,83$), $F = 7,85$, значимость $F = 0,02$, уравнение значимо в целом.

Для сравнительного анализа применения PLS-метода выделим, используя таблицу коэффициентов корреляции (табл. 3), переменные (X_1 , X_3 , X_7 , X_9) с наибольшим влиянием на эндогенную переменную Y . С учетом параметров прогноза на 2021 г. проводим регрессию по этим переменным.

Ее уравнение имеет вид:

$$Y = -123,43 + 0,164X_1 - 0,003X_3 + 2,48X_7 + 0,209X_9,$$

$R^2 = 0,89$; $F = 7,86$; значимость $F = 0,03$; уравнение значимо в целом.

Коэффициенты при X_7 и X_9 значимы.

Использование данного инструментария позволяет проводить статистический анализ динамики и прогноз урожайности зерновых в Республике Беларусь.

Полученные результаты способствуют принятию управленческих решений с учетом наиболее значимых показателей, выделенных методом PLS.

Литература

Белько И.В., Криштапович Е.А., Лапко В.В. 2013. Применение математических моделей для анализа показателей урожайности зерновых культур. *Экономика, моделирование, прогнозирование*. Сборник научных трудов. Минск: НИЭИ Министерства экономики Республики Беларусь. Выпуск 7. С. 127–131.

Белько И.В., Криштапович Е.А., Сапун О.Л. 2021. Статистический анализ урожайности в Республике Беларусь методом главных компонент. *Экономика, моделирование, прогнозирование*. Сборник научных трудов. Минск: НИЭИ Министерства экономики Республики Беларусь. Выпуск 15. С. 270–276.

Елисеева И.И. (ред.). 2009. *Эконометрика*. Учебник. Москва: Проспект. 288 с.

Esposito Vinzi V., Chin W.W., Henseler J., Wang H. 2010. *Handbook of Partial Least Squares (PLS)*. Concepts, Methods and Applications. Springer Handbooks of Computational Statistics. Springer, Berlin, Heidelberg. 850 p.

Henseler J., Ringle C.M., Sinkovics R.R. 2009. *The use of partial least squares path modeling in international marketing*. New Challenges to International Marketing. Vol. 20. P. 277–319.

Статья поступила 22.02.2022 г.