DOI: 10.12731/2658-6649-2022-14-2-401-426
UDC 633.1:631.53.04

EFFECT OF FARMING TECHNOLOGIES INTENSIFICATION ON WHEAT YIELD IN THE DRY STEPPE OF SIBERIA: AN ANALYSIS BASED ON THE METHOD OF PROPENSITY SCORE MATCHING

K.O. Tarasov, E.V. Ponkina, A.B. Nugumanova

The study examines the effect of crop production technology intensification on wheat yield under drought conditions. The causal effect of fertilizers and herbicides usage was quantified based on a dataset of reports from 196 farms located in the Kulunda Steppe in Altai Krai. To quantify causal effects, the various methods were tested – Two Means Comparison tests and Linear regression. Due to the not experimental nature of the data, Propensity Score Matching was used to balance the sample based on the main indicative signs (covariates).

We refined the effects of crop technologies intensification based on the balanced sample. The results showed that, according to all considered methods, intensification of agriculture even in the dry climatic conditions resulted in the expected growth of wheat yield. The average causal effect of intensive farming was +2.02 dt/ha. The results also showed that it is possible to obtain more correct estimates of causal effects based on balanced samples. Using simple approaches like Two Means Comparison methods lead to underestimation or overestimation. Finally, we highlighted some limitations and peculiarities of the Propensity Score Matching method. However, the Propensity Score Matching can be considered a good and prospective tool for developing digital services in agricultural analytics.

Keywords: *causal effect; two means comparison; propensity score matching; wheat yield; intensive farming; steppe; Siberia*

For citation. *Tarasov K.O., Ponkina E.V., Nugumanova A.B. Effect of Farming Technologies Intensification on Wheat Yield in the Dry Steppe of Siberia: an Analysis Based on the Method of Propensity Score Matching. Siberian Journal of Life Sciences and Agriculture, 2022, vol. 14, no. 2, pp. 401-426. DOI: 10.12731/2658-6649-2022-14-2-401-426*

ВЛИЯНИЕ ИНТЕНСИФИКАЦИИ ТЕХНОЛОГИЙ ЗЕМЛЕДЕЛИЯ НА УРОЖАЙНОСТЬ ПШЕНИЦЫ В СУХОЙ СТЕПИ СИБИРИ: АНАЛИЗ НА ОСНОВЕ МЕТОДА СОПОСТАВЛЕНИЯ ОЦЕНОК СКЛОННОСТЕЙ

К.О. Тарасов, Е.В. Понькина, А.Б. Нугуманова

В работе изучены подходы к оценке влияния элементов интенсификации технологии растениеводства на урожайность пшеницы в засушливых климатических условиях. Причинный эффект от применения удобрений и гербицидов был количественно рассчитан на основе данных отчетов 196 хозяйств, расположенных в Кулундинской степи Алтайского края. Для количественной оценки был использован набор методов статистического анализа – тесты парных сравнений среднего и линейная регрессия. Поскольку данные не относятся к экспериментальным, выборка была сбалансирована по основным индикативным признакам (ковариатам) методом сопоставления оценок склонностей и получены уточненные оценки эффектов интенсификации технологии земледелия. Результаты показали, что согласно всем рассмотренным методам интенсификация земледелия даже в засушливых климатических условиях привела к увеличению урожайности пшеницы. Средний эффект, выраженный величиной прироста урожайности от использования удобрений и гербицидов составляет +2,02 ц/га. Также показано, что более корректные оценки причинного эффекта могут быть получены на основе сбалансированных выборок, а использование простых методов парных сравнений приводит к недооценке или переоценке величины эффекта. В завершении работы проанализированы некоторые ограничения и особенности метода псевдорандомизации как одного из инструментов потенциально интересных для развития цифровых сервисов в аграрной аналитике.

Ключевые слова: *причинный эффект; парное сравнение; сопоставление оценок склонности; урожайность пшеницы; интенсивные технологии; Кулундинская степь; Сибирь*

Для цитирования. *Тарасов К.О., Понькина Е.В., Нугуманова А.Б. Влияние интенсификации технологий земледелия на урожайность пшеницы в сухой степи сибири: анализ на основе метода сопоставления оценок склонностей // Siberian Journal of Life Sciences and Agriculture. 2022. Т. 14, № 2. С. 401-426. DOI: 10.12731/2658-6649-2022-14-2-401-426*

Introduction

Despite being a region with the largest wheat sowing area, Altai Krai is lags behind in comparison with wheat yield from such regions as Krasnodar Krai and Stavropol Krai, because it is one of the lowest grain yields in the country [9]. Though Altai farmers managed to achieve the 16 dt/ha of wheat yield in 2018 and 2019 it was significantly lower compared to the region of Kuban (where the yield leader – Krasnodar Krai – is situated) – 62.9 dt/ha in 2019 [5]. It is worth mentioning that the climatic conditions in Kuban and Krasnodar Krai are more favorable for wheat production. However, according to long-term field experiments and farmers' experience, the potential wheat yield (depending on varieties and climatic zones of Altai Krai) can reach more than 25.2-39.8 dt/ha [3]. It is also worth mentioning that growing the optimized-for-region varieties of wheat makes it possible to raise it to more than 60 dt/ha [6]. Due to heavy draughts, wheat yield dramatically dropped to 5-10 dt/ha in arid and semi-arid zones of the Altai Krai in 2012 [2]. In 2020, after another heavy draught, wheat yield results might set a new record low since the already reported wheat yield in the regions most affected by the drought was less than 5 dt/ha [7]. Therefore, there is a wheat yield gap in the Altai Krai due to many reasons, and one of them is ineffective or extensive farming practices.

Since the collapse of the Soviet Union, usage of intensive crop production technologies became less widespread in Altai Krai. The input of mineral fertilizers has declined from 22 kg per hectare in 1990 to 4 kg in 2012 [10]. Since 2013, mineral fertilizers usage has increased to 15.3 kg per hectare. However, it is still 7 and 9 times lower than in Stavropol and Krasnodar regions. Moreover, it is 4 times lower than on average in Russia. Adoption of modern technologies aimed at improving soil fertility, preventing soil erosion, and increasing crop yield, is still not widespread in Altai Krai. Sustainable farming practices like "No-Till" and "Strip-Till" have been adopted only at 418 thousand hectares (8% of the total sown area) in 2019 [4]. Farmers often avoid fertilizers and herbicides usage due to different reasons, one of which is the uncertainty of possible positive effects on crop yield. That is why it is crucial to estimate the effect of using chemical treatments on wheat productivity based on current farming practices in the region. Finally, we can consider methods of the effect estimation as an additional effective tool for its implementation in digital services for agricultural analytics.

The effect of an intervention is quantified and measured according to the theory of causality. It is based on relations between causes and effects when an action (an intervention) directly causes an effect. Causal inference is a way to analyze the effect of an intervention or a cause on the resultant signs (target indicators). It helps to conclude the truth of hypotheses by finding differences between facts and hypothetical situations (counterfactuals). The fundamental basis for the study of causal effects is the Theory of Potential Outcomes (TPO) proposed by J. Neyman and D. Rubin [23]. The central problem of the TPO is an estimation of the causal effect as a result of an intervention.

According to TPO, a *causal effect* is a difference between two potential outcomes for a unit – Y(1) and Y(0). Y(1) represents an measurement of the features resulting from the intervention (intervention = TRUE) while Y(0) is an evaluation of the resultant indicator in a case of no intervention (intervention = FALSE) [16; 18; 23]. If we can observe two outcomes simultaneously, Y(1)-Y(0) is presented the True Causal Effect at a unit level.

The fundamental problem of causal inference is the problem of missing values – we cannot observe resultant signs with and without intervention for each individual unit simultaneously. That is why the True Causal Effect at a unit level can not be measured. To solve this problem, researchers supposed that the causal effect could be quantified by comparing the resultant features measured for two groups: a *Treated Group* (TG) – the group of units treated by an intervention, a *Control Group* (CG) – the group had no intervention. We should note that units in TG and CG have to be homogeneous and be identically distributed. It is usually achieved in experimental studies with a random sampling procedure. However, an assessment of the causal effect is often based on unbalanced data samples (in terms of control features such as age, respondent's education, farm size, etc.) that are not represented by randomized experimental data. That leads to biased estimates of the causal effect (Bias error) [23] if we apply traditional statistical methods like two means comparison or regression estimations.

To bring the conditions of a study based on observational data closer to an experimental study, G.W. Imbens and D. Rubin [17] proposed the Rubin-Neyman Causal Model. It allows pseudorandomizing data, reducing bias in the causal effect estimates and bringing them closer to the True Causal Effect. The data of agricultural enterprises describing farming practices and resulting crop yield refer specifically to observational data and require balancing. Therefore, we considered a Propensity Score Matching as a data balancing method to estimate the effects of crop production technology intensification.

Therefore, the key research questions of the study were:

- *What is the expected growth of spring wheat yield resulting from fertilizers input under drought conditions in Altai Krai?*
- *What is the expected growth of spring wheat yields resulting from crop protection agents (chemical treatments) under drought conditions in the Altai Krai?*
- *What are the differences among causal effects assessed by various methods?*
- *What methods are better to use for causal effects estimating?*

The results contribute to effect farming technology estimation methods within regional studies and comparison studies within field experiments. They also contribute to developing applied digital services for farm management to analyze various effects of farming and tillage systems on crop productivity.

Materials and methods

The area under study. The Kulunda steppe zone is situated in the south-eastern part of Western Siberia and stretches from the center to the south of Altai Krai. In 2011, 576 agricultural organizations, 1325 peasant farms, and individual entrepreneurs (IP) operated in the Kulunda steppe zone, where spring wheat is primary commodity production. Approximately 34% of agricultural organizations and farms grew the spring wheat on less than 2000 ha of arable land, 28% of them – on 2000-6000 ha, and 38% – on more than 6000 ha [21].

Kulunda has a continental climate with long-time average temperatures of -18 °C in the coldest month (January) and +19 °C in the warmest month (July) [1]. The Kulunda Steppe area is characterized by a dry climate, with the amount of precipitation during the growing season being the main limiting factor for spring wheat yields. The climatic condition of the 2012 growing season was drought and characterized by low rainfall and the record low spring wheat yield. Frequent droughts in Altai Krai significantly reduce crop yields and increase climatic risks for agricultural producers [13; 20]. The wheat yield varies considerably depending on soil-climatic and agro-ecological subzones identified in the Kulunda steppe. In 1996-2011, the average wheat yield was 15 dt/ha in the forest-steppe zone, in typical steppe – 11 dt/ha, and in dry steppe zone reached 8 dt/ha [11]. In 2012, the average wheat yield plunged to the lowest value over many years – 7.7 dt/ha due to the drought. The significantly lower productivity of wheat in 2012 is explained by low precipitation – 86 mm on average over the territory (from April to August).

Climatic conditions in this area are critical for achieving high wheat yields. Therefore, it is essential to assess the effect of fertilizers and chemical treatment usage under the drought conditions of 2012.

Data collection. The data comes from the farmers' survey conducted in 2013 [21] within project Kulunda. The survey was aimed at revealing the most critical determinants of the wheat yield based on current farming practices in the study area. The questionnaire consisted of information about agroecological and climatic conditions, farm's and farm manager's characteristics, agronomical parameters of wheat production technologies. The survey was done in a face-to-face interview with managers and specialists of agricultural enterprises to assure truthfulness and completeness of answers. The list of interviewed farmers was formed as a randomized sample. Farmers were chosen taking their distribution among 3 agroecological subzones of the Kulunda steppe into account. On average, each interview lasted 2–2.5 hours, including filling out the questionnaire.

Respondents were mainly managers of agricultural enterprises or heads of private farms. Missing data were excluded from the raw data sample. We selected the target and control variables based on survey results (67 interviews). Finally, we enriched the survey dataset with data from official statistical reports of agricultural enterprises in 2012.

The dataset consisted of 196 agricultural enterprises with different legal forms, of which 111 had the legal form "Obshchestvo s ogranichennoy otvetstvennost'u" – OOO (Limited Liability Companies – LLC) and "Akzionernoe obshchestvo" – AO (Stock Companies), and others 85 were the private peasant farms or individual interpreters. The average area of agricultural land was 9,333 ha, and the cultivated area was 7,150 ha. The 2012 vegetation season showed a low average wheat yield of 6.01 dt/ha in 2012, with 9.36 dt/ha for 2008-2012.

Table 1 shows the descriptive statistics of the dataset, where the entire sample was divided into two groups. Group 1 represents farmers who applied herbicides and/or fertilizers within wheat production technologies in 2012. Group 2 represents farmers who applied neither intervention. Descriptive statistics revealed that only 24% of farmers put in mineral fertilizers, while 69% used crop protection agents.

However, 30% of the surveyed farmers used neither fertilizers nor chemical treatments. We also found the differences in average wheat yield between these two groups. Thus, the average yield of farmers who practiced intensive technologies was 8.0 dt/ha, while farmers from Group 2 reported only 5.4 dt/ha in 2012. Notably, such difference can be observed in the 5-year average wheat yield. We also should note that the Shapiro-Wilk test showed that the normal distribution hypothesis was rejected for all continuous variables.

Table 1.

Descriptive statistics of the selected variables, Altai Krai, 2012

Variable name	Variable description	Statistics	Entire sample	Group 1: Used intensive technology	Group 2: Not used intensive technology
General information:					
Type_Ownership	A legal form: *1 – LLC* *0 – others*	Freq. (%)	111 (57) 85 (44)	30 (64) 17 (36)	81 (54) 68 (46)
Agricultural_land	Agricultural land, ha	Mean (SD)	9,333 (7,072)	10,419 (6,801)	8,990 (7,143)
Sown_area	Total sown area, ha	Mean (SD)	7,150 (5407)	8,340 (5400)	6,775 (5373)
Labor_ha_Persons	Amount of arable land per an employee, ha	Mean (SD)	152 (140)	129 (119)	112 (88)
Structure of production:					
Wheat_Share	Share of wheat in the sown area, %	Mean (SD)	45 (21)	42 (19)	46 (22)
Fallow_Share	Fallow share in the sown area, %	Mean (SD)	15 (16)	18 (15)	14 (17)
Farm_Specialization	Share of crop production in total revenue, %	Mean (SD)	54 (31)	55 (29)	53 (31)
Wheat growing technology (treatment variables):					
Fertilizers_Use	Input fertilizers - *Yes* *No*	Freq. (%)	47 (24) 149 (76)	47 (100) 0 (0)	0 (0) 60 (100)
Herbicides_Use	Application of chemical treatments - *Yes* *No*	Freq. (%)	136 (69) 60 (31)	136 (100) 0 (0)	0 (0) 60 (100)
Wheat Productivity:					
Yield_Wheat_2012	Wheat yield in 2012, dt/ha	Mean (SD)	6.0 (3.74)	8.0 (4.30)	5.4 (3.31)
Av_Yield_Wheat	Average wheat yield for 2008-2012, dt/ha	Mean (SD)	9.4 (3.87)	12.2 (4.24)	8.5 (3.26)
Farm manager's characteristics:					
Age	Age, years	Mean (SD)	53 (8.95)	53 (9.86)	52 (8.66)
Education	Farm managed graduated special agricultural education - *Yes* *No*	Frequency (%)	131 (67) 65 (33)	28 (60) 19 (40)	103 (69) 46 (31)
Climatic conditions:					
Precipitation	Sum of precipitation from April to August in 2012, mm	Mean (SD)	86 (38)	95 (39)	88 (40)

Methods. The Propensity Score Matching (PSM) method and weighted linear regression were employed to quantify the causal effects of intensive farming practices on wheat yield. The PSM method was developed by Donald Rubin in collaboration with Paul Rosenbaum in 1983 in order to reduce the bias associated with confounders or unrandomized samples, which can be found when calculating a causal effect of an intervention. A confounder (confounding variable) is a factor that influences both an independent variable and the outcome (dependent variable). They create a difference between an estimated causal effect and the actual causal effect, causing a spurious association and conclusion. As the distribution of treatments on a sample in observational studies does not meet the requirements of randomization, the problem of bias requires a specific solution to be found to reduce or eliminate it. The PSM method allows the simulation of a randomized sample using the Propensity Score and then various matchmaking techniques to obtain better results than matchmaking techniques on an untreated controlled sample. PSM can utilize the Stable Unit Treatment Value Assumption (SUTVA) [22] if we are not sure all covariates were found. Additionally, SUTVA assumes that the outcomes of one unit are not affected by the treatment assignment of another unit.

There are five main steps to implement PSM.

Let Y be the resultant indicator of the intervention/non-intervention, and D be a treatment binary variable (1/0).

The *first step* is data collection and preliminary data analysis. After data collection, it is necessary to understand whether the PSM method should be applied. Therefore, it is essential to analyze the balance between the original control and treatment data groups, clean outliers, drop missing data and compare distributions between TG and CG.

The *second step* is the selection of covariates (the components of a vector X) to add them to the compliance model. Covariates are factors that predict receiving the treatment for units. They are selected to improve a balance of observations (surveyed farms) between TG and CG in order to reduce the bias. Thus, covariates are similarity features that allow matching a pair of units from TG and CG. They can also be related to the target variable (wheat yield) and the treatment variables (usage of fertilizers or herbicides). After balancing across a set of selected covariates, data consistency should be improved; if this has happened, the covariates have been chosen correctly.

The *third step* is to select a model for computing Propensity Score estimates. There are several ways, but the most popular is logistic regression. This model describes the probability of a unit i to get treatment (D_i=1) given X_i:

$$\Pr(D_i = 1 \mid X_i) = \frac{1}{1 + e^{-aX_i}} \quad ,(1)$$

where $aX_i = a_1 x_{i\,1} + a_2 x_{i\,2} + \ldots + a_n x_{i\,n}$, X_i – vector covariates.

A unit *i*'s propensity score is indicated the propensity of the unit to get the treatment (Y_i=1):

$$\text{Score}_i = \Pr(D_i = 1 \mid X_i). \quad (2)$$

The *fourth step* is a selection of a matching method. The choice of a matching method largely determines the composition of treatment and control groups. Therefore, multiple procedures are often used in one study. The choice of matching method is based on comparing results of data balancing. There are several methods – *Exact Matching* (for discrete covariates), *Subclassification*, *Nearest Neighbor Matching*, *Optimal Matching*, *Full Matching,* and *Genetic Matching* [12; 14]. All of these methods are available in various software packages. Using a distance metric, the matching procedure finds a pair of units – one from TG and one from CG to estimate the causal effect on the outcome variable. In order to obtain improved estimations, we tested two methods – Full Matching and Genetic Matching.

The Full Matching method finds a pair for each unit based on a distance metric:

$$\text{Distance}(i, j) = |\text{Score}_i - \text{Score}_j|,\ i, j=1,\ldots,N. \quad (3)$$

The Genetic Matching method (GenMatch algorithm) operates the Generalized Mahalanobis Distance (GMD) [12], which has an additional weight matrix *W*:

$$GMD(i, j, W) = \sqrt{(X_i - X_j)^T (S^{-\frac{1}{2}})^T W S^{-\frac{1}{2}} (X_i - X_j)}, \quad (4)$$

where *W* is a positive specific weight (balance) matrix of size and is the Cholesky decomposition of a parameter *S*, the variance-covariance matrix of *X*. All elements of *W* are limited to zero, except for those located on the main diagonal, which consists of *k* parameters to be selected. When a formula for propensity scores estimation is applied, *X* is replaced by Z. Z is a matrix composed of propensity scores and primary covariates X. It is beneficial to combine PSM and Mahalanobis distance [12].

The *fifth step* is a data balancing analysis. There are different ways to quantify the efficiency of data balancing – QQ plot and Love plot, draw the data distribution for TG and CG before and after matching [26]. The QQ plot is a scatter plot created by placing two sets of quantiles opposite each other. The Love plot visually inspects the quality of data balancing by comparing the absolute standardized mean difference in covariates calculated for TG and CG before and after matching. If an average absolute standardized difference is close to 0, the data is balanced well, indicating minor differences between the control and treatment groups in the resulted sample. Software packages often also provide comparison using parameter tables with averages for groups with calculated percentage improvements.

The main problem with the PSM method is that it is possible to increase bias after balancing if implemented incorrectly. This phenomenon has been called *the Propensity Score Matching Paradox* [19]. Therefore, it is necessary to use this method with utmost care and compare the residual bias with the original bias, which will show that pseudo randomization does not worsen the data distribution.

After completing these five steps, it is possible to calculate the causal effect, which can be done using various statistical methods.

Some matching procedures like the genetic and nearest neighbors matching results in the reduced matched sample for that the average causal effect can be estimated as an *average treatment effect on the treated* (*ATT*) [16]:

$$ATT = E(Y(1) - Y(0) \mid D = 1, X), \tag{5}$$

where D – is a treatment variable, X – is a vector of covariates. The ATT is calculated as an expected value of differences between Y(1) and Y(0) employed all matched units (D – 1) from TG and their pairs from CG.

The full matching procedure leads to an estimation of an *average treatment effect* (*ATE*) [16] utilizes all matched units from TG and CG:

$$ATE = E(Y(1) - Y(0) \mid D, X). \qquad (5')$$

One way of causal effect calculation is the generalized linear regression model [15] that allows us to quantify the effect of an intervention D_i (fertilizers or herbicides) on an outcome variable (Y_i – wheat yield) given other influencing factors X_i:

$$Y_i = aD_i + bX_i + c + \varepsilon_i, \tag{6}$$

where ε_i – an error term, a – a value of ATE or ATT (depending on a method), X_i – a vector of independent variables or covariates, Y_i – an outcome variable.

We should note that a confidence interval (CI) of ATE (or ATT) estimations highlights an amplitude of the causal effects.

To compare the performance of PSM and highlight various ATE estimations, we tested the following methods:

Option 1. Unbalanced sample:

- *t-test* – the Student's t-test [25];
- *U-test* – the Mann-Whitney U-test [27];
- *GLM* – the generalized linear regression approach [15] according to (6).

Option 2. Balanced sample:

• PSM with the full matching techniques, ATE estimation according to (6).

• PSM with the genetic matching techniques, ATT estimation according to (6).

All calculations were performed in R (the free software) with the additional packages MatchIt and Matching [24, 16].

Results and Discussion. To quantify causal effects of wheat production technology intensification, we analyzed the following variants of interventions:

– Fertilizers usage;

– Herbicides usage;

– Application of both treatments.

The concept of calculation is given in Table 2.

Table 2.

The variants of the causal effect estimations

The intervention option	*Treatment Group*	*Control Group*
Fertilizers usage	Fertilizers_Use = 1 Herbicides_Use = 1 or 0	Fertilizers_Use = 0 Herbicides_Use = 1 or 0
Herbicides usage	Herbicides_Use = 1 Fertilizers_Use = 1 or 0	Herbicides_Use = 0 Fertilizers_Use = 1 or 0
Fertilizers & Herbicides	Fertilizers_Use = 1 Herbicides_Use = 1	Fertilizers_Use = 0 Herbicides_Use = 0

To balance environmental conditions, we used the amount of precipitation from April to August in 2012 (*Precipitation*). The farm's characteristics were balanced via information about the scale of crop production (*Sown_area*), farm's specialization (*Farm_Specialization, Labor_ha_Persons*), a legal form of an enterprise (*Type_Ownership*) as well as features of a crop production technology (*Fallow_Share, Wheat_Share*). The sown area shows an overall farm size, and the share of wheat in the sown area indicates how much an enterprise is specialized in wheat production. The indicator *Labor_ha_Persons* showing production load on workers (in hectares), is significantly different for crop and livestock enterprises. Considering that current farming practices like fertilizers and herbicides usage depend on the personal characteristics of farm managers, we also included information about farmers' age and the presence of agricultural education. Table 3 represents logit model estimations to balance the sample in both cases – fertilizers usage (Model F) and herbicides usage (Model H) separately. We also considered that the studied treatments could influence each other. All selected variables are involved with the farm's performance and consequently related to the wheat yield.

Interestingly, as fertilizer application was not a determinant of crop chemical protection while herbicides were not a factor in applying mineral nutrients. Moreover, the shares of wheat and fallow are not indicators of farm propensity to the studied treatments. The presence of agricultural education also is not a significant factor as well. Finally, the farmer's age played minor role.

Table 3.

Determinants affecting intensive crop technology according to logit regression estimates

Variable	**Model F** (Fertilizers usage)		**Model H** (Herbicides usage)	
	Coeff.	St. Error	Coeff.	St. Error
Fertilizer_Use	-	-	0.28	0.39
Herbicides_Use	0.31	0.38	-	-
Type_Ownership	0.61*	0.36	-0.56	0.37
Sown_area	9.7e-05**	3.7e-05	1.1e-04**	4.2e-05
Labor_ha_Persons	-4.0e-03**	1.7e-03	-2.5e-03*	1.2e-03
Farm_Specialization	0.64	0.70	1.90**	0.69
Wheat_Share	0.63	0.95	0.01	0.88
Fallow_Share	2.0	1.0	-2.89	1.1
Precipitation	6.6e-03*	4.1e-03	8.9e-03*	4.7e-03
Age	-9.4e-03	1.9e-02	5.1e-03	1.9e-02
Education	-0.71	0.37	0.21	0.37
Intercept point	*-2.0*	*1.2*	*-1.5*	*1.2*
AIC	*246.4*		*246.9*	
McFadden's pseudo R^2	*0.10*		*0.11*	

Note: *, ** – p-value < 0.1, 0.05 according to z-test in model F or H, respectively.

Considering the statistically significant variables highlighted in logit models and selecting covariates that could improve data balance, we balanced the data in TG and CG using the full and genetic matching procedures. Exact matching options applied for binary covariates like *Fertilizers_Use*, *Herbicides_Use*, *Education,* and *Type_Ownership*. Quality of data balancing is graphically represented by Love plots (Appendixes A, B, and C). In addition, quantitative analysis of data balancing represented by Percent Balance Improvement and values of average standardized mean differences can be found in Appendixes D and E. The full matching gave better data balance than the genetic matching technique.

As shown in Appendixes A and D, *Farm_Specialization* and *Wheat_Share* covariates worsened data balance slightly in the case of fertilizers treatment

(Model F). However, the overall balance of the sample is good, and excluding these variables worsened the result significantly. The full matching showed an excellent balance of data for herbicides usage treatment (Model H) – the balance was improved for all covariates (Appendixes D and E). The genetic matching technique was inconsistent for all covariates, except *Fertilizers_Use* matched according to the exact matching. Thus, the results of causal effect estimation according to genetic matching for herbicides usage treatment can not be considered satisfactory. Data balance in the case of causal effect estimation for both treatments (Appendix C) was not perfect and showed a similar quality for both methods. However, the initial balance was improved for all selected covariates. Restriction of the sample by considering only observations with equal treatment status (Fertilizers_Use=Herbicides_Use=1 for TG; Fertilizers_Use=Herbicides_Use=0 for CG) resulted in reducing the sample to 93 observations – 46 in TG and 46 in CG.

Assessments of the causal effects of production technology intensification on yield wheat are given in Table 4. A positive effect of intensification can be seen from the wheat-growing practices in the Kulunda steppe of Altai Krai. According to the expected value of causal effect, the results showed that crop chemical treatment and fertilizers might increase the wheat yield by 1.0-3.5 dt/ha even under drought conditions in the Kulunda Steppe of Altai Krai. It could have brought an additional profit when an average selling price per ton of grain of 8,500 rubles in 2012 [8]. The cost of herbicides would have been around 280 rubles per hectare, while the cost of fertilizers would have been 328 rubles per hectare on average.

In general, according to the full matching, the causal effects within the 95% confidence interval (CI) resulting from fertilizers input ranged from 0 to 2.0 dt/ha, from herbicides application varied from 0.7 to 2.8 dt/ha. Application of both treatments led to increased wheat yield by 0.6-3.4 dt/ha. The considered interventions, among other factors, were statistically significant variables that explained the difference of wheat yield between TG and CG (p-value < 0.1). We should also note that all methods showed statistically significant differences in wheat yield between TG and CG.

The results also showed that the calculation of causal effect does not satisfy the additivity criterion. Thus, it confirmed that the sum of effects (fertilizers and herbicides usage) does not equal the causal effect of both treatments.

We also confirmed that the simple statistical techniques and the genetic matching routing, which gave insufficient data balance, estimated the causal effect higher than the full matching procedure. Comparative evaluations in agriculture on large-scale areas such as the Kulunda steppe have to be done considering various factors met on farm level (scale of production, legal form, climatic conditions, farm specialization, etc.) due to heterogeneity of production units and variability of environmental conditions. Therefore, we concluded that the full matching estimations were more confident compared with others.

Table 4.

Results of the causal effect estimations of crop technology intensification on wheat yield, Kulunda steppe, Altai Krai, Russia, 2012

Method	Statistics	Wheat production technology intensification		
		Fertilizers usage	*Herbicides usage*	*Herbicides & Fertilizers*
Means Comparison:				
T-test	Expected value 95% CI p-value	1.85 (0.68; 3.02) 0.002	2.11 (1.09; 3.13) 0.001	3.47 (2.02; 4.51) 0.001
U-test	Expected value 95% CI p-value	1.60 (0.60; 2.6) 0.002	2.00 (1.10; 2.90) 0.001	3.20 (2.00; 4.50) 0.001
Regression analysis:				
Simple linear regression	Expected value 95% CI p-value	1.85 (0.79; 2.91) 0.001	2.11 (1.05; 3.18) 0.001	3.47 (2.00; 4.94) 0.001
Multiple linear regression	Expected value 95% CI p-value	1.50 (0.48; 2.58) 0.007	2.06 (0.94; 3.18) 0.001	3.12 (1.34; 4.91) 0.001
Propensity Score Matching:				
Full matching method	Expected value 95% CI p-value	0.96 (-0.03; 1.95) 0.060	1.75 (0.72; 2.77) 0.001	2.02 (0.61; 3.42) 0.006
Genetic matching method	Expected value 95% CI p-value	1.29 (-0.22; 2.80) 0.097	2.12 (0.63; 3.60) 0.006	2.71 (0.85; 4.60) 0.005

Notes: CI – the confidence interval.

Conclusion

This study investigated the causal effect of chemical crop protection agents (including herbicides) and fertilizers usage on wheat yield in the Kulunda steppe in Altai Krai under drought conditions of 2012. Based on the dataset collected from the farmers' survey and annual reports, we revealed that intensification of wheat production technology is one of the possible ways to reduce the yield gap in Altai

Krai. All methods showed an improvement in wheat yield, which would cause the average wheat yield to increase about 18-64% for farmers who declined chemical treatment and fertilizers usage due to various reasons. In fact, they had only 5.4 dt/ha grain wheat reduced their net profit. We should note that we highlighted positive expected wheat yield growth without clarifying dozes and types of plant protection products as well as kinds of used fertilizers. The laters significantly define yield variability within a treated group of farmers. We could not include such information in the dataset due to the enormous variability of treatments in current farming practices. That creates considerable difficulties in detecting binary forms of interventions. For instance, we did not exclude crop rotation as a treatment variable of crop production technology for the same reasons. Moreover, such causal effects (nutrient doze or chemical plant protection product) can be estimated correctly within the field experiments under local climatic and soil conditions under the same soil tillage system.

Increasing crop productivity, particularly wheat, is a core problem of agricultural development and directly affects the profitability of agricultural enterprises. We believe that this study can raise the question of the importance of crop production intensification, especially in the light of global warming occurring, effects of which have already been seen in Altai Krai. Droughts as one in 2012 occurred more often in the last years. We believe that intensification of crop production in line with the introduction of modern digital services in agriculture, growing acclimatized wheat varieties would lead to a significant improvement in wheat yield in Altai Krai, thus improving the local economy. Therefore, we propose that additional funding from the local and state governments to develop some digital agricultural services, which have analytical functions and can help farmers quantify the causal effect on their cultivated fields, would help drive the process of intensification. Additionally, it is helpful to implement such services at a regional level to assess the current effect of different crop production strategies on a regional scale.

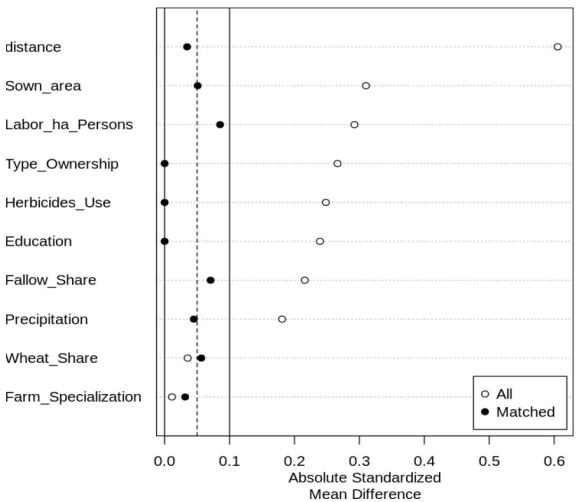
The advantages of the PSM method over other methods are primarily in applying pseudorandomization of data, which makes it possible to consider the resulting causal effect assessment close to the true causal effect and reduce the bias. This method also makes it possible to carry out studies close to experimental conditions on datasets from large areas where experimental research is impossible or extremely difficult to conduct, such as agriculture. Balancing data via the PSM method provides unit comparisons and causal effect estimation with greater correctness than simple statistical methods. However, it is worth mentioning that this method sometimes worsens a data balance [18; 19]. Therefore, it is essential to follow the implementation procedure and check the data balance at the crucial stages. It is also important to double-check the list of covariates since missing even one causes more bias, thus rendering all data balancing made after this useless.

In light of widespread digital services implementation in agriculture, PSM algorithms can be adopted as an element of an applied digital tool in smart agricultural systems. It is possible to create a cloud-based digital system, which can suggest the best amount of fertilizers and herbicides used, based on combining methods and data from various digital soil moisture and climatic monitoring systems. Implementing such technologies in the Kulunda steppe, including areas in Kazakhstan, would allow farmers to prevent profit loss and minimize the yield gap under uncertain climatic conditions. We believe that is the future direction of digital technology development in agriculture.

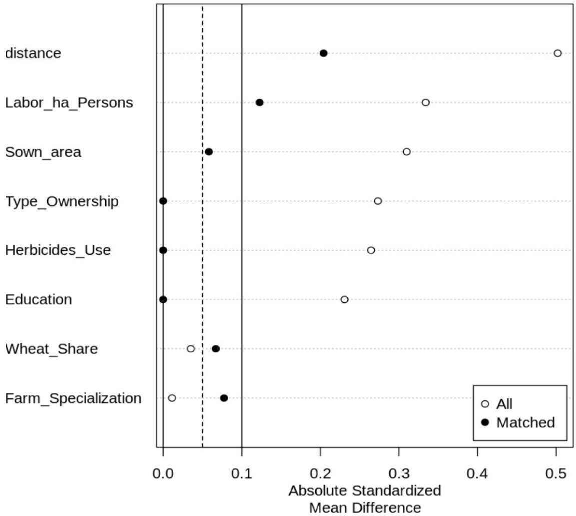
Acknowledgments

The authors are thankful to Prof. Alexander V. Prishchepov from Copenhagen University, Prof. Valentina Kundius, and Prof. Irina Kovaleva from the Altai State Agricultural University for supporting the questionnaire design and data collection in Altai Krai. The study has been funded by The Grant of the Science Committee in the Ministry of Education and Science of the Kazakhstan Republic AR 09259379.

Appendix A. Results of the data balancing for fertilizer usage treatment.



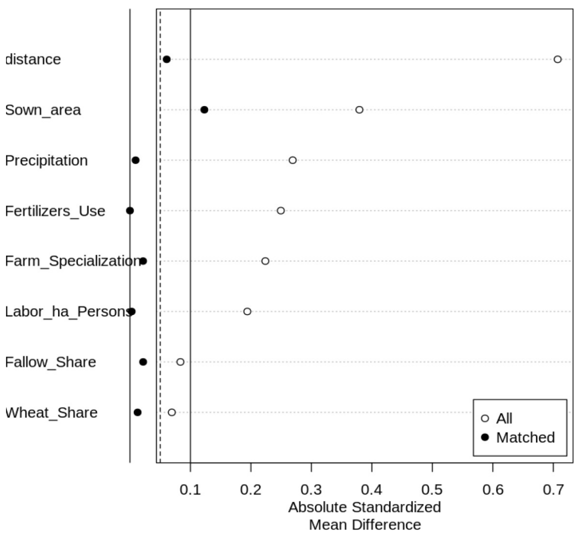
a) Full Matching method



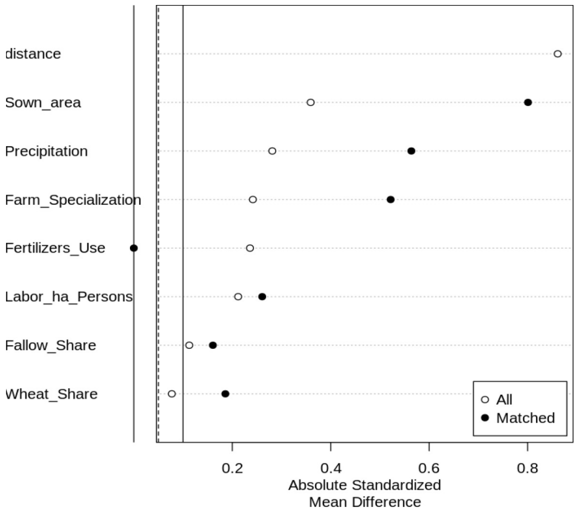
b) Genetic Matching method

Source: Authors' calculations.

Appendix B. Results of the data balancing for herbicide usage treatment.

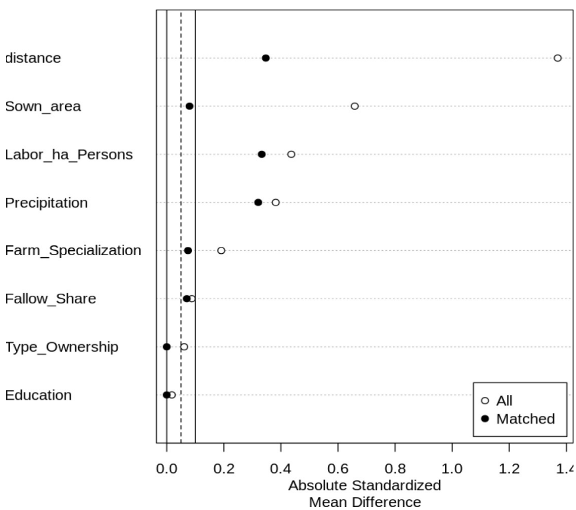


a) Full Matching method.

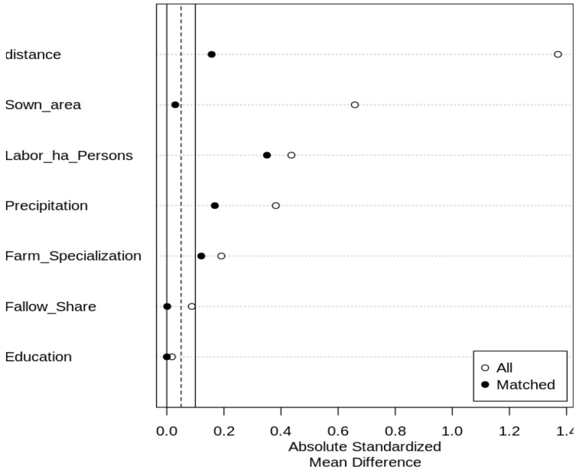


b) Genetic Matching method.
Source: Authors' calculations.

Appendix C. Results of the data balancing for both treatment usage (Fertilizers & Herbicides).



a) Full Matching method



b) Genetic Matching method.
Source: Authors' calculations.

Appendix D. Data balance improvement, fertilizer usage treatment.

Variable	Initial balance (All data)			Matched data			Percent Balance Improve-ment
	Mean TG	Mean CG	Std. Mean Diff.	Mean TG	Mean CG	Std. Mean Diff.	
Full matching							
Distance	0.36	0.29	0.60	0.31	0.31	0.03	94.3
Sown_area	8297	6632	0.31	7468	7195	0.05	83.6
Labor_ha_Persons	125	164	-0.29	149	160	-0.08	70.8
Farm_Specialization	0.53	0.54	-0.01	0.54	0.55	-0.03	-180.8
Wheat_Share	0.44	0.45	-0.04	0.44	0.46	-0.06	-59.2
Fallow_Share	0.17	0.14	0.21	0.14	0.15	-0.07	67.2
Precipitation	90.8	84.1	0.18	87.0	85.3	0.04	75.2
Type_Ownership	0.65	0.52	0.26	0.57	0.57	0	100
Education	0.59	0.70	-0.24	0.66	0.66	0	100
Herbicides_Use	0.77	0.66	0.25	0.69	0.69	0	100
Obs. *Unmatched*	*61*	*135*	-	*61* *0*	*135* *0*	-	-
Genetic Matching							
Distance	0.34	0.29	0.50	0.35	0.33	0.26	48.0
Sown_area	8297	6632	0.31	8375	7736	0.11	61.6

End of a table

Labor_ha_Persons	125.3	164.3	-0.33	109.8	125.5	-0.13	59.6
Farm_Specialization	0.53	0.54	-0.01	0.52	0.50	0.07	-524.1
Wheat_Share	0.44	0.45	-0.03	0.45	0.44	0.06	-88.1
Type_Ownership	0.65	0.52	0.27	0.62	0.62	0	100
Education	0.59	0.70	-0.23	0.64	0.64	0	100
Herbicides_Use	0.77	0.65	0.26	0.75	0.75	0	100
Obs. *Unmatched*	*61*	*135*	-	*56* *5*	*56* *79*	-	-

Source: Authors' calculations.

Appendix E. Data balance improvement, herbicide usage treatment.

Variable	Initial balance (All data)			Matched sample			Percent Balance Improvement
	Means Treated	Means Control	Std. Mean Diff.	Means Treated	Means Control	Std. Mean Diff.	
Full matching							
Distance	0.73	0.62	0.71	0.70	0.69	0.06	91.4
Sown_area	7757	5775	0.38	7279	6637	0.12	67.6
Labor_ha_Persons	143.6	171.5	-0.19	158.3	158.7	-0.00	98.5
Farm_Specialization	0.56	0.49	0.22	0.54	0.55	-0.02	90.3
Wheat_Share	0.44	0.46	-0.07	0.44	0.44	-0.01	81.7
Fallow_Share	0.14	0.16	-0.08	0.14	0.13	0.02	73.9
Precipitation	89.3	79.0	0.27	87.8	88.2	-0.01	96.5
Fertilizers_Use	0.35	0.23	0.25	0.31	0.31	0	100.0
Obs. *Unmatched*	*60*	*136*	-	*60* *0*	*136* *0*	-	-
Genetic Matching							
Distance	0.72	0.61	0.86	0.82	0.62	1.64	-90.2
Sown_area	7757	5774	0.35	10193.9	5774.7	0.80	-122.9
Labor_ha_Persons	143.5	171.5	-0.21	137.1	171.6	-0.26	-23.1
Farm_Specialization	0.56	0.49	0.24	0.64	0.49	0.52	-115.7
Wheat_Share	0.44	0.46	-0.07	0.42	0.46	-0.19	-140.8
Fallow_Share	0.14	0.16	-0.11	0.14	0.16	-0.16	-42.6
Precipitation	89.3	79.0	0.28	99.7	79.0	0.56	-100.4
Fertilizers_Use	0.34	0.23	0.23	0.23	0.23	0	0
Obs. *Unmatched*	*60*	*136*	-	*60* *0*	*60* *76*	-	-

Source: Authors' calculations.

References

1. *Agroekologicheskiy atlas Rossii i sopredel'nykh stran 2008 g.* [Agroecological Atlas for Russia and Transborder Countries, 2008]. URL: http://www.agroatlas.ru/ru/index.html (accessed June 03, 2020)
2. *V Altayskom kraye podveli itogi raboty APK za 2019 god 2019 g.* [The results of work in agro-industrial complex of Altai Krai for 2019 were summarized, 2019]. URL: https://www.kommersant.ru/doc/4196634. (accessed June 03, 2020)
3. Grigor'eva E.S. *Chto Dolzhen Znat' Spetsialist ob Osobennostyakh Biologii Grecheskikh, Sudovykh Trav i Mnogoletnikh Trav i Tekhnologiyakh Vozdelyvaniya Kul'tur* [What Have to Know a Specialist About Features of Crop Biology and Technologies of Its Growing]. Barnaul: Ministry of Agriculture of the Altai Krai, 2001, 233 p.
4. *Doklad o khode i registratsii realizatsii v 2018 godu, 2019 g.* [Report about the progress and results of the implementation in 2018 of state programs in the field of agricultural development of the Altai Krai]. URL: https://altagro22.ru/activity/analytics/doklad-o-khode-i-rezultatakh-realizatsii-v-2018-godu-gosudarstvennykh-programm-v-sfere-razvitiya-sel/ (accessed October 09, 2020)
5. *Kuban' sobrala rekordnyy urozhay zerna. Srednyaya urozhaynost' zernovykh v strane pochti sravnyalas' s proshlogodney, 2019 g.* [Kuban harvested a record amount of grain, 2020]. URL: https://www.agroinvestor.ru/regions/news/32115-kuban-sobrala-rekordnyy-urozhay-zerna/ (accessed September 23, 2020)
6. *"Lider 80" zadayet ton Altayskiye selektsionery sozdali novyy sort pshenitsy, 2020 g.* [Lider 80, Altai breeders created a new variety of wheat, 2020]. URL: https://rg.ru/2019/12/17/reg-sibfo/altajskie-selekcionery-sozdali-novyj-sort-pshenicy.html (accessed September 23, 2020)
7. *Meneye 5 tsentnerov s gektara. V Altayskom kraye ubrali chast' zernovykh, 2020 g.* [Less than 5 centners per hectare. Part of grain was harvested in Altai Krai, 2020]. URL: https://www.politsib.ru/news/39560-menee-centnerov-s-gektara-v-altajskom-krae-ubrali-cast-zernovyh (accessed September 23, 2020)
8. *Monitoring tsen na pshenitsu, 2012 g.* [Monitoring of wheat prices by organizations, 2012]. URL: https://agro-bursa.ru/prices/wheat/01-10-2012/ (June 06, 2020)
9. *Regiony lidery po sboru zerna, 2016 g.* [Leader regions in grain harvesting, 2016]. URL: https://www.agroinvestor.ru/regions/article/22996-regiony-lidery-po-sboru-zerna/ (accessed September 23, 2020)
10. *Federal'naya sluzhba gosudarstvennoy statistiki, 2020 g.* [Federal State Statistics Service, 2020]. URL: https://www.fedstat.ru/ (accessed October 09, 2020)

11. Ponkina, E., Beljaev, V., Frühauf, M., Bugay, Y., Kasarjyan, M., Meinel, T., Bavorova, M., Kovaleva, I., Kozhanov, N., & Prishchepov, A. *Ekonomicheskiye, Ekologicheskiye, Tekhnologicheskiye Faktory i Rezul'taty Deyatel'nosti Sel'skokhozyaystvennykh Predpriyatiy v Usloviyakh Kulundinskoy Stepi* [Economic, Ecological, Technological Factors and Performance of Agricultural Enterprises in Kulunda Steppe Conditions]. Barnaul: Altai State University Publ., 2014, 140 p.
12. Diamond A., Sekhon J.S. Genetic matching for estimating causal effects: a general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, 2013, vol. 95, no. 3, pp. 932-945. https://doi.org/10.1162/REST_a_00318
13. Dronin N.M., Bellinger, E.G. *Climate dependence and food problems in Russia, 1900-1990: The interaction of climate and agricultural policy and their effect on food problems.* Budapest: Central European University Press, 2005, 383 p.
14. *Exact Matching*, 2010. URL: https://r.iq.harvard.edu/docs/matchit/2.4-15/Exact_Matching.html (accessed June 14, 2020)
15. Generalized linear models / Rindskopf D. // *APA handbook of research methods in psychology*, *Vol 3: Data analysis and research publication*. [In eds H. Cooper, P.M. Camic, D.L. Long, A.T. Panter, D. Rindskopf, K.J. Sher]. New York: American Psychological Association, 2012, pp. 191-206. https://doi.org/10.1037/13621-009
16. Ho D.E., Imai K., King G., Stuart E.A. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 2007, vol. 15, no. 3, pp. 199-236. https://doi.org/10.1093/pan/mpl013
17. Imbens G.W., Rubin D.B. *Rubin causal model*. London: Palgrave Macmillan, 2008. 10 p. https://doi.org/10.1057/978-1-349-95121-5_2469-1
18. King G., Lucas C, Nielsen R.A. The balance-sample size frontier in matching methods for causal inference. *American journal of political science*, 2016, vol. 61, no. 2, pp. 1-17. https://doi.org/10.1111/ajps.12272
19. King G., Nielsen R. Why propensity scores should not be used for matching. *Political analysis*, 2019, vol. 27, no 4, pp. 435-454. https://doi.org/10.1017/pan.2019.11
20. Kogan F., Guo W. Early twenty-first-century droughts during the warmest climate. *Geomatics, natural hazards and risk*, 2016, vol. 7, no. 1, pp. 127-137. https://doi.org/10.1080/19475705.2013.878399
21. Prishchepov A.V., Ponkina E., Sun Z., Müller D. Revealing the determinants of wheat yields in the Siberian breadbasket of Russia with Bayesian networks. *Land use policy*, 2019, vol. 80, pp. 21-31. https://doi.org/10.1016/j.landusepol.2018.09.038

22. Rubin D.B. Bias reduction using mahalanobis-metric matching. *Biometrics*, 1980, vol. 36, no. 2, p. 293. https://doi.org/10.2307/2529981
23. Rubin D.B. *Basic concepts of statistical inference for causal effects in experiments and observational studies*. Cambridge: Harvard University, 2004, 140 p.
24. *Statistics and Causal Inference*, 2014. URL: https://imai.fas.harvard.edu/talk/files/Taiwan14.pdf (accessed June 12, 2020)
25. Student's t-Tests / Kalpić D., Hlupić N., Lovrić, M. // *International encyclopedia of statistical science*. [In ed M. Lovric]. Berlin: Springer, 2011. Pp. 1559-1563. https://doi.org/10.1007/978-3-642-04898-2_641
26. *Understanding Q-Q Plots. University of Virginia Library Research Data Services + Sciences*, 2015. URL: https://data.library.virginia.edu/understanding-q-q-plots/ (accessed May 28, 2020)
27. Wilcoxon–Mann–Whitney Test / Neuhäuser M. // *International encyclopedia of statistical science*. Berlin: Springer, 2011, pp. 1656-1658. https://doi.org/10.1007/978-3-642-04898-2_615

Список литературы

1. Агроэкологический атлас России и сопредельных стран, 2008 г. URL: http://www.agroatlas.ru/ru/index.html (дата обращения: 03.06.2020)
2. В Алтайском крае подвели итоги работы АПК за 2019 год, 2019 г. URL: https://www.kommersant.ru/doc/4196634 (дата обращения: 03.06.2020)
3. Григорьева Е.С. Что должен знать специалист об особенностях биологии гречихи, суданской травы и многолетних трав и технологии возделывания культур. Б.: Министерство сельского хозяйства Алтайского края, 2001. 233 с. https://rusneb.ru/catalog/000199_000009_004742743/
4. Доклад о ходе и результатах реализации в 2018 году государственных программ в сфере развития сел, 2019 г. URL: https://altagro22.ru/activity/analytics/doklad-o-khode-i-rezultatakh-realizatsii-v-2018-godu-gosudarstvennykh-programm-v-sfere-razvitiya-sel/ (дата обращения: 09.10.2020)
5. Кубань собрала рекордный урожай зерна. Средняя урожайность зерновых в стране почти сравнялась с прошлогодней, 2019 г. URL: https://www.agroinvestor.ru/regions/news/32115-kuban-sobrala-rekordnyy-urozhay-zerna/ (дата обращения: 23.09.2020)
6. "Лидер 80" задает тон Алтайские селекционеры создали новый сорт пшеницы, 2020 г. URL: https://rg.ru/2019/12/17/reg-sibfo/altajskie-selekcionery-sozdali-novyj-sort-pshenicy.html (дата обращения: 23.09.2020)
7. Менее 5 центнеров с гектара. В Алтайском крае убрали часть зерновых, 2020 г. URL: https://www.politsib.ru/news/39560-menee-centnerov-s-gektara-v-altajskom-krae-ubrali-cast-zernovyh (дата обращения: 23.09.2020)

8. Мониторинг цен на пшеницу, 2012 г. URL: https://agro-bursa.ru/prices/wheat/01-10-2012/ (дата обращения: 06.06.2020)
9. Регионы лидеры по сбору зерна, 2016 г. URL: https://www.agroinvestor.ru/regions/article/22996-regiony-lidery-po-sboru-zerna/ (дата обращения: 23.09.2020)
10. Федеральная служба государственной статистики, 2020 г. URL: https://www.fedstat.ru/ (дата обращения: 09.10.2020)
11. Экономические, экологические, технологические факторы и результаты деятельности сельскохозяйственных предприятий в условиях кулундинской степи / Понькина Е., Беляев В., Фрюхауф М., Бугай Ю., Касарджян М., Мейнель Т., Баворова М., Ковалева И., Кожанов Н., Прищепов А. Б.: Издательство Алтайского государственного университета, 2014. 140 с. https://www.researchgate.net/publication/262866647_EKONOMICESKIE_EKOLOGICESKIE_TEHNOLOGICESKIE_FAKTORY_I_REZULTATY_DEATELNOSTI_SELSKOHOZAJSTVENNYH_PREDPRIATIJ_V_USLOVIAH_KULUNDINSKOJ_STEPI
12. Diamond A., Sekhon J.S. Genetic matching for estimating causal effects: a general multivariate matching method for achieving balance in observational studies // Review of Economics and Statistics, 2013, vol. 95, no. 3, pp. 932-945. https://doi.org/10.1162/REST_a_00318
13. Dronin N.M., Bellinger, E.G. Climate dependence and food problems in Russia, 1900-1990: The interaction of climate and agricultural policy and their effect on food problems. Budapest: Central European University Press, 2005. 383 p. http://www.gbv.de/dms/zbw/478845847.pdf
14. Exact Matching, 2010. URL: https://r.iq.harvard.edu/docs/matchit/2.4-15/Exact_Matching.html (accessed June 14, 2020)
15. Generalized linear models / Rindskopf D. // APA handbook of research methods in psychology, Vol 3: Data analysis and research publication. [In eds H. Cooper, P.M. Camic, D.L. Long, A.T. Panter, D. Rindskopf, K.J. Sher]. New York: American Psychological Association, 2012, pp. 191-206. https://doi.org/10.1037/13621-009
16. Ho D.E., Imai K., King G., Stuart E.A. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference // Political Analysis, 2007, vol. 15, no. 3, pp. 199-236. https://doi.org/10.1093/pan/mpl013
17. Imbens G.W., Rubin D.B. Rubin causal model. London: Palgrave Macmillan, 2008. 10 p. https://doi.org/10.1057/978-1-349-95121-5_2469-1
18. King G., Lucas C, Nielsen R.A. The balance-sample size frontier in matching methods for causal inference // American journal of political science, 2016, vol. 61, no. 2, pp. 1-17. https://doi.org/10.1111/ajps.12272
19. King G., Nielsen R. Why propensity scores should not be used for matching // Political analysis, 2019, vol. 27, no 4, pp. 435-454. https://doi.org/10.1017/pan.2019.11

20. Kogan F., Guo W. Early twenty-first-century droughts during the warmest climate // Geomatics, natural hazards and risk, 2016, vol. 7, no. 1, pp. 127-137. https://doi.org/10.1080/19475705.2013.878399
21. Prishchepov A.V., Ponkina E., Sun Z., Müller D. Revealing the determinants of wheat yields in the Siberian breadbasket of Russia with Bayesian networks // Land use policy, 2019, vol. 80, pp. 21-31. https://doi.org/10.1016/j.landusepol.2018.09.038
22. Rubin D.B. Bias reduction using mahalanobis-metric matching // Biometrics, 1980, vol. 36, no. 2, p. 293. https://doi.org/10.2307/2529981
23. Rubin D.B. Basic concepts of statistical inference for causal effects in experiments and observational studies. Cambridge: Harvard University, 2004, 140 p.
24. Statistics and Causal Inference, 2014. URL: https://imai.fas.harvard.edu/talk/files/Taiwan14.pdf (accessed June 12, 2020)
25. Student's t-Tests / Kalpić D., Hlupić N., Lovrić, M. // International encyclopedia of statistical science. [In ed M. Lovric]. Berlin: Springer, 2011. Pp. 1559-1563. https://doi.org/10.1007/978-3-642-04898-2_641
26. Understanding Q-Q Plots. University of Virginia Library Research Data Services + Sciences, 2015. URL: https://data.library.virginia.edu/understanding-q-q-plots/ (accessed May 28, 2020)
27. Wilcoxon–Mann–Whitney Test / Neuhäuser M. // International encyclopedia of statistical science. Berlin: Springer, 2011, pp. 1656-1658. https://doi.org/10.1007/978-3-642-04898-2_615

ДАННЫЕ ОБ АВТОРАХ

Тарасов Кирилл Олегович

Алтайский Государственный Университет
просп. Ленина, 61, г. Барнаул, 656049, Российская Федерация
kirilltarasov99@gmail.com

Понькина Елена Владимировна

Алтайский Государственный Университет
просп. Ленина, 61, г. Барнаул, 656049, Российская Федерация
ponkinaelena77@mail.ru

Нугуманова Алия Багдатовна

Восточно-Казахстанский университет им. Сарсена Аманжолова
ул. 30-й Гвардейской дивизии, 34, г. Усть-Каменогорск, Республика Казахстан
yalisha@yandex.kz

DATA ABOUT THE AUTHORS
Kirill O. Tarasov
Altai State University
61, Lenin Ave., Barnaul, 656049, Russian Federation
kirilltarasov99@gmail.com
ORCID: https://orcid.org/0000-0003-0691-5351

Elena V. Ponkina
Altai State University
61, Lenin Ave., Barnaul, 656049, Russian Federation
ponkinaelena77@mail.ru
ORCID: https://orcid.org/0000-0001-7604-6337

Aliya B. Nugumanova
Sarsen Amanzholov East Kazakhstan University
34, 30 Gvardeiskoi divisii Str., Ust-Kamenogorsk, 070002, Republic of Kazakhstan
yalisha@yandex.kz
ORCID: https://orcid.org/0000-0001-5522-4421

Поступила 23.02.2022
После рецензирования 02.03.2022
Принята 04.03.2022

Received 23.02.2022
Revised 02.03.2022
Accepted 04.03.2022