Bulgarian Journal of Agricultural Science, 29 (Suppl. 1) 2023
Agricultural Academy

Agricultural land price through the focus of NSI, Eurostat, BAALO data

Mihaela Mihailova* **and Stefan Asenov**

Agricultural academy, Institute of agricultural economics
**Corresponding author:* m.mihailova92@gmail.com

Abstract

Mihailova, M. & Asenov, S. (2023). Agricultural land price through the focus of NSI, Eurostat, BAALO data. *Bulg. J. Agric. Sci., 29* (Supplement 1), 39–48

Land prices are integral for a variety of stakeholders, ranging from urban planners to real estate investors. Land prices play a pivotal role in economic theory, influencing a myriad of decisions from urban planning to investment strategies. The tree datasets we are focusing on are from separate sources present information on a land prices in Bulgaria, it becomes essential to determine their consistency. Linear regression serves as a powerful tool to achieve this and a second method of MANOVA is applied to understand the tree datasets. The aim of this study is to prove the reliability of the information on the agricultural resource.

Keywords: land price; MANOVA; linear regression

Introduction

Land prices are integral for a variety of stakeholders, ranging from urban planners to real estate investors. Land prices play a pivotal role in economic theory, influencing a myriad of decisions from urban planning to investment strategies. The valuation of land is not just a reflection of its physical attributes but also encapsulates the economic, social, and political dynamics of a region. This essay delves into the significance of land prices in economic theory, drawing from various scholarly sources. Gregory (1992) delves into the intricacies of rents and land prices, emphasizing their role in economic theory. He provides a historical perspective, shedding light on the evolution of land prices and their implications for economic structures and societal norms (Gregory, 1992).

Misztal et al. (2022) research underscores the interplay between land prices, economic factors, and sustainable development in the EU context (Misztal et al., 2022). Khitov (2020) discusses the economic interactions between developing and developed countries within the EU framework, taking Bulgaria as a case study. The research highlights the economic dynamics and their impact on land prices in Bulgaria and the broader EU context (Khitov, 2020). Investing in agricultural land involves considering a variety of factors, including the level of subsidisation of agricultural production, the profitability of the agricultural products produced, the country's tax system, access to credit resources, inflationary processes in the economy, etc (Kirechev, 2022). However, for further research of land prices and factors for their change in Bulgaria the reliability of land price data can vary based on its source and it is important for us to further study the tree available sets for this study. When tree datasets from separate sources present information on a similar subject matter, it becomes essential to determine their consistency. Linear regression serves as a powerful tool to achieve this and a second method of MANOVA is applied to understand the tree datasets. The aim of this study is to prove the reliability of the information on the agricultural land prices collected by the statistical offices of the NSI, Eurostat and the Bulgarian Association of Agricultural Land Owners and use this sets for further studies of land prices.

Methodology

For the purpose of this study, we as researchers aim to check if the tree datasets, we want to use form different sources and different methodology for extraction of the data, processing and sampling, are statistically different. The tree datasets

are BAALO data, NSI data and Eurostat. We are testing the hypothesis that there are differences between the sets but is not significant. Firstly, will do it using linear regression and R^2, if there is no difference the two datasets tested would correlate. In the context of two datasets from different sources, the fact is that if R^2 is less than 100% it does suggest differences between the datasets. The sound thing we will do is MANOVA analysis that requires several steps. Test for normal distribution is performed and a test for autocorrelation between the sets is done. The study is for period of 13 years (N13). As for this period there is years where there is no data for these years the data is substituted using the linear model of forecast in Excel.

We are going to go through several steps to determine if our assumption/hypothesis that the tree datasets are different is true and then proceed with the MANOVA.

1. Step descriptive statistic of the data of the distribution of the data
2. Check if the data is normally distributed using the Kolmogorov-Smirnova Shapiro-Wilk.
3. After checking the distribution chose a parametric or nonparametric test for the correct distribution and connected samples.

The statistics table provided offers insights into the pricing of a hectare of land based on three different data sources – "BAALO data" (Price per hectare – BAALO), "Eurostat" (Price Eurostat), and "NSI" (Price NSI). Each of the sources presents the same number of valid observations (N = 13), indicating that the data is complete, with no missing values in any of the three datasets.

The program used for processing the datasets is SPSS.

Firstly, we will present the descriptives of the data and preform a test of Statistical Normality Tests: Kolmogorov-Smirnov and Shapiro-Wilk is made. For the KS test and the SW test, the null hypothesis (H0) states that the sample data comes from a normally distributed population. The alternative hypothesis (H1) asserts that the sample data does not come from a normally distributed population (Shapiro & Wilk, 1965; Massey, 1951).

The Kolmogorov-Smirnov Test begins by comparing maximum absolute difference between the EDF and CDF is calculated, and this difference, D, is the test statistic (Massey, 1951).

The Shapiro-Wilk Test, on the other hand, measures the linear correlation between the ordered sample data and the corresponding expected values under the normal distribution and the W, is calculated from this correlation (Shapiro & Wilk, 1965). If the calculated p-value is less than the chosen significance level, the null hypothesis is rejected, suggesting that the data do not follow a normal distribution. It's important to note that these tests are sensitive to large sample sizes. Even minor departures from normality can result in significant p-values when the sample size is large. Therefore, researchers should also consider graphical methods, like Q-Q plots, and measures of skewness and kurtosis, to assess the normality of data (Razali & Wah, 2011).

This methodology section is dedicated to the application of the Multivariate Analysis of Variance (MANOVA), a statistical technique used to examine the differences in two or more continuous dependent variables by an independent grouping variable (Tabachnick & Fidell, 2013).

1. Hypotheses Formulation

Before conducting the MANOVA, the researcher needs to formulate the null and alternative hypotheses. The null hypothesis (H0) states that there is no significant difference in the mean vectors of the dependent variables across the levels of the independent variable. The alternative hypothesis (H1) asserts the opposite: there is a significant difference in the mean vectors of the dependent variables across the levels of the independent variable (Rencher, 2002).

2. Assumption Checking

Prior to running a MANOVA, several assumptions should be checked: multivariate normality, linearity, homogeneity of variance-covariance matrices, and absence of multicollinearity and singularity (Tabachnick & Fidell, 2013).

3. Performing MANOVA

MANOVA is conducted using statistical software SPSS. Once the MANOVA is performed, the researcher needs to check the overall test statistics, Wilks' Lambda, Pillai's Trace, Hotelling's Trace, or Roy's Largest Root, depending on the design and assumptions of the study (Tabachnick & Fidell, 2013). A significant result (p-value less than the chosen significance level, typically 0.05) rejects the null hypothesis, suggesting there is a significant difference in the mean vectors of the dependent variables across the levels of the independent variable.

If the overall MANOVA is significant, univariate ANOVAs are typically conducted as follow-up tests to explore which dependent variables contribute most to the group differences (Rencher, 2002).

Results and Discussion

Firstly, to understand more about the tree sets we are comparing we are using descriptive statistics.

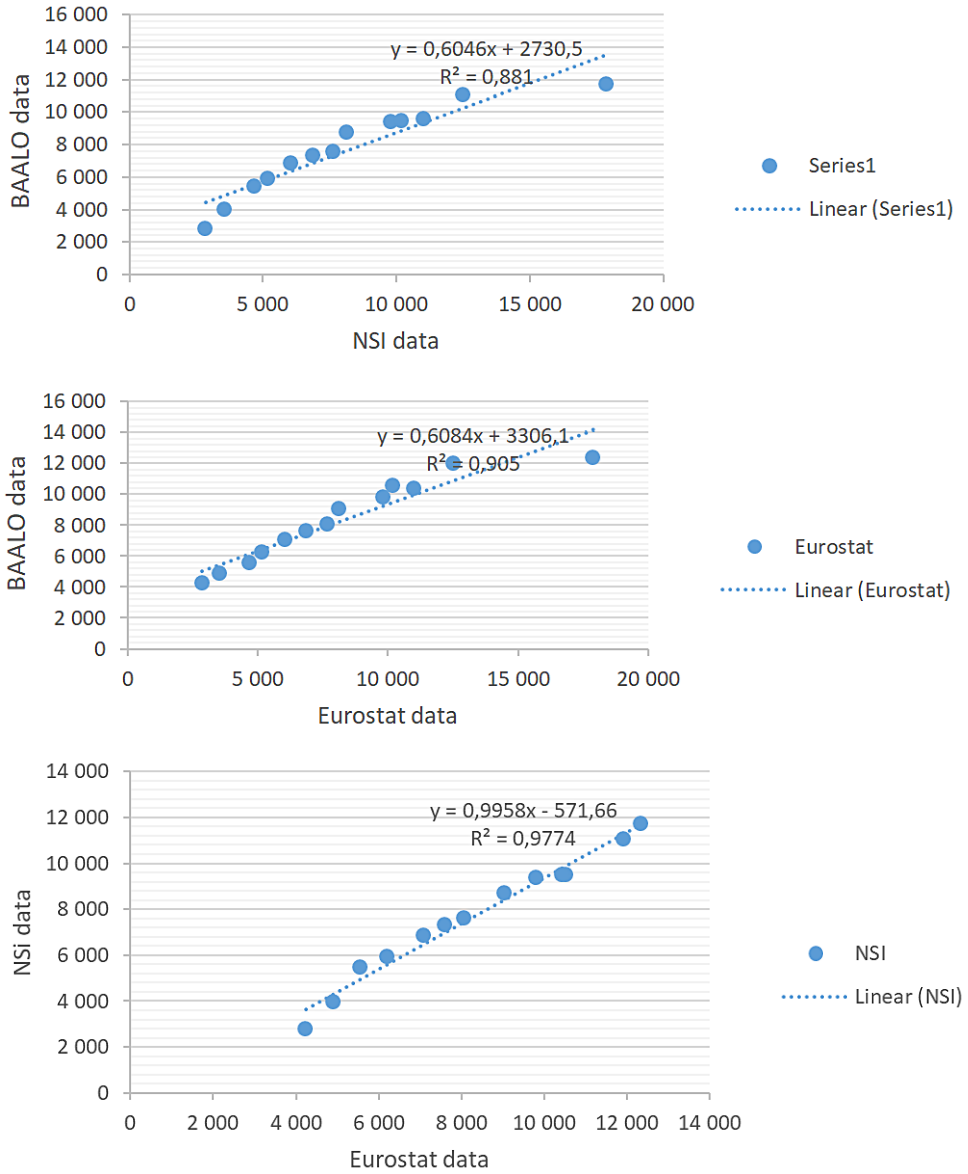
Eurostat has the smallest standard error of mean (731.924), suggesting the highest precision in estimating the mean compared to the other two sources. The median values for the three datasets are fairly close to their corresponding mean values, suggesting a somewhat symmetrical distribution. The BAALO data set has the largest standard

deviation, suggesting a wider spread of data around the mean compared to the other two sources. The variance is highest for the BAALO dataset source, reinforcing the notion of a wider data spread for this source.

BAALO data source and Eurostat data have positive skewness (.971 and .017 respectively), while NSI has a slightly negative skewness (-.325), suggesting a slight left skew in its distribution. BAALO data source displays a leptokurtic distribution, while Eurostat and NSI show platykurtic distributions. "Range" represents the difference between the highest and lowest values in a dataset. BAALO data source has the greatest range, suggesting a wider spread of data.

Linear regression

The R^2 value, also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent



Graph. 1. Linear regression of NSI, Eurostat and BAALO data

Source: NSI, Eurostat and BAALO data

variable(s). It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.

Given the R^2 values for the three datasets:

BAALO data and NSI $R^2 = 0.881$
BAALO data and Eurostat $R^2 = 0.905$
NSI and Eurostat $R^2 = 0.977$
Let's interpret them:

- BAALO data and NSI $R^2 = 0.881$: This suggests that 88.1% of the variance in the NSI dataset can be explained by the BAALO data. This is a high R^2 value, indicating a strong linear relationship between the two datasets.
- BAALO data and Eurostat R^2 0.905 Here, 90.5% of the variance in the Eurostat dataset can be explained by the BAALO data. Again, this is a high R^2 value, suggesting a strong linear relationship.
- NSI and Eurostat $R^2 = 0.977$ an impressive 97.7% of the variance in the Eurostat dataset can be explained by the NSI data. This is very close to 1, indicating an extremely strong linear relationship between NSI and Eurostat datasets.

How similar are the three datasets?

Given the high R^2 values, it can be inferred that there is a strong linear relationship between each pair of datasets. The strongest relationship is between NSI and Eurostat, followed closely by BAALO data with Eurostat, and then BAALO data with NSI. However, if two datasets are being compared to check for consistency (as a proxy for reliability), a high R^2 value suggests that they are consistent with each other in a linear manner. In this context, given that all R^2 values are greater than 0.75, it can be inferred that the datasets are consistent with each other, but it's a leap to say they are "equally reliable" without additional information on the sources, data collection methods, and other factors that might influence reliability. Here we are saying they are reliable because of two datasets NSI and Eurostat that are from reliable sources and the high similarity of the 3rd set is prove of the reliability.

MANOVA

To start the MANOVA we are going to take several steps.

To understand better why we have chosen MANOVA several different parameters have to be fulfilled. Understanding why we use method for related samples is first. If the data from different sources corresponds to the same subjects or entities, then the samples are related. Second is test of normal distribution of the data.

Table 1 shows the results of two statistical tests (Kolmogorov-Smirnov and Shapiro-Wilk) that are used to check if data is normally distributed. The tests are applied to three different datasets: "BAALO data", "Eurostat", and "NSI". Here's what each part of the output means:

The Kolmogorov-Smirnov test is a non-parametric test that compares your data with a reference probability distribution (in this case, the normal distribution). The Shapiro-Wilk test is another test for normality; it is generally considered to be more powerful than the Kolmogorov-Smirnov test, especially for small sample sizes.

The value of the test statistic will depend on the specific test and the data. df: This stands for degrees of freedom, which is the number of values in the final calculation of a statistic that are free to vary. Sig.: This is the p-value for the test. The p-value is the probability of obtaining the observed data (or data more extreme) if the null hypothesis is true. A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis. In this case, the null hypothesis is that the data comes from a normal distribution. In our output, all the p-values are greater than 0.05, which means there is not enough evidence to reject the null hypothesis that the data comes from a normal distribution. Therefore, we would conclude that our data does not significantly deviate from normality.

The note about "Lilliefors Significance Correction" indicates that the p-values for the Kolmogorov-Smirnov test have been adjusted using Lilliefors' method, which is a correction for the fact that the parameters of the normal distribution (mean and standard deviation) are estimated from the data. The asterisk next to the p-values for the Kolmogor-

Table 1. Test of Normality

Tests of Normality						
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
BAALO data	,121	13	,200*	,941	13	,472
Eurostat	,101	13	,200*	,962	13	,787
NSI	,127	13	,200*	,972	13	,916

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Table 2. Model Summary

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,997[a]	,994	,992	,353	2,710

a. Predictors: (Constant), NSI, BAALO data, Eurostat
b. Dependent Variable: YEAR, not periodic

ov-Smirnov test indicates that these are lower bounds on the true significance level, due to the Lilliefors correction.

The Model Summary table (table 2) provides several statistics that can be useful in understanding the fit of our regression model. The multiple correlation coefficient, which measures the strength of the linear relationship between the dependent variable and the set of independent variables is 0.997, which indicates a very strong linear relationship. The coefficient of determination R^2, is 0.994, which means that 99.4% of the variance in the dependent variable can be predicted from the independent variables. Adjusted R Square is 0.992, which is very close to the R Square value, indicating that all of your predictors are contributing to the model. The standard error of the regression, which measures the standard deviation of the residuals (the difference between the observed and predicted values) is it's 0.353.

The Durbin-Watson statistic ranges from 0 to 4. A value of 2 means there is no autocorrelation in the sample. Values from 0 to less than 2 indicate positive autocorrelation and values from 2 to 4 indicate negative autocorrelation. In your case, it's 2.710, which suggests that there is no significant autocorrelation in your data. Even if we can find autocorrelation in our case, the Durbin-Watson statistic is 2.710. This value is above 2, suggesting that there might be negative autocorrelation.

The predictors in your model are "NSI", "BAALO data", and "Eurostat", and the dependent variable is "YEAR, not periodic".

This output is from a multivariate test in a repeated measures ANOVA. The multivariate test is used when you have more than one dependent variable, and it tests whether the entire set of means for your within-subject factor is different across the levels of the factor (land price). The analysis aims to determine how variations in land price impact the dependent variables being measured. All four tests (Pillai's Trace, Wilks' Lambda, Hotelling's Trace, Roy's Largest Root) show a significant effect, as indicated by the F-statistics (13,833 for each) and very low p-values (0.001 for each). This suggests that changes in factor 1 (price) have a statistically significant impact on the dependent variables as a group. The fact that all four tests yield similar results adds robustness to this conclusion. Pillai's Trace Value of 0.716: This high value suggests a strong relationship between factor1 and the dependent variables. Wilks' Lambda Value of 0.284: The complement of Pillai's Trace, this also indicates a strong relationship. Hotelling's Trace and Roy's Largest Root of 2.515: These values further confirm the significant impact of land price. Hypothesis df and Error df: The degrees of freedom for the hypothesis (2) and for the error (11) are used in calculating the F statistic. These values suggest that the model is based on a comparison of two groups or

Table 3. Descriptive statistics

Descriptive Statistics			
	Mean	Std. Deviation	N
BAALO data	8188,93	4126,415	13
Eurostat	8288,18	2638,991	13
NSI	7681,65	2658,059	13

Table 4. Multivariate test

Multivariate Tests[a]									
Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[c]
factor1	Pillai's Trace	,716	13,833[b]	2,000	11,000	,001	,716	27,666	,987
	Wilks' Lambda	,284	13,833[b]	2,000	11,000	,001	,716	27,666	,987
	Hotelling's Trace	2,515	13,833[b]	2,000	11,000	,001	,716	27,666	,987
	Roy's Largest Root	2,515	13,833[b]	2,000	11,000	,001	,716	27,666	,987

a. Design: Intercept
Within Subjects Design: factor1
b. Exact statistic
c. Computed using alpha = ,05

conditions in relation to land price. The significance value (p-value) is 0.001 for each test, which is much lower than the typical alpha level of 0.05. This indicates a very low probability that the observed results are due to chance. The value of 0.716 for partial eta squared is quite high, indicating that a substantial proportion (about 71.6%) of the variance in the dependent variables can be explained by the variation in land price.

Table 5. Within – Subjects factors

Within-Subjects Factors (Land price)	
Measure: MEASURE_1	
Price	Dependent Variable
1	BAALO
2	Eurostat
3	NSI

This refers to factors (or independent variables) where each level of the factor has been applied to each subject or case. In other words, every participant or case in the study has been measured under each condition or level of the factor. This is in contrast to between-subjects factors, where each participant is only exposed to one level of the factor. In an SPSS repeated measures analysis with this setup, you would be examining whether there are significant differences in the "Land price" (dependent variable) across the three different sources or conditions (BAALO data, Eurostat, and NSI). Since it's a within-subjects design, each participant or case in the dataset has a price measurement from all three sources, and the analysis would determine if these measurements significantly differ from one another.

MANOVA is used because we have multiple dependent variables and want to test if groups differ on a combination of these variables. Table interpretation :

This section provides various test statistics to determine if there are significant group differences on a linear combination of dependent variables.

Land Price: This is the independent variable or factor you are testing. In this case, it seems to be the "Land Price" variable, which might represent different levels or conditions of price.

Test Statistics: There are several test statistics provided in MANOVA, each with its own assumptions and interpretations:

Pillai's Trace: This is one of the multivariate test statistics. The value provided (0.716) represents the proportion of variance in the dependent variables that is accounted for by the independent variable (Land Price). The Wilks' Lambda statistic represents the proportion of the total variance in the combination of dependent variables that is unaccounted for by the independent variable. A value closer to 0 indicates that the groups differ, while a value closer to 1 indicates that they are similar, the value here is 0.284. Hotelling's Trace value (2.515) indicates the difference between groups relative to the variance within groups. A large F-value suggests that the means are different. Here, the F-value is 13.833 for all the tests. Hypothesis df: high F-value of 13.833 for all tests suggests that the variance between the different levels or conditions of land price is significantly greater than the variance within these levels. This is an indicator that the mean values of the dependent variables are not all equal across the different levels of land price.This is the degrees of freedom for the denominator of the F-statistic. Here, it's 11.000.

The p-value or sig. associated with the F-statistic. It tests the null hypothesis that the group means are equal. A p-value less than 0.05 (common significance level) suggests that you reject the null hypothesis. Here, the p-value is 0.001 for all the tests, indicating that there are significant group differences on the combination of dependent variables.Based on the provided output, there are significant differences in the dependent variables across the levels or conditions of the "Price" variable. All the multivariate test statistics are significant with a p-value of 0.001.

Mauchly's Test of Sphericity:

Sphericity is an assumption that the variances of the differences between all possible pairs of within-subject con-

Table 6. Multivariate test

Multivariate Tests[a]						
Effect		Value	F	Hypothesis df	Error df	Sig.
Land Price	Pillai's Trace	,716	13,833[b]	2,000	11,000	,001
	Wilks' Lambda	,284	13,833[b]	2,000	11,000	,001
	Hotelling's Trace	2,515	13,833[b]	2,000	11,000	,001
	Roy's Largest Root	2,515	13,833[b]	2,000	11,000	,001

a. Design: Intercept
Within Subjects Design: Price
b. Exact statistic

Table 7. Mauchly Test of Sphericity

Mauchly's Test of Sphericity[a]							
Measure: MEASURE_1							
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Price	,130	22,442	2	,000	,535	,545	,500
Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.							

a. Design: Intercept
Within Subjects Design: Price
b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

ditions (levels) are equal. Violation of this assumption can inflate the Type I error rate. Mauchly's test checks this assumption.

Measure: MEASURE_1:

This indicates the dependent variable or measure being analyzed.

This is the p-value associated with the chi-square test. A p-value less than 0.05 typically indicates a violation of the sphericity assumption. Here, the p-value is 0.000, suggesting that the assumption of sphericity has been violated. Based on the provided output, the assumption of sphericity has been violated for the "Price" within-subjects factor (p-value = 0.000). Therefore, when interpreting the results of the repeated measures ANOVA, you should use the corrected F-statistic values (using one of the epsilon corrections) from the "Tests of Within-Subjects Effects" table.

Given that the assumption of sphericity has been violated for the "Price" within-subjects factor, it's crucial to interpret the results using the corrected values. Here's the interpretation:

Interpretation of the Tests of Within-Subjects Effects:

1. Sphericity Assumed: This row provides the results under the assumption that sphericity is met. However, since we know from Mauchly's Test of Sphericity that this assumption has been violated (p-value = 0.000), we should be cautious about interpreting results from this row.
2. Greenhouse-Geisser:
   – This correction is often used when the sphericity assumption is violated.
   – The F-statistic for the "Price" factor after applying the Greenhouse-Geisser correction is 1.190.
   – The associated p-value is 0.300. Since this p-value is greater than the typical significance level of 0.05, we fail to reject the null hypothesis. This suggests that there are no significant differences in the dependent variable "MEASURE_1" across the levels or conditions of the "Price" factor when using the Greenhouse-Geisser correction.
3. Huynh-Feldt:
   – This correction is less conservative than Greenhouse-Geisser and might be used when the violation of sphericity is not too severe.
   – The F-statistic remains 1.190.
   – The associated p-value is 0.301, which is also greater than 0.05. Thus, the conclusion remains the same:

Table 8. Test of Within-subjects effects

Tests of Within-Subjects Effects						
Measure: MEASURE_1						
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Price	Sphericity Assumed	2751869,335	2	1375934,668	1,190	,321
	Greenhouse-Geisser	2751869,335	1,070	2572983,562	1,190	,300
	Huynh-Feldt	2751869,335	1,089	2526855,994	1,190	,301
	Lower-bound	2751869,335	1,000	2751869,335	1,190	,297
Error(Price)	Sphericity Assumed	27740842,594	24	1155868,441		
	Greenhouse-Geisser	27740842,594	12,834	2161462,001		
	Huynh-Feldt	27740842,594	13,069	2122712,050		
	Lower-bound	27740842,594	12,000	2311736,883		

no significant differences across the levels of the "Price" factor.

4. Lower-bound:
   – This is the most conservative correction and is often used when the sample size is small or when the violation of sphericity is severe.
   – The F-statistic is still 1.190.
   – The p-value is 0.297. Even with this conservative correction, the conclusion is consistent with the other corrections: there are no significant differences in the dependent variable "MEASURE_1" across the levels or conditions of the "Price" factor.

Given the violation of the sphericity assumption, it's essential to rely on the corrected values for interpretation. Across all the corrections (Greenhouse-Geisser, Huynh-Feldt, and Lower-bound), the results consistently indicate that there are no significant differences in the dependent variable "MEASURE_1" across the different levels or conditions of the "Land Price" factor. In summary, based on these results, there is no significant difference in the dependent variable (MEASURE_1) attributable to changes in Price. The F-value is relatively low, and the p-values are all above the commonly used threshold of 0.05, indicating that any differences observed in the dependent variable across the different levels of Price are not statistically significant.

This table provides the results of the ANOVA tests for the within-subjects factor "Price" by breaking down the contrasts into linear and quadratic components. Let's interpret the table:

The Tests of Within-Subjects Contrasts provides the results of the repeated measures ANOVA for the within-subjects factor "Price," specifically focusing on the linear and quadratic contrasts.

MEASURE_1 indicates the dependent variable or measure being analyzed.

This column lists the source of variation, which in this case is "Land Price" (the within-subjects factor). The contrasts (Linear and Quadratic) are used to test specific patterns of differences across the levels of the "Land Price" factor.

1. Linear Contrast:
   – This tests for a linear trend across the levels of the "Price" factor. Essentially, it checks if there's a consistent increase or decrease across the levels.
   – The F-statistic for the linear contrast is 0.955.
   – The associated p-value is 0.348. Since this p-value is greater than the typical significance level of 0.05, we fail to reject the null hypothesis for the linear contrast. This suggests that there isn't a significant linear trend in the dependent variable "MEASURE_1" across the levels of the "Price" factor.
2. Quadratic Contrast:
   – This tests for a quadratic (or U-shaped) trend across the levels of the "Price" factor.
   – The F-statistic for the quadratic contrast is 1.926.
   – The associated p-value is 0.190. Since this p-value is also greater than 0.05, we fail to reject the null hypothesis for the quadratic contrast. This indicates that there isn't a significant quadratic trend in the dependent variable "MEASURE_1" across the levels of the "Land Price" factor.

Based on the provided output, neither a linear nor a quadratic trend is observed in the dependent variable "MEASURE_1" across the different levels or conditions of the

Table 9. Tests of Within-Subjects Contrasts

Tests of Within-Subjects Contrasts						
Measure: MEASURE_1						
Source	Price	Type III Sum of Squares	df	Mean Square	F	Sig.
Price	Linear	1672619,948	1	1672619,948	,955	,348
	Quadratic	1079249,387	1	1079249,387	1,926	,190
Error (Land Price)	Linear	21016640,504	12	1751386,709		
	Quadratic	6724202,090	12	560350,174		

Table 10. Tests of Between-Subjects Effects

Tests of Between-Subjects Effects					
Measure: MEASURE_1					
Transformed Variable: Average					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	2529130515,922	1	2529130515,922	87,985	,000
Error	344941333,268	12	28745111,106		

"Land Price" factor. Both contrasts (linear and quadratic) are not statistically significant at the typical 0.05 significance level.

Based on the provided output, the overall mean of the dependent variable "MEASURE_1" is significantly different from zero. This is indicated by the significant F-statistic for the Interceptation. In summary, this table is showing the results of a between-subjects ANOVA analysis on a measure with the transformed variable being the average. The Intercept has a significant effect on the dependent variable with a very low p-value, suggesting that the model is explaining a significant amount of the variability in the data.

Conclusion

The method of linear regression shows that the datasets are different but not significantly different since all of the R^2 are bigger than 0, 75. Based on the outputs from the MANOVA it appears that there is no significant difference between the three datasets. While the multivariate tests suggest some differences when considering the combination of dependent variables, the specific analyses focused on the "Price" factor (which seems to represent the datasets) do not show significant differences. The within-subjects effects and contrasts, which are more directly related to the question of differences between the datasets, consistently indicate no significant differences for the dependent variable "MEASURE_1" across the datasets. Here's why:

Mauchly's Test of Sphericity indicated a violation of the sphericity assumption, which means that the variances of the differences between all possible pairs of within-subject conditions are not equal. This led us to consider the corrected values in the subsequent analyses.

Multivariate Tests from the MANOVA showed significant results, suggesting that there are differences between the groups on the combination of dependent variables. However, this doesn't directly indicate differences in the specific variable of interest across the datasets.

Tests of Within-Subjects Effects showed no significant differences in the dependent variable "MEASURE_1" across the levels or conditions of the "Price" factor. The p-values for all corrections (Greenhouse-Geisser, Huynh-Feldt, and Lower-bound) were greater than 0.05.

Tests of Within-Subjects Contrasts also indicated no significant linear or quadratic trends across the levels of the "Price" factor.

Tests of Between-Subjects Effects showed that the overall mean of the dependent variable "MEASURE_1" is significantly different from zero, but this doesn't provide information about differences between the datasets.

Autocorrelation: The Durbin-Watson statistic (2.710) suggests that there is no significant autocorrelation in your data. This means that the residuals (the differences between the observed and predicted values) are not correlated. This is a good thing, as autocorrelation can violate the assumptions of many statistical tests and models.

Given the violation of the sphericity assumption, it's essential to rely on the corrected values for interpretation. Across all the corrections (Greenhouse-Geisser, Huynh-Feldt, and Lower-bound), the results consistently indicate that there are no significant differences in the dependent variable "MEASURE_1" across the different levels or conditions of the "Land Price" factor.

Based on these results, it seems that there is a difference between the three datasets. However, the specific nature of these differences (i.e., which datasets are significantly different from each other, and in what way) would require further investigation, such as post-hoc tests or pairwise comparisons. The difference is not significant and based on the two methodologies we can safely assume that the tree sets of data are reliable sources of information and can be used for further analysis of the state of agricultural land markets in Bulgaria. The similarity with official sources like Eurostat and NSI is a positive indicator. We conclude that the third dataset aligns well with the official sources across multiple considerations and there are no known issues or biases and with that it can be deemed reliable for the aspects of further analyses.

Acknowledgements

We are grateful for the financial aid this research article is created from NRF and project КП-06-Н35/2 from 18.12.2019. Project Land relationships and European policy: synergy and perspectives for Bulgarian agriculture.

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Gregory, P. (1992). Rents, land prices and economic theory. In Economy and Society in Russia and the Soviet Union, 1860–1930, 6–23. Palgrave Macmillan UK.

Kirechev, D. (2022). Efficiency of investments in purchase for agricultural land through financial instruments. *Izvestiya na syyuza na uchenite – Varna. Seriya Ikonomicheski nauki, 11*(3), 126-138.

Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association, 46*(253), 68-78. doi:10.1080/01621459.1951.10500769.

Misztal, A., Kowalska, M. & Fajczak-Kowalska, A. (2022). The impact of economic factors on the sustainable develop-

ment of energy enterprises: The case of Bulgaria, Czechia, Estonia and Poland. *Energies, 15*(18), 6842.

Razali, N. M. & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21-33.

Rencher, A. C. (2002). Methods of Multivariate Analysis (2nd ed.). Wiley.

Rodgers, J. L. & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician, 42*(1), 59-66.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591-611. doi:10.2307/2333709.

Tabachnick, B. G. & Fidell, L. S. (2013). Using multivariate statistics (6th ed.). Pearson. https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1541229.

Received: **October, 13, 2023**; *Approved*: **November, 16, 2023;** *Published:* **December, 2023**