EFFECTS AND CHARACTERISTICS OF THE IMPLEMENTATION OF THE 1921 AGRARIAN REFORM IN IAȘI COUNTY – AN ECONOMETRIC APPROACH

Sebastian DOBOȘ
"Gh. Zane" Institute of Economic and Social Research, the Romanian Academy, Iași Branch, 700481, Romania
sebastian.d.dobos@gmail.com

Abstract
Notwithstanding the multiple perceived deficiencies of the 1918-1921 agrarian reform pointed out by several authors, it can be firmly stated that it had considerable impact both in the County of Iași as well as in most other territorial units of interwar Romania, fact which is recorded in most of the papers written on the topic of the reform. Some of the most important effects of the vast process of redistribution of the land fund - perhaps the most extensive one ever recorded in the world or in Europe at least, according to the opinion of many reputed historians and economists were in the end expropriation of the big agricultural landowners in the county and the appropriation of peasants with individual plots. The discrepancy between the positive perception of the interwar period by a large share of contemporary Romanian society and the critical assessments and analyses in the works of some historians, mainly foreign ones, which are based on historical reports, data and information, can be justified and better understood on making use of statistics and statistical validation of the economic performance of the Kingdom of Romania between 1918 and 1939. Notwithstanding any critique of Statistics as a universal panacea with inevitable potential shortcomings, most authors agree that such assessments are highly relevant.

Key words: *economic history, econometrics, historiography, interdisciplinarity, rural economics*

JEL Classification: *B23, C01, N00, N30, N50, Q01.*

I. INTRODUCTION

In the interwar period, the 1921 agrarian reform, identified by the phrase "the agrarian issue" or alternately "the agrarian question", was a permanent topic of heated debate and ideological confrontation within the political and socio-economic circles, given the widely differing interests of the social categories at the end of the 19th and the beginning of the 20th centuries. This confrontation, a major one for the subsequent evolution of society became generalized, eventually acquiring the status of "a national problem". Thus, at least in the *"Vechiul Regat"* (Wallachia and Moldavia), the ideological divergences increasingly focused on the argument of the necessity of expropriation of land ownership and, implicitly, the subsequent endowment of peasants with individual plots, this solution being strongly advocated following the generalized rebellion of peasants in 1907, which made headlines abroad as a result of their amplitude and, above all, the significant loss of life (Axenciuc, 1997; Axenciuc, 2012; Dropu, 2011, p. 125-126; Doboș, 2018, p. 84; Hitchins, 2007, p. 114; Murgescu, p. 109; p. 126; p. 129; p. 130; p. 139).

II. METHOD AND MATERIALS

A number of historians, sociologists, economists, academics and archivists who have taken an interest in history in general and in the topic of the agrarian reform in particular have proven the particular importance of this topic. Starting from both meanings of the term "historiography" the topic of the agrarian reform of 1918-1921, also known in the literature as "The Agrarian Question" has been approached in several manners, with the political context being a major factor. The paper is intended to be a detailed analysis of some of the factors leading to the current state of the overall development of the economy in the rural areas of present-day Romania. The present study also aims at revealing the specific elements that can be used as a means of solving the numerous issues affecting the current state of the local social and economic environment in the country. Given the overwhelming number of scientific papers and studies on the topic of the agrarian reform legal foundations of which were established in 1918, a selection of the important results obtained by previous research proved necessary (Chivu & Ioan-Franc, 2019, p. 908-916). An in-depth systematic analysis of the effects of the implementation of the 1921 agrarian reform in the Romania should focus on the economic issues, which are the most important, in the general political, social and cultural context. Objective identification of the relevant economic indicators which allow a correct assessment of the practical effects is a difficult enterprise, considering the fact that despite the official statements proclaiming 1926 as the year of completion of its application, the reform was nevertheless repeatedly hindered and delayed. Notwithstanding such statement by state authorities and politicians, the historical records show that in fact various technical operations implied by the reform were still in progress after the beginning of the Second World War, for a number of different reasons: litigation, cadastral operations, administrative issues and so on. An in-depth systematic analysis of the effects of the implementation of the 1921 agrarian reform in the county of Iași should focus on the economic issues, which are the most important, in the general political, social and cultural context (Șandru, 1975; Șandru, 1975). Objective identification of the relevant economic indicators which allow a correct assessment of the practical effects in the county of Iași of the laws published in "Monitorul Oficial nr. 82/17" in July of 1921 is a difficult enterprise, considering the fact that despite the official statements proclaiming 1926 as the year of completion of its application, the reform was nevertheless repeatedly hindered and delayed. Notwithstanding such statement by state authorities and politicians, the historical records show that in fact various technical operations implied by the reform were still in progress after the beginning of the Second World War, for a number of different reasons: litigation, cadastral operations, administrative issues and so on (Doboș, 2018, p. 99; Șandru, 1975).

Present day Economics highlights the still little-known and used role of historical studies (Andrei, J.V. et al. (2020), p. 15-16; p. 49-54). Whereas economists stand out because of their obvious tendency to over-simplify, historians, on the other hand, analyze current events in all of their complexity. It is often said that you cannot be a good economist without mathematical training; nowadays, it is ever more stressed that one cannot be an economist in the absence of some historical training. Encouraging empirical, multi- and interdisciplinary research calls for a reassessment of the scale of values in the estimation of economists. Thus, it is necessary to give greater importance to the faculties of observation, than to those of abstraction. In other words, the perspicacity/insight of the historian should precede the rigor of the mathematician. The methodology of research specific to the discipline of Economics is deeply and constantly involved with both progress as well as inadequacies of Economics on the whole.

Historians have used the historical statistical method for a variety of purposes. A first goal would thus be to check for the existence of statistical regularities of mass phenomena, in which case the estimation method has a major role. It is about reviewing the data that is known to be recorded in a source as well as investigating it when there is no clear information about it in a particular source. The second objective of historical statistics is to suggest hypotheses regarding the causal dependencies between facts and the formulation of so-called statistical laws. The analysis of correlation indices in evaluating the causal dependencies and the degree of influence of each factor is a relevant example in this regard. A third objective of this research method is to facilitate the description of the facts which show some features that allow for classification according to criteria of choice. Currently, statistics is mostly used as a descriptive tool, rather than as an induction method. In other words, its function is to summarize or help find patterns for information from broad historical databases. The reason for this change is, first and foremost, that most statistical methods operate with data samples with a statistical distribution that historical data does not meet. A second reason is that historians are reluctant to make generalizations for the entire population, starting from the statistical results collected on the basis of the samples.

The first set of approaches is descriptive statistics and statistical inference, namely logistic regression. An important quantitative method of research is the Time Series Analysis used in historical research. In Economic History, time series analyses have been used particularly in the last decades, mainly for purposes of comparisons at the international level, such as industrialization, economic growth, and unemployment rates. The analysis of statistical time series involves the reconstruction and study of the dynamics of either one or more variables in time, employing methods of both descriptive and inferential statistics. Drawing the graphs of time series provides researchers with the possibility of visualizing clearly the chronology of the increases and/or decreases of variables, as well as seasonal fluctuations and long-term trends. These graphs aided by additional calculations yield rates of growth and acceleration or decline rates.

III. RESULTS AND DISCUSSION

The techniques of statistical association for the identification, isolation, and measurement of the extent of association between two or more variables, irrespective of their being a time series or not, are often used in social and historical studies. Statistical methods of correlation and regression provide mathematical evidence of the existence and corresponding degree of intensity of a possible link between two or more variables. Researchers resorting to statistics to support their various hypotheses or theories ought to make careful use of this kind of information, and check it against further references or indeed by applying additional methods. Adapting statistical methods to historical purposes and interpretations is aimed at gathering and disseminating information. In many cases however the quantitative component substitute for qualitative accounts which resist quantification (Centoni and Cubadda, 2015, p. 415–434; Gavrilă, 2009, p. 31; p. 33; p. 76-98; p. 98-126; p. 133-160). By resorting to available software such as *gretl* (G. R. E. T. L., 2018 edition) one in is able to load the *"Agrarian Reform of 1921 in the former Kingdom of Romania"* Excel (.xlsx) data set (Microsoft Excel, 2019) that I created and obtain several relevant average values for various variables of interest. When tackling time series data, the field literature acknowledges that there is one appropriate manner to insert and save the data in chronological order – with the first time period recorded as the first scientific observation and the most recent one as the last. On resorting to an Excel spreadsheet, one has the possibility to compute the requested values according to one`s interests and needs by resorting to the “Statistical Menu”, which offers numerous functions required in order to compute the results of interest. As is now widely known, spreadsheets (MS Excel ones included) facilitate data processing, for example arranging or calculating averages, among others. At present, however, this feature is less significant since I am using or resorting to a statistical software which enables sophisticated data storage and handling. Numerous statistical software such as *gretl* (G.R.E.T.L.) 2018 edition, for example, can be included in this category. After using a spreadsheet for initial data input, the dataset obtained can then be exported or loaded in a format that the statistical software can analyze and interpret. (Wooldridge, 2012, p. 681). Irrespective of the issue or topic addressed, computing a table of summary statistics containing minimum and maximum values, as well as standard deviations for each variable, is also quite useful. By making use of such a table allows one to better study, interpret and understand the computed coefficient estimates aiming at emphasizing and highlighting the measurement units of the variables of choice. According to theoretical econometrics field literature, for binary variables, the only required summary statistic (e.g. mean, median, standard deviation and other values) is only a part of the ones in the sample of choice, matching the sample mean. It is often useful and illustrative to compute the average growth rate in a variable over the years in my sample (Wooldridge, 2012, p. 690).

Table 1. The Agrarian Reform of 1921 in the Former Kingdom of Romania.DES

1911	11.7	1365013	547963	3132000	6135000	0
1912	11.7	1266511	518774	3198000	6135000	0
1913	11.7	1375637	531123	3250000	6135000	0
1914	11.7	1067813	557988	3435000	6135000	0
1915	11.7	1269009	548389	3490000	6135000	0
1919	9.2	1997256	1253951	7081000	11477000	1
1920	9.1	1876138	1389425	7102000	11477000	1
1921	7.2	1729691	1620146	7188000	11477000	1
1922	9.2	2098718	1829676	7299000	11477000	1
1923	9.1	2232591	1824652	7407000	11477000	1
1924	7.3	1965388	1775355	7515000	11477000	1
1925	9	2291741	1724592	7626000	12550000	1
1926	12.1	2818406	1776312	7736000	12550000	1
1927	8.6	2332947	1788714	7837000	12550000	1
1928	8.7	2323914	1804238	7948000	12550000	1
1929	13.6	3099258	1860014	8061000	12550000	1
1930	12.1	2918897	1817394	8252000	13123000	1
1931	12.3	3065240	1556051	8302000	13123000	1
1932	9.9	2531927	1806730	8421000	13123000	1
1933	11	2721375	1852568	8524000	13123000	1
1934	8.6	2383199	1954130	8644000	13123000	1
1935	9.9	2643428	1815384	8723000	13706000	1
1936	12.1	2994482	1707485	8829000	13706000	1
1937	11	2824300	1657368	8927000	13706000	1
1938	11.8	3000382	1754937	9026000	13706000	1
1939	12.3	3076191	1721830	9110000	13706000	1

Source: My own computations based on Axenciuc, V. *Evoluţia Economică a României: Cercetări Statistico-Istorice, 1859-1947*, Vol. II Agricultura, Editura Academiei Române, Bucureşti, 1992, p. 516-521; p. 645-654

For instance, I have included in my time series data set of choice the 1911-1915 interval *(before the First World War)* and also the interwar period 1919-1939, the one which basically represents my time span of particular interest. Thus, researchers and individuals interested in the matter of the agrarian reforms in Eastern Europe or in the one carried out in Romania in particular only, can easily see that the statistical observations (*26 years in total*) are saved in the data set I have managed to compile in the file (.xlsx format), as seen in the table above. One should remember and keep in mind, needless to say, that the 1914-1919 interval represents a challenge, in the sense that because of the losses during WWI there is often very little and rather untrustworthy statistical data, if any whatsoever. Thus, researchers and any individuals interested in the agrarian reforms implemented in Eastern Europe at the end of the 19th century and at the beginning of the 20th, as is the case with the one carried out in Romania, can easily notice that the statistical observations (*26 years in total*) are recorded and saved in the dataset that I have managed to compile in the .xlsx format file, as aforementioned. One can easily create time trends, for instance, irrespective of the annual indicators. By resorting to *gretl* (G.R.E.T.L.) 2018 edition software, one is able to compute the statistical results below, which are necessary for my scientific endeavor. When using *gretl* (G.R.E.T.L.) 2018, for instance, one has at hand the summary statistics submenu, which easily and automatically computes the requested values *(Mean, Median Minimum, etc.)* for the "v2" dependent variable *(**cerealtotoutput** - yearly volume of cereals - total output harvested)*, my indicator of interest. The results provided by *gretl* (G.R.E.T.L.) 2018 are the following:

Table 2. Summary statistics, using the observations 1914 – 1939 for the variable "v2-cerealtotoutput" (26 valid observations)

Mean	10.485
Median	11.000
Minimum	7.2000
Maximum	13.600
Standard deviation	1.7322
C.V.	0.16521
Skewness	-0.26411
Ex. kurtosis	-1.0459
5% percentile	7.2350
95% percentile	13.145
Interquartile range	2.8000
Missing obs.	0

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package

My econometric approach centers on the Multiple Linear Regression (O.L.S.) method of estimation for my particular time series dataset. One can notice that this is not necessarily a weakness, as O.L.S. is still the most widely used econometric method. Obviously, one can still question whether any of the variants of O.L.S. - such as weighted least squares or correcting for serial correlation in a time series regression are warranted (Wooldridge, 2012, p. 690). I have resorted to the total harvested cereal output according to recorded statistics as a relative proxy of the overall level of welfare of the population in the interwar *"Kingdom of Romania"*, with the caveat represented by the post 1919 border changes, a fact which makes it much more difficult to determine the effects of various variables and the subsequent evolution of different socio-economic indicators.

Table 3. Full descriptive statistics, using observations: 1914 - 1939

Variable	Mean	Median	Minimum	Maximum
v1	1925,9	1926,5	1911,0	1939,0
v2	10,485	11,000	7,2000	13,600
v3	2,2796e+006	2,3284e+006	1,0678e+006	3,0993e+006
v4	1,4998e+006	1,7398e+006	5,1877e+005	1,9541e+006
v5	7,1563e+006	7,7865e+006	3,1320e+006	9,1100e+006
v6	1,1401e+007	1,2550e+007	6,1350e+006	1,3706e+007
v7	0,80769	1,0000	0,00000	1,0000
Variable	Standard deviation	C.V.	Skewness	Ex. kurtosis
v1	8,5179	0,0044227	-0,22834	-1,0576
v2	1,7322	0,16521	-0,26411	-1,0459
v3	6,3891e+005	0,28027	-0,41204	-1,0226
v4	4,9881e+005	0,33258	-1,2928	-0,040819

v5	2,0099e+006	0,28086	-1,2405	0,0054507
v6	2,7326e+006	0,23968	-1,2713	0,013617
v7	0,40192	0,49761	-1,5614	0,43810
Variable	5% Perc.	95% Perc.	Interquartile range	Missing obs.
V1	1911,3	1938,7	13,500	0
v2	7,2350	13,145	2,8000	0
v3	1,1374e+006	3,0912e+006	1,0084e+006	0
v4	5,2310e+005	1,9212e+006	4,6033e+005	0
v5	3,1551e+006	9,0806e+006	1,4573e+006	0
v6	6,1350e+006	1,3706e+007	1,6460e+006	0
v7	0,00000	1,0000	0,00000	0

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package

In order to justify the O.L.S. Method, it is therefore important to make a compelling argument that the main O.L.S. statistical assumptions for the model of choice are fulfilled. As highlighted and debated by numerous academics in the field one of the first issues one has to check is whether or not the error term (μ) is uncorrelated with any of the explanatory variables included in the proposed model. In an ideal situation, one has been able to control for few enough factors to presume that those that which are included in the error term (μ) are not in any form connected or related to the computed regressors (Wooldridge, 2012, p. 683). As is emphasized in the field literature, time series applications require special care with many questions arising. Should the equation be estimated in levels? If that is the case does one need to resort to using time trends? If one needs to resort to distributed lag dynamics, amongst others, one important issue and question that arises is determining the precise number of the lags that need to be included in my proposed model (Wooldridge, 2012, p. 684). Given my particular topic of study I have considered it necessary, based on my intuition, to take into account some lags (*five in total*), keeping in mind that this topic is by and large an empirical subject. More specifically, my time lags cover the 1911-1915 interval before World War I, as a reference time span in order to facilitate further potential comparisons.

I can thus reasonably assume the fact that my model might have several potential misspecifications, such as omitted variables. According to the field literature, if this is the case, researchers resorting to the O.L.S. method ought to try to resort to any form of misspecification in the proposed regression analysis with the aim of determining, on the basis of sound or rational assumptions, the presence and effect of any bias/slant in the estimators. It can be noted, by consulting various works in the field literature, that generally, empirical social sciences often resort to what is widely known as *sensitivity analysis*. This emphasizes the fact that one should try to estimate the initial model and change it in such a way that it seems plausible. The end goal of such an endeavor is that the most relevant conclusions are not subject to major changes.

Even if one attempts to be as cautious as possible in tackling the subject matter of choice, in developing the proposed model, gathering the necessary data needed to be subjected to statistical analysis it is still very likely that one will often obtain confusing results. If this is the case, the best scientific approach is to test out new models, other estimation methods, or even different datasets, until the statistical software computed results correspond more accurately to what was expected from the proposed model. The field literature indicates that most researchers and academics usually try time and again before selecting and choosing the Best Linear Unbiased Estimator (B.L.U.E.) model (Wooldridge, 2012, p. 685). However, as with many other case studies, the very data gathering process contradicts the initial assumptions I made with my econometric study. The tests of the unbiasedness of O.L.S. and especially the commonly used t and F statistical distributions that I have computed for the hypothesis test presume that I have observed a statistical sample following the population model and measured the proposed computed econometric model once only. The approach that involves the repeated estimation of several other proposed models that are basically other versions of the initial one undermines that assumption, given that I am basing my computations on the very same dataset when trying to determine a particular issue of interest. Basically, according to field literature, it is common practice to make use of the computed statistical results provided by the software of choice to reconsider my proposed model aiming at rewriting it as well as measuring its *goodness-of-fit*. However, most authors in the field of econometrics agree that the computed estimates and various other tests run on different variants of proposed models are not really that different (Wooldridge, 2012, p. 685-686).

Thus, I then try to estimate my proposed econometric model: **cerealtotoutput** = *β0 + β2 * agrioutputvegetablemass + β3 * agrioutputanimalorigin + β4 * ruractpop + β5 * agrisurf + Δ0 * AgriReform1921 (β6) + μ*, and then insert the computed results into my equation of choice. According to the usual econometric analysis one must try to interpret the intercept (*constant*). More specifically one has to check whether it has a useful meaning.

Table 4. Model 1: OLS, using observations 1914-1939 (T = 26), Dependent variable: cerealtotoutput

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
Const	10.8950	1.32570	8.218	<0.0001	***
Agrioutputvegetables	5.26033e-06	4.50715e-07	11.67	<0.0001	***
agrioutputanimal	−1.91523e-06	7.74066e-07	−2.474	0.0224	**
Ruractpop	4.70478e-08	6.19380e-07	0.07596	0.9402	
Agrisurf	−8.13170e-07	5.09580e-07	−1.596	0.1262	
dummyvar	−0.736564	1.28981	−0.5711	0.5743	

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package; **N.B.** Excluding the constant, p-value was highest for variable no. 5 *("AgriReform1921")*

Table 5. Summary statistics for Model 1: OLS, using observations 1914-1939 (T = 26), Dependent variable: cerealtotoutput

Mean dependent var	10.48462	S.D. dependent var	1.732211
Sum squared resid	5.046375	S.E. of regression	0.502313
R-squared	0.932727	Adjusted R-squared	0.915909
F(5, 20)	55.45959	P-value(F)	4.90e-11
Log-likelihood	−15.57986	Akaike criterion	43.15972
Schwarz criterion	50.70830	Hannan-Quinn	45.33343
Rho	−0.161097	Durbin-Watson	2.275492

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package; **N.B.** Excluding the constant, p-value was highest for variable no. 5 *("AgriReform1921")*

Table 6. Model 1: OLS, using observations 1914-1939 (T = 26), Dependent variable: cerealtotoutput, (LM Test and White`s Test)

LM test for autocorrelation up to order 1 -	White's test for heteroskedasticity -
Null hypothesis: no autocorrelation	Null hypothesis: heteroskedasticity not present
Test statistic: LMF = 0.54236	Test statistic: LM = 18.2764
with p-value = P(F(1, 19) > 0.54236) = 0.470452	with p-value = P(Chi-square (18) > 18.2764) = 0.437586

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package; **N.B.** Excluding the constant, p-value was highest for variable no. 5 *("AgriReform1921")*

As can be noted, the value of the ***R-squared*** (0.932727) is very high. Thus, it can be stated that due to the significance level recorded, the proposed econometric model is very suitable, or fits my data set. I then try to estimate the proposed theoretical econometric model: ***welfare/cerealtotoutput =*** $\beta_0 + \beta_2$ ** cerealtotoutput +* β_3 ** agrioutputvegetablemass +* β_4 ** agrioutputanimalorigin +* β_5 ** ruractpop +* β_6 ** agrisurf +* Δ_0 ** AgriReform1921* $(\beta_7) + \mu$, and write the results in the proper equation form.

In order to determine the numerical results for the previous question, one has at hand the *Ordinary Least Squares* (O.L.S.) method for *Time Series*, which offers the estimated linear econometric model. Thus, one can obtain the O.L.S. method results of the loaded dataset by resorting to the *Ordinary Least Squares* (O.L.S.) function/submenu in *gretl* (G.R.E.T.L.) *2018* edition. Following computation, one can obtain several relevant results. In order to try to adequately estimate my proposed theoretical econometric model ***cerealtotoutput =*** $\beta_0 + \beta_2$ ** agrioutputvegetablemass +* β_3 ** agrioutputanimalorigin +* β_4 ** ruractpop +* β_5 ** agrisurf +* Δ_0 ** AgriReform1921* $(\beta_6) + \mu$, one has to look at the coefficient column, in the output window provided by the gretl (G.R.E.T.L.) 2018 software edition. Thus, one can observe and make use of the $\beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6,$ parameter values, estimated for the 26 statistical observations *(years)* provided by the loaded dataset. In my econometric approach I consider that the β_1 parameter represents each year of observation in the data set (1911-1915;1919-1939). By introducing the coefficient values in the simple linear regression, one will obtain the following estimated equation: ***cerealtotoutput*** = 10.8950 + 5.26033e-06 * *agrioutputvegetablemass* - −1.91523e-06 * *agrioutputanimalorigin* + 4.70478e-08 * ruractpop + −8.13170e-07 * agrisurf − 0.736564 * ***AgriReform1921.***

As to the issue of the O.L.S. coefficients, one should note that the **10.8950** value of the intercept (***β_0 parameter***) tries to offer a measurement, in other words to quantify to some extent the average effect or the results registered on the whole period of time that I have taken into account, whilst holding all other unobservable factors fixed.

Table 7. Frequency distribution for dependant variable "cerealtotoutput", obs 1-26

interval	*midpt*	*frequency*	*rel.*	*cum.*
< 7.7333	7.2000	2	7.69%	7.69% **
7.7333 - 8.8000	8.2667	3	11.54%	19.23% ****
8.8000 - 9.8667	9.3333	5	19.23%	38.46% ******
9.8667 - 10.933	10.400	2	7.69%	46.15% **
10.933 - 12.000	11.467	8	30.77%	76.92% ***********
12.000 - 13.067	12.533	5	19.23%	96.15% ******
>= 13.067	13.600	1	3.85%	100.00% *

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package; **N.B.** number of bins = 7, mean = 10.4846, sd = 1.73221

As to the test for the null hypothesis of the normal distribution, the output of gretl (G.R.E.T.L.) 2018 provides us with the following numerical values and graph: Chi-square (2) = 2.159 with p-value 0.33977.

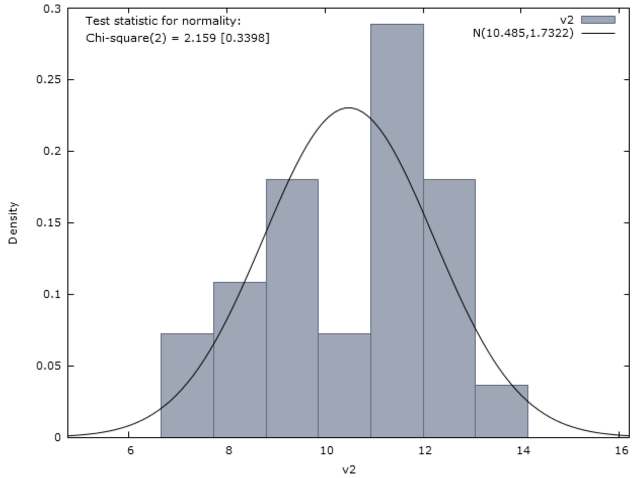

Figure 1. Test Statistic for Normality – *"Cerealtotoutput"* dependent variable
(*Source:* My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package)

Irrespective of the manner in which one selects the base group the most important issue one has to tackle with is not making any mistake with respect to the initial base. Many papers in the field of applied econometrics state that at least several researchers choose to make do without the β_0 overall intercept due to various reasons and opt instead for dummy variables for each selected group of focus. However, it is widely-known that testing for a difference in the intercepts requires a great deal of effort and is also quite difficult to grasp. Moreover, there is no general consensus on the best method of computing *R-squared* values in regressions in the absence of an intercept. Therefore, I choose to include an β_0 overall intercept for my base group of choice. (Wooldridge, 2012, p. 490-492). Factors of a qualitative nature frequently take the form of binary information. The field literature acknowledges the fact that useful or highly relevant information can be included by resorting to what can be described as a zero (0) - one (1) variable. Although the term is not very suggestive, variables of a binary nature are usually referred to as dummy variables. (Wooldridge, 2012, p. 227). When defining a dummy variable, one must decide which significant event is assigned the value ***1*** and which is assigned the value ***zero***.

For instance, in my proposed case study, of the overall effects of the agrarian reform that was implemented on the ground as of the year 1919 in Romania, one can define the *"Agrarian Reform of 1921 in the former Kingdom of Romania"* to be a variable a of binary nature taking on the value ***one*** (1) for *the years it underwent effective implementation* and the value ***zero*** (0) for the sample years it did not. The suggestive name I tried to assign in this case aims at indicating the historical event for Romania, at least, with the value one (**1**). I then tried to incorporate binary information into my proposed regression model, as in many other simple similar cases, by making use of just one dummy explanatory variable, by simply adding it as an independent variable in the econometric model or equation. For instance, one can take into account the following rather uncomplicated econometric model comprised in an attempt to try to encompass the level of the overall state of welfare of the population, as an effect/consequence of the ***Reform of 1921*** "variable" influence: ***"welfare"*** = $\beta_0 + \beta_2 * cerealtotoutput + \beta_3 * agrioutputvegetablemass + \beta_4 * agrioutputanimalorigin + \beta_5 * ruractpop + \beta_6 * agrisurf + \Delta_0 * AgriReform1921\ (\beta_7) + \mu$.

I have used the $\mathbf{\Delta_0}$ as the parameter on the ***AgriReform1921*** variable in order to highlight the interpretation of the parameters multiplying dummy variables. In my proposed O.L.S. model there are only ***five*** observed factors that affect the overall results on the level of the total output of cereal production, as a proxy for the welfare of the mostly rural population at that time in Romania: *agricultural output of vegetable mass*, *agri-food products of animal origin*, *rural active population* and *agricultural areas* (Wooldridge, 2012, p. 228-230).

Because the ***AgriReform1921*** = 1 for the years during which the Reform of 1921 was implemented (after World War I), and the ***AgriReform1921*** = 0, for the years during which the Reform of 1921 was an absent process (before World War I), the parameter $\boldsymbol{\Delta_0}$ has the following interpretation: $\boldsymbol{\Delta_0}$ is the difference in yearly overall quantitative difference/change that occurred *(relative change between the old value and the new one)*, during the years in which the Reform of 1921 was implemented and the period of time in which the reform was out of the question, given the same amount of observed factors such as total *agricultural output of vegetable mass*, *agri-food products of animal origin*, overall *rural active population* and geographic span of *agricultural areas* (**and the same error term** μ). Thus, the coefficient $\boldsymbol{\Delta_0}$ determines whether there is a relative quantitative change: if $\boldsymbol{\Delta_0}<0$, then, for the same level of other factors (*caeteris paribus*) the ***AgriReform1921*** variable had no effect whatsoever, for the interwar period of study, on the level of the welfare of individuals ("**cerealtotoutput"** dependent variable**)**, on average. In terms of expectations, if one assumes the *Zero Conditional Mean* (Z.C.M.) *Assumption* or criteria as being valid (Wooldridge, 2012, p. 229): E (u|AgriReform1921, **cerealtotoutput**) = 0, then Δ_0 = *E(cerealtotoutput* | ***AgriReform1921 = 1***, *cerealtotoutput, agrioutputvegetablemass, agrioutputanimalorigin, ruractpop, agrisurf) - E(cerealtotoutput* |***AgriReform1921 = 0***, *cerealtotoutput, agrioutputvegetablemass, agrioutputanimalorigin, ruractpop, agrisurf).*

Given that **AgriReform1921** = 1 corresponds to the period of time in which the reform was implemented and **AgriReform1921** = 0 corresponds to the time span in which the agrarian reform was still a highly debated proposal *(before World War I)*, one can write this more simply as follows: $\boldsymbol{\Delta_0}$ = E(cerealtotoutput | **AgriReform1921 = 1**, cerealtotoutput, agrioutputvegetablemass, agrioutputanimalorigin, ruractpop, agrisurf) - E(cerealtotoutput | **AgriReform1921**, cerealtotoutput, agrioutputvegetablemass, agrioutputanimalorigin, ruractpop, agrisurf). The key issue in the proposed econometric model is that the overall volume of cereal production harvested at countrywide level, as according to historical statistical records, is the same in both expectations. Thus, the difference measured by the *Δ_0 parameter* is due to the effective implementation of the *countrywide agrarian reform only*. From a historical perspective, one can note that the concept of a stationary process has been considered of great significance in the study and use of time series. One of the main assumptions and features of this particular concept is that if one takes any set of random variables in the series or sequence and afterwards moves that same sequence of interest ahead say ***h*** periods of time, the combined probability distribution should remain unaltered (Wooldridge, 2012, p. 381).

Given that stationarity is feature of an underlying stochastic process it may be quite difficult to determine whether the collected data were the result of a potentially stationary process. However, according to the field literature it is sometimes easy to identify some sequences that are nonstationary. A process with a time trend is obviously nonstationary: *at a minimum, its mean changes over time* (Wooldridge, 2012, p. 382). However, both theoretical and applied econometrics show that if one wants to get a grasp or better understand of the relationship between two, three or more variables taken into account by using the methods of regression analysis, one needs to assume the presence of some form of stability over time. If one accepts the possibility that a relationship between two or more variables of interest to alter randomly in each given period, one cannot expect to understand much about how the change in one variable affects the others under the circumstances of having only one time series at hand. The field literature acknowledges that when one tries to draft or further develop a linear equation, in other words a multiple linear regression model for time series, one must assume from the very beginning the presence of some form stationarity. Furthermore, with respect to time series analysis, including the present case study, the ***Homoskedasticity*** (TS.4) and ***No Serial Correlation*** (TS.5) criteria presume that the variation of the error process is constant over time and that the association between errors in two adjacent time intervals is equal to ***zero*** (Wooldridge, 2012, p. 382; p. 419).

In both field literature and practice most authors agree that processes that evince deterministic trends which are weakly dependent can be resorted to in regression analysis, if time trends are included in the proposed model. The same applies to processes with seasonality. When the time series subjected to a particular analysis are *highly persistent,* in other words, they have unit roots, one must be very cautious in using them directly in regression models - unless one is convinced of the *Classical Linear Model* (C.L.M.) assumptions/criteria. Due to recent advances in both theoretical and applied econometrics one can note that several more complex methods for using I(1) variables have become available and as a result, researchers can more easily note the presence of complete dynamics. In other words, interested individuals can more easily determine when no further lags are needed for any of the variables included in the model of choice. However, according to numerous papers tackling the issue of distributed lag models the errors will still bear signs of what is considered serial correlation. (Wooldridge, 2012, p. 621).

Step 1: testing for a unit root in cerealtotoutput

*Augmented Dickey-Fuller test for **cerealtotoutput***, including one lag of (1-L) **cerealtotoutput**, sample size 24, **unit-root null hypothesis**: **a = 1**

test with constant

model: (1-L)y = b0 + (a-1)*y(-1) + ... + e, estimated value of (a - 1): -0.622713;
test statistic: tau_c(1) = -2.49342;
asymptotic p-value 0.117;
1st-order autocorrelation coeff. for e: -0.012.

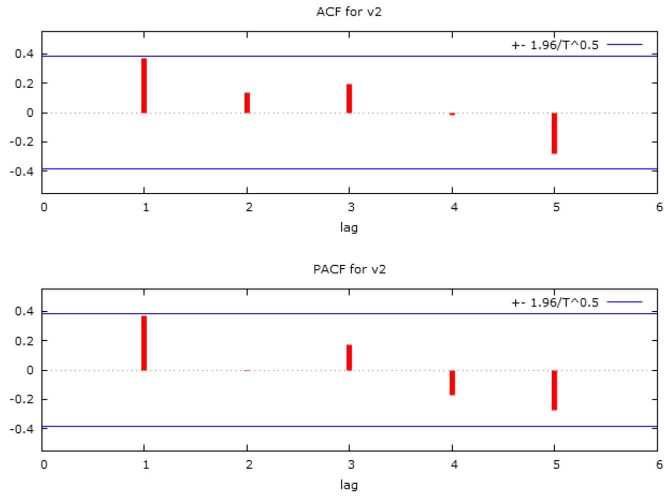

Figure 2. ACF Test for *"Cerealtotoutput"* dependent variable
(*Source:* My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package)

Table 8. Autocorrelation function for the *"Cerealtotoutput"* variable

LAG	ACF	PACF	Q-stat.	[p-value]
1	0.3714 *	0.3714 *	4.0160	[0.045]
2	0.1341	-0.0044	4.5614	[0.102]
3	0.1969	0.1723	5.7889	[0.122]
4	-0.0153	-0.1718	5.7967	[0.215]
5	-0.2784	-0.2762	8.4827	[0.132]

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package; **N.B.** ****, **, * indicate significance at the 1%, 5%, 10% levels using standard error 1/T^0.5*

Given the "Asymptotic" *Gauss-Markov* set of criteria for *Time Series Regression*, one has to check if these criteria are met in order to perform the inference required for time series regressions. One has to keep in mind this very important set of assumptions given that when discussing the time series variants of the classical linear model of regressions the main initial criteria are in a lot of cases not met, particularly the one usually known as the *strict exogeneity* condition, the *no serial correlation* criteria as well as the *normality distribution* (Wooldridge, 2012, p. 645; p. 686; p. 714; p. 828). A fundamental issue in this particular case is that some type of weak dependence is needed so as to make sure that the C.L.M. applies. When one further checks the *TS.4' and TS.5'* (after the initial *TS.1' through TS.3`* consistency check) then one can resort to the normal *confidence intervals, t-test results, and F-test results* as being roughly valid in large samples. According to the field literature, the usual initial assumptions needed in *Time Series Analysis* are basically the correspondents of the well-known Gauss-Markov set of criteria that let researchers, scholars or interested individuals to apply *statistical standard inference* (Wooldridge, 2012, p. 518; p. 646; p. 812).

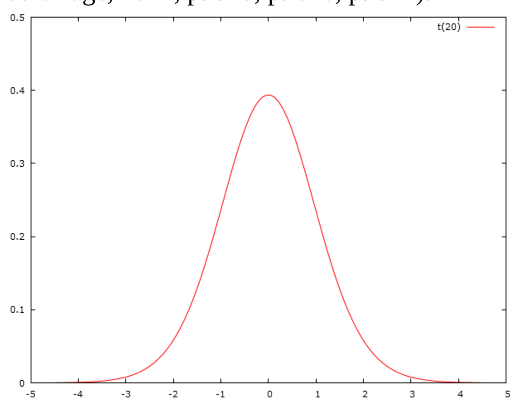

Figure 3. Distribution of the *"Cerealtotoutput"* dependent variable

(*Source:* My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package)

In my case study of the ***Reform of 1921***, one can observe that the sample analysis can only be rather limited due to the small size dataset taken into account and computed, given the general context and several methodological precise reasons. Thus, one has to check, amongst others, if the normality assumption condition or criteria is met.

Step 2: testing for a unit root in agrioutputvegetables

Augmented Dickey-Fuller test for ***agrioutputvegetables*** including one lag of (1-L) agrioutputvegetables; sample size 24; unit-root null hypothesis: **a = 1.**

test with constant, model: (1-L)y = b0 + (a-1) * y(-1) + ... + e;
estimated value of (a - 1): -0.155527;
test statistic: tau_c(1) = -1.33689;
asymptotic p-value 0.6145;
1st-order autocorrelation coeff. for e: -0.124.

By consulting available references in field literature one can furthermore observe the importance of the issue of what is called serial correlation in the errors of multiple regression models. It is widely known that positive correlation between adjacent errors is common occurrence, mainly in what is described as *finite distributed lag models*, amongst others. This rather important issue often causes the typical O.L.S. method standard errors and statistic results to mislead or confuse one. Typically, the O.L.S. standard errors do not estimate correctly the real uncertainty level in the computed parameter estimates. By taking into account this commonly known issue as a starting point, it is then more or less at hand for one to test for the presence of AR(1) serial correlation by resorting to the what is known as O.L.S. residuals (Wooldridge, 2012, p. 539; p. 540; p. 542; p. 574; p. 636; p. 637; p. 639; p. 640; p. 641; p. 649). Moreover, current field literature as well as applied Econometrics papers clearly show one that an asymptotically valid *t-test statistic* can be achieved by regressing the O.L.S. residuals computed on the lagged residuals, if and only if the regressors of interest comply with the strict exogeneity condition and the homoskedasticity criteria holds.

Step 3: cointegrating regression

Table 9. Cointegrating regression – O.L.S., using observations 1914-1939 (T = 26). Dependent variable: cerealtotoutput (t statistic results in short)

	coefficient	std. error	t-ratio	p-value
const	8.90525	1.26492	7.040	2.80e-07 ***
agrioutputvegeta~	6.92826e-07	5.35048e-07	1.295	0.2077

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package

Table 10. Summary statistics, Cointegrating regression – O.L.S., using observations 1914-1939 (T = 26). Dependent variable: cerealtotoutput

Mean dependent var	10.48462
Sum squared resid	70.11534
R-squared	0.065301
Log-likelihood	−49.78899
Schwarz criterion	106.0942
rho	0.438735
S.D. dependent var	1.732211
S.E. of regression	1.709232
Adjusted R-squared	0.026356
Akaike criterion	103.5780
Hannan-Quinn	104.3025
Durbin-Watson	1.070979

Source: My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package

Like many aforementioned concepts in the field, most researchers are also aware of the fact that computed *R-squared values* in time series regressions are frequently very high. However, the question one should ask is does this mean that one observes more about the factors that can exert some type of influence on the y variable in time series data? Apparently in the opinion of most academics and scholars - both normal and adjusted R-squared values for regressions involving time series can be in fact artificially high when the dependent variable shows signs of what is known as the `trending` phenomenon (Wooldridge, 2012, p. 646; p. 818).

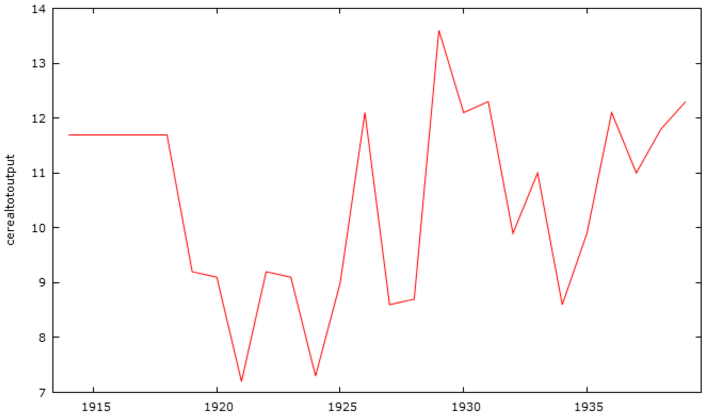

Figure 4. *"Cerealtotoutput"* variable trend
(*Source:* My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package)

One has to clearly remember the fact that R-squared value computed is in fact a means of determining how large the error variance in relation to the variance of the y variable. The field literature also shows one that when the dependent variable of interest meets the criteria for *linear, quadratic, or any other polynomial trends*, then it is quite easy to obtain what is widely known as the `*goodness-of-fit*` measure that in the first place "nets out" the influence of any time trend on the y_t variable (Wooldridge, 2012, p. 370).

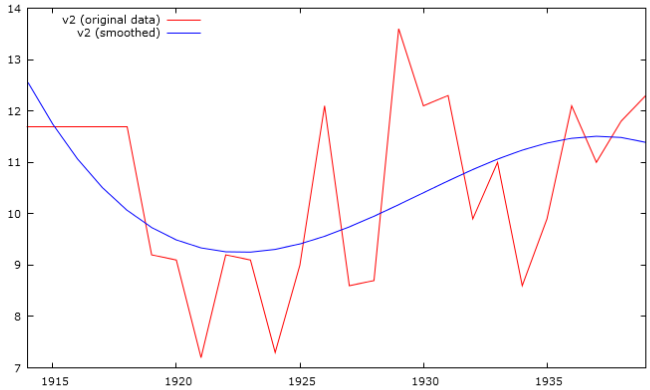

Figure 5. *"Cerealtotoutput"* dependent variable - polynomial trend (smoothed)
(*Source:* My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package)

Many time series of an economic nature exhibit a typical tendency of growing with time, as can be noted in the present case study. However, it must be acknowledged that certain series include a time pattern in order to draw a causal inference via time series data. Disregarding the fact that two sequences are going in the same or opposite directions will lead us to wrongly assume that changes in one variable are directly induced by changes in another. In many such cases, two time series processes seem to be bear signs of correlation only due to the fact that they are both trending over the course of time due to *causes related to other unobserved types of factors* (Wooldridge, 2012, p. 11; 363; 391; 660). Interpreting and keeping track of explanatory variables which show signs attributable to the trending phenomenon is more or less useful and quite straightforward, particularly in *Time Series Analysis*. In the first place, nothing related to trending variables must be taken to undermine the *classical linear model criteria from TS.1 to TS.6*. I must nevertheless allow for the fact that unobserved trending factors that exert an influence on the y_t variable could be correlated with the explanatory variables of interest. Finding a relationship between two or more trending variables, simply because each is increasing in time, represents an instance of what is well known as a *"spurious regression problem"*. According to the field literature, the classic approach with respect to this particular issue is adding a time trend in order to eliminate this particular issue (Wooldridge, 2012, p. 395; 498; 632).

Step 4: testing for a unit root in uhat
Augmented Dickey-Fuller test for uhat, including one lag of (1-L) uhat,
sample size 24, **unit-root null hypothesis: a = 1**
model: (1-L) y = (a-1) * y(-1) + ... + e, estimated value of (a - 1): -0.581256
test statistic: tau_c (2) = -2.60703

asymptotic p-value 0.2347
1st-order autocorrelation coeff. for e: -0.027
Source: My own computations based on the gretl 2018 edition statistical package

More often than not, I can include a set of seasonal dummy variables to interpret the seasonality phenomenon in the dependent variable, the independent variables, or both. Given the issue of strength and extent of *temporal correlation* present in most time series data, one must make assumptions with respect to how the errors might be related to the explanatory variables in all time periods while also checking for temporal correlation in the errors themselves (Wooldridge, 2012, p. 373; p. 375; p. 539; p. 654; p. 683; p. 825; p. 857).

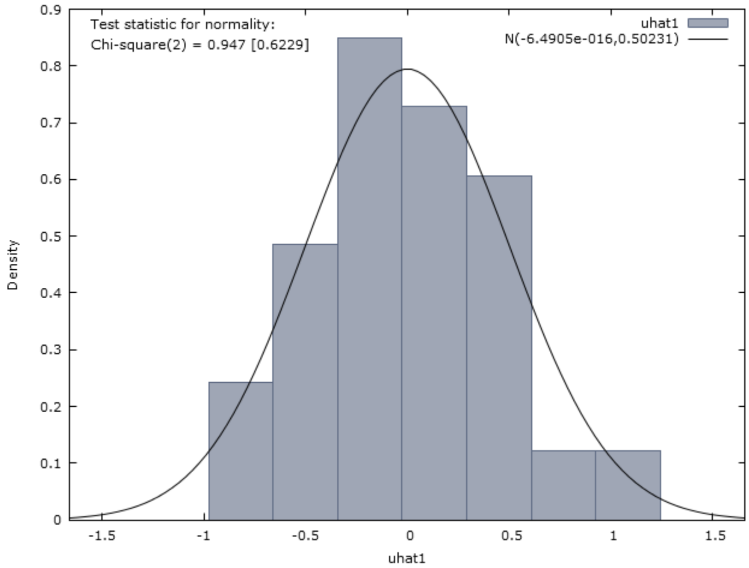

Figure 6. *"Cerealtotoutput"* dependent variable – polynomial trend (smoothed)
(*Source:* My own computations based on the gretl (G. R. E. T. L.) 2018 edition statistical package)

For models with a lagged dependent variable, like in my case, *the standard t-test on ˆu is still valid*, provided all independent variables are included as regressors along with **ˆu.** I can use an *F test or an L.M. statistic to test for higher order serial correlation*. In models with strictly exogenous regressors, one can attempt to use a *feasible G.L.S. procedure - Cochrane-Orcutt or Prais-Winsten* - to correct for A.R. (1) serial correlation, for instance. This provides interested individuals with estimates that are at variance with the O.L.S ones: the F.G.L.S. estimates are derived from O.L.S. on quasi-differenced variables. All of the usual test statistics from the transformed equation are asymptotically valid. Statistical regression software, such as the gretl (G.R.E.T.L.) 2018 edition, have integrated features for estimating models with A.R. (1) errors. An alternate way to tackle the issue of serial correlation, particularly when the strict exogeneity assumption might not hold, is not to resort to the O.L.S. method but to obtain and interpret the serial correlation - robust standard errors. The *Breusch-Pagan* and *White* tests, for example, can be directly used, keeping in mind that the errors ought not to be serially correlated. **There is evidence for a cointegrating relationship if:** (a) The unit-root hypothesis is not rejected for the individual variables, and (b) the unit-root hypothesis is rejected for the residuals **(uhat)** from the cointegrating regression, according to my computations based on the gretl (G.R.E.T.L.) 2018 edition output. (Wooldridge, 2012, p. 287; 289; 440; 843).

Any studies of historiography, both the existing and potential ones, necessary for a better understanding of the rural space, imply a series of major and minor inherent objectives. One important difficulty is that the sources of information are inconsistent, incompletely preserved, ambiguous or unclear. In other cases, the information available contains varying units of measurement, a situation that requires additional conversions. Further problems are related to conflicting information as found in the available sources, forcing researchers to rely on or to resort to mere estimations of data series, calculations and/or information which cannot be easily or even at all verified. Such difficulties and similar ones have often led to the discouragement of researchers attempting to study the rural space (Doboș, 2020). In spite of these shortcomings, the available research on the Economic History, the History of Statistics and Historical Sociology of the Romanian rural space provides a general assessment of both advantages and drawbacks, for example as a result of the agrarian reforms and following the strategies and measures adopted by authorities with a view to improve the living standards of the national rural areas (Axenciuc, 2012, p. 9-33). Moreover, irrespective of any inherent deficiencies and/or limitations, the general assessment is bound to be an important source of information for future researchers, authorities or indeed anyone interested. The available literature (primary, secondary, tertiary) in libraries and archives also makes for the effective elimination of potential reiterations of the former errors, aleatory or systematic, of past generations.

IV. CONCLUSIONS

As is well-known, the interwar period in Romania was an era marked by political, economic, social and cultural mutations. Various economic interests that became imperative in order to achieve the goals of modernization, Europeanization and the overall development of society, aimed to turn Romania into a country in which the free movement of goods, competition, freedom of transactions, alongside private property and the entire bourgeois-liberal legal system were to be established after the Western model. Independence and the economic liberal principles were the main requisites for achieving fully-fledged individuality as well as economic and political independence. Despite the apparently large expanse of available land (over 6 million hectares), the average size of peasant property that resulted from the redistribution of available land failed in many cases to reach the 5 hectares stipulated, which were at that time considered to be the minimum threshold required for the subsistence of family members within a typical traditional rural household.

In the entire Kingdom of Romania, including the former interwar county of Iaşi, changes to the land ownership structure were constant in the post-legislative period of the 1921 agrarian reform, the attribute of the dynamics of this process being the inherent result of the relatively constant division of the land into smaller and smaller plots due to the generalized phenomenon of the failure of villagers` payments, situation which in many cases resulted in the end in the actual alienation of the lots awarded. The moral justification or fairness of the reform was intensely discussed at the time, and remains even at present, at least in some authors` view, to some extent controversial. On the other hand, the peasantry had fully acquired the right to property, from a moral point of view, be it only for their undeniable sacrifices in the trenches of the First World War.

Many authors view the 1921 land reform as a state-mediated process of wide-scale sale of expropriated landed properties to the peasants who did not own any land or whose land was insufficient to ensure subsistence. The actual implementation of the provisions of the decree-laws issued in the period 1918-1921 triggered the substantial decrease of the large landed properties exceeding 100 hectares, according to estimates, to 15-17% of the total arable land in Romania, and to 27-28% of the total agricultural areas. Peasantry (the "small agricultural holdings") thus became prevalent in the primary sector of the country, including in terms of the volume of agricultural production. Out of the over documented 2,300,000 peasants who were either unappropriated or whose land was insufficient in terms of subsistence and recorded as such, only 1,479,000 or 64% were eventually appropriated, which means that a large number of villagers across the country remained unappropriated in spite of the expended efforts and initial targets.

Although there are inevitable quantitative differences in terms of the effects produced, it can be safely said that the 1921 agrarian reform improved the material situation of the peasantry, to a smaller or larger extent, in accordance with the historical circumstances in the various regions of the country. This conclusion is confirmed by the numerous archival records and relevant statistics from the interwar period, as well as by the monographs of villages in the county of Iași, which clearly prove that the implementation of the 1921 agrarian reform improved the economic situation of a large number of household heads, especially those previously deprived of any form of land ownership. The potential results of an analysis of a series of social and economic parameters can quantitatively provide further proof of the favorable impact of the agrarian reform on the peasants' standard of living following the 1921 agrarian reform.

V. REFERENCES

1. Andrei, J.V., Chivu, L., Ioan-Franc, V., Sima, V. (2020). Practici și Exigențe în Scrierea Lucrărilor Științifice, Editura Expert, București www.researchgate.net/publication/343336664_PRACTICI_SI_EXIGENTE_IN_SCRIEREA_LUCRARILOR_STIINTIFICE
2. Axenciuc, V. (1997). *Introducere în istoria economică a României,* Editura Fundației „România de Mâine", București
3. Axenciuc, V. (1992). *Evoluţia Economică a României: Cercetări Statistico-Istorice, 1859-1947*, Editura Academiei Române, Bucureşti
4. Axenciuc, V. (2012). *Produsul intern brut al României - 1862-2000. Serii statistice seculare și argumente metodologice,* Vol. II, Editura Economică, București
5. Centoni, M. & Cubadda, G. (2015). "Common Feature Analysis of Economic Time Series: An Overview and Recent Developments". SSRN Electronic Journal. 10.2139/ssrn.2669527 or https://dx.doi.org/10.2139/ssrn.2669527
6. Chivu, L. & Ioan-Franc, V. (2019). „Cercetarea economică din România", în Iancu, A.G. et al. (2019). *Economia României după Marea Unire Vol. 2: Economia sectorială*, Editura Academiei Române, București https://www.researchgate.net/publication/335858231_Cercetarea_economica_din_Romania
7. Doboș, S. (2020). "THE IMPLEMENTATION OF THE AGRARIAN REFORM OF 1921 IN THE FORMER "COPOU" UNIT OF THE COUNTY OF IAȘI - NEW FINDINGS". Ecoforum Journal, North America, 9, jan. 2020. Available at: <http://www.ecoforumjournal.ro/index.php/eco/article/view/1055/646>. Date accessed: 18 Sep. 2020.
8. Doboş, S. (2018). *Economia rurală din judeţul Iaşi după reforma agrară din 1921: Evoluţii, particularităţi şi învăţăminte*, Editura Terra Nostra, Iași
9. Dropu, C. (2011). *Un secol de frământări agrare și drama țăranului român,* Editura Academiei Române, București

10. Gavrilă, I. (2009). *Cercetarea istorică asistată de calculator,* Editura Oscar Print, București
11. Hitchins, K. (2007). Rumania: 1866-1947, Clarendon Press, Oxford, Reprinted
12. Gretl 2018 official website (2020). gretl: Gnu Regression, Econometrics and Time-series Library for Microsoft Windows ("cross-platform software package for econometric analysis, written in the C programming language") http://gretl.sourceforge.net/
13. Microsoft Corporation. (2019). *Microsoft Excel*. Retrieved from https://office.microsoft.com/excel
14. Murgescu, B. (2010). *România și Europa: Acumularea decalajelor economice (1500-2010)*, Editura Polirom, Iași
15. Șandru, D. (1975). *Reforma agrară din 1921 în România*, Editura Academiei R.S.R., București
16. Șandru, D. (1980). *Populația rurală a României între cele două războaie mondiale*, Editura Academiei R.S.R., Iași
17. Wooldridge, J.M. (2012). *Introductory Econometrics: A Modern Approach*, Cengage learning, Australia