TWO YEARS ANALYSIS OF ROMANIAN COUNTIES FROM AGRICULTURAL ACTIVITIES AND GDP PER CAPITA POINT OF VIEW

Ionela-Cătălina Zamfir[1]
[1)] *The Bucharest University of Economic Studies, Romania*
E-mail: ionela.zamfir@csie.ase.ro

Please cite this paper as:

Zamfir, I.C., 2020. Two Years Analysis of Romanian Counties From Agricultural Activities and GDP per Capita Point of View. In: R. Pamfilie, V. Dinu, L. Tăchiciu, D. Pleșea, C. Vasiliu eds. *6th BASIQ International Conference on New Trends in Sustainable Business and Consumption.* Messina, Italy, 4-6 June 2020. Bucharest: ASE, pp. 864-870

Abstract
Data analysis methods are used mostly to identify hidden patterns between data in a multidimensional approach, so that the main purpose and hypothesis to be tested in this research is the existence of a relationship between GDP per capita and agriculture activities. Methods like principal components analysis to lower how many variables are used, as well as clustering techniques for grouping Romanian counties were applied. The purpose is to identify if counties with high agriculture activities (in terms of production and cultivated areas) also have a similar level for GDP per capita. Confirming this hypothesis, the connection between GDP per capita and agriculture is obvious. Romanian counties were analyzed considering two recent years (2017 and 2018) and 43 indicators that reflect the agriculture activity and the results are that, in both years, there is a relationship between the level of GDP per capita and the level of agriculture activities (in average).

Keywords
Principal components analysis, fuzzy clustering, agriculture variables, Romanian Counties

JEL Classification
C38

Introduction
There are many interactions, correlations and dependencies among the macroeconomic variables and sometimes a slightly change of one variable can lead to major changes of other variables (if macroeconomics is seen as a cybernetic system). The main objective of this research is to analyze the Romanian counties from agricultural activities point of view (cultivated area and production), in relation with GDP per capita. A two years analysis provides an idea about the changes that occur in different areas and what is the impact (if there is any impact) in GDP per capita.
The main hypotheses that the research rely on are (the hypothesis that the selected variables are reliable is considered true by default): there is a relation between the production and

cultivated areas of cereals, vegetables, fruits and GDP per capita; this relation is better described by a clustering model applied on Romanian counties and is "visible" in time (for a period of at least two years). Taking into consideration that, in Romania, the agriculture, forestry, and fishing, value added is over 4% of GDP (World Bank, n.d.) in the last 2 years with available data (period 2017-2018), there is a high probability to confirm the hypotheses from above.
The research is divided in sections like: after the literature review part, the next section presents the methodologies used, the variables and observations considered, followed by the results of applied methods and discussions about the results. The final part represents the conclusions, the limits of this study and further research.

Literature review

In 2015 (Tudorache (Zamfir), 2015) a study that uses 2014 data from agriculture provides a general view of Romanian counties. In this study, the classification of Romanian counties is made using 33 variables from agriculture, without including macroeconomic variables. In another study from 2015 (Muhammad, Saba and Ghulam, 2015), data from 1975 to 2012 is used to identify the influence of agriculture in GDP by regression model. The results show that "the parameters estimate for Industry, trade and agriculture are showing positive and significant relation with GDP growth rate" (Muhammad, Saba and Ghulam, 2015). Another similar study from Nigeria (Olajide, Akinlabi and Tijani, 2012) show a strong relation between agriculture output and GDP taking into account a regression model applied on data from 1970 to 2010. For Nigeria, another study (Odetola and Etumnu, 2013) analyze data from 1960 to 2011 with methods like Granger causality test and demonstrate that "the agriculture sector has contributed positively and consistently to economic growth in Nigeria, reaffirming the sector's importance in the economy" (Odetola and Etumnu, 2013).
A study from 2019 (Toacă and Olărescu, 2019) show that agriculture production has a significant influence for other economic sectors and is considered for prediction of main macroeconomics indicators, using econometric models.
In 2020 (Mehta, 2020) a study from India with data from 1961 to 2016 (and econometric models) reveals that there is a strong relation between the production from agriculture and economic growth in India and shows that the "Economic Growth leads to Agriculture Production but Agriculture Production does not lead to Economic Growth in the long run" (Mehta, 2020). Another recent study (Ayuda and Pinilla, 2020) for Spain consider agriculture exports to have a "positive although moderate" impact in economic growth.

Methodologies, datasets and variables

Seeking a relation between a group of variables and another variable in order to group and describe a set of observations is a problem that can be solved in many different ways. The regression analysis is one of the most used methodologies to analyze the relations between variables, but for grouping a set of observations, cluster methods are required, like: partitioning algorithms, hierarchical clustering, density-based methods or soft clustering (like fuzzy clustering). Also, the dimensionality of datasets represents an issue. Among methods to reduce the number of variables, the principal components analysis provides the best results. This type of analysis is preferred in high dimensionality datasets, because, using a maximization problem (the maximization of variance contained by new variables), it reduces the number of variables to several new variables named principal components. The way that the maximization problem is applied provides several properties for principal components, like: principal components are uncorrelated one with each other and the first principal component takes more information from original variables than the rest of the components. On the other side, for unsupervised pattern recognition, clustering methods are used in general to group a set of forms so that the variability within clusters is minimum and the variability between classes is maximum. There are hard-clustering methods, when each observation is allocated to one class and soft-clustering methods (like fuzzy) when each observation belongs to all classes, but with different probabilities. One of the most known and used fuzzy clustering algorithm is FCM - fuzzy c-means, where "the centroid of a cluster is calculated as the mean of all points, weighted by their degree of belonging to the cluster" (source: https://www.datanovia.com/en/lessons/fuzzy-clustering-essentials/). In R, this algorithm is implemented in cluster package (fanny function).

The main source of the dataset is the National Institute of Statistics (site: http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table) available statistics (most variables), as well as the available data from the National Commission for Strategy and Prognosis (the GDP per capita at county level, site: http://www.cnp.ro/ro/prognoze). The variables considered are grouped in four categories: I. Cultivated area (hectares) of: pepper, potatoes, onion, beans, sunflower, wheat, orchards, green peas, barley, oat, corn, root vegetables, colza, soy, tomatoes, garlic, cabbage, eggplants; II. Production (tones) of: pepper, beans, sunflower, wheat, field vegetables, solar vegetables, garden vegetables, green peas, barley, root vegetables, colza, rye, soy, tomatoes, garlic, cabbage, eggplants; III. Production (tones) of: apricots, cherries and sour cherries, apples, nectarines, nuts, pears, peaches, plums; IV. Indicators like: GDP per capita in euro. All these variables are considered at counties level (41 counties were selected) and for years 2017 and 2018. The GDP per capita for all years is provided by National Commission for Strategy and Prognosis reports (because the analyzed reports are from May 2018 and December 2019, the values for GDP per capita in euro are considered prognosis values). The use of prognosis estimations is for computing and comparing clusters centroids, and for grouping counties. Finally, there are 2 major datasets (one for each considered year), each having 43 indicators from agriculture and one macroeconomic indicator (GDP per capita in euro).

Results and discussions

In this research, the principal components analysis is used on each dataset that represent the agriculture results. The table no. 1 shows the principal components analysis results in terms of total information contained by principal components.

Table no. 1 Principal components analysis results

PCA	%of info (3 PC)		
	cultivated area	cereals & vegetables production	fruits production
2017	68,24%	69,87%	78,69%
2018	67,89%	70,41%	79,84%

Source: Author's computations

In order to keep the comparability between years (but also, taking into consideration the scree plot and Kaiser criteria for choosing the best number of principal components), three components were considered from each dataset. From 18 variables that represent the cultivated area with vegetables, cereals and fruits, 3 principal components synthesize about 68% of information. Using 17 variables that represent the production of cereals and vegetables and keeping 3 principal components, the total information considered is about 70%. Also, from 8 variables that represent the fruits production, three principal components have between 79% and almost 80% of information.

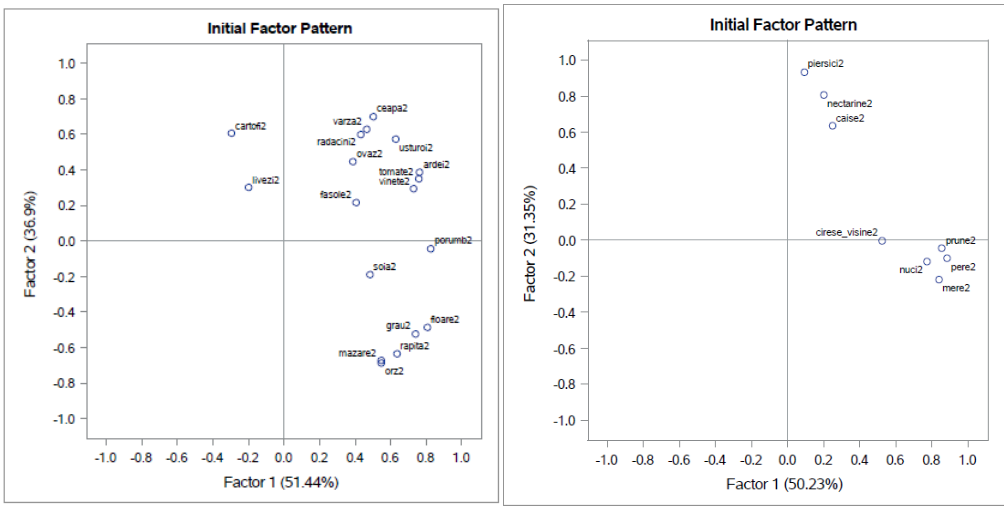

Fig. no. 1 Factor pattern (2018 dataset)
Source: Author's computations using SAS

The factor pattern matrix (graphically represented above - fig. no. 1 - for 2018, cultivated areas dataset in the left side and production of fruits dataset in the right side; SAS output) is the correlation between original variables and principal components. Using these correlations, each principal component may be named like: cereals_ha, vegetables_ha, beans_soy_orchards_ha for the dataset that represents the cultivated area for cereals, vegetables and fruits; vegetables_tones, cereals_tones, beans_rye_tones for the second dataset with the production of cereals and vegetables, and autumn_fruits, summer_fruits1, summer_fruits2 for the dataset that represents the production of fruits. In the first dataset (the cultivated area), the first principal component is correlated with most cereals (and few vegetables), so the name is cereals_ha, while in the last dataset, the new variable autumn_fruits is highly correlated with fruit production of apples, nuts, pears and plums, the second principal component (summer_fruits1) is highly correlated with the production of apricots, nectarines and peaches, while the last principal component is strong correlated with the production of cherries and sour cherries. The figure no. 2 shows the silhouette graph for 2017 dataset.

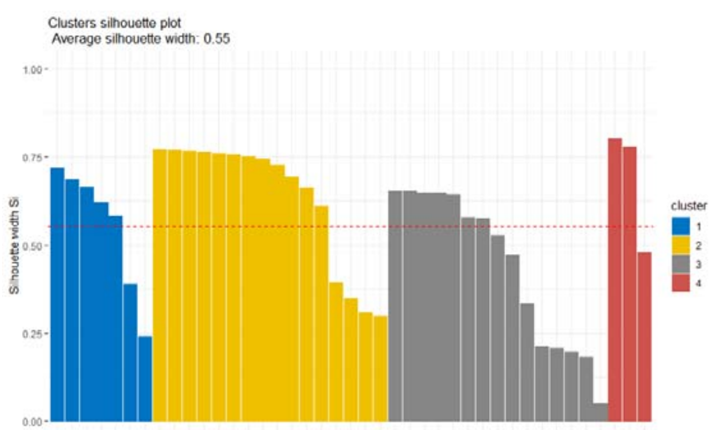

Fig. no. 2 Silhouette graph (2017 dataset)
Source: Author's computations using R

A negative value for silhouette width means misclassification for that county, while a closer to unit positive value represents a good allocation to class. No county from 2017 dataset is misclassified. Also, the nbclust function in R suggest that either 2 or 4 classes is the best number of clusters, according to the majority rule, so that 4 classes were selected, to obtain a more detailed analyze.
The table no. 2 is the classes structure for each dataset.

Table no. 2 Classes structures

Class	Year/number of counties	
	2017	2018
CLS1	7	7
CLS2	16	17
CLS3	15	14
CLS4	3	3

Source: Author's computations

There is a high similarity between 2017 and 2018, according to classes structure. Using each class average vector, it is possible to characterize classes as:
- one class have (in average) high cultivated areas with cereals, but low cultivated areas with orchards and low cultivated areas with vegetables, beans and soy. The production of cereals, beans and rye is among the highest, the production of vegetables, autumn fruits is high, the production of cherries and sour cherries is high, while the production of apricots, nectarines and peaches is low. The GDP per capita for this class is the lowest (in average). This is class 3 in 2017 and 2018 datasets.
- another class have counties with small cultivated areas for vegetables, large cultivated areas for cereals, the highest production for vegetables, cereals and apricots, nectarines and peaches, as well as the highest GDP per capita (in average). This is class 4 in 2017 and 2018 datasets.
- the other two classes have (in average) lower cultivated areas for cereals than for vegetables, as well as for beans and soy; a low production (in average) for vegetables, cereals, beans and rye, big production (for class 2) for autumn_fruits (apples, nuts, pears and plums) and the average value of GDP per capita between the two classes described above. These are classes 1 and 2 for 2017 and 2018 datasets.
The maps (fig. no. 3) represent the conclusion of fuzzy clustering.

Fig. no. 3 Clustering results (2017 - left and 2018 - right datasets)
Source: Author's computations

The counties colored in light yellow correspond to the first class described above, the green color is for class 4 in 2017 and 2018 datasets, while light blue and light red are for the classes from the middle. There is a high similarity between 2017 and 2018 datasets in terms of classes computed by fuzzy c-means method. The only notable difference is Covasna county that changed class from yellow color to light red color, meaning an increase of GDP per capita, the rest of production and cultivated areas having low changes from one year to another (except the fruits production, that is significantly higher in 2018 than 2017).

Conclusions

The main results of the study show the confirmation of the hypotheses proposed, that there is a relationship between GDP per capita and agriculture activities (cultivated areas and production). This relation is described in time (for two consecutive years) by comparing the GDP per capita (in average) with the main indicators that show (in average) the agriculture activities. Classes with counties having the most intense agriculture activity (in average), also have, in average, the highest GDP per capita (for both years), so the relation between agriculture and the macroeconomic indicator is positive.

But, taking into consideration the limitations for this study, few ideas should be considered like: the natural endowment of each county (some counties that have mountain terrain are expected to produce more fruits than cereals) and the weather from each year that has a high impact on agriculture activities (for example late winter affects the production of summer fruits, the lack of precipitations affects cereals production).

As further research, in order to identify how intense is the relationship between GDP and agriculture at county level and to compute the impact of agriculture activities in GDP (or economic growth), other methodologies could be applied, like econometric methods (maybe time-series).

References

Ayuda M.I., Pinilla, V., 2020. Agricultural exports and economic development in Spain during the first wave of globalization. *Documentos de Trabajo de la Sociedad Española de Historia Agraria 2001*, Sociedad Española de Historia Agraria.

Data Novia, n.d. *Fuzzy Clustering Essentials*, [online] Available at: <https://www.datanovia.com/en/lessons/fuzzy-clustering-essentials/> [Accessed at 24 February 2020].

INSSE, n.d. *Baze de date statistice*, [online] Available at: <http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table> [Accessed February 2020].

Mehta, S. N., 2020. Does Agriculture Production Matter For Economic Growth? Empirical Evidence from India. *Our Heritage*, 68(1), pp.7367-7373.

Muhammad, M.A., Saba, F. and Ghulam, Y.K. 2015. Agriculture sector performance: An analysis through the role of agriculture Sector share in GDP. *Journal of Agricultural Economics, Extension and Rural Development*, 3(3), pp.270-275.

Olajide, O.T., Akinlabi, B.H. and Tijani, A.A., 2012, Agriculture Resource and Economic Growth in Nigeria. *European Scientific Journal*, 8(22), pp.103-115.

Toacă, Z. and Olărescu, Z., 2019. Model Estimates of the Macroeconomic Indicators of the Republic of Moldova for the Period 2019-2022. Culegere de lucrări ştiinţifice ale Conferinţei Ştiinţific Internaţional*”Competitivitate şi Inovare în economia cunoaşterii*”, Ediţia a XXI-a, Chişinău, Moldavia, 27-28 septembrie 2019, pp.669-675.

Tolulope, O. and Chinonso, E., 2013, Contribution of Agriculture to Economic Growth in Nigeria. *The 18th Annual Conference of the African Econometric Society (AES) Accra*, Ghana, July 2013.

Tudorache (Zamfir), I.C., 2015, The Classification of Romanian Counties from Agricultural Point of View. In *Proceedings of the 4th International Conference Competitiveness of Agro-Food and Environmental Economy, Competitiveness of Agro-Food and Environmental Economy* (CAFEE) 2015, București, România 12 noiembrie 2015, pp.190-201.

World Bank, n.d. *Indicators,* [online] Available at: <https://data.worldbank.org/indicator> [Accessed February 2020].