

Developing an Appropriate Methodology for Farm Sampling in Order to Build a Representative Farm Database for Agricultural Policy Analysis Purpose. The Romanian Case.

Ionel Mugurel JITEA, Cristian MERCE, Diana Elena DUMITRAȘ

University of Agricultural Sciences and Veterinary Medicine, Faculty of Horticulture, Department of economic analysis and business administration, 3-5 Mănătur Street, Cluj Napoca, Romania;
mjitea@usamvcluj.ro

Abstract. The Common Agricultural Policy is foreseen to suffer important reforms after 2013. They may greatly affect the farm agricultural production because the way of accessing the direct payments will suffer important changes. In this context, impact studies at the farm level are very important. To do that, appropriate data about farms techniques, yields, costs, revenues and alike are needed. This paper investigates if the available Romanian statistical data are appropriate to be used in agricultural policy impact analyses. The lack of important information about Romanian farm's techniques is evident in the official sources. Moreover, the aggregation biases can alter the agricultural policy impact analysis based upon simulative means. To overcome this, an appropriate methodology to collect data from farm level, at the North-West Romanian development region, was proposed. The farm population was firstly divided in homogenous strata. The first one deals with the farm type, the second one count for the geographical position and the third one with the farm size. Then, using official statistical data and the selective research methodology, a stratified random sample was proposed. It comprised 285 mixed cropping farms, 140 dairy and 200 sheep farms. The farmer's professional organisation serves as a way of randomly identifying them into the survey. Individual data were collected based upon a face-to-face interview with the questionnaire as instrument.

Keywords: farms; random stratified sampling; Romania;

INTRODUCTION

Agricultural policies played an important role in the human society development along time. Usually, they represent a set of mandatory regulations that deal with the domestic agriculture production reglementation and the imports of foreign agricultural products into the internal market (Norton, 2004). In 2007, Romania became the 26th member of the European Union. It engaged to apply the European legislation framework in all economical fields. Agriculture is not an exception. Between 2007 and 2013, the Romanian farmers receive different types of annual subsidies within the first CAP pillar (Common Agriculture Policy) such as single area payment scheme (SAPS), compensatory national payments for the vegetal sector or compensatory national payments for livestock breeders. But, the actual agricultural policy measures, are foreseen to suffer important reforms after 2013 (European Commission, 2010). In this context is important to seek appropriate answers to the question: how these reforms will affect the Romanian agriculture in the future?

Such answers were assessed at different economic levels along time. At the farm level, especially by the means of simulative models (Dent and Blackie, 1979), different types of investigation were carried out. Buisse *et al.* (2007) or Matthews *et al.* (2006) chose to build different types of simulative farm models in order to empirically measure the 2003 CAP's

reform effects at the farm level. Regional models (Therod, 2011; Sofria and Degre, 2010) or general equilibrium models (Ianchovichina Elena *et al.*, 2011; Gotor Elisabeta *et al.*, 2010) were also extensively used in practice. The different type of employed tools depends on the available data, the research objectives and the cost of data collecting. Thus, farm models which are usually more accurate use individual farm data, whereas the regional and equilibrium models employ aggregate data issue from different types of official database such as FDAN (Farm Accountancy Data Network) or periodically official general surveys in agriculture. One objective of this paper is to analyse if the existing Romanian official agricultural data can be appropriate for policy purpose analysis. Then, a second objective is to identify an appropriate methodology to collect the unavailable data from the farm level. This is presented in details in the third part of the paper.

MATERIALS AND METHODS

Agricultural Romanian statistical data can be found in different types of official sources. One is represented by the Romanian National Institute of Statistics database, which annually presents official aggregate agricultural figures at national, regional and local level (<http://www.insse.ro>). Furthermore, periodically, special studies are carried out in agriculture. Starting with 2002, two Structural Investigations (2005, 2007) and two General Agricultural Census (2002, 2010) were realised. Even these data permit to draw out a good image of the Romanian agriculture; they say little or almost nothing about the Romanian individual farm techniques: revenues; costs; yields; yearly crop rotation; etc. Employing aggregate data in simulative models will produce inappropriate results due to aggregation biases (Arfini *et al.*, 2001). Other official data are represented by the FDAN database. It is set into practice by the Romanian Minister of Agriculture and Rural Development (<http://www.madr.ro>) by sampling techniques. A report is presented annually in which regional averages are delivered. The access of individual data is restricted by a set of laws that deal with personal data access, making the database inappropriate for farm level simulation. Except the Romanian sources, there are also international databases like Eurostat, Faostat, World Bank. They summarise the national official sources without adding other valuable information.

All these findings allow understanding that in order to simulate the CAP's impact at the farm level, the Romanian official sources should be completed with data taken directly from farm by sampling techniques. It is the method that presumes to extract from the entire farm population only a finite number of items forming a representative sample (Poate and Daplyn, 1993). In order to apply this methodology, several issues should be firstly addressed (Mead *et al.*, 2003). The first one presumes to proper define the universe population of the sampling. Thus the answer to the questions What is a farm in Romania? and In what conditions such a structure is relevant to be considered for sampling in the CAP's reform analysis? are very important. Ahearn Mary *et al.* (2009), stated that a farm represents the base unit for analysis in agriculture. It consists of a land area allocated to crop or livestock production. Such a definition presumes to include in the analysis the small subsistence or hobby farms too. Moreover, the farm definition varies depending on the analyzed system. O'Donoghue *et al.* (2009) shows that in USA a farm represents „the place where at least 1,000 dollars of agricultural products were produced and sold or which should be yearly sold”. At the European level, Ahearn Mary *et al.* (2009) extensively deal with this problem. Thus, in the European statistics, a farm represents a technical economical unit that has its own management and that is engaged in the agricultural production.

Such definitions are too general in the context of the *a priori* analyse of the CAP's reform in Romania. To properly define the universe population, the actual conditions for obtaining the direct CAP's subsidies were analysed. A farm will be entitled for subsidies if it develops agricultural activities on the Community territory regardless its juridical status. Furthermore, several minimum thresholds were imposed in the legal Romanian agricultural framework (Romanian Government, 2006, 2007): minimum 1 hectare size for crop production but with parcels bigger than 0.3 hectares in order to receive direct payment under the SAPS scheme; a minimum 3 heads in the cattle production (the age of the cattle should be at least 6 months); minimum 50 female sheep and 25 goat. So, all farms that fulfilled these minimum size levels represent the universe sampling population.

The second step consists of defining the research area and to identify the data base used for farm segmentation. Due to the important budgetary costs implied by data collection, the research area was restricted to the North Western Romanian development region (NUTS 2) (outlined with black in figure 1). It has six departments and it covers around 14.3% of the entire Romanian territory. Concerning the database used in segmentation there can be two options. The first one presumes to construct a special one within each research. It is both time and financial costly and that is why it is not recommended (Poate and Daplyn, 1993). The second option means to use the existing official data. It is used on a large scale in similar researches. For instance, Petrick, (2001) used a database constructed by the Polish extension service in the analysis of the inability of Polish farms to access credits. In order to do the farm segmentation the last Structural Investigation (2007) made by the National Institute of Statistics might be an appropriate one because it summarised the most up-to-date statistics about Romanian farms (results for 2010's General Agricultural Census were not available).

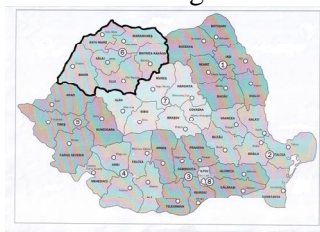


Fig. 1 "The research area"

The third step in collecting data by the means of selective research is to choose the population sampling methods. They can be classified as being either probability or nonprobability methods (Levy and Lemeshow, 2008). In probability samples, each farm has a known non-zero probability of being selected in the sample and the results for the sample can be representative for the entire population with a known probability. Furthermore this technique includes random sampling, systematic sampling, and stratified sampling procedures. The last one is commonly used because it reduces sampling error. A stratum represents a subset of population that shares one or more common characteristics. For example, in agriculture, strata can be defined as being the region, the farming type, the farm size, the farm juridical status and alike. Because this research method proved to be highly applied in other similar researches (Stevens and Olsen, 2004) it will be applied in Romania, such as to obtain the necessary data from the farm level.

RESULTS AND DISCUSSION

In order to collect individual farm data appropriate for the simulation of the post 2013 CAP's reform by simulative means, several farm strata were defined. The first one deals with the farm type. In this research, three farm strata were defined: mixed cropping, dairy and

sheep production. A farm is considered to be part of the first type if it has no animal production. The second one presumes that the farm has dairying cow production and the third one, sheep production (at least 2/3 of the revenues count for these specific products). This stratum was limited only to these general classes because the first CAP's reform pillar is foreseen to concern especially these farming types. Moreover, Romanian official statistical data do not allow dividing the Romanian farms into more detailed farm types. On the other hand, extending the strata to all nine classes as in the FDAN's regulations will produce unsustainable expenses for data collecting.

Tab. 1

The strata considered in the farm sampling

First level	Second level	Third level
Farm type: (1) mixed cropping; (2) dairying production; (3) sheep production;	Department: (..1) Bihor – BH; (..2) Bistrița Năsăud – BN; (..3) Cluj – CJ; (..4) Maramureș -MM; (..5) Satu Mare – SM; (..6) Sălaj – SJ;	Farm size (heads/farm)*: (...1) 50 – 99; (...2) 100 – 199; (...3) 200 – 499; (...4) above 500; *only for sheep farms

The second stratum type is formed by the six departments of the North-Western development region (tab1). They were included into survey segmentation to take into account the geographical particularities that might greatly influence the farm's techniques: higher altitude with mountainous and hilly areas in Maramureș, Bistrița Năsăud, Sălaj and Cluj counties; lower altitude in Bihor and Satu Mare counties. The last stratum deals with the farm's physical size. All the farms that did not reach the minimum threshold that makes them eligible for direct payments into the CAP (1 hectare for agricultural land; 3 cows and 50 sheep heads) were excluded from sampling. The mixed cropping farms are further analyzed into 4 size classes: 1-2 ha, 2-10 ha, 10-50 ha, over 50 ha; the dairy farms, also in 4 size classes: 3-5 heads, 6-19 heads, 20-49 heads, over 50 heads and the sheep farms – as in the above table.

Tab. 2

Farm number distribution in the N-W Romanian region according to the sample stratum

Department	Total farm number	Mixed cropping	Cattle farms total	Dairy farms > 3 heads	Sheep farms total	Sheep farms >50 heads
BH	122,492	27,194	32,222	1,461	2,102	731
BN	72,127	16,248	26,053	3,896	9,811	1,517
CJ	106,276	28,284	24,281	4,121	9,209	1,217
MM	95,294	27,189	42,667	3,585	8,896	883
SM	72,260	17,484	20,946	2,169	1,664	444
SJ	65,318	13,799	16,266	1,358	3,198	600
Total	533,767	130,198	162,435	16,590	34,880	5,392

Source: (Agricultural Structural Survey, 2007);

The optimal survey size was determined by the help of the official statistical data (tab.2). From the total number of 533,767 farms at the region level, 130,198 recorded only crop production. 149,806 represent cattle farms from which 16,590 farms had more than 3 heads of dairy cows. In the sheep production, there are 34,880 farms from which 5,392 farms had more than 50 sheep. These data permit to construct an alternative variable for each farm type. The first counts if farms employed only mixed crop production, the second one if the cattle farms are bigger than 3 milking cows and for the third one, if the sheep farms are bigger than 50 heads (tab.3) The distribution probability (p_i) for these variables was obtained by

dividing the total number of farms that respect the specific criteria (size, type), to the entire class population number.

Tab.3

The distribution probability of farm alternative variables

Description	Mixed cropping		Daring farm > 3 heads		Sheep farms >50 heads	
	1	0	1	0	1	0
Farm number	130,198	403,569	16,590	145,845	5,392	29,488

Source: own calculations;

The optimal survey strata size when applying a stratified sample without replacement within an alternative variable is (Levy and Lemeshow, 2008):

$$n = \frac{z_{\alpha/2}^2 \overline{\sigma^2}}{\Delta_{x_{st}}^2 + \frac{z_{\alpha/2}^2 \overline{\sigma^2}}{N}} \quad (1)$$

- $\overline{\sigma^2}$: the strata variance, equal for alternative variable with $p_i(1-p_i)$;

- $z_{\alpha/2}$: the multiplication coefficient for the $\Phi(z)$ function;

Where:

- N: the universe farm population;

- n: the sampling size (number of farms);

- $\Delta_{x_{st}}$: the error limit;

Thus for an error limit established at 5% and a probability to guarantee the results of 95% ($z_{\alpha/2}=1,96$), the optimal sample size became:

$$n_{sheep\ farms} = \frac{1.96^2 \times 0.1306}{0.05^2 + \frac{1.96^2 \times 0.1306}{34,880}} \approx 200\ farms$$

By the means of the same methodology, the sample is form by 285 cropping farms and 150 dairy farms. Moreover, the farm distribution into second and third level stratum was realised proportionally. If the number was smaller than 2, at least two farms were considered for face to face interviews in every stratum (tab.4).

Tab.4

The number of farm per stratum. Example for sheep farms.

	50-99 heads	100-199 heads	200-499 heads	> 500 heads	Total
BH	9	8	7	3	27
BN	39	15	3	2	58
CJ	18	14	11	3	46
MM	18	12	3	2	35
SM	4	7	4	2	18
SJ	8	7	5	3	23
Total					207

Source: own calculations.

CONCLUSIONS

Results shows that the *a priori* analysis of the potential effects of the CAP post 2013 reforms at the Romanian farm level need appropriate farm data that are not available in official sources. Based upon the sampling techniques and the official statistics the paper presents a representative stratified sample. It is form by 285 mixed cropping farms, 140 daring farms and 200 sheep farms. They were extracted randomly from the professional association's data bases in order to collect individual farm data base a face-to-face interview having the questionnaire as a working instrument.

Acknowledgments. This research is part of a scientific project financed by the Romanian National Council of Scientific Research (post doctoral fellows PN II - project number 190/2010).

REFERENCES

1. Ahearn Mary, K. Poppe, Cristina Salvioni, K. Boone, A. Roest (2009). The family farm in a flat world: implications for farm household data collection. WYE City Group on Statistics, Rural Development and Agricultural Household Income. Italy, Rome: 1-20;
2. Arfini F., Cristina Brasili, R. Fanfani, M. Mazzocchi, Elisa Montresor, Q. Paris (2001). Tools for evaluating EU agricultural policies: an integrated approach. *Statistical Methods & Applications*. 10: 191-210;
3. Buysse J, Fernagut B, Harmignie O, Frahan BH, Lauwers L, Polome P, Van Huylenbroeck G, Van Meensel J (2007). Farm based modelling of the EU sugar reform: impact on Belgian sugar beet suppliers. *European Review of Agricultural Economics*. 34(1): 21-52.
4. Dent, J.B., M.J. Blackie (1979). *Systems simulation in agriculture*. Applied Science Publishers Ltd. London. UK.
5. European Commission (2010). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: The CAP towards 2020: meeting the food, natural resources and territorial challenges of the future. European Commission, Brussels.
6. Gotor Elisabeta, M.E. Tisigas (2010). The impact of the EU sugar trade reform on pour household in developing countries: A general equilibrium analysis. *Journal of Policy Modelling*. 33: 568-582.
7. Ianchovichina Elena, R. Darwin, R. Shoemaking (2011). Resource use and technological progress in agriculture: A dynamic general equilibrium analysis. *Ecological Economics*. 38: 275-291;
8. Levy P.S., S. Lemeshow (2008). *Sampling of populations: methods and applications*. John Wiley & Sons, USA.
9. Matthews K.B., I.A. Wright, K. Buchan, D.A. Davis, G. Schwarz, (2006). Assessing the options for upland livestock systems under CAP reform: Developing and applying a livestock whole farm system analysis. *Agricultural Systems*. 90: 32-61.
10. Mead R., R.N. Curnow, A.M. Hated (2003), *Statistical methods in agriculture and experimental biology*. Chapman & Hall, USA.
11. Norton, R. (2004). *Agricultural Development Policy Concepts and experiences*. John Wiley Sons Lts. Chichester. UK.
12. O'Donoghue E., R. Hoppe, D. Banker, P. Korb (2009). Explaining alternative farm definitions. U.S. Department of Agriculture, Economic Research Service. 49: 1-33;
13. Petrick M. (2001). Documentation of the Poland Farm Survey. Institute of Agricultural Development in Central and Eastern Europe, Discussion Paper (36): 1-52;
14. Poate C.D., P.F. Daplyn (1993). *Data for agrarian development: Wye studies in agricultural and rural development*, Cambridge, Cambridge University Press.
15. Romanian Government (2006), Governmental Ordinance 125/2006 for the approval of the direct and complementary national direct payments, which are given in agriculture since 2007: 1-6;
16. Romanian Government (2007), Ministerial Decree 295/2007 for the establishment of specific conditions and eligibility criteria in the application of complementary national direct payment schemes in the livestock sector : 1-8;
17. Sofria C., A. Degre (2010). Modelling the effects of the current policy measure in agriculture: A unique model from field to regional scale in Walloon region of Belgium. *Environmental Science and Policy*. 13 (2010): 754-765.
18. Stevens D.L, A.R. Olsen, (2004). Balanced sampling of natural resources. *Journal of the American Statistical Association*. 99 (465): 262 -278;
19. Therond O., H. Hengsdijk, E. Casellas, D. Wallach, Myriam Adamb, H. Belhouchette, R. Oomen, G. Russell, F. Ewert, J.E. Bergez, S. Janssen, J. Wery, M. K. Van Ittersum (2011). Using a cropping system model at regional scale: Low-data approaches for crop management information and model calibration. *Agriculture, Ecosystems and Environment*. 142:85-94.