ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume LXI 46 Number 2, 2013
http://dx.doi.org/10.11118/actaun201361020405

USE OF THE CLUSTER ANALYSIS FOR ASSESSMENT OF ECONOMIC SITUATION OF AN ENTERPRISE

Martin Maršík, Daniel Kopta

Received: September 6, 2012

Abstract

MARŠÍK MARTIN, KOPTA DANIEL: *Use of the cluster analysis for assessment of economic situation of an enterprise.* Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2013, LXI, No. 2, pp. 405–410

The aim of the paper was to discuss the disadvantages of enterprises in dry areas compared to enterprises farming in similar production area outside a rain shadow. The analysis was based on the sample of 45 enterprises; twelve of which farming in the area of a rain shadow. In the first step, enterprises were sorted by the cluster analysis into groups farming in the same area, at a similar altitude, with the same structure in a similar manner, and under comparable financial conditions – (such as debt ratio, liquidity and activity ratio). The results of this step showed a different method of farming within enterprises in disadvantaged areas. Such enterprises have created two distinct, separate clusters differing from the average in the use of fixed assets, technical equipment of labour, labour productivity and income structure. In the second step, the way how the return on assets of such enterprises is different from the average profitability of the enterprise was assessed. Testing differences in mean values of profitability was performed using the Student t–test. Due to the high variability of the Return on assets (ROA), the difference between standard and disadvantaged enterprises has not been proved.

financial analysis, dry areas, enterprise evaluation, agriculture

The aim of the paper was to prove how much are enterprises farming in dry areas handicapped compared to farms in similar production area outside a rain shadow. Our working hypotheses suggested that profitability of farms in dry areas is lower and official price based on the BPEJ (translated as valuated soil ecological unit or estimated pedologic-ecological units) system is overrated. This overestimation would influence taxes paid by an enterprise or rate of subsidies etc. Economic subjects farming in dry areas would therefore be handicapped. The problem was discussed by Němec (2004). Němec described and explained methods of valuation of agricultural land in the Czech Republic. The influence of a rain shadow is discussed in Culek (1995) and Kurpelova *et al.* (1975). They specified methods of defining rain shadows in the Czech Republic and former Czechoslovak Socialist Republic. However, the above mentioned works do not discuss the disadvantage of enterprises farming in the rain shadow compared to enterprises of similar production area outside the rain shadow.

Support of enterprises farming in less favoured areas (LFA) has been discussed in the Common Agricultural Policy (CAP) of the European Union. In the nineties, the McSharry reform and the Agenda 2000 defined three types of disadvantaged areas:

a) mountain areas – defined by a high altitude, slopes at a lower altitude and climatic conditions;
b) other – areas with poor productivity or difficult cultivation resulting in limited productivity as well as areas with a low or dwindling population predominantly dependent on agricultural activity;
c) areas affected by specific handicap – where farming should be kept in order to protect the environment, the coastline, to maintain the countryside; to conserve the environment or to preserve the tourist potential of the areas.

The Czech Republic sees an LFA support as an important part of the Common Agricultural Policy as approximately 50% of Czech agricultural land is defined as the LFA according to the above mentioned criteria. In 2004, the European Union proposed a new scheme of the LFA classification aimed at natural disadvantages only, such as land productivity or climatic conditions.

The proposal had been discussed in 2005 however no agreement was found. In 2006, an evaluation led by independent specialists was performed and member states discussed the situation. Based on the discussion a communication was adopted in 2009 identifying 8 criteria for a classification of the above mentioned areas. A legislative proposal is supposed to be launched after assessing the new classification by simulation models undertaken by member states. Fajmon (2010) discussed the problem. The new methodology could help eliminate the discussion about profitability of enterprises farming in similar areas with different precipitation. This would result into fairer dividing of subsidies.

MATERIALS AND METHODS

BPEJ – agricultural lots are those written in the cadastre of real estates of the Czech Republic such as arable land, hop gardens, vineyards, gardens, orchards, meadows or grazing land (Němec, J., 2001).

Basic price of such lots is evaluated by the income approach (under §11 of the Act of Property Valuation of the Czech Republic) using the BPEJ units. The BPEJ is five-digit code that expresses soil and climate conditions influencing productivity of agricultural land. The BPEJ values can be found at land offices or in the archive of the VÚMP (Research Institute for Soil and Water Conservation) in Praha. The institute also updates the BPEJ units. The BPEJ value is also stated at the extract of cadastre of real estates. Analysing the BPEJ code reveals that the 1st number expresses a climatic region. The most important are the following criteria – the sum of average daytime temperature above 10°C together with average annual temperatures, average precipitation, probability of dry vegetation period occurrence and moisture security.

The paper presumes that the precipitation amount and the probability of dry vegetation period were not included sufficiently into the BPEJ adaptation.

To complete the BPEJ analysis, the interpretation of remaining numbers follows: The 2nd and the 3rd number reveals the main pedological unit; the 4th number reveals the slope and exposure and the 5th number reveals soil stoniness and soil rooting depth.

Analysed database – the analysis was based on the sample of 45 enterprises; 12 of the sample were farming in rain shadow areas. The following records were available: basic financial statements, harvest statement and a questionnaire for farms describing average natural conditions in farming areas including calculation data and indicators of operating activities.

Statistical methods – the analysis itself was divided into two steps. In the first part, enterprises were sorted by the cluster analysis into groups farming in the same area, at a similar altitude, with the same structure in a similar manner, and under comparable financial conditions, (debt ratio, liquidity and activity ratio). The influence of the rain shadow is not considered in this step.

In the second step, an analysis of profitability of the above mentioned groups was planned. Possible inner-cluster variance of the income would be a result of disadvantage due to dry weather. The analysis itself revealed that differences in profit of farms are so significant that two of four clusters consist absolutely or almost absolutely of farms from disadvantaged areas. Remaining two clusters consisted of farms from areas that were not handicapped by dry weather. Due to this, it was possible to compare the inner cluster variance of the return on assets. A null hypothesis that the profitability of farms in dry areas is not significantly different from standard farms (the alternative hypotheses supposed that the profitability of these groups is different).

Testing was performed by comparing mean values based on two-sample Student's test at the level of significance of $\alpha = 0.05$. Before the test itself it was necessary to prove if the variances of both samples are significantly different or not. To prove it, the F-test was performed so that we were able to define test statistic of the Student's test. We followed methods from literature (Lukasová, A., Šarmanová, J., 1985), (Hebák, P., Hustopecký, J., 2005), (Kahounová, J., 1994), (Ajvazan, S., Bežejevová, Z., 1981).

The test statistic for equal variance $\sigma A = \sigma B$ is calculated as follows:

$$t_e = \frac{\left|\overline{x}_A - \overline{x}_B\right|}{\sqrt{s_p^2\left(\frac{1}{n_A - 1} + \frac{1}{n_B - 1}\right)}}. \quad (1)$$

$$s_p{}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \quad (2)$$

The test statistic for unequal variance $\sigma A \neq \sigma B$ is calculated as follows:

$$t_e = \frac{\left|\overline{x}_A - \overline{x}_B\right|}{\sqrt{\frac{s_A{}^2}{n_A} + \frac{s_B{}^2}{n_B}}}. \quad (3)$$

The critical value of was $\upsilon = nA + nB - 2$ degrees of freedom was subtracted from tables of critical values of the Student's t-distribution $tk(\alpha, \upsilon)$. In

case of te ≤ tk H0 is not rejected and estimated mean values are not significantly different. On the other hand, if te > tk we will reject the H0 and accept the alternative H1.

Features and characteristics:

1. Farm size can be measured by the size of farmed land in hectares as well as its economic power given by assets in the balance sheet. Authors wanted to use both features, however because of the mutual dependency of these variables, cluster analysis requires independent input variables, data were transformed to assets/ha ratio. As an advantage, this ratio is able to show the rate of investment of an enterprise.
2. Natural conditions – the BPEJ is based on an altitude, precipitation amount, slope and soil types. All features are correlated to a considerable extent. Fertile black earth can be found in lowland river basins; on the other hand stony soil with low quality is typical for mountain and sub mountain regions of a higher altitude. Slope could be defined similarly. Due to their mutual relation the cluster analysis included only data of average altitude, as the data that were the most easier to find out and to aggregate. The average precipitation was used secondary when analysing cluster features.
3. To characterize economic features standart financial ratios were employed, debt ratio, liquidity and activity ratio. Profitability indicator ration were not considering in the cluster analysis; they were employed secondary in the assessment of clusters. Regarding the activity ratios, both were employed considering different features of fixed and current assets, fixed assets increase the production base, on the other hand current assets freeze means of production. The asset turnover was calculated for both cases (calculated as the ration of revenues and appropriate assets). Regarding debt management ratios, the debt ratio calculated as the ratio of total debt to total assets was employed. Regarding liquidity ratios, the current ratio calculated as the ratio of current assets to current liabilities was used. This ratio measures how many times a company is able to pay back its short-time liabilities by turning out all its current assets into cash.
4. The production structure was assessed by two variables. Economic structure of production was defined as the ratio of plant production revenue to total revenue. Ecological production structure was calculated as the percentage of arable land.

Used methods of clustering – it was necessary to standardize data before the clustering itself. After that a dendrogram was prepared in the Statistica program so that an optimum number of clusters could be decided. It revealed that the best choice would be to divide enterprises into four groups. Similarities of farms were calculated according to a Euclidian distance. To classify an enterprise into a specific group the Ward's method was used. Unlike the method of the nearest or the furthest neighbour, the Soal-Sneath's method of the Gower's method, the criterion for connecting clusters is the increase of internal sum of squares of variances of the cluster average. The increase expressed as the sum of squares in a new cluster is reduced of sums of squares in both disappearing clusters as seen in the first formula:

$$\Delta C_1 = \sum_{i=1}^{g}\sum_{j=1}^{p}(x_{gij} - \overline{x_{gj}})^2 - \sum_{i=1}^{h}\sum_{j=1}^{p}(x_{hij} - \overline{x_{hj}})^2 - \sum_{i=1}^{h'}\sum_{j=1}^{p}(x_{h'ij} - \overline{x_{h'j}})^2 , \quad (4)$$

where *xgij* is the value of the *j*th feature of the *i*th item of the g-cluster that was formed by connecting the *h* and *h'* clusters. The g value is the number of items in the new cluster.

The value of *xgj* with a bar is the mean value of the *j*th feature in the *g*th cluster. The use of this method minimises the trace of a matrix of the inside cluster E-variability and maximizes the trace of the B-variability between clusters matrix while total T-variability remains constant.

The E-variability inside a cluster is revealed in the second formula and the matrix of the B-variability between clusters is shown in the third formula. The total T-variability is expressed as the fourth formula.

$$E = \sum_{h=1}^{k}\sum_{i=1}^{nh}(X_{hi} - \overline{\overline{X}}_h)(X_{hi} - X_h)' \quad (5)$$

$$B = \sum_{h=1}^{k} n_h(\overline{X}_h - \overline{X})(\overline{X}_h - \overline{X})' \quad (6)$$

$$T = \sum_{h=1}^{k}\sum_{i=1}^{nh}(X_{hi} - \overline{X})(X_{hi} - \overline{X})' , \quad (7)$$

where *k* stands for total number of clusters; n_h is the number of items in the *h*th cluster; X_{hi} is the *i*th item in the *h*th cluster; X_h with a bar is a vector mean for the *h*th cluster. X with a bar is a vector mean for the whole sample. If we employ the matrix of the Euclidian distance, as in this paper, it will be possible to use the following formula:

$$D'_{gg} = \frac{1}{n_h + n_{h'} + n_g}((n_h + n_g)D_{hg'} + (n_{h'} + n_{g'})D_{h'g'} - n_{g'} - n_{g'}D_{hh'}) . \quad (8)$$

The Dgg' value is the distance between g and g' cluster and nh and ng' is the number of items in each cluster. Dispersion quality was assessed by the F-criterion value that can be defined as the ratio of a mean square of the variance between clusters and a variance inside a cluster. The sufficient significance level ranges between 1and 5%.

RESULTS AND DISCUSSION

Assessment of the dispersion quality – Tab. I shows the results of the dispersion quality assessment. It measures the mean instance of enterprises inside a cluster (scatter inside cluster column) with a mean distance of cluster centres (scatter of clusters column). High variability between clusters and low variability inside a cluster appeared within the most important features in clustering. The test F-statistic is therefore high and low at a level of significance p up to 5%. These requirements are fulfilled by the following features: ratio of assets per ha, technical equipment of labour, labour productivity, turnover ratio, debt ratio and arable land percentage. The above mentioned features had significant difference between clusters. The dispersion quality of liquidity ratio and the debt management was under average, however still satisfactory (P = 0.025, resp. 0.010).

On the other hand it was not possible to prove the difference between clusters for area of land, altitude, and ratio of plant production and total assets.

The absence of differences between clusters of the size and altitude could have been presumed. The size of farms was given by historical development of the fifties when cooperatives were created regardless economic relations. Regarding altitude, enterprises farming less than 450 meters above sea level were included in the control sample so that the sample was rather homogenous.

Considering the classification of variables into groups the best results were reached within the economic activity. It is possible to say that farms in the same cluster are farming in the same way.

Regarding the production structure, different clusters have the same share of plant production but surprisingly they differ in the way of land use (lower share of arable land and greater share of hop gardens could be presumed within some types of farms). However, as unconvincing variables were found within secondary indicators of the structure and size only and important indicators were at the appropriate level of significance the total dispersion quality is acceptable.

Cluster assessment – cluster 1 features – this cluster included 12 farms only 3 of them were not of a dry area. Calculated as the relative value, more than 75 percent of farms farming in the rain shadow were classified into the cluster.

Farms within this cluster have lower area in the average (1,415 ha; that is statistically insignificant) and above average assets calculated per ha (92 thousand CZK). The above average state of long-term assets (TVP of 1,012 thousand CZK per worker) is considerable. On the other hand these assets are intensively used. The labour productivity amounted to 1.133 thousand CZK with 1.12 of the fund efficiency. Greater use was found for the current assets as well (1.94 turnovers/years). Sources for purchase of long-term assets are supposed to come from foreign capital with more than 57% of the indebtedness (that is approximately 10% above the average of the sector). Lower share of arable land is somehow interesting. It could be presumed that an important part of farm land consists of hop gardens. See Tab. II.

Cluster 2 features – the cluster included 16 farms none of which in the rain shadow. It includes large farms with the average size of 1,912 ha, great share of arable land (86%) and lower share of plant production (51%).

There is 80.3 thousand CZK of assets per each hectare. The fund efficiency is approximately 1.09 with the turnover ratio of current assets of 1.71. Technical equipment of labour (TVP) reached to 902 thousand CZK with the labour productivity of 980 thousand CZK of revenue per worker. Other features are similar to the average farm of the same production area with greater share of plant production.

Cluster 3 features – the cluster included farms outside the rain shadow as well. There are 14 farms in the cluster that differs from the previous cluster by low share of arable land (75.3%) and the lowest revenues from plant production (43.4%).

Regarding the economy, the cluster has the worst activity ratios – with the fund efficiency of 1.03 CZK of revenue per 1 CZK of long-term assets. The number of turnovers is low as well (1.69),

I: *Assessment of the dispersion quality (standard values)*

Indicator	Scatter of clusters	Scatter inside the cluster	F – calculated	Signif. P
Area	0.43	2.81	0.02	0.12
Assets/ha	1.58	2.20	0.09	0.00
Altitude	0.26	2.76	0.01	0.29
TVP	1.91	2.12	0.12	0.00
Productivity	1.88	1.97	0.13	0.00
Turnover ratio	0.99	2.25	0.06	0.00
Debt ratio	0.70	2.52	0.04	0.02
Liquidity	0.60	3.19	0.03	0.00
Arable land (%)	1.62	3.37	0.07	0.01
Return (%)	0.54	3.10	0.02	0.08

Source: financial statements of sample farms; authors' works

II: *Financial ratios in different clusters*

Average values	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Num. of farms	12	16	14	3
Indicator	9	0	0	3
Area (ha)	1415	1912	1716	1205
Assets/ha	92.1	80.3	75.6	107.4
Altitude	395	405	415	376
TVP	1012.0	902.0	805.0	1236.0
Productivity	1133.4	983.2	829.2	1532.6
Turnover ratio	1.94	1.71	1.69	2.01
Debt ratio	52.9	43.1	41.9	57.1
Liquidity	3.20	3.41	3.72	3.16
Arable land (%)	81.1	86.0	75.3	82.2
Return (%)	51.5	48.7	43.4	63.2

Source: Financial statements of sample farms; authors' works

III: *Return on assets in each cluster*

Average values	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Number of farms	12	16	14	3
Number of farms in dry areas	9	0	0	3
Average return on assets	2.82	2.40	2.01	1.59
Variability return on assets	4.08	3.96	3.70	1.76
Proven signif. dif. compared to other clusters	no	no	no	no

Source: Financial statements of sample farms; authors' works

similarly to the labour productivity (829 thousand CZK of revenue/worker). On the other hand, farms within this cluster were in the lowest danger of indebtedness (the share of foreign capital to total liabilities of 46.9% only). These results revealed that the cluster consists of farms that are specialized to more extensive animal production (compared to the second cluster).

Cluster 4 features – the cluster included 3 farms from the area of the rain shadow. It is possible to say that the characteristics of the cluster are the same as the characteristics of the first group with more distinctive features and more extreme financial values. These farms are of the under average size (1,205 ha) and above average capital (107.4 thous. CZK/ha) as well as the technical equipment of labour (1236). The assets are extensively used – the fund efficiency of 1.24; labour productivity of 1,532 thous. CZK/worker. The current assets turnover is overreaching 2. The indebtedness is above 63%. The fourth group therefore consists of farms that differs from the average due to above average investment in the majority of economic indicators and therefore it is very difficult to compare them with the rest of the sample.

Differences in profitability – Tab. III reveals the average return on assets in each cluster and its variability. Clusters of handicapped areas reached the return on assets of 2.82 and 1.59%. Farms of normal areas reached the return on assets ranging from 2.01 to 2.40%. High variability of the indicator is problematic with the standard deviation of more than 3.7% in all primal clusters with the exception of the small cluster 4. The above mentioned revealed that it was not possible to prove any differences in the profitability of farms.

CONCLUSION

The initial presumption supposed that enterprises in the rain shadow are farming in the similar way as enterprises that are not handicapped and differences are expressed as differences in the profitability, lower in dry areas. Based on this presumption, our hypotheses were formulated and an appropriate methodology was created. None of these presumptions was proved. The cluster analysis classified handicapped farms into two separate clusters differing from the average by greater share of long-term assets (measured by the share in total assets and its volume per ha), increased technical equipment of labour and significantly greater labour productivity. High indebtedness proves that the increase of the labour equipment was financed by the foreign capital mainly. The difference in the production structure was proved as well. Revenues from plant production prevailed in all cluster; however farms from dry areas had lower share of arable land, it decreased in favour of hop gardens.

The above mentioned features were evident in both clusters with more extreme values within the fourth cluster. On the other hand, the difference between disadvantaged and standard farms has not been proved due to great variability of the ROA.

SUMMARY

The aim of the paper was to prove how much are enterprises farming in dry areas handicapped compared to farms in similar production area outside a rain shadow. Our working hypotheses suggested that profitability of farms in dry areas is lower and official price based on the BPEJ system is overrated. The analysis was based on the sample of 45 enterprises; 12 of the sample were farming in rain shadow areas. The analysis itself was divided into two steps. In the first part, enterprises were sorted by the cluster analysis into groups farming in the same area, at a similar altitude, with the same structure in a similar manner, and under comparable financial conditions. The influence of the rain shadow is not considered in this step. In the second step, an analysis of profitability of the above mentioned groups was planned. Features and characteristics: 1) Farm size can be measured by the size of farmed land in hectares as well as its economic power given by assets in the balance sheet. Data were transformed to assets/ha ratio, 2) Natural conditions – the BPEJ is based on an altitude, precipitation amount, slope and soil types, 3) To characterize economic features standard financial ratios were employed (debt ratio, liquidity and activity ratio), 4) The production structure was assessed by two variables. Economic structure of production was defined as the ratio of plant production revenue to total revenue. Ecological production structure was calculated as the percentage of arable land. It was necessary to standardize data before the clustering itself. After that a dendrogram was prepared in the Statistica program so that an optimum number of clusters could be decided. It revealed that the best choice would be to divide enterprises into four groups. The cluster analysis classified handicapped farms into two separate clusters differing from the average by greater share of long-term assets, increased technical equipment of labour and significantly greater labour productivity. High indebtedness proves that the increase of the labour equipment was financed by the foreign capital mainly. The difference in the production structure was proved as well. Revenues from plant production prevailed in all cluster; however farms from dry areas had lower share of arable land (it decreased in favour of hop gardens). The above mentioned features were evident in both clusters with more extreme values within the fourth cluster. On the other hand, the difference between disadvantaged and standard farms has not been proved due to great variability of the ROA.

REFERENCES

NĚMEC, J., 2001: *Bonitace a oceňování zemědělské půdy České republiky.* Praha: VÚZE, Praha, 255 s. ISBN 80-85898-90-X.

NĚMEC, J., 2004: *Pozemkové právo a trh půdy v České republice.* Praha: SNTL, Praha, 391 s., ISBN 80-86671-12-7.

CULEK, M., 1995: *Biogeografické členění České republiky.* Praha: ENIGMA, Praha, 347 s., ISBN 80-85368-80-3.

KURPELOVA, M., COUFAL, L. a ČULÍK, J., 1975: *Agroklimatické podmienky ČSSR. Bratislava*: Příroda – vydavatelstvo knih a časopisov HMU, Bratislava, 270 s.

FAJMONT, H., 2010: *Současnost a budoucnost českého zemědělství v EU, 2010, (www.fajmon.eu/doc/).*

LUKASOVÁ, A., ŠARMANOVÁ, J., 1985: *Metody shlukové analýzy.* Praha: SNTL, Praha, 245 s.

HEBÁK, P., HUSTOPECKÝ, J., 2005: *Vícerozměrné statistické metody.* Praha: Informatorium, 135s. ISBN 80-7333-036-9.

KAHOUNOVÁ, J., 1994: *Měření podobnosti struktur.* Praha: Vysoká škola ekonomická v Praze, 135 s. ISBN 80-7079-256-6.

AJVAZAN, S., BEŽEJEVOVÁ, Z.: *Metody vícerozměrné analýzy. Praha*: SNTL, 1981, 252 s.

StatSoft, Inc.: *Manuál k programu Statistica*. 2004.

StatSoft, Inc. (20011). STATISTICA (data analysis software systém), version 10. www.statsoft.com.

Address

Ing. Martin Maršík, Ph.D., Ing. Daniel Kopta, Ph.D., Department of Accounting and Finances, Faculty of Economics, University of South Bohemia in České Budějovice, Studentská 13, 370 05 České Budějovice, Czech Republic, e-mail: mmarsik@ef.jcu.cz, kopta@ef.jcu.cz