Maksym Matskul
Odesa National Economic University, Ukraine

A MULTIVARIATE FACTOR ANALYSIS: APPLICATION FOR ENTERPRISES COMPETITIVENESS ESTIMATION

The aim of the paper is to develop the reliable method of enterprises competitiveness estimation. It this paper it is proposed to consider variety of indexes of enterprise functioning as objective competitiveness indicators. It is suggested to apply modern methods of multivariate statistical analysis namely PCA-PM/R (Principal Component Analysis-Path Modeling/Regression) or Batch Analysis of Principal Component, PLS-PM/R (Projection on Latent Structures-Path Modeling/Regression), and cluster analysis to evaluate the competitiveness of enterprises. It is advisable to use methods of consolidation of uncertain (latent) indexes of enterprise competitiveness. The author has selected 24 food industry enterprises of Odesa region (Ukraine) to define and calculate the competitiveness of food industry enterprises.

Keywords: enterprises competitiveness, enterprises of the food industry, multivariate factor analysis, cluster analysis, principal component analysis, regression (projection) on latent structure.

Introduction. Activity of any enterprise is characterized by a lot of indicators that influence various aspects of its activity. Some of these indicators cannot be measured directly (so-called latent indicators) but they appear as the result of acquisition of specific numerical values for several interrelated indicators-symptoms. One of these latent indicators (indexes) is level of competitiveness of economic entities: regions, enterprises and others. The concept of competitiveness is obviously relative integral characteristic. It provides the comparison of the enterprise with its competitors through the lot of their activities indicators.

Literature review. Therefore, it is necessary to apply multivariate statistical analysis methods and models to quantitative investigation of these integral indicators. Many papers were dedicated to research competitiveness of enterprises, we note only papers with usual methods of multivariate statistical analysis of quantify competitiveness: monograph[1] and the paper[2].

Concept description. PCA-PM/R (PLS-PM/R) is a recent technique that generalizes and combines features from principal component analysis and multiple regression. It is mathematical procedure that aims to represent a set of (possibly correlated) multivariate variables b smaller number of uncorrelated variables known as principal components. It is a multivariate projection method developed to extract systematic variation and relationships among the variables of a data set. This transformation (projection) often simplifies the analysis at hand while also alleviating the worse symptoms of high dimensionality that arises at a large number or variables.

The goal of this method is to predict a matrix Y of dependent variables by means of a matrix X of independent variables (i.e., predictors) and to describe their common structure (see, for example[3]). When Y is a vector and X is full rank, this goal could be accomplished using ordinary multiple regression. When the number of predictors is large compared to the number of observations, X is likely to be singular and the regression approach is no longer feasible (because of multicollinearity). Several approaches have been developed to cope with this problem. One approaches is to eliminate some predictors another one (called principal component) is to perform a PCA of the matrix X and then use the principal component of X as regressors on Y. The orthogonality of the principal component eliminates the collinearity problem. But,

[1] Янковий, О.Г. (2013). *Конкурентоспроможність підприємства: оцінка рівня та напрями підвищення*. Одеса: Атлант.
[2] Кендюхов, А.В., Толкачев, Д.О. (2013). Использование метода главных компонент для оценки конкурентоспособности машиностроительных предприятий. *Маркетинг і менеджмент інновацій, 4,* 219-227.
[3] Lewis-Beck M., Bryman,A., Futing T. (Eds.) (2003). *Encyclopedia of Social Sciences Research Methods*. Thousand Oaks (CA): Sage.

the problem of choosing an optimum subset of predictors remains. A possible strategy is to keep only a few of the first components. But they are chosen to explain X rather than Y, and so, nothing guarantees that the principal components, which explain X, are relevant for Y. By contrast, PCA-PM/R (PLS-PM/R) finds components from X that are also relevant for Y. This method decomposes both X and Y as a product of a common set of orthogonal factors and a set of specific loadings. So, the independent variables are decomposed as $X = TP^T$ with $T^T T = I$ where I is the identity matrix (some variations of the technique do not require T to have unit norms). By analogy with PCA T (with columns t_j) is called the score matrix, and P (with rows p_i) the loading matrix (in PCA-PM/R method the loadings are not orthogonal). Likewise, Y is estimated as $\tilde{Y} = TBC^T$ where B is diagonal matrix with the regression weights (coefficients) as diagonal elements. The columns of T are latent vectors. When their number is equal to rank of X, they perform an extract decomposition of X. Note, that they only estimate Y (in general $\tilde{Y}$ is not equal to Y).

The latent vectors could be chosen in a lot of different ways. In fact in the previous formulation, any set of orthogonal vectors that present the column space of X could be used to play the role of T. In order to specify T, additional conditions are required. The purpose of this method is to find two sets of weights w and in order to create (respectively) a linear combination of the columns of X and Y such that their covariance is maximum. Specifically, the goal is to obtain a first pair of vectors $t = Xw$ and $u = Yc$ with the constraints that, $w^T w = 1, t^T t = 1$ and $t^T u$ is maximal. When the first latent vector is found, it is necessary to subtract it from both X and Y , and the procedure is re-integrated until X becomes a null matrix.

Algorithm NIPALS (Nonlinear Iterative Partial Least Squares) does not calculate all the principal components at once. The first step is to compose two matrices: $E = X$ and $F = Y$. These matrices are then column centered and normalized (i.e., transformed into Z – scores). The sum of squares of these matrices is denoted by SS_X and SS_Y. Before starting the iteration process, the vector u is initialized with random values. Results all operations transformed into Z – scores (centered and normalized).

Step 1 (estimate X weights): $w = E^T u$;

Step 2 (estimate X factor scores): $t = Ew$;

Step 3 (estimate Y weights): $c = F^T t$;

Step 4 (estimate Y scores): $u = Fc$.

If t has not converged, then go to Step 1, if t has converged, then compute the value of b which is used to predict Y from t as $b = t^T u$, and compute the factor loadings for X as $p = E^T t$. Now subtract (i.e., partial out) the effect of t from both E and F as follows $E = E - tp^T$ and $F = F - btc^T$. The vectors t, u, w, c, p are then stored in the corresponding matrices, and the scalar b is stored as a diagonal element of B. The sum of squares of X (respectively Y) explained by the latent vector is computed as $p^T p$ (respectively b^2), and the proportion of explained variance is obtained by dividing the explained sum of squares by the corresponding total sum of squares (i.e., SS_X and SS_Y). If E is a null matrix, then the whole set of latent vectors has been found, otherwise the procedure can be re-iterated from Step 1 on.

Results of research. Consider the actual data of 24 food industry enterprises activity in Odesa region during 2012-2016 by the following indicators (calculated by author according[1] and useful activities of this enterprises): x1 – material consumption (%); x2 – coefficient of the manufacturing cost; x3 – gross income ratio; x4 – capital-labor ratio; x5 – capital productivity ratio; x6 – coefficient of mobility; x7 – the share of productive capacity in assets (%); x8 – coefficient of wear and tear; x9 – coefficient of suitability; x10 – coefficient of renewal; x11 – coefficient of withdraws; x12 – coefficient of growth fixed assets; x13 – critical point in sales; x14 – operation leverage; x15 – specific wage; x16 – glut of finished product; x17 – share of long-term assets in total assets (%); x18 – share of fixed assets in total assets (%); x19 – share of fixed assets in long-term assets (%); x20 –coefficient of labor productivity; x21 – coefficient of internal

[1] Головне управління статистики в Одеській області. <http:www.od.ukrstat.gov.ua/>.

level of competitiveness; x22 – coefficient of competitiveness margin; x23 – coefficient of financial stability margin; x24 – turnover ratio of current assets; x25 – inventory turnover ratio; x26 – asset turnover ratio; x27 – turnover ratio of receivables; x28 – days sales outstanding; x29 – coefficient of accounting payable turnover ratio; x30 – payable turnover in day; y31 – return of sales (%); y32 – profitability of production (%); y33 – operation return on working capital (%); y34 – return on equity (%); y35 – return on fixed assets (%); y36 – return on assets (%); y37 – maneuverability equity ratio; y38 – financial stability coefficient; y39 – debt to equity of equity ratio; x40 – ratio of receivables and payables; x41 – absolute liquidity ratio; x42 – quick ratio; x43 – current ratio; x44 – equity ratio.

Thus, we use the following data matrix to evaluate of the latent factor – competitiveness of enterprises C_1 – C_24 during 5 years:

A) The independent variables are indicators x_j $X_{120\times 38} = \left(x_{ij}\right); i = \overline{1,120}; j = \overline{1,30;37,44};$

B) The dependent variables are indicators y_j $Y_{120\times 6} = \left(y_{ij}\right); i = \overline{1,120}; j = \overline{31,36}.$

Considering that competitiveness is primarily determined by the resulting indicators of enterprise activity we carry out a classification (ranking) for enterprises-objects by aggregating the values of dependent variables-indicators y_j, $j = \overline{31,36}$ for all 24 enterprises behind 2012-2016 years. Hereinafter we use the program Statistica[1] for computing. Note that the indicators are closely related:

Table 1

The Correlation Matrix between dependent variables

Variable	*Correlation (Worksheet.sta) Marked correlations p-level p<,05000 N=120 (Missing data were casewise deleted)*						
	y31	*y32*	*y33*	*y34*	*y35*	*y36*	
y31	*1,00*	***0,93***	***0,73***	*0,03*	***0,50***	***0,70***	
y32	***0,93***	*1,00*	***0,82***	*0,02*	***0,52***	***0,69***	
y33	***0,73***	***0,82***	*1,00*	*-0,05*	***0,64***	***0,79***	
y34	*0,03*	*0,02*	*-0,05*	*1,00*	*0,11*	*0,04*	
y35	***0,50***	***0,52***	***0,64***	*0,11*	*1,00*	***0,90***	
y36	***0,70***	***0,69***	***0,79***	*0,04*	***0,90***	*1,00*	

Source: calculated by the author

Then we use the Mahalanobis metric (or so called 1-Pirson r) as a measure of similarity of objects (or distance d_{ps} between p-th and s-th enterprises-objects): $d_{ps} = \left(z_p - z_s\right) r^{-1} \left(z_p - z_s\right)^T$, where z_p, z_s are row vectors p and s in the space of standardized indications $\left(z_k = \dfrac{y_k - \overline{y_j}}{\sigma_j},\ j = \overline{31,36}\right)$, $\overline{y_j}$ are averages, σ_j are standard deviations of all objects from aggregate for j-th indications, and r^{-1} is inverse matrix to matrix of paired coefficients between the indications.

Applying a hierarchical – agglomerative procedure by criterion of Ward (Ward's Method) we'll perform the distribution of our aggregate of 24 enterprises on groups as dynamic clusters. The analysis by the Ward's method is based on study increment of the intragroup variances of indicators for all possible variants union of the clusters (we get the clusters of about the same size).

[1] Statistica. Системный подход к анализу данных. <http://www.statsoft.ru/>.

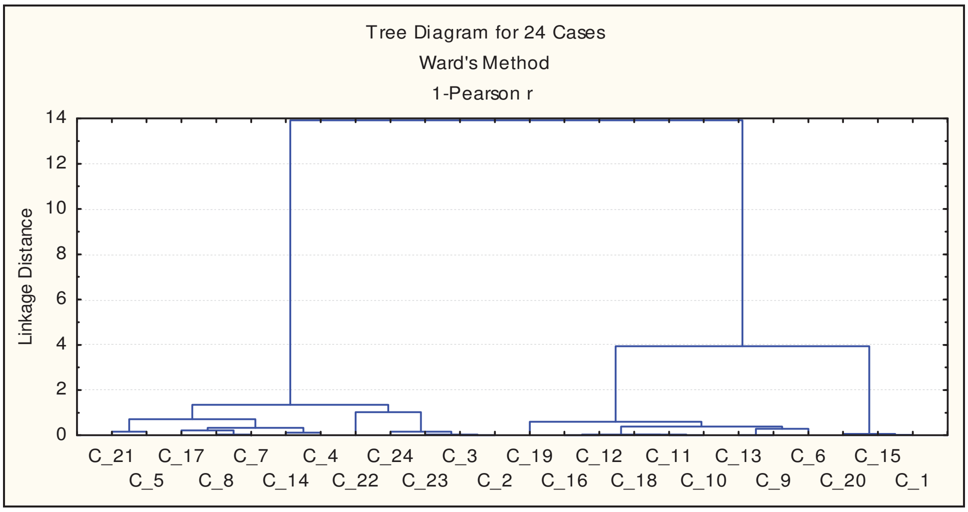

Figure1. Dendrogram of distribution of aggregate of enterprises into clusters based on the results of 2016

Source: prepared by the author

Then, using the method of K-means, we obtain four clusters from our set of 24 enterprises:

Table 2

Distribution the aggregate of enterprises into clusters based on the results of 2016

№ Cluster – categorical variable group { }	*The set of enterprises objects that form a cluster*
group {1}	*Dolinka PAO (C_11), Odessa Sparkling Wine Company PAO (C_9), Yuzhnyi PAO (C_12), Artsyzkyy Zavod prodtovariv PAO (C_16), Plant" Illichivsk" PAO "(C_18), Odessa Baby Food Cannery PAO (C_19), Leather Goods Factory PAO(C_20)*
group {2}	*Odesskiy Karovay PAT (C_1), Dniestrovskiy PAT (C_10), Vinogradar PAT (C_13), Odesavinprom PAT (C_15), Belgorod-Dniestrovskiy Kombinat Hliboproductiv PAT (C_6)*
group {3}	*Lyubashivskiy Elevator PAT (C_2), Odessa Cannery VO (C_17), Reni Meat PAT(C_21), Odessa Meat PAT(C_22), Combinat ATZT(C_23) Razdel'nyanskiy Elevator PAT (C_5), Zaplazke Cereal PAT (C_7)*
group {4}	*Balta Cereal PAT (C_3), Odessa Cognac Factory PAT (C_8), Harchovyk PAT (C_14), Yantar PAT (C_24), Aliyahske Cereal PAT (C_4)*

Source: calculated by author

Next, we consider all activity indicators for all 24 enterprises in 2012-2016 (120 cases). Also, we treat the categorical indicator "group" that takes four values (group {1}, group {2}, group {3}, and group {4}) and defines that enterprise belongs to a particular Cluster). There are fairly close the correlation's relationships among all variables (indicators). In such special cases, the correlation matrix is indeterminate (its eigenvalues are closely approximated by their values) and is essentially difficult to extract principal components (factors) by ordinary methods. Therefore, we apply algorithm NIPALS of principal components method (PCA-PM/R, PLS-PM/R) of program Statistica to determine the structure of variables. Before the analysis, we assume that four groups (factors, principal components) of indicators can be selected from the space of all indicators:

F_1 is level of competitiveness (previously, as resultative indicators y_j, $j = \overline{31,36}$);

F_2 is production potential (previously, as indicators x_j, $j = \overline{1,23}$);

F_3 is business activity (previously, as indicators $x_j,\ j=\overline{24,30}$);

F_4 is financial condition (previously, as indicators $x_j,\ j=\overline{37,44}$).

The following result of Batch method of principal components shows that the factors identified above are really the principal components, since they explain over 90% of the total variance.

Table 3

Characteristic numbers (eigenvalues) and percentage contribution of Principal component at the total variation of all indicators

Components	*PCA – Eigenvalue (Table) Number of components =4 PCA – Sum of variance 7,0000*			
	Eigenvalues	*% Total variance*	*Cumul. Eigenvalue*	*Cumul. %*
1	*2,678720*	*39,76355*	*2,678720*	*39,76355*
2	*1,499577*	*26,66196*	*4,178296*	*66,42552*
3	*1,303247*	*14,48052*	*5,481544*	*80,90604*
4	*0,873215*	*9,70239*	*6,354759*	*90,60843*

Source: calculated by the author

One more result of the analysis shows the importance of indicators:

Table 4

The importance of indicators in regard to the allocated principal components (fragment)

Variable	*Importance of variable Number of components = 4*			
	No variable	*Score of category*	*Power*	*Importance*
group{2}	*45*	*2*	*0,995723*	*1*
group{4}	*45*	*4*	*0,989380*	*2*
group{1}	*45*	*1*	*0,981636*	*3*
group{3}	*45*	*3*	*0,978979*	*4*
y36	*44*		*0,737372*	*5*
y33	*41*		*0,600278*	*6*
x42	*36*		*0,590040*	*7*
y31	*39*		*0,586155*	*8*
x43	*37*		*0,571269*	*9*
y35	*43*		*0,560632*	*10*
y32	*40*		*0,555401*	*11*
x21	*21*		*0,547783*	*13*
y34	*42*		*0,114465*	*35*
x11	*11*		*0,024741*	*47*
x19	*19*		*0,013352*	*48*

Source: calculated by the author

Analysis of the results leads to the conclusion that the partition of aggregate of enterprises by four groups - clusters that we performed above confirms the high significance for as of categorical variable "group {}" so and for the resultative indicators $y_j,\ j=\overline{31,36}$, except of the indicator y_{34}. The final factor solution farther we make use the method "Equamax normalized" of factor rotation. The results of this method are presented in the following table:

Table 5

The factor loadings (fragment)

Variable	*Factor loadings (Equamax normalized)* *Extraction: Principal components* *(Marked loading are >0,500000)*			
	Faktor 1	*Faktor 2*	*Faktor 3*	*Faktor 4*
x1	*-0,050176*	*-0,550176*	*-0,076872*	*0,129953*
x2	*-0,068090*	*-0,606809*	*0,007235*	*0,399105*
x21	*0,858389*	*0,118978*	*0,086623*	*0,045770*
x22	*0,858389*	*0,118978*	*0,086623*	*0,045770*
y31	*0,784943*	*0,043039*	*0,170138*	*0,338086*
y32	*0,800242*	*-0,000651*	*0,144431*	*0,564303*
y33	*0,641017*	*0,035728*	*0,142188*	*0,632655*
y35	*0,656406*	*0,001401*	*0,061183*	*0,223350*
y36	*0,702674*	*0,002066*	*0,147798*	*0,242046*
Total variance	*7,107102*	*4,773123*	*4,384467*	*4,087629*
Total ratio	*0,384252*	*0,208480*	*0,199647*	*0,192901*

Source: calculated by the author

Thereby we obtaine:

– the resultative indicators $y_j,\ j=\overline{31,36}$ (except of the indicator y_{34}), and also the indicators – symptoms (predictors) x_{21}, x_{22} pretty much carry the loading on latent resultative factor F_1 (level of competitiveness), as expected. At that, we have the approximate equality:

$$F_1 \approx 0,78y_{31} + 0,80y_{32} + 0,64y_{33} + 0,65y_{35} + 0,70y_{36} + 0,85x_{21} + 0,85x_{22};$$

– indicators - symptoms $x_j,\ j=\overline{1,6},8,13,14,15,17,18,20$ carry a significant loading on the latent factor F_2 (production potential). Moreover, the signs of factor loadings are adjusted with the meanings of the corresponding economic indicators:

$$F_2 \approx -0,55x_1 - 0,60x_2 + 0,65x_3 + 0,89x_4 + 0,71x_5 + 0,85x_6 - 0,66x_8 - 0,51x_{13} + \\ +0,76x_{14} - 0,53x_{15} + 0,87x_{17} + 0,64x_{18} + 0,78x_{20}.$$

It should be noted that the indicators $x_j,\ j=4,5,6,14,17,20$ have the greatest positive influence on the factor F_2 from the symptoms $x_j,\ j=1,2,8,13,15$;

– significant stimulating influence of indicators is present: turnover ratio of current assets (x_{24}), assets (x_{26}) and accounts payable (x_{29}), and de-stimulating influence of days sales outstanding (x_{28}) and payable turnover in day (x_{30}) for the third factor F_3 (business activity):

$$F_3 \approx 0,67x_{24} - 0,59x_{28} + 0,61x_{29} - 0,70x_{30};$$

– fourth factor F_4 (financial condition) is appreciably loaded influent by indicators $x_{38,40,42,43}$ among which we note de-stimulating influence of the ratio of payables and receivables (x_{40}) and of the financial stability coefficient (x_{38}) as most stimulating indicator:

$$F_4 \approx 0,74x_{38} - 0,56x_{40} + 0,57x_{42} + 0,52x_{43}.$$

Independent latent factors are listed in following order, in accord with the power of influence on the level of competitiveness (resulting latent factor F_1): F_4, F_3, F_2.

The analyzed aggregate of enterprises in terms of competitiveness can be divided into four clusters: "leaders" as group {2}; "above average" as group {4}; "below average" as group {1}; and "outsiders" as group {3}.

Conclusions and prospects for further research. We developed and mathematically proved a new method for evaluating the competitiveness of enterprises, based on method PCA-PM/R, PLS-PM/R, which does not contain the subjective estimations and it takes into account as many different indicators of activity of enterprises as possible. The developed method was applied for predicting the competitiveness of enterprises and other economic entities.

References:

1. Yankovyj, O.G. (2013). *Konkurentospromozhnist pidpryyemstva: ocinka rivnya ta napryamy pidvyshhennya* [Competitiveness of the enterprise : assessment of the level and directions increase]. Odesa: Atlant. [in Ukrainian].
2. Kendjuhov, A.V., Tolkachev, D.O. (2013). Ispol'zovanie metoda glavnyh komponent dlja ocenki konkurentosposobnosti mashinostroitel'nyh predprijatij [Application of the method of the main components for the evaluating the competitiveness of engineering companies]. *Marketyng i menedzhment innovacij* [Marketing and management innovations], *no. 4,* 219-227. [in Russian].
3. Lewis-Beck, M., Bryman, A., Futing, T. (Eds.) (2003). *Encyclopedia of Social Sciences Research Methods.* Thousand Oaks (CA): Sage.
4. Golovne upravlinnya statystyky v Odeskij oblasti [The main department of statistics in Odesa region]. <http:www.od.ukrstat.gov.ua/>. [in Ukrainian].
5. Statistica. Sistemnyj podhod k analizu dannyh[Statistics. A systematic approach to data analysis]. <http://www.statsoft.ru/>. [in Russian].