

# Methodology to determine the business size of SMEs through financial ratios

Ma. de Lourdes Elena García Vargas<sup>1\*</sup>, Gerardo Reyes Ruiz<sup>2</sup>

<sup>1</sup>Universidad Autónoma del Estado de Hidalgo's, Escuela Superior Tepeji, Ave del Maestro S/N, Col Noxtongo II, C.P.42855, Tepeji del Río, Hidalgo.

<sup>2</sup>Instituto Politécnico Nacional (CIECAS), Lauro Aguirre 120 esquina Sor Juana Inés de la Cruz, Col. Agricultura, Del. Miguel Hidalgo, C.P. 11360 Cd. de México, Email: greyesruiz@hotmail.com

## Abstract

The present empirical work shows a methodology to determine the business size of Medium and Small Enterprises (SMEs), through the financial ratio analysis. The objective is to expose a business size measurement criterion useful to collect taxes, to evaluate the conditions for representing as a job generator company, to measure the innovative potential, and to analyze whether the company is suitable for technology or equipment finance. Multivariate analysis approach is utilized in this work in order to achieve the above objectives. This approach primarily generates synthetic variables that will further be incorporated into a secondary analysis that explains the business size of a SMEs through a discriminant analysis. In the discriminant analysis, the input synthetic dimensions generated through the Principal Component Analysis are considered as a variable which is based on a battery of financial ratios. The proposed methodology uses first-hand information to explain the use of the categorical variable size in a conglomerate of companies through its financial performance. The assessment of the classification obtained was more than acceptable. A categorization was obtained for some Spanish companies (micro, small and medium enterprises) because it requires a better classification and clearly define their rights and obligations). However, the use of other techniques, such as the artificial neural network analysis, is recommended to confirm the results obtained.

**Keywords:** Size enterprises, Financial ratio analysis, Discriminant analysis, SMEs.

## 1. INTRODUCTION

For considering a company as a Catalan Small Medium Enterprises (SMEs), it shouldn't meet the following requirements: total number of employees of the company should be less than 250, to have an annual turnover less than 50 million euros and to have a fixed asset of no more than 43 million euros. Most important characteristics of Catalan SMEs are, no supply of marketable securities, owners are generally risk-prone entrepreneurs, no investment portfolios (Ferrer and Tresierra, 2009) and there is a very limited access to the capital market and long-term bank financing. The Catalan SMEs also rely more on trade credits without generating tax savings (their profit is earned by the interest payment generated by indebtedness) and show a lower life expectancy and a greater probability of bankruptcy. Due to the lack of specialization and business training provided by their owners, some of the SMEs are led to failure owing to their flow of limited assets. Furthermore, when the business owner and the manager position are taken as a one role, it generates distrust when applying for bank loans increasing the financial risk of the debt. According to Ferrer and Tresierra (2009), the capital structure of the Catalan SMEs focuses on tax deductions and its application in the company, growth opportunities, size, profitability, age of business, and their fixed assets.

For this reason, it is important to study models that present companies' classification according to their size and their characteristics to classify them into their corresponding category. In this way, they can have access

to better opportunities and comply with the tax authorities. The company classification also allows to evaluate whether the company generates employment, if its innovation potential can be assessed, and the type of loan in technology, machinery or equipment the company. In a sample of 12,658 SMEs, a battery of 22 financial ratios is defined to generate synthetic variables that will later be incorporated into a second analysis that will build a measure of business size to improve the tax collection. As it has already been mentioned, this methodology is able to diagnose whether the organization provide jobs; if its innovative potential can be analyzed, and the type of loan in technology, machinery or equipment the company applies for. The study shows that different techniques of data analysis serve to classify a conglomerate of companies into small and medium-sized companies.

In the first stage of the methodology, the Principal Component Analysis (PCA) was applied to the battery of 22 financial ratios to make the information more manageable and facilitate the interpretation of data. Subsequently, the principal components were obtained by diagonalizing the correlation matrix and these were the input of a discriminant analysis (DA). The results were obtained by means of a PCA and later from a DA. In the PCA, a matrix of communalities was obtained by extracting six components or latent variables showing the PCA as more enriching than a factor analysis since it presented a greater number of latent variables. The number of components or synthetic variables was based on the use of the correlation matrix as the number of components had a value greater than one. Tests such as the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy, the Bartlett's test for Sphericity, a Varimax rotation with Kaiser normalization, the Jarque-Bera test, the Box's M test, and the Wilks' lambda test were applied to these six components.

The purpose of this empirical work is to show a classification of companies based on their size through the financial ratio analysis. The objective<sup>1</sup> focuses on defining a criterion for the measurement of the company size (small or large). Therefore, the classical approach which refers mainly to the number of employees will be omitted (Galindo, 2005), and instead the approach of the so-called multivariate techniques will be applied. This will first serve to generate synthetic variables that will later be incorporated to a second analysis to build a business measure. This approach was chosen as the definition of quantitative limits in terms of variables does not respond, in general, to the use of any method that discriminates the reason for a company to belong to a category of micro, small or medium size.

## 2. MATERIAL AND METHODS

The database<sup>2</sup> is composed of 12,658 SMEs and for each of them, 22 financial ratios were defined; these are shown in Table 1. Subsequently, in the second stage of the analysis, there were used the variables regarded as *size* and *number of employees*.

<sup>1</sup>The objective of this research is not quantitative; rather, this study shows that different techniques of data analysis serve to classify a conglomerate of companies into micro, small and medium size.

<sup>2</sup>The total number of companies is based on different sectors, and as the purpose of this research is to measure the business size, the difference presented is not considered by sectors or productive activities, nor by categories of independent professionals, small retail businesses, bars, restaurants, etc. This, undoubtedly, is a topic for future work and research.

No.	Concepto	Etiqueta	Ratio
1	Activo circulante	act_circ	Act circ explot / Actiu net
2	Activo neto	act_net_	Act net explot / Actiu net
3	Rentabilidad de las ventas	altres_d	Altres desp explot / Ing explot
4	Liquidez	cash_flo	Cash Flow / creditors a c/t
5	Operación	consums_	Consums explot / Ing explot
6	Créditos	creditor	Creditors c/t / Passiu remunerat
7	Gastos de personal por empleado	desp_per	Desp personal / n treballadors
8	Fondos propios	fons_pro	Fons propis / passiu remunerat
9	Financiación ajena del inmovilizado	immob_ne	Immob net explot / Actiu net
10	Margen	marge	Marge
11	Endeudamiento	rec_alie	Rec aliens ll/t / passiu remunerat
12	Recursos generales	recursos	Recursos Generats
13	Rentabilidad económica	rendib1	Rendibilitat econòmica
14	Rentabilidad financiera	rendibil	Rendibilitat financera
15	Cobertura de Gastos financieros	rene_d	Rene / Desp finances
16	Coberetura de efectivo	rene_i	Rene / Ing explot
17	Rotación	rotació	Rotació
18	Solvencia	tresorer	Tresoreria/Ingressos explot
19	Productividad del activo	vab_ac	VAB / Act net explot
20	Productividad del inmovilizado	vab_im	VAB / Immob net explot
21	Productividad del ingreso	vab_in	VAB / Ing explot
22	VAB (valor añadido bruto) por ocupante	vab_nom	VAB/ Nombre treb

Table 1. Financial ratios. Source: Pymes data.

## 2.1 Methodology.

For the first stage of this study, the Principal Components Analysis (PCA) was applied to the battery of financial ratios (chosen from Table 1). The main objective of this multivariate method was to convert a very complex statistical information problem (many variables of quantitative type calibrated in different units of measurement) in another almost equivalent but more manageable problem (few new variables) without losing relevant information (E. Pla, 1986). The use of the PCA is very effective because on the one hand, it allows to represent high-dimensional space observations in a smaller dimensional space. (The PCA is the first step to identify possible latent, or unobserved variables that generate data); on the other hand, it also allows the transformation of the correlated variables, into new, uncorrelated variables facilitating the interpretation of the data (Fisher, 1936).

By using the multivariate analysis, it is possible to obtain linear combinations of the original variables that are orthogonal to each other. This can generate a subset of new variables that are independent from one another but are highly correlated with the original variables (Principal Components); when these components are ordered in decreasing order according to their variance, it is possible to comprehend the

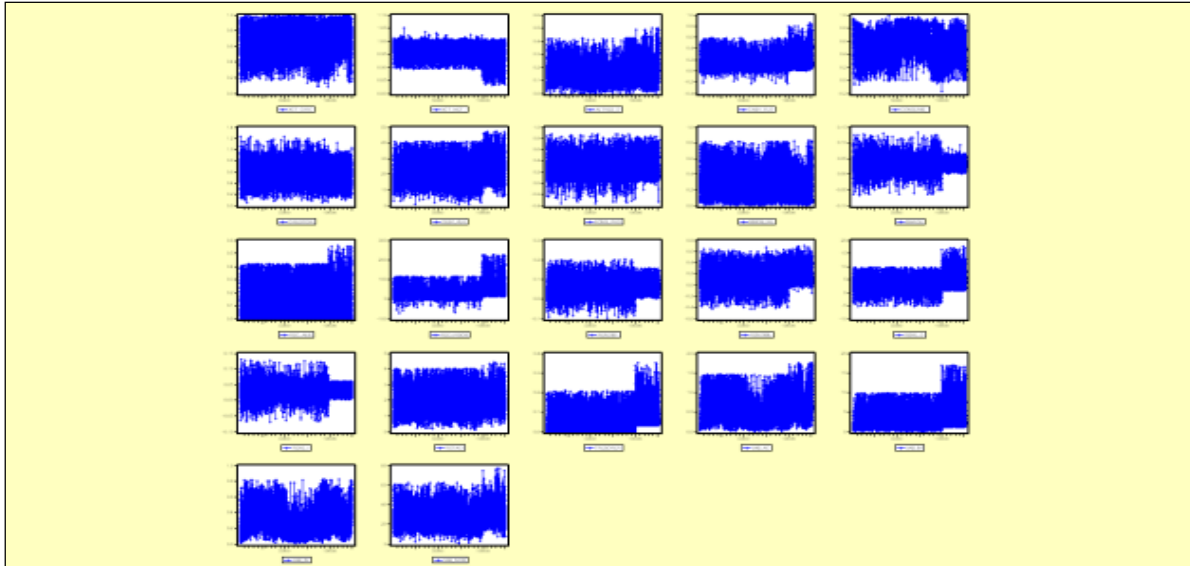
percentage of data variability for each component (Montanero, 2008). In consequence, every observation of the sample obtains a score in each of the main selected components. This process allows the observations to be ordered based on multivariate data.

In this study, the principal components were obtained by diagonalizing the correlation matrix since the ratios have different units of measurement and each variable participates with the same degree of importance. Once the latent variables were generated, these were the input in a Discriminant Analysis (DA) that could explain with a synthetic measure the business size of the Catalan SMEs. Therefore, this analysis pretends to find a pattern to separate a group of companies in relation to their size. The DA analysis was used as it is a statistical technique that classifies objects into different groups according to observations of some characteristics of these objects (Shih-Wei and Shih-Chieh, 2009). Basically, this analysis consists of obtaining a linear combination of some independent variables (this is guaranteed by the PCA previously carried out), that maximize the differences between the classification groups and minimize the differences between each of these groups (maximum value of the quotient defined by variability between groups/intra group variability). The discriminant function is a weighted linear combination of the values that acquire the observed characteristics of the object. Those weightings represent the relative importance and impact of the characteristics that are part of the discriminant function. Thus, the object is classified into a single group or class based on the score obtained by the discriminant function. However, the DA is subject to many hypotheses, for instance: each group must follow a normal multivariate distribution, the covariance matrices of each group should be identical and, the covariance matrices, the a priori probabilities as well as the errors of classification must be known. If these hypotheses are not met, the results obtained may be inaccurate. Hence, it is worth pointing out that the DA was used since this method selects the variables that have greater discriminating power and have previously been generated by an PCA.

## 2.2. Financial Ratio Analysis.

Working with financial ratios will always be complicated; owing to their construction, it is very common to encounter problems of linear dependence between them. For the ratios that were proposed in Table 1, the analysis was the following: Due to the ratio construction, four groups were appreciated. However, the first problem shown by these ratios is that they do not follow a behavior of a univariate normal distribution, so at this point, any parametric statistical inference is discarded - such as the Analysis of Variance (ANOVA); the Pearson correlation coefficient, which assumes normality in its calculation is also discarded since any result would lead to erroneous conclusions. If normality is assumed in the ratios, it would be obtained an ANOVA that firstly would imply, the lack of equality of variances between the groups of the ratio *tamany* or size; and secondly, that the ratios susceptible to be omitted at a certain moment would be: *rotació*, *rec\_alie*, *immob\_ne* and *act\_circ*. The reason is that ratios illustrate significant differences between the different subgroups in which the *tamany* or size factor has been separated.

The next problem was the high degree of multicollinearity presented in these ratios; it was verified when performing a PCA with all the ratios and therefore, there was shown a synthetic correlation matrix that was not positively defined. This implies the discrimination of some ratios to obtain more robust results. Graphically, it is difficult to make a value judgment because in Graph 1, direct *a priori* relationships are not perceived.



Graph 1. Multiple representation of the 22 financial ratios. Source: Elaboration based on Eviews.

The problem of linear dependence can be verified by doing a multiple linear regression without the constant term and using any of the 22 ratios as an endogenous variable and the rest of them as exogenous variables. The result will always be the same, it means that this regression cannot be carried out because the message "Near Singular Matrix" would appear. In other words, some of the ratios are an exact linear combination of other financial ratios which is in fact, something logical. This last linear relationship was seen omitting a variable; for example, the ratio called *act\_circ*; then, when explaining the ratio *altres\_d* using the rest of the ratios, the result of Table 2 could be obtained.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CASH_FLO	-9.63E-14	1.10E-14	-8.744889	0.0000
CONSUMS_	-1.000000	1.14E-14	-8.75E+13	0.0000
CREDITOR_	1.000000	3.26E-14	3.07E+13	0.0000
DESP_PER	-5.32E-15	3.21E-16	-16.58892	0.0000
FONS_PRO	1.000000	3.29E-14	3.04E+13	0.0000
IMMOB_NE	4.27E-15	7.61E-15	0.561940	0.5742
MARGE	-1.52E-11	2.35E-13	-64.64626	0.0000
REC_ALIE	1.000000	3.28E-14	3.05E+13	0.0000
RECURSOS	-2.97E-16	2.77E-17	-10.69459	0.0000
RENDIB1	1.61E-12	6.23E-14	25.88477	0.0000
RENDIBIL	-8.71E-14	8.24E-15	-10.57439	0.0000
RENE_D	6.05E-15	4.43E-16	13.64236	0.0000
RENE_I	1.21E-11	2.17E-13	55.69137	0.0000
ROTACI	4.21E-14	2.53E-15	16.61078	0.0000
TRESORER	4.56E-14	1.72E-14	2.650537	0.0080
VAB_AC	-2.53E-13	7.97E-15	-31.74186	0.0000
VAB_IM	-9.69E-16	5.08E-16	-1.904872	0.0568
VAB_IN	-1.000000	1.98E-14	-5.06E+13	0.0000
VAB_NOM	3.46E-15	2.43E-16	14.25310	0.0000
ACT_NET_	-2.78E-12	3.21E-14	-86.60744	0.0000
R-squared	1.000000	Mean dependent var	0.143915	
Adjusted R-squared	1.000000	S.D. dependent var	0.080202	
S.E. of regression	9.40E-14	Akaike info criterion	-57.15134	
Sum squared resid	1.12E-22	Schwarz criterion	-57.13958	
Log likelihood	361730.9	F-statistic	4.85E+26	
Durbin-Watson stat	1.963626	Prob(F-statistic)	0.000000	

Table 2. Linear dependence between the ratios. Source: Elaboration based on Eviews.

Based on this analysis, a very high linear relationship was observed among almost all the ratios involved; it is because a coefficient of determination equal to one was obtained and almost all the  $t$  statistics were greater than two; therefore, they turned out to be significant. In addition, because Durbin-Watson's  $d$  test is not valid in the regression model without the constant term, value judgments about the first-order correlation could not be made. As it was not possible to estimate the normality assumption for an ANOVA, it was proceeded to use non-parametric tests, and although they are less powerful, they allowed to complete a discrimination between the ratios and, most importantly, they did not make any assumption about normality among the data. Nevertheless, when verifying the independence of these ratios by means of runs tests, it was observed that they did not justify the hypothesis of randomness nor, the hypothesis of independence. This last result implied that it was not a random sampling of companies and that the non-parametric tests can lead to erroneous conclusions. However, as an example, if the independence of these ratios was assumed when performing a Kruskal-Wallis test for  $k$  independent samples, it would be established that the ratios called *rotació*, *rec\_alie*, *immob\_ne*, *act\_circ* and *tresorer* show significant differences between the arithmetic means for the groups of the ratio *tamany* or size; this could suggest excluding such ratios from the study. On the other hand, when performing a median test, there could be obtained the following ratios with significant differences between the groups of the ratio *tamany* or size: *rotació*, *rec\_alie*, *immob\_ne*, *act\_circ*, *altres\_d*, *tresorer* and *rene\_d*. This last analysis could well lead to a poor selection of ratios.

Therefore, for the discrimination of all ratios, it was used a distance analysis between the ratios of each group with respect to the ratio *tamany* (entrepreneurial size). For this analysis, it was used the Chebyshev distance which is standardized by its regular deviation and absolute values for each ratio; thus, it was observed a better appreciation of distance. For the nature and construction of financial ratios four groups were appreciate and his analysis was the next:

1. In block A: the ratio *rec\_alie* showed a greater distance with respect to ratio *tamany* or size, so the ratio *rec\_alie* can be omitted for the corresponding PCA.
2. In block B: the ratio *act\_net* presented the greatest distance with respect to ratio *tamany* or size; so, it can also be omitted for the PCA.
3. In block C: there was an unexpected result; the ratio *rene\_i* also showed the gratest distance with respect to ratio *tamany* or size. However, this ratio did not provide any information to the PCA as it was perceived a correlation matrix that was not positively defined. On the other hand, when the ratio *altres\_d* (show the shortest distance with respect to ratio *tamany* or size), was omitted, more solid results in the PCA could be seen. Therefore, based on these results, it was chosen to omit the ratio *altres\_d*.
4. In block D: This set of financial ratios integrated with a base with the little likeness that they showed with respect to ratio named *tamany* (entrepreneurial size). As well this pudding stone of financial ratios did not show a good correlation with the Sperman coefficient rank.

With this analysis, the battery of selected financial ratios that will subsequently be the input for the PCA, is presented in Table 3.



	No.	Concepto	Etiqueta	Ratio
A	1	Fondos propios	fons_pro	Fons propis / passiu remunerat
	2	Créditos	creditor	Creditors c/t / Passiu remunerat
B	3	Financiación ajena del inmovilizado	immob_ne	Immob net explot / Actiu net
	4	act_circ	act_circ	Act circ explot / Actiu net
C	5	Operación	consums_	Consums explot / Ing explot
	6	Productividad del ingreso	vab_in	VAB / Ing explot
	7	Coberetura de efectivo	rene_i	Rene / Ing explot
	8	Solvencia	tresorer	Tresoreria/Ingressos explot
D	9	Rentabilidad financiera	rendibil	Rendibilitat financera
	10	Rentabilidad económica	rendib1	Rendibilitat econòmica
	11	Margen	marge	Marge
	12	Rotación	rotació	Rotació
	13	Productividad del inmovilizado	vab_im	VAB / Immob net explot
	14	Recursos generales	recursos	Recursos Generats
	15	Liquidez	cash_flo	Cash Flow / creditors a c/t
	16	Cobertura de Gastod financieros	rene_d	Rene / Desp finances
	17	Productividad del activo	vab_ac	VAB / Act net explot
	18	VAB (valor añadido bruto) por ocupante	vab_nom	VAB/ Nombre treb
	19	Gastos de personal por empleado	desp_per	Desp personal / n treballadors

Table 3. Grouping of the selected financial ratios. Source: Author's elaboration.

### 3. RESULTS

This section of the study shows the main results obtained by means of the PCA and subsequently by the DA. The PCA is a technique used to extract factors and form independent linear combinations of the observed variables. The first component has the maximum variance. The successive components progressively explain minor proportions of the variance and are not correlated with each other. The PCA is used to acquire the initial factorial solution; it can also be used when a correlation matrix is singular. In the PCA that was carried out for the ratios of Table 3, there was observed the following matrix of communalities with the extraction of 6 components (see Table 4) or latent variables:

Comunalidades

	Inicial	Extracción
Rendibilitat financiera	1,000	,722
<b>Rendibilitat</b>	1,000	,878
Marge	1,000	,919
Rotació	1,000	,934
Fons propis / passiu remunerat	1,000	,844
Creditors c/it / Passiu remunerat	1,000	,840
Immob net explot / Actiu net	1,000	,934
Act circ explot / Actiu net	1,000	,933
Consums explot / Ing explot	1,000	,922
VAB / Ing explot	1,000	,956
VAB / Immob net explot	1,000	,799
VAB / Act net explot	1,000	,943
Desp personal / n treballadors	1,000	,923
VAB/ Nombre treb	1,000	,937
Tresoreria/Ingressos explot	1,000	,269
Recursos Generals	1,000	,382
Cash Flow / creditors a c/it	1,000	,735
Rene / Ing explot	1,000	,911
Rene / Desp financers	1,000	,523

Método de extracción: Análisis de Componentes principales.

Table 4. Matrix of communalities for the selected financial ratios. Source: Base in SPSS.

It is worth pointing out that the PCA was more enriching than a Factorial Analysis (FA) in the sense that it showed a greater number of latent variables. This can be explained since the AF generated only four latent factors that explained the 64.3% of the total variance of the selected ratios. Furthermore, if the main objective had been to reduce the dimension as much as possible, then the FA would have given better results; the reason is that with the use of four latent factors, the same test statistics as in a PCA would have been obtained (see Table 5).

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,698
Prueba de esfericidad de Bartlett:	Chi-cuadrado aproximado	260998,4
	gl	171
	Sig.	,000

Table 5. KMO and Bartlett's test for an FA. Source: Own elaboration based on SPSS.

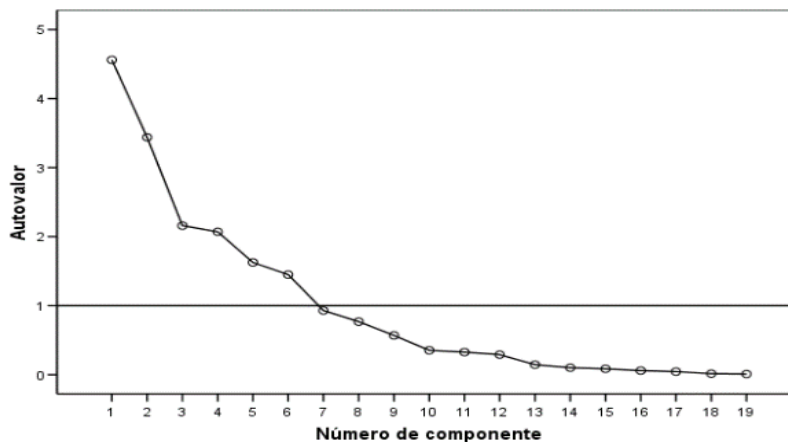
It is of vital importance to highlight that the principal axis factoring (Factor Analysis) is a method used for the extraction of factors that starts from the original correlation matrix with the squares of the multiple correlation coefficients inserted in the main diagonal as initial estimates of the communalities. The resulting factor saturations are used to estimate the communalities one more time and replace the previous estimates in the diagonal matrix. The iterations continue until the change in the communalities (from one iteration to the next one), satisfies the convergence criterion for the extraction. However, the work done in this study was the following; by means of a greater number of possible synthetic variables, the objective was to find the highest level of association of these synthetic variables with regards to the size of the companies



selected. Therefore, it was decided to obtain the results from the PCA; they had a mean of 0 and a variance equal to the square of the multiple correlation between the estimated scores and the real values. These scores can also be correlated when the factors obtained are orthogonal.

It is interesting to mention the following: as the correlation matrix is calculated using the Pearson correlation coefficient, which is a measure of linear association, two variables can be perfectly related; but, if this relationship is not linear, the coefficient of Pearson's correlation will not be an adequate statistic to measure its association. Furthermore, for the calculation of this coefficient, it should be assumed normality in the ratios. For this reason, and to assess the level of agreement with the Spearman correlation coefficient, (for its calculation does not presuppose normality in the ratios), it was calculated the difference that exists between the Pearson correlation matrix and the correlation matrix of Spearman; Subsequently, it was calculated the determinant of such a difference to know the level of "error" developed in the measurement.

Nonetheless, it is also worth saying that if the Spearman correlation coefficient is used, it will be not feasible to make value judgments about cause-effect signs between the ratios since this statistic is non-parametric. Thus, the determinant calculated for the difference matrix was  $1.02763 \times 10^{-30}$ , that basically says that the distance between these matrices is very close to zero; it means that the use of the Pearson correlation coefficient does not have greater impact on the PCA. Therefore, for the ratios in Table 3, it was carried out a PCA through which synthetic variables were generated. Subsequently, the scores of the components were used as input in a DA to obtain a synthetic measure for the business size. In other words, the PCA aims to reduce the size of the problem by creating highly correlated variables that, in turn, could be a factor to differentiate one company from another in terms of its size. Then, according to the results of Graph 2, there was a selection of 6 components or synthetic variables. The number of components was based on the correlation matrix used and the number of components that had a value greater than 1.



Graph 2. Sedimentation line. Source: Own elaboration with SPSS results.

In the PCA, the statistics of Table 6 were also obtained to contrast their representativeness.

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,698
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	260998,4
	gl	171
	Sig.	,000

Table 6. KMO and Bartlett's test. Source: Own elaboration with SPSS results.

The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy, contrasts if the partial correlations between the variables are small; hence, for this study it was obtained a coefficient very close to 70%, which implied a good level of association between the selected ratios. Bartlett's test of sphericity contrasts whether the correlation matrix is an identity matrix that indicate that the model is inadequate. However, for this study, it is rejected the hypothesis that states the determinant of the correlation matrix equal to one; this is because a significance of less than 5% was obtained. These last results, together, imply that the model is adequate.

With these six components, it can be explained the 80.5% of the total variance for these ratios (see Table 7) and although it is not relevant to label these synthetic variables, it was used a PCA through a Varimax rotation with Kaiser normalization (see Table 8).

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	4,559	23,996	23,996	4,559	23,996	23,996	3,695	19,449	19,449
2	3,439	18,098	42,094	3,439	18,098	42,094	2,837	14,933	34,382
3	2,159	11,362	53,456	2,159	11,362	53,456	2,684	14,125	48,507
4	2,070	10,894	64,350	2,070	10,894	64,350	2,445	12,867	61,374
5	1,626	8,556	72,906	1,626	8,556	72,906	2,048	10,781	72,155
6	1,450	7,630	80,536	1,450	7,630	80,536	1,592	8,381	80,536
7	,929	4,891	85,427						
8	,769	4,047	89,474						
9	,569	2,995	92,469						
10	,354	1,862	94,330						
11	,327	1,721	96,052						
12	,290	1,529	97,580						
13	,145	,766	98,346						
14	,101	,530	98,876						
15	,085	,448	99,324						
16	,059	,311	99,635						
17	,044	,234	99,869						
18	,015	,080	99,949						
19	,010	,051	100,000						

Método de extracción: Análisis de Componentes principales.

Table 7. Total variance explained. Source: Authors own elaboration.

The results for the component matrix vs. the rotated component matrix were the following:

	a							a					
	Componente							Componente					
	1	2	3	4	5	6		1	2	3	4	5	6
Rendibilitat financera							Rendibilitat financera	,779					
Rendibilitat econòmica	,660						Rendibilitat econòmica	,883					
Marge	,739						Marge	,909					
Rotació						,789	Rotació						,949
Fons propis / passiu remunerat							Fons propis / passiu remunerat				,912		
Creditors c/t / Passiu remunerat							Creditors c/t / Passiu remunerat					-,869	
Immob net explot / Actiu net		-,679					Immob net explot / Actiu net				-,918		
Act circ explot / Actiu net		,675					Act circ explot / Actiu net				,915		
Consums explot / Ing explot							Consums explot / Ing explot		-,916				
VAB / Ing explot							VAB / Ing explot		,945				
VAB / Immob net explot							VAB / Immob net explot				,824		
VAB / Act net explot					,692		VAB / Act net explot			,823			
Desp personal / n treballadors			,615		,600		Desp personal / n treballadors						,934
VAB/ Nombre treb							VAB/ Nombre treb						,949
Tresoreria/Ingressos explot							Tresoreria/Ingressos explot						
Recursos Generats							Recursos Generats						
Cash Flow / creditors a c/t,754							Cash Flow / creditors a c/t				,622		
Rene / Ing explot						,738	Rene / Ing explot	,908					
Rene / Desp finances							Rene / Desp finances	,635					

Método de extracción: Análisis de componentes principales.  
a.6 componentes extraídos

Método de extracción: Análisis de componentes principales.  
Método de rotación: Normalización Varimax con Kaiser.  
a. La rotación ha convergido en 6 iteraciones.

Table 8. Component matrix vs. the rotated component matrix. Note: correlations less than 0.6 in absolute value have been suppressed. Source: Authors own calculation.

With these last results, the dimension of the problem has been reduced and now 6 independent variables will be used instead of 19. For the 6 latent components or variables, the factor scores were calculated, so these could be standardized to guarantee their homogeneity in the Discriminant Analysis. Thus, the variables that were used as an input for the DA are defined in Table 9.

No.	Etiqueta	Variable
1	F01	Factor 01
2	F02	Factor 02
3	F03	Factor 03
4	F04	Factor 04
5	F05	Factor 05
6	F06	Factor 06
7	numero_e	Número empleados 2002
8	tamany	Tamaño empresarial

Table 9. Synthetic variables for the DA. Source: Own elaboration with SPSS results.

When the Jarque-Bera test was performed, problems of univariate normality were observed in the six synthetic variables obtained through the PCA and the variable that refers to the number of employees, (see Table 10).

	FACTOR1	FACTOR02	FACTOR03	FACTOR04	FACTOR05	FACTOR06	EMP_2002
Mean	-2.31E-12	-1.81E-12	9.08E-13	1.03E-12	-4.74E-13	8.53E-13	10.70398
Median	-0.031284	-0.104791	0.056657	-0.120248	-0.113794	-0.050867	8.000000
Maximum	4.022478	3.421737	3.510298	4.032051	3.996283	4.347965	245.0000
Minimum	-4.535290	-2.336568	-2.917688	-3.200732	-2.946192	-4.499070	1.000000
Std. Dev.	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	10.67817
Skewness	-0.142793	0.416481	-0.117654	0.484072	0.485847	0.213561	4.709507
Kurtosis	4.260840	2.562898	2.690508	3.084496	3.070916	2.981031	57.15522
Jarque-Bera	881.4587	466.7019	79.72179	498.1149	500.6331	96.40785	1593593.
Probability	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Observations	12658	12658	12658	12658	12658	12658	12658

Table 10. Results from the Jarque-Bera test. Source: Own elaboration based on Eviews.

On one side, the variable that presented a greater linear relationship with regards to the variable *tamany* or size was the variable of the number of employees; its correlation coefficient amounted to just over 70% (see Table 11).

Correlation Matrix								
	FACTOR1	FACTOR02	FACTOR03	FACTOR04	FACTOR05	FACTOR06	EMP_2002	TAMANY
FACTOR1	1.000000	-6.27E-12	-2.14E-12	1.87E-12	-2.44E-12	2.19E-12	-0.038403	-0.024366
FACTOR02	-6.27E-12	1.000000	2.98E-12	-5.31E-12	1.88E-12	8.07E-13	0.146135	0.116763
FACTOR03	-2.14E-12	2.98E-12	1.000000	4.56E-13	3.87E-12	-3.52E-12	0.028377	0.019847
FACTOR04	1.87E-12	-5.31E-12	4.56E-13	1.000000	-3.22E-12	-2.28E-12	0.058828	0.070179
FACTOR05	-2.44E-12	1.88E-12	3.87E-12	-3.22E-12	1.000000	-6.08E-12	0.045620	0.085520
FACTOR06	2.19E-12	8.07E-13	-3.52E-12	-2.28E-12	-6.08E-12	1.000000	0.049876	0.049326
EMP_2002	-0.038403	0.146135	0.028377	0.058828	0.045620	0.049876	1.000000	0.713159
TAMANY	-0.024366	0.116763	0.019847	0.070179	0.085520	0.049326	0.713159	1.000000

Table 11. Correlation Matrix. Source: Own elaboration based on Eviews.

This last result could be confirmed with the structure matrix (see Table 12). In this table the variable referring to the number of employees in 2002 presented more information for the discriminant function 1 and this also retained 99.4% of the total explained variance.

	Función	
	1	2
Número empleados 2002	,998 *	-,049
REGR factor score 5 for analysis 1	,051	,687 *
REGR factor score 4 for analysis 1	,040	,626 *
REGR factor score 6 for analysis 1	,034	,272 *
REGR factor score 2 for analysis 1	,097	,154 *
REGR factor score 1 for analysis 1	-,026	,136 *
REGR factor score 3 for analysis 1	,018	-,021 *

Correlaciones intra-grupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas  
Variables ordenadas por el tamaño de la correlación con la función.

\*. Mayor correlación absoluta entre cada variable y cualquier función discriminante.

Table 12

In the DA, the variable called *tamany* or size was the *grouping* variable for which three subgroups were defined; 1 (micro), 2 (small) and 3 (medium). Table 13 shows the integration of the grouping variable.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	7642	60,4	60,4	60,4
	2	4901	38,7	38,7	99,1
	3	115	,9	,9	100,0
	Total	12658	100,0	100,0	

Table 13

When performing the multivariate DA technique to the variables in Table 9, it was observed that, for the test of equality of means, all the variables were found to be significant. However, with the Box's M test, which contrasts the equality of the covariance matrices, it was found that these variables showed a clear heterogeneity between their variances. This last result could be verified by the contrast shown in Table 14.

M de Box		12025,123
F	Aprox.	212,305
	gl1	56
	gl2	259671,731
	Sig.	,000

Table 14. Results for the Box's M test. Source: Elaboration with SPSS results

This last result implies that there are substantive differences between the subgroups analyzed, with regards to their variance to explain the business size; it is a logical analysis because the dispersion of the data of the subgroup that is integrated by the micro companies does not have to be homogeneous with respect to the subgroup that is composed of medium-sized companies. Another important result obtained with this analysis was the Wilks' Lambda for discriminating canonical functions; when a level of significance of less than 5% is obtained for each of them, it means that the two discriminating functions (when defining three size groups for SMEs, two linear functions can be obtained that discriminate these groups) are statistically significant. This implies that discrimination is possible and there is a reason for it. Thus, as seen in Table 15, the discriminant function 1 absorbs the highest percentage of variability, whereas the discriminant function 2 does not show much association between the discriminant scores and the groups involved. It means that the discrimination is not so clear to distinguish a company from subgroup 2 (small) to one from subgroup 3 (median).

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	1,596 <sup>a</sup>	99,4	99,4	,784
2	,009 <sup>a</sup>	,6	100,0	,095

a. Se han empleado las 2 primeras funciones discriminantes canónicas en el análisis.

Table 15. Results for the eigenvalues. Source: Own elaboration with SPSS results.

This last result can be verified with Table 16 which refers to the discriminant scores for the centroids of the subgroups:

Tamany	Función	
	1	2
1	-,839	-,044
2	1,103	,087
3	8,775	-,747

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

Table 16. Functions for centroids. Source: Own elaboration with SPSS results.

The above table shows that function 1 results in a better discrimination than function 2. Nevertheless, the two discriminant functions turned out to be significant, and this can be verified by means of the Wilks' lambda test (see Table 17). When a statistical significance (p value) of less than 5% is obtained for each discriminant function, it implies that each individual function, is statistically significant.

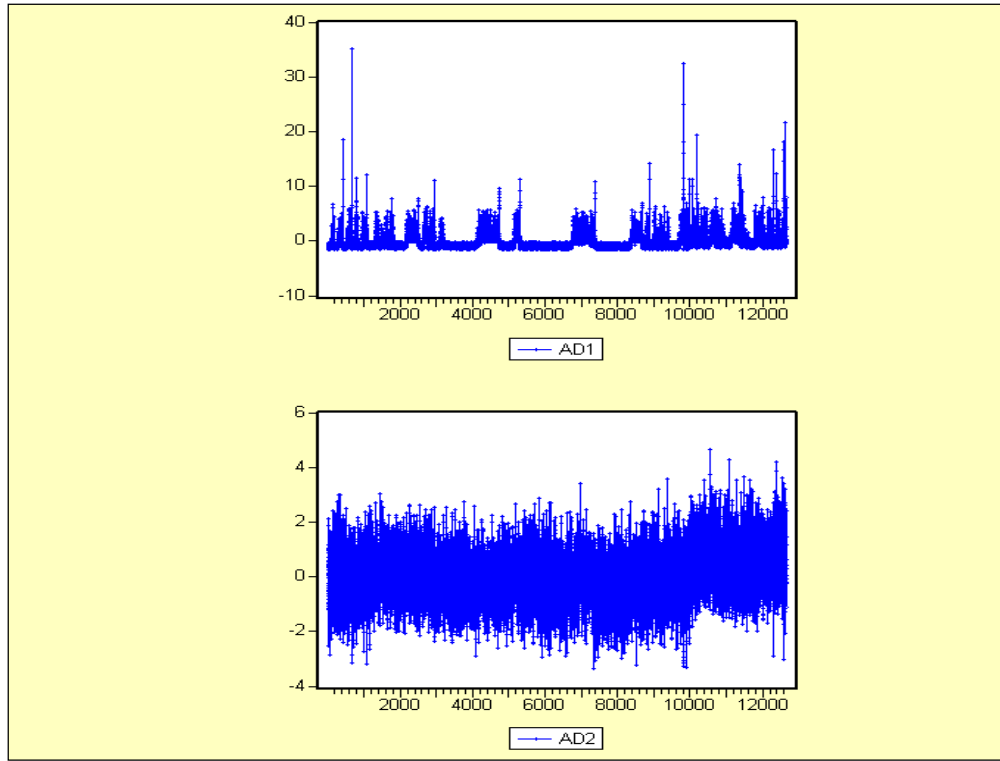
Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 2	,382	12183,980	14	,000
2	,991	115,488	6	,000

Table 17. Results for Wilks' Lambda. Source: Own elaboration with SPSS results. Graph shows the scores calculated by a DA for the two discriminant functions

The Graph 3 show the points AD for discriminant functions

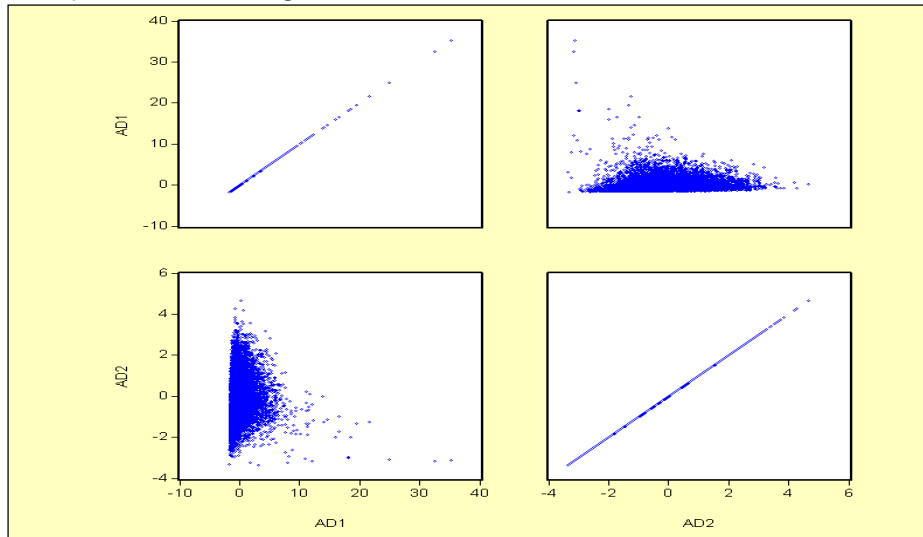


Graph 3. Scores using a DA for the discriminant functions. Source: Own elaboration based on Eviews



Graph 4 Show the scatter plot calculated through a DA for the two discriminant functions.

Graph 4. Score scatter plot calculated through a DA for the two discriminant functions. Source: Elaboration based on Eviews.



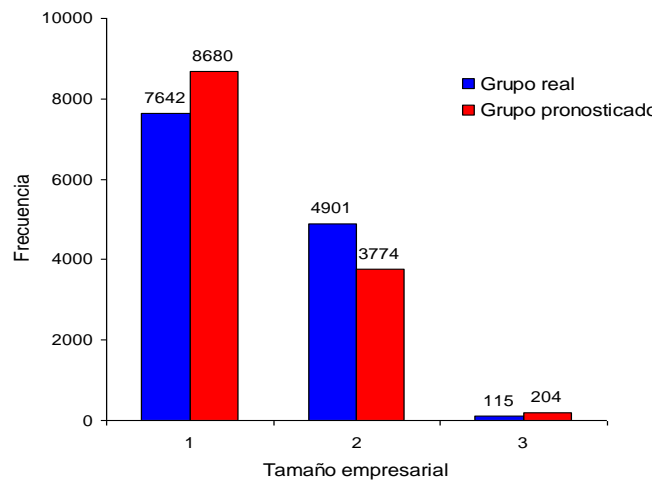
It is important to note that by using the multivariate analysis DA with a cross variation, it was obtained the 89.9% of correct classification. Similarly, Table 20 shows the frequency statistics for the variable *tamany* or size and its corresponding predicted group. Finally, Graph 5 illustrates the classification assessment.

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos 1	7642	60.4	60.4	60.4
2	4901	38.7	38.7	99.1
3	115	.9	.9	100.0
Total	12658	100.0	100.0	

Table 18.

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos 1	8680	68.6	68.6	68.6
2	3774	29.8	29.8	98.4
3	204	1.6	1.6	100.0
Total	12658	100.0	100.0	

Table 19. Real and predicted business size. Source: Own elaboration with SPSS results.



Graph 5. Classification. Source: Own elaboration.

#### 4. CONCLUSIONS

The results obtained in this empirical study demonstrate that the classification of a company is possible when it is done by means of its accounting-financial information. The immediate implication of these results could result in a better tax take because some companies classified in other classic criteria such as *micro* could be in the small business conglomerate. In turn, if this methodology, detects when a company has a "larger"

size, it could be considered as a potential employment generating company. On the other hand, and far from what has been obtained in this research, the importance of the results obtained in recent decades for the SMEs has been justified by economic growth and the innovation potential of a specific country or region.

The ratios selected for the PCA allow an acceptable DA, even though these ratios present statistical problems from their construction. To verify the results obtained with the selected financial ratios in PCA, it would be advisable to carry out independent samples (such as the systematic random sampling, as the range of observations is already known) to verify the consistency of these results. There is statistical proof that allow the estimation of two linear functions to discriminate the three groups of companies. Furthermore, the discrimination of the business size for 2002 based on the synthetic variables generated by an PCA and the variable that refers to the number of employees in 2002, shows a level of reliability of 89.9% through the cross-validation technique. With this discrimination it was possible to differentiate a micro company from a small company. However, between a small company and a medium-sized company, the discrimination was not very clear.

Regarding the variable used in the DA referring to the number of employees, it provides the most information to perform a discrimination. Therefore, it is recommended to reinforce the results with an account analysis. The synthetic measure that explains the business size could be biased due to the nature of the ratios selected; it suggests a transformation of ratios or a collection of a more enriched database. Thus, an alternative to this would be choosing an appropriate selection of other more significant financial ratios for the same sector to validate the results obtained. The search for a concept or business measure from numerical data using traditional statistical procedures did not limit the purpose of the study. However, due to the nature of the data, it would be helpful to use other techniques for the analysis, such as a Probit model (Logit) with three categorizations, or a neural network analysis. The latter would strengthen the ratio relationship and would facilitate the construction of other synthetic ratios that could provide more information about the business size.

## REFERENCES

- [1] E. Pla, L. (1986), *Análisis Multivariado: Método de Componentes Principales*, Departamento de Asuntos Científicos, Secretaría General de la Organización de los Estados Americanos, monografía 27, págs. 94.
- [2] Ferrer, M.A, y Tresierra, A. (2009). *LasPyMES y las Teorías modernas sobre Estructura de Capital*, Compedium 22, pags.19.
- [3] Fisher, R. A. (1936). *The use of multiple measurements in taxonomic problems*, Annals of Eugenics, 7: 179-188.
- [4] Galindo L., A. (2005), *El tamaño empresarial como factor de diversidad*. Universidad de Málaga. Grupo SEJ-366.
- [5] Montanero F., J. (2008), *Análisis Multivariante*, Universidad de Extremadura Servicio de Publicaciones, 59.
- [6] Shih-Wei, L. y Shih-Chieh, C. (2009), *A particle swarm optimization approach for enhancing classification accuracy rate of linear discriminant analysis*, Applied Soft Computing, 9: 1008-1015.
- [7] Gerardo Reyes Ruiz thanks for the support the National Polytechnic Institute (CIECAS) through its program of national postdoctoral stays CONACYT

\*Corresponding author: [maria\\_garcia10476@uaeh.edu.mx](mailto:maria_garcia10476@uaeh.edu.mx)