

The Parameters A Statistical Methodology to Determine the Profit of Business Activities

Ursula Herr, Sogei Spa, Rome, Italy
Giovambattista Giangrieco, Sogei Spa, Rome, Italy
Fabio Zaccaria, Sogei Spa, Rome, Italy
Danilo Ballanti, Sogei Spa, Rome, Italy

Abstract

Sogei, the Agency of the Ministry of Finance for data computing, has created, within its fiscal matters research program, a system of parameters to determine the profits, the compensations, and the volume of business made by companies, and individual in business.

The parameters (also known as "Indexes of Profit") are based on a methodology that determines the amount of profit a person in business would be likely to make under normal market conditions.

The estimation of the parameters is calculated on the basis of the income tax returns filed with the Tax Registry (Internal Revenue), by selecting, for each economic activity, a sample of tax payers who are "economically correct" based on economic conditions, and economic-accounting indexes.

To construct the parameters, homogeneous groups of tax payers in the same type of business are identified.

After determining the amount of profit or compensation for each group, their capacity to make profits or earn compensation is linked to other economic variables.

To apply afterwards the methodology to the universe of tax payers, we use a discriminating analysis that calculates the probability that a person in business could belong to a homogeneous business group.

Moreover, we determine a "greater profit or compensation" for each tax payer on the basis of the estimated profit, and his/her own confidence interval.

The statistical analyses are carried out using mainframe computers, and applying the module SAS/IML, and the procedures PRINCOMP, FASTCLUSTER, STEPWISE, REG and DISCRIM of SAS/STAT.

How to construct the parameters

An analysis strategy that combines several statistical methods has been designed in order to identify the possible differences between various economic activities and inside the same economic activity, as

well as local and territorial differences and different management choices.

In such a complicated context, we have considered that varied processing of information guarantees results which are nearer to reality, more than would be the case with an analysis based on one computing process.

The analysis strategy used to construct the parameters was, therefore, as follows:

- identification of an "economically correct" tax payer sample, which is used as the data base;
- identification of the variables on which to base the analysis (economic variables checked on the income-tax returns);
- identification of the homogeneous groups of tax payers inside each economic activity;
- identification of a "profit function" for each homogeneous group of tax payers to establish a connection between profit and the other variables;
- identification of functions which will permit the application of the structures obtained in the sample, to the universe of tax payers.

Before selecting the sample of income-tax returns on which to construct the methodology, it was necessary to eliminate all the irregular cases from the universe of tax payers, as these would not have permitted the determination of the real economic links between the accounting variables which were used.

The methodology for the construction of the parameters was not applied to those activities which the low number of tax payers, or the particular productive process, did not permit the determination of a standard of behaviour.

How to select the sample

It is necessary to base the analyses on samples of tax payers which are "economically correct", on the base of economic conditions and economic-accounting indexes. Typical economic indexes can be found for each type of business.

For example, for business in trade, a typical index is "mark-up", that is the connection between selling price and cost of goods.

Those tax payers who have, in their tax returns, indicated that index between the 20th and the 90th percentile are defined as “economically correct”.

Other indexes are:

- inventory turnover;
- employees productivity, that is profits related to employees;
- yield on working capital, that is profits related to working capital;
- yield on production surface, that is profits related to the size of production area.

Certainly, each index works when it has an economic sense.

Therefore, for each sector of economic activity and for every index, the percentile distribution is calculated using the SAS PROC UNIVARIATE .

The tax returns of those tax payers who come between the 20th and the 90th percentile of each significant index are then included in the sample.

In this way it is possible to ensure that the sample includes only the tax returns that are “economically correct”.

How to identify the homogeneous groups

The selected sample of tax payers is used for the construction of homogeneous groups within the same type of business.

To identify the homogeneous groups of tax payers two statistical methodologies are used: the principal component analysis and the cluster analysis (using the PRINCOMP and the FASTCLUSTER procedures of the SAS/STAT module).

The combined use of the two methods is better than a direct application of the cluster analysis.

The higher the number of variables, the more complex is the procedure of clustering.

Before using the cluster analysis, the variables must be standardized: variables with large variance have more effect on the resulting cluster than those with small variance.

The principal component analysis is a variable reduction procedure similar to factor analysis.

It is applied to reduce the number of observed variables to a smaller number of artificial variables (named “principal components”), which account for the largest part of the variance of the observed variables.

In the case of companies the observed variables are:

- cost of sales;
- consumption of utilities;
- assets;
- depreciation;
- personnel costs;
- earning per share.

In the case of individuals in business the observed variables are:

- professional fees;

- consumption of utilities;
- assets;
- depreciation;
- personnel costs;
- other expenses.

Technically, a principal component can be defined as a linear combination of optimally-weighted observed variables.

Generally, only the first few components will account for meaningful amounts of variance and the later components will tend to account for only trivial variance.

For cluster analysis we have considered those components for which the cumulative percentage of variance accounts at least 80% of the total variance.

The cluster analysis, using the principal components as input variables, guarantees that the distances observed within the same cluster are less than the distances observed in different clusters.

Each cluster includes at least 30 tax payers.

Some similar activities, which include a low number of tax payers, have been aggregated.

How to identify the "Function of profit"

For each homogeneous group of tax payers, "function of profit" is estimated applying the *Multiple Regression*.

This method interpolates data using a statistical-mathematic model that describes how to link independent variables to the dependent variable (profit).

For companies, independent variables used in the regressive model are:

- cost of goods or cost of the production of services;
- purchase of services;
- assets;
- fees (personnel costs).

For individuals in business, independent variables used are:

- consumptions;
- assets;
- other costs;
- fees (personnel costs).

The variables used in the regression model have been selected with the STEPWISE procedure of the SAS/STAT module, forward selection method, establishing the following criteria:

- regression response through the origin;
- use of non linear functional forms including the square term of the cost of goods for all the activities where this choice improved the estimation quality;
- in case of grouped activities, to identify the possible different productive peculiarities, dummy variables were applied to the independent variables.

After the selection of the variables, REG procedure of the SAS/STAT module was used to perform a regression analysis.

To avoid that “outliers” can influence the estimates of the parameters and make the resulting analysis less useful we have considered those tax payers with a residual value (R of Student) included between -3 and +3.

Moreover, in presence of heteroschedastic disturbances, the "profit function" was calculated with the generalized least squares method, using again the REG procedure on the observations proportional inversely weighted to the variances of residuals.

How to apply the methodology to all tax payers.

In order to apply the methodology to all tax payers , the discriminant analysis was used.

This methodology enables each tax payer to be put in one of the homogeneous group in which each economic activity is classified.

The discriminant function is created using the SAS/STAT PROC DISCRIM on the basis of the variables used in the principal components.

The standard way of the discriminant analysis, on which every subject is classified in one group with the higher probability, has not been used, because we have considered the probability that the subject could be classified in any group.

For each tax payer we estimate his own profit by applying the related coefficients of the regression function in every homogeneous group.

In every homogeneous group, the profit of each tax payer is estimated by applying the related coefficients of the regression function.

The predicted profit is calculated using the average of profits obtained in each homogeneous group weighted with the related probabilities.

When the single tax payer ,in his tax return, declares a profit higher than the estimated one, his return is considered “correct”; otherwise a “ greater profit” is assigned.

To this greater profit a “reducing factor” is applied, so that the error probability in the predicted profit may be regarded within 95% of the confidence interval.

This last step establishes the greater profit of the tax payer.

This “reducing factor “ has been introduced to consider that:

- the analyses are based only on data contained in the tax returns;
- the effect of different economic situations and territorial influences.

The procedure

The whole data processing procedure is realized in mainframe environment using SAS programs.

Considering the data base dimension (about 4 million income-tax returns) and that the parameters have been processed, both for companies and individuals in business, on the basis of the Istat (Italian National Statistics Institute) classification in 920 types of economic activity, it has been essential to use the SAS MACRO LANGUAGE for the automatic realization of the entire procedure.

The DISCRIM and REG procedures of the SAS/STAT module output has been later processed using the IML procedure to prepare files and matrixes to use in the methodology application.

Therefore a software program for Personal Computers was prepared and distributed to the Financial Administration offices, the Associations and the tax payers, to permit to verify the congruence of the proceeds declared from companies and individuals in business for the tax years 1994, 1995, 1996 and 1997.

This procedure will be operating untill a new one, which is already being studied , will be ready.

The new procedure, which will be named “Sector’s Studies”, will include, not only the economic variables but, also the territorial and the structural variables.

Mrs. U. Herr
Sogei Spa - V. Carucci 99
00143 ROME (Italy)
phone n. 039-6-50252425
fax n. 039-6-50957224