POSSIBILITIES OF QUANTITATIVE STUDY OF SOME QUALITATIVE PROCESSES IN ECONOMY

Cristina COCULESCU[1]
Aura-Marilena DIN[2]

Abstract:

Qualitative variables have a big importance in economy, being associated to a big number of economic processes and phenomena. In economic-mathematic modeling process, the main problem which needs to be solved is to express numerically the level of such variables, especially if their importance is big, or if qualitative effects worth to be considered. Measuring of qualitative variables give the chance of input them in quantitative kind models that would help mostly to reduce the degree of indetermination and therefore, to a fuller explanation of the evolution of economic phenomena.

In this work, we'll refer to the opportunities to measure qualitative variables encountered till nowadays in dedicated literature, giving examples for several cases through processes and phenomena frequently encountered in economic practice.

Key words*:* **q**ualitative variables, quantitative modeling, economic processes

JEL classification*:* C1, C5.

1. INTRODUCTION

Qualitative or attributive variables refer to characteristics, qualities and categories etc. whose dimension are expressed using attributes or names about categories (classes, cases). In economy, these variables have a big importance, being associated to a big number of economic processes and phenomena. Therefore, the demand is conditioned not only by the size of income, price etc. but also by the degree of buyer's satisfaction or his/her thoughts etc. for a given

[1]Associate Professor at the Romanian-American University, Bucharest, Romania, e-mail: cristina_coculescu@yahoo.com
[2] Associate Professor at the Romanian-American University, Bucharest, Romania

product. In what concerns the offer, it depends on numerical variables (budget, number of wage earners etc.) and also on the degree of production management, stimulation politics, level of wage earners' training (qualification), their creative capacity, social climate etc.

In economic-mathematic modeling process, the main problem that is set on is to numerically express the level of such variables, especially if their contribution is important or if qualitative-kind effects are interesting. After measuring problem is solved, it appears the chance of putting qualitative variables in quantitative-kind models, which would mainly contribute to the diminution of indetermination degree and, thus, to a more complete explanation of the development of economic phenomena.

2. QUALITATIVE PROCESSES "REPRESENTED" BY NUMERICAL VARIABLES

Measuring level of variables is a criterion of classification of them, of a big importance for statistic study. We can consider four measuring levels (nominal, ordinal, of interval or of rapport), depending on three criteria:

- the chance to arrange variable values;
- the equality of intervals between the values of the variable (the existence of an measure unit);
- the existence of a "reference point" of the variable or, with other words, of an "absolute zero" [1].

Nominal measuring level supposes classification of attributes, characteristics, phenomena etc. in categories that must be different, mutual exclusive and exhaustive. This kind of variables (i.e. the scales used for measuring) shows only that there is a qualitative difference between studied categories but not the range of this difference. At the limit, we can consider these variables as typologies [9]. Some examples of variables measured at nominal level are: human's working status (peasant, wage earner, small businessman, unemployed etc.), religion (Orthodox, Romano-catholic, Greco-Catholic etc.), nationality (Romanian, Hungarian, Gipsy etc.), residence environment (rural, urban etc.) etc. The values of this kind of variables can't be arranged or, with other words, there is no hierarchy (but possibly considering intrinsic criteria) and therefore, "distance" problem and small values interval one, can't be considered. So much the less we can approach the existence of an "absolute zero".

Ordinal measuring level implies not only the classification of the elements in categories but also the chance of arranging them from minimum to maximum. However, at this level of measuring, no information is given about the "distance between the values of measuring scale". Otherwise saying, the difference between first value and the second one can be different from that between forth value and fifth [3]. Examples of variables measured at ordinal level are scholar marks (having the values: "insufficient", "sufficient", "good" and "very good"), the satisfaction about certain aspects (having the values: "very unsatisfied", "unsatisfied", "satisfied", "very satisfied") etc.

Interval-level measuring gives additionally to prior level (that ordinal) also information about the distance between scale values and it is characterized by the existence of equal intervals. However, at this level, of measuring there is no absolute zero, bat better said, a conventional one [1].

Measuring at rapport level incorporates all the features of prior levels (arrangement and equal intervals) and the existence of a "reference point" or absolute zero. This thing allows building sentences containing proportion terms (rapports between). Examples of variables measured at this level are: age, weight, height, distance, number of children for a house etc.

The correct identification of used measuring level is very important for the choice of statistical procedures of analysis. As how we can view in above description, for each level there are allowed mathematical operations and not allowed ones. Thus, at first level, that nominal, there are allowed neither arrangement nor addition/subtraction nor multiplication/division. At ordinal level, only arrangement is allowed, at the interval one, are additionally allowed addition/subtraction operations and finally, at the last level, that of rapport, all operations are allowed. Depending on measuring level, we'll speak about variables measured at nominal level, at ordinal level etc. or, shortly, nominal variables, ordinal ones, of interval and of rapport variables [2, 9].

Reducing the four classes to two, we can tell about qualitative variables (nominal and ordinal level) and quantitative ones (interval and rapport). Grace of "hierarchic" and cumulative character of measuring levels (from many restrictions to no restriction about allowed operations or from "qualitative" to "quantitative") we'll be always able to consider a variable from an "upper" level of measuring as how would have been measured at a "lower" level. Now, grace of the development of new techniques of analysis, especially conceived for "qualitative" measuring levels, there are more and more frequently the cases when a variable of a "lower" level is considered as one "upper" situated in hierarchy [10].

The expression using numbers of the presence of qualitative variable can be often indirectly made, using a basic numerical variable or a variable wherewith it is strongly linked to. This evaluation method, frequently used, supposes the replacement of qualitative variable with a representative one.

The existence of a variable which can be considered only with two aspects which are excluded each other, as existence / non-existence, acceptance / refusal, urban / rural, masculine / feminine, make possible the transfer into numerical space using the size 1 for one of the cases and 0 for the other. The encoding of the values of such variable, with 0 and 1, brings to a variable which practically measures the presence or the absence of studied feature. Such cases are generally frequent both in economy and in social domain (the success of a development program, intention to vote etc.) [5].

Either they refer to a given case (see above mentioned alternatives) or they are the result of the kind wherein the problem is set on, or they refers to a dichotomy generated by a past event, dichotomist or binary variables, i.e. the variables with only two possible values are very useful in statistical analysis practice. In these conditions, it is important to know how numerical characteristics are computed, especially the average and standard deviation of these variables.

If we start with computing formula of the average for grouped data, then we'll have:

$$\bar{X} = \frac{\sum_{j=1}^{k} f_j x_j}{N} = \frac{f_0 \times 0 + f_1 \times 1}{N} = \frac{f_1}{N} = p$$

where f_0 is absolute frequency of rising for 0 and $f_1$is absolute frequency for the apparition of.

With other words, the average of a dichotomist variable is right relative frequency of getting the value 1 (the number of observations which gave the value 1 divided to the total number of observations from data series) that is, relative frequency (marked here as *p*) of the cases wherein studied characteristic is shown [10].

If the analysis is made for a population, not for a sample then we say that average of a dichotomist variable is even the likelihood of rising studied feature (but only if that variable is encoded 0/1).

The same as the demonstration for average case, we can show that the formula of standard deviation for a dichotomist variable is:

$$\sigma = p(1-p)$$

where p is relative frequency of the presence of studied feature.

We discern the following cases:

a) *one-factorial case, wherein the variable x is of binary kind* ("dummy")

As example, in the linear case:

$$\boldsymbol{y = \hat{a}_0 + \hat{a}_1 x + u}$$

where: y= investments;
x = property form, having the possibilities:
- state ownership (1);
- particular property (0).

b) *two-factorial case wherein explanation variables are binary*

For example, in the case of linear model:

$$\boldsymbol{y = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 z + u}$$

where: y = national income;
x = kind of economy
- centralized-planned (1);
- of market (0);

z = the status - of peace (1);
- of war (0).

c) *The case wherein the model contains a qualitative factor and a numerical one*

For example, in the case of the model:

$$\boldsymbol{y = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 z + u}$$

where: y = computer demand;
x = income;
z = environment
- urban (1);
- rural (0).

In this case:

- for $\boldsymbol{z = 0 \Rightarrow y = \hat{a}_0 + \hat{a}_1 x + u_1}$
- for $\boldsymbol{z = 1 \Rightarrow y = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 + u_2 = (\hat{a}_0 + \hat{a}_2) + \hat{a}_1 x + u_2}$

That supposes that between the two cases, only differences in what concerns starting point can be reached, i.e. in the first case, this is $\boldsymbol{\hat{a}_0}$ and in second case, it is $(\boldsymbol{\hat{a}_0 + \hat{a}_2})$, the slope measured by the parameter$\boldsymbol{\hat{a}_1}$being the same. Graphically, the case is drawn by two parallel lines [6].

$$\boldsymbol{y = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 z + \hat{b} xz + u}$$

In this case:

- for $\boldsymbol{z = 0 \Rightarrow y = \hat{a}_0 + \hat{a}_1 x + u}$
- for $\boldsymbol{z = 1 \Rightarrow y = (\hat{a}_0 + \hat{a}_2) + (\hat{a}_1 + \hat{b})x + u = \alpha + \beta x + u}$

The comparison between the case when $\boldsymbol{z = 0}$ and that when $\boldsymbol{z = 1}$ marks both the different starting level ($\boldsymbol{\hat{a}_0}$if $\boldsymbol{z = 0}$, $(\boldsymbol{\hat{a}_0 + \hat{a}_2})$ if $\boldsymbol{z = 1}$), and that the lines have unequal slopes. ($\boldsymbol{\hat{a}_1 \neq \beta}$), the case of parallelism observed before, being not again.

A kind of model global checking, when besides an explanation numerical variable we reach one or several binary variables, can run the stages of test *F*, case when $\boldsymbol{F_{computed}}$results from:

$$F_{computed} = \frac{\frac{\sum u_x^2 - \sum u_{x,s}^2}{s}}{\frac{\sum u_{x,s}^2}{n-k}} \geq F_{\alpha;s;n-k}$$

where:

$\sum u_x^2 =$ sum of the squares of errors in the case$y = f(x)$;

$\sum u_{x,s}^2$ = sum of the squares in the case when we include besides x one or severalbinary variables;

s =number of input binary variables;

n–k= number of cases minus number of parameters, in the case that includes the binary variables.

The model if confirmed for $F_{computed} > F_{\alpha;s;n-k}$.

.

d) *The case when the effect variable is binary kind.*

For example, in the model:

$$y = \hat{a}_0 + \hat{a}_1 x + u,$$

where: y = family equipage with a car;
x = family income.

In this case, we distinguish the situations:

- for $y = 0$, the family has no car, it results: $-\hat{a}_0 - \hat{a}_1 x = u$;
- for $y = 1$, the family has a car, it results: $1 - \hat{a}_0 - \hat{a}_1 x = u$,

That implies perturbation heteroskedasticity, whereat there is added that its repartition isn't normal. In consequence, the evaluations for parameters, especially for small samples, are outside the possibilities of usual checking [8].

3. NUMERICAL EXPRESSION OF MULTI-VALUES VARIABLES

When qualitative variable has more than two possibilities, we give them marks or, depending on the case, numbers of order, respective to their status wherein they act. We discern the fallowing cases:

1) *Qualitative variable has a lot of states, without imply an arrangement of them respective to intensity*. Such variables are: nationality, religion, category, season etc. In such cases, we can proceed the same as in the case of binary variable, to giving equal distance values: 0, 1, 2, 3 ... for the different possibilities.

2) *Qualitative variable has different intensities, in a gradual shape of action, reason wherefore numbers of order are set to them, or the computing of proportional marks to intensity is done*. Such cases are on, especially in marketing studies, sociology, politology, wherein the opinion, satisfaction, confidence, training etc. can differ in intensity [9].

Giving values for some of variables is released because observation unity is framed in a category respective to training level, value of results, equipage, generation etc. So we meet: persons who passed a learning step, computers of a certain generation etc. In such cases, giving ordinal numbers which to follow the arrangement of categories wherein observation unity is framed would represent a solving of quantization problem. The lapse consists of that the values so given are step-by-step, that doesn't always correspond to the distances between categories of quality [4].

In what concerns the intensity of the dependence for such variables, using range (ordinal numbers) correlation quotient is recommended that can be get so (Spearman choice):

$$r_s = 1 - \frac{6 \cdot \sum_i d_i^2}{n \cdot (n^2 - 1)}$$

where:

$\boldsymbol{d_i}$ = difference between the ranges of variant "i" of the behavior of analyzed variables.

We remark that so quotient, computed for absolute values $|\boldsymbol{u_i}|$ corresponding to $\boldsymbol{x_i}$, is also useful for checking the hypothesis about equal error spreading and error independence respective to x. So, if $\boldsymbol{r_s}$is unimportant respective to *t* test, the hypothesis is verified as true. In the case of signification, we say that heteroskedasticity is present.

4. QUALITATIVE VARIABLES JOINING

In practice, many times the problem of checking the existence of a link between two qualitative variables is considered. First step we must made for getting answer to such challenge, is building a two-input table, also named contingency table wherein the values of one variable are on columns and the values of second one appear on rows. Within the cells of such table, we can have four kinds of information [9]:

a) absolute frequencies (number of elements from each cells and also the total number of elements that are included in different categories of variables);
b) relative frequencies on rows (which show what percent from the category which defines a row is also of the category which defines the column);
c) relative frequencies on columns (which show what percent from the category which defines a column is also of the category which defines the row);
d) relative frequencies from the whole (which show us what percent of the amount of analyzed types simultaneously belong to row category and column one).

Such a table gives information about two kinds of distributions; edge distributions and conditioned ones. Edge distributions are practically frequencies distributions of variables. Comparing conditioned distributions to edge distribution, we can have a first view about the relation between the two variables. When conditioned distributions are much different from marginal ones, we can expect to have an association relation between the two variables.

In the following part, we'll see how we can check the existence of such relation. The test $\chi2$ of independence, used for testing if two variables are or are not linked each other, give only information about the existence/non-existence of linking relation between two variables but not about the intensity of this relation when it exists.

For answer to the question “How strong is the association relation between two variables?” we need specific measures. Two of these we’ll be shown in the followings.

In the case of nominal variables, the (lambda) quotient is used. As how it is build, this quotient can be situated only between 0 and 1 (0 meaning the absence of any relation between variables, that is, independence and 1 meaning maximum intensity of association, that is strong association). The advantage of this quotient consists of relatively easy and intuitive computing kind. The main lack of this measure is that when a category of a variable has a big number of types, λ can be 0 even if the two variables are not independent [6, 8].

In the case of ordinal variables, there is the chance of arranging the values of variable and in consequence, there is the possibility to give ranges to types, depending on the values that they have for one variable. In consequence, in the analysis of this kind of variable we’ll be able to tell about a sign of association (or the sense of association). The measures of association of ordinal values can have values between -1 and 1. Generally speaking, a measure of linking between two ordinal variables, we’ll be positive if a type having a big range for X variable aim to have a big range also for variable Y, and the persons having small ranges on X variable, also have small ranges on Y. Negative association appears when the types having big range for variable X tends to have small ranges for Y and at reverse. If a measure of association between two ordinal variables takes the value 0, that we say that the two variables are independent. The more a linking relation between two ordinal variables will be stronger, the better association measure will be bigger in absolute value (nearer to 1)[9]. All the properties which evaluated values of a regression model have them, are also preserved when one or several independent variables are dichotomist (variables which take two values). The consequences of this property are important because they allows not only the evaluation of the effects of some variables that are usually dichotomist over the dependent variable and also the incorporation in a regression analysis of some nominal or ordinal variables that have more than two categories. This thing is possible after turning a variable having n categories into n-1dichotomist variables.

CONCLUSIONS

The systemic concept in approaching economic processes and phenomena demands the study of all factors categories, which through several kinds of connections, are present in the results of economic activity. Qualitative factors are those of the same kind as analysis object, differing from phenomenon through extension degree. Quantitative factors are the material wearers of those qualitative, the basic and absolutely necessary condition for the action of those qualitative. The factors establish the shaping and changing of an effect, of a result. They run, as a rule, not alone but interdependently, jointly in a solid system of links. Their finding need the exact knowledge of shaping way for the result, of causal or inside links if result as analyzed phenomenon.

The whole measuring of the action of every factor over analyzed result (effect) has the role to give end of size and sense for causal links, to put in evidence the factors having more important action over the results and to appreciate the measure wherein internal resources have been used.

Quantitative methods that can be used in economic-mathematic modeling vary depending on the aim of analysis, on the information sources and on the kind of causality relations between factors. In using quantitative methods of analysis basing on certain models, there is necessary to care about the contradictions which made these methods not to give whole satisfaction in knowledge economic phenomena: between essential and phenomenological, between causal and stochastic, between rational and empirical.

Bibliography

[1]Andrei, T. 2004.*Statistică și econometrie.* Editura Economică, București

[2] Baumol, W. J. 1985. *Economic Theory and Operations Analysis*. Prentice Hall

[3] Coman, Gh. 2007. *Statistică*: *teorieși aplicații*. PIM

[4] Gourieroux, C. 2000.*Econometrics of Qualitative Dependent Variables.* Cambridge University Press

[5] Greene, W. 2000. *Econometric Analysis.* Prentice Hall International, Inc.

[6]Iacob, A.I., Tănăsoiu, O. 2005. *Modele econometrice.* Editura ASE, București

[7]Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H. Lee, T. C. 1985. *The Theory and Practice of Econometrics*. John Wiley&Sons, New York

[8] Pecican, E., Tănăsoiu, O., Iacob, A.I. 2001. *Modele econometrice*. Editura ASE, București

[9] Pop, L. 2001. *Statistică.* Universitatea București

[10] Voineagu, V., Țițan, E. 2007. *Teorie și practică econometrică.* Editura Meteor Press, București.