Proceedings of the *International Conference on Theory and Application of Mathematics and Informatics ICTAMI 2005* - Alba Iulia, Romania

USING ECONOMETRIC MODELS FOR SOLVING ECONOMIC ISSUES. THE COMPUTER ASSISTED APPROACH

Remus Joldeş, Emil Olteanu and Horea Joldeş

Abstract. This paper presents some aspects as concerns the econometric models for solving economic issues by using computer-assisted methods. A hybrid assisted method, which combines programs written by user, and specialized software used for symbolic calculus has been proposed. The contents and mathematical expression of an econometric model is presented, the nature of the endogen, exogenous and residual variables and considerations on the elaboration and using of the econometrics model, the phases of an econometric model, an example of an econometric model and further approaches have been presented as well.

Keywords: Econometrics, econometric models, econometric algorithms, econometric models and computer-assisted methods.

1. Introduction

In the econometric models three types of variables are used: endogen variables, exogenous variables and residual variables.

The endogen variables named as explicated variables express the studied economic phenomena by a symbolical form. In most of the cases these variables present a certain complexity level thus their variation can be explained by one or more mathematical factors. Usually, the endogen variables are aleatory because they depend by a ε which has an aleatory value.

The exogenous variables named as variables of the model, explain the main part, which belongs to the variation of the endogen variables. These variables represent the main factors of influence of the endogen variables and the exogenous variables are considered as determinist variables or predetermined variables (e.g. the income of an i individual in the reference period of the consumption is expressed by the exogenous variable v_i).

The residual variables ε have an aleatory character and they synthesize the effect (influence) of other factors, which have not been assessed in the model by aleatory variables.

As an example, we will present the classical econometric model:

$$c_i = f(v_i) + v_i$$

where: c_i = consumption of a certain good achieved in an individual household i in a given period of time, v_i = the income in the same period of time, achieved in the household i and ε_i = the concentrated effect of the others from the household i who have an influence upon the consumption.

Definition 1. *Econometric model is considered as specified when the model has a mathematical definitive form, the mathematical form of function f is given and the probability law for the variation of ε, as well.*

Definition 2. *The ensemble of parameters, which occur in the model and completely define it, represents the model's structure.*

By applying these two definitions to the classical example which has been presented above and considering the correlation between c and v as a linear one,$f(v) = E(c/v) = av+b$, where $E(c/v)$ = the mathematical hope of variable c conditioned by v, therefore $c_i = av_i + b + \varepsilon_i$ for every $i = 1, 2, \ldots, n$, in the literature ([1], [3] page. 15 and [4]) is given $a = 0.1, b = 15$ (f being linear in v), and for the residual value ε is given a normal distribution with $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma_\varepsilon^2$ and $\sigma_\varepsilon = 0.02$.

The most general form of an econometric model, named also aleatory model, is given by the following system ([3] page. 14):

$$\begin{cases} f_1(y_1, ..., y_m, x_1, ..., x_m, \varepsilon_1) = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ f_i(y_1, ..., y_m, x_1, ..., x_m, \varepsilon_i) = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ f_m(y_1, ..., y_m, x_1, ..., x_m, \varepsilon_1) = 0 \end{cases}$$

where: $y_1, ..., y_m$ are endogen variables, $x_1, ..., x_m$ are exogenous variables and $\varepsilon_1, \ldots, \varepsilon_m$ are residual variables of the econometric model. This system can be often in an explicit form ([2] and [5]).

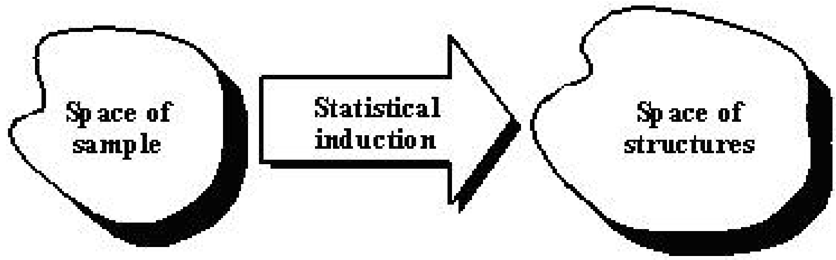

Figure 1: Paradigm of the statistical induction (inference)

For the establishment of the econometric model's structure the starting point is the aleatory sample of the measurements applied to the variables of the model, represented by the endogen variables and their exogenous pairs. This is the first step for establishing the econometric model's structure. In the above-mentioned example, the starting point is the consumption C and the income V.

In the first step it is used a unique working hypothesis in the econometrics practice and it is sustained by the paradigm of the statistical induction (statistical inference, figure.1).

The starting point being the unique source of information represented by the sample of the achievements of the endogen and exogenous variables of the model, it has been used the paradigm of the statistical induction which means the determination of a procedure which permits the transition from the space of the sample to the space of the structures. This procedure doesn't change the chosen model and estimates the parameters of the model (a and b) by confidence intervals and by determining the σ deviation, combined with tests of statistical hypothesis with regards the parameters of the econometric model.

The goal of the achievement of an econometric model with a well determined structure is to be used in future predictions or in a given situation (static model) of the values of the endogen variables in the case when the exogenous variables have been given.

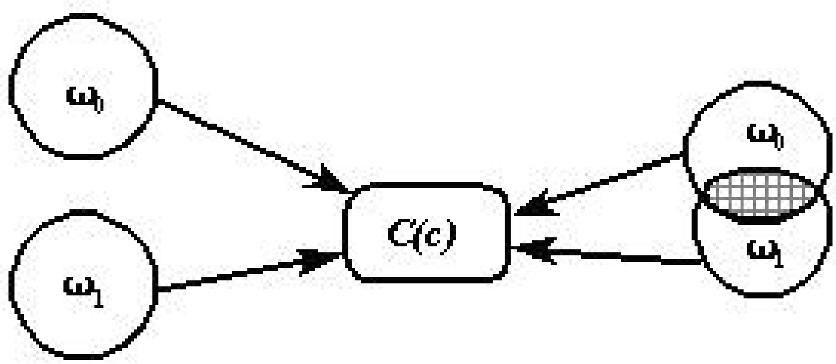

Figure 2: Two structures of the same econometric model incomplete identified

2. What happens if the procedure conducts to two structures of the model?

If we suppose that the model $c_i = av_i + b + \varepsilon_i$ (for each $i = 1, 2, \ldots, n$) conducts by the processing of the sample (c_i, v_i) not two a unique structure, but to two structures of the econometric model ([3] page. 20), structures like:

$$\omega_0 = \left\{ \begin{array}{c} a_0 \\ b_0 \\ \sigma_0 \end{array} \right\} \quad \text{and} \quad \omega_1 = \left\{ \begin{array}{c} a_1 \\ b_1 \\ \sigma_1 \end{array} \right\}$$

In this case the probability law defined for the residual variable ε express the variation law of the endogen variable C of the econometric model. If we consider the values of each structure and the given values of the exogenous variables, it will result a single law of $C(c)$. In this case will result two possible cases presented in Figure 2.

- If ω_0 and ω_1 are disjunctive and we cannot chose ω_0 excluding ω_1, we have the case of an unidentified econometric model and the values of the model's parameters can not be determined;
- If ω_0 and ω_1 aren't disjunctive, the ω_0 and ω_1 structures facilitate the partial identification of the model's parameters, of those parameters, which belong to the intersection of structures $\omega_0 \cap \omega_1$.

These two cases are equivalent because they do not permit the complete identification of the econometric model.

3. Aspects as concerns the elaboration and using of the computer assisted models

Any econometric research requires the achievement of models, which permit to the decisional factors of the economic agents to intervene in an effective manner.

For the achievement of an econometric model the following steps are required:

a. **Choosing and acceptance** of representative samples for data sets which will be the basic data of the model;

b. **Choosing of theoretical models** which are known and checked by practice;

c. **Specification of the econometric model** by its formalization and choosing the forms of the functions which describe the model;

d. **Selection and measurement** of the variables of the model;

e. **Validation of the econometric model**.

The complete automate treatment of these steps is almost impossible to be achieved but we are convinced that the accomplishment of specialized software which to assist the specialist in the econometric modeling is perfectly achievable and absolutely required.

In the first step the representative data sets are chosen and accepted for the construction of the future econometric model. This is an extremely important step because the non-representative samples or statistics affected by errors can conduct to an invalid model. This is a relatively simple step from the informatic point of view; it can be transposed in an input and validation program of the required numeric series. For example, if n represents the number of items of the series which define the endogen and exogenous variables, respectively, then the simplified algorithm used in this step can be:

```
int i,n
dim x[i],y[i]
read n
foreach (i) {read x[i],y[i]}
call selstat ; chose the statistical distribution for the residual variable
```

; by a ”case” structure and calculate the ε_i variables

The **selstat** procedure assists the econometrics specialist in choosing the statistical distribution for the achievement of the econometric model, by proposing several statistical distributions from which he will chose one. Here a “case” type logical structure has been applied.

In the second step, the known theoretical models, which are checked by practice, are called. This step is consisted of the formulation of the ensemble of hypothesis, which ground the real model. For the given example, in the economic practice [3] exist three fundamental sentences:

- The incomes and consumptions are bound variables;
- The volume of private investments and the rate of interests are dependent;
- There is an autonomous public investment;
- $PN = C + I_{pv} + I_{pu}$ where: $PN =$ is the National Product, $I_{pv} =$ Private Investment and $I_{pu} =$ Public investment.

For each economic problem such fundamental sentences can be formulated, sentences which can represent the basic starting point for the establishment of the econometric model.

From the informatics point of view the second step can be computer assisted by creation of a database and an expert system for the management of all fundamental sentences of the economic processes.

The third step deals with the specification of the econometric model by formalization and choosing of the form of functions, which describe the model. Starting from the fundamental sentences which have been previously enunciated the relationships of the model can be formalized as follows:

- Consumption C is a function $f(y)$ depending of the income having $f' > 0$;
- Private investment $I_{pv} = g(d)$ depends of the applied interest d which have $g' < 0$;
- We consider that exists a previous public investment at least;
- PN the National Product or the National Income is equal with the sum between the consumption $C$and the investment (private and public):

$$PN = C + I_{pv} + I_{pu}$$

equivalent with:

$$y = c + I + I \ .$$

In this phase there is no mention on the particular form of functions f and g. Considering the sign of the two derivates we can state that there are a set of functions with different forms but with identical derivates (e.g.: $f(y) = a_0 + a_1 y$ and $f(y) = a_0 y^{a_1}$). But, the functions from this set don't reflect the same behavior for the studied system. For the previous example the following explicit model can be written:

$$\begin{cases} c = a_0 + a_1 y + \varepsilon > 0, \quad \text{where } a_0 > 0, 0 < a_1 < 1 \\ I = b_0 = b_1 d + e \text{ where } b_0 > 0, b_1 < 0 \\ y = c + I + I \end{cases}$$

The explicit model is formed by two equations and an identity.

The selection and measurement of the variables of the specified model is an issue of retaining the main exogenous variables and of establishing the means of measurement and collecting of data for the respective model. There are two types of data, which make the model to be dynamic or static.

The dynamic model is mathematically expressed by form (3). For the variables of the dynamic model the chronological series have been set up. These chronological series have been obtained by statistical observations at equal time spans (e.g.: years, months, decades, weeks etc.).

$$\begin{cases} c = a_0 + a_1 y_{t-1} + \varepsilon_t > 0, \quad \text{where } a_0 > 0, 0 < a_1 < 1, t = 1, 2, ..., T \\ I = b_0 = b_1 d_t + e_t \text{ where } b_0 > 0, b_1 < 0, t = 1, 2, ..., T \\ y = c + I_t + I_t \end{cases}$$

The static model is mathematically expressed by form (4). For the variables of the static model the set of observed data is elaborated at the same time, which is considered referential time and these data are viewed as a sample.

$$\begin{cases} c = a_0 + a_1 y_i + \varepsilon_i > 0, \quad \text{where } a_0 > 0, 0 < a_1 < 1, i = 1, 2, ..., n \\ I = b_0 = b_1 d_i + e_i \text{ where } b_0 > 0, b_1 < 0, i = 1, 2, ..., n \\ y = c + I_i + I_i \end{cases}$$

In the static model c_i, y_i and d_i are variables which belong the units $i = 1, 2, \ldots, n$ of the sample, recorded at the same time.

Finally, the last step is the validation of the econometric model which means the checking of the whole ensemble of hypothesis with regard the forms

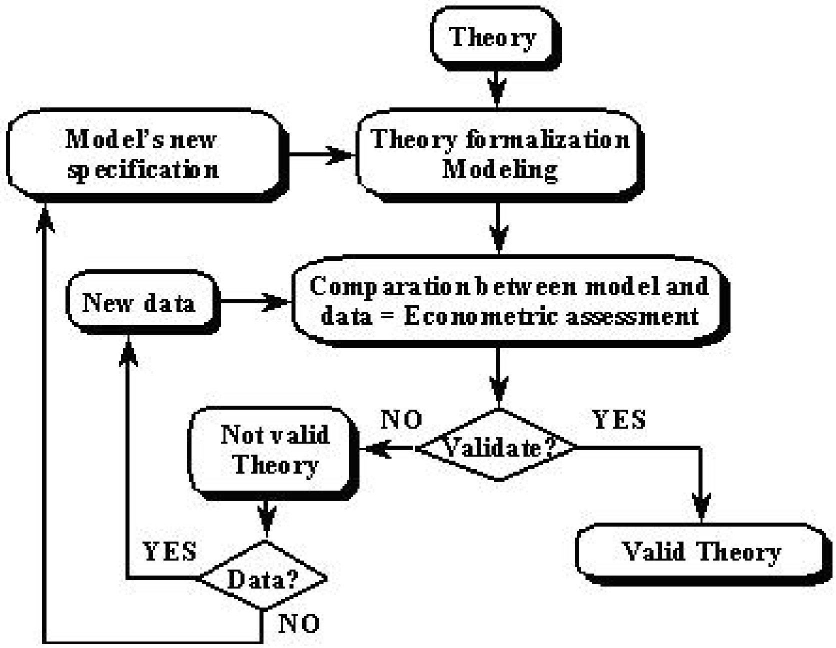

Figure 3: Validation chart of the theory, which ground the econometric model

and the variables included into the model. For the validation of the model a set of inference statistical techniques are used, namely statistical tests for the validation of a certain econometric model. If the model has been validated, then it is given to the economist who will check if a theory is valid or not by using the chart presented in Figure 3.

The steps of specification of the econometric model by using formalization and choosing the form of functions which describe the model, the selection and measurement of variables in the specified model and the validation of specified econometric model can be computer assisted by using specialized software in symbolic calculus, brand name products.

4. Results of the illustrative example

By applying the above mentioned steps and considering for the residual variables a normal distribution with $E(\varepsilon) = 0$ and $V(\varepsilon) = \sigma_\varepsilon^2$ and $\sigma_\varepsilon = 0.02$ an econometric model has been obtained for the consumption of 21 households depending on the incomes with the following mathematical expression: $c_i = 0.1102v_i + 14.901 + \varepsilon_i$ (with $i = 1, 2, \ldots, 21$), values which are relatively close to the models presented in [3] pag.15. For the symbolic calculus DERIVE product has been used.

5.Conclusions and further developments

The elaboration of computer assisted econometric models is possible and absolutely required considering the multitude of the current and future economic issues. The econometric models, econometrics in generally have to be seen not only as validation instruments but as analytical and investigating tools, as well. The decision factors in the economic organizations can achieve predictions by using econometric models with the goal of anticipation and if possible, to react on the economic environment. This aspect is one of the advantages offered by econometrics assisted by computer.

I am sure that this paper is a beginning for my future works. I have to continue with the treatment of the linear model of the simple regression, the generalized linear model, the general linear model with correlated errors, autoregressive processes, models with phased or displaced delay and of course, non-linear econometric models.

References

[1] Bourbonais, H., *Econometrie - cours et exercices corriges*, Editura Dunod, Paris 1993.

[2] Buiga, A., *Metodologie de sondaj şi analiză a datelor în studiile de piaţă*, Presa Universitară Clujeană, Cluj-Napoca 2001.

[3] Florea, I., *Econometrie*, Editura Universităţii din Oradea, Oradea 2003, pp.14-46.

[4] Florea, I., (Coordinator), Bâldea, P., Dragoş, C., Gabor, M., *Culegere de modele econometrice*, Editura Muntele Sion, Cluj-Napoca 2000.

[5] Purcaru, I., Berbec, F., Sorin, D., *Matematici financiare şi decizii în afaceri: teorie, cazuri, soluţii,* Editura Economică, Bucureşti 1992.

Remus Joldeş, Emil Olteanu and Horea Joldeş
Departament of Informatics, "1 Decembrie 1918" University of Alba Iulia,
Nicolae Iorga Street, 11-13, 2500, Alba Iulia, Romania,
e-mail: rjoldes@uab.ro