MULTI-STAGE SELECTION ALGORITHMS IN SIMULATION

M.R. Motzev and A.A. Marchev
Economic University "Karl Marx", Dept. of
Systems Analysis and Management
Sofia - 1000 G.Dimitrov str. 53
Bulgaria

Abstract : The present paper represents a part of the research* done in the Laboratory of Systems Analysis and Computers Aided Management in the Economic University "Karl Marx", Sofia. The main result of this research is a software package and a detailed instruction for automated design of simulation models. The program languages used are PL/I for computer IBM 4331 under VM-370 operation system and QUICKBASIC for IBM PC.

Synthesis of simulation models

The characteristics of the the method we used (4) for automated construction of simulation models are:

1) The design of simulation models in the form of a system of simultaneous equations.

2) A synthesis of the model using a multi-stage selection procedure.

3) The construction of a line of step by step improved models.

4) A high level of automatization in the process of building the model.

5) Evaluation of adequacy of the model using comparison with real data for the modeled system.

Multi-Stage Selection Procedure

The main idea in our work is that the synthesis of the model is based on a multi-stage selection procedure. At each stage of selection a variety of hypotheses are generated. Each generated hypothesis is verified and evaluated and then it competes with other possible hypotheses "fighting for survival". After that only a few of them are selected as "best" in the sense of a predefined selection criteria.

The so selected hypotheses are used for generating more complicated hypotheses at the next stage. The process of selection is started again etc... The selection procedure ends with the achievement of a certain condition, limits or results.

Working permanently in an interactive mode the user has the possibility for final choice of the best model and for some additional considerations (but after having the guarantee of evaluating a great number of possible models and the choice of a small number of good ones).

Multi-Stage Selection Procedure for Synthesizing Simulation Models in the Form of Simultaneous Equations

In the Multi-stage selection procedure (MSSP) which we designed (3) for synthesizing simulation models in the form of a system of simultaneous equations (SSE), we can differentiate three main parts:

First part: On the basis of existing real data, grouped in a table of observations of the researched system we make a first choice of significant factors, which will be included in the model. Then:

1) The generating of the variety is done by including non-linear transformations and taking into account the pre-history of the factors through euristic considerations by the researcher and automatically using a computer program.

2) Each competing hypothesis represents a hypothesis of inclusion or non-inclusion of an additional factor in the table, that is, a hypothesis on the fact is this factor potentially important for the mathematical description of the research system.

3) The limit choice and the limiting of the variety is done using the criteria "correlation with a dependant variable", but can be supplemented with euristic solutions given by the researcher. In all cases the preservation of the initial factors - "protection of the variables" is guaranteed.

Second part: It represents a procedure for multi-stage selection of each equation participating in the model. The form of the wanted equation is not predetermined. The task of finding the form and coefficients of each equation is divided in many tasks for determining the coefficients of equations with two variables - the principle of grouping variables.

In this way, we make certain of obtaining statistically important coefficients, based on undersized samples of data. The form of the equation is determined by a few consecutive stages of selection:

1) The generation of the variety is done by introducing new intermediate variables for each stage and the creation of a new generation of intermediate equations - functions with two variables which include indirectly more and more complex combinations of the initial factors.

2) Each generated equation is considered as the potential description of a given connection in the model, which competes with the other possible descriptions "fighting for survival".

3) The estimation of the coefficient in each intermediate equation is done using Ordinary Least Squares method. The existing set of data is divided into two parts: a "teaching" set which is used for estimating the coefficients in the intermidiate equations, and the control set used for evaluating the adequacy of the obtained equations - the principle of "Cross Validation Choice" of a good model.

4) After each selection stage, there is no choice of a unique equation, but a limited choice of a predetermined number of good equations is done - the principle of non-finalized solutions. This gives us the possibility to obtain a set of alternative good equations.

5) The chosen equations in a given generation are used for generating new, more complex equations in the next generation. From these a predetermined number are chosen as good etc.

6) The selection procedure ends when the following condition is fulfilled: "a predetermined by the researcher number of selection stages has been reached".

*This project has been completed with the financial support of the Committee for Science at the Council of Ministers under contract No. 396/1987.

7) At the end of the second part of the MSSP automatically the full form of the best equations is restored. Each equation describing a given connection in the model has a given number of alternative variations.

The so-described procedure is a typical multistage selection algorithm for the synthesis of separate regression equations. It is worth noting that in their general appearance, the selection procedures (1) can not be directly applied for synthesizing simultaneous equation models. In this connection, an extra third part has been worked out in the present MSSP, and the resulting algorithm has been organised in the manner of an iterative step procedure (fig.1).

At the First step is done the synthesis of equations in the reduced form of SSE, and here all predetermined variables are used as exogenous variables. The theoretical values of the endogenous variables are calculated using these equations.

At the Second step is done the structural and parametrical identification of the equations in the structural form of SSE. Here, for every equation the researcher inputs in the computer all potential decisive variables (here the endogenous variables do not participate with their absolute values, but with the theoretical values calculated at the Firts Step), while the type and value of the coefficients is automatically determined in the process of MSSP.

<u>Third part</u> : In it from the obtained equations a synthesis of the simulation model is done in the form of SSE:

1) The variety of alternative variants of the model, which represent SSE is obtained by combined, already chosen, best equations.

2) Each of the competing hypotheses represents a hypothesis for inclusion in the model of a given variant of the separate equation.

3) Each generated SSE is considered as a potential model which imitates the analyzed system.

4) The evaluation of these competing models is complexly done, using a great number of statistical tests and criteria - coefficient of correlation, variation coefficient, average relative error, precision in forecasting and others.

5) If the results are unsatisfactory (shifted values of the regression coefficients, low accuracy of the equations, etc.) the MSSP is continued. SSE is solved in respect to the endogenous variables and their theoretical values are again calculated.

6) These newly-acquired values of the endogenous variables are used for a new synthesis of the structural equations in the Second part of the MSSP. From the so-obtained new best equations are formed new SSE in the Third part, then these pass through selection, etc. The iterations end when satisfactory results are obtained.

7) The final choice of the best model is made by the researcher, who has the possibility to introduce some additional non-formalized considerations, but after having the guarantee of evaluating a great number of possible models and choosing one of the best synthesized.

We have to remark, that in difference to existing traditional methods (2), where the structural and parametrical identification of SSE is done separately, in the described MSSP, this is done in a unified, highly automated procedure. In it, we have the possibility for imposing predetermined limits on SSE

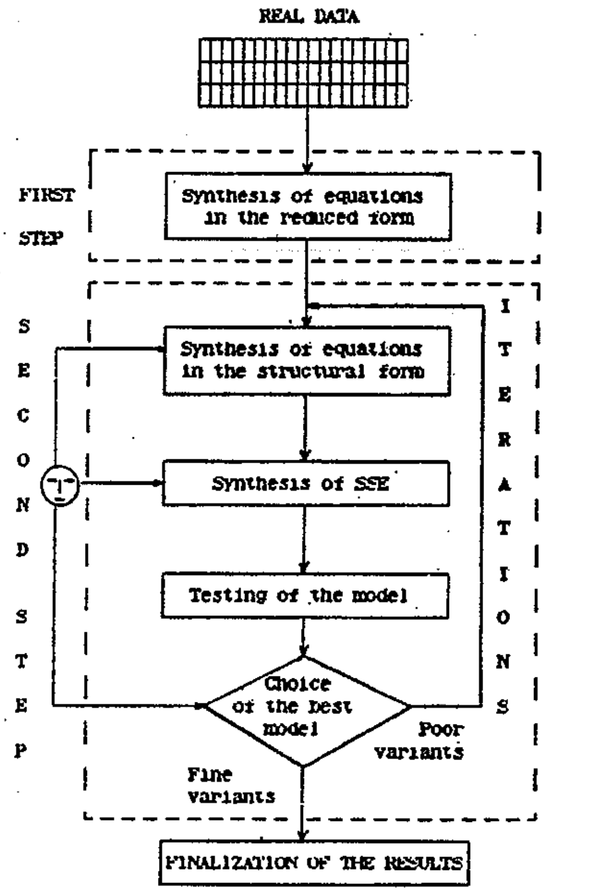

Figure 1. Procedure for automated synthesis of SSE

coefficients as in traditional methods. The difference is that an additional structural identification is done during multi-stage selection.

During this structural identification automatically some of the variables are removed and in the synthesized model remain only those, which are important. In this way, simultaneously the form of each equation is determined, as well as the important variables participating, and an estimation of the coefficients is done.

In addition the SSE coefficients can be obtained as time functions, which makes the MSSP applicable to the non-stationary systems describing most economic processes.

<u>Software</u>

For the described procedure for automatic synthesis of simulation models, a great number of algorithms and programs have been designed. These are organized in the following program sets together with additional programs, applied program sets (APS):

1) APS "ANALYSIS" - used for preliminary analysis of the dynamic role of variables in the designed model and the relation between them.

2) APS "SELECTION" - it is used for synthesizing the equations in the designed model, using the multi-stage selection procedure.

3) APS "TUNNING" - it is used for automatic tunning of undefined coefficients in the designed model and for automatic choice of equation blocks in SSE with

TABLE 1. The family of models " S I M U R "

Model No. (Year of design)	Main purposes	Specific characteristics
SIMUR No. 0 (1977 - 1978)	First step of the SIMUR project. Traditional methods used.	A one-product macroeconomic model in the form of SSE with 5 equations. Contains 5 endogenous, 5 lag and 1 exogenous variables.
SIMUR No. I (1978 - 1980)	Analysis of possibilities for automatic synthesis of SSE during the run of MSSP.	A one-product macroeconomic model in the form of SSE with 5 equations. Contains the same set of variables.
SIMUR No. II (1981 - 1982)	Design and experimenting of a programming system for automatic holding of simulation experiments with SSE. Analysis of different criteria for the precision's estimating of SSE.	Aggregated macroeconomic model in the form of 12 interdepending simultaneous equations. Contains 12 endogenous, 5 exogenous and 26 lag variables with lag of up to 3 years.
SIMUR No. III (1983 - 1985)	Improving the MSSP for synthesis of SSE with many equations. Simulating and forecasting of main macroeconomic indexes.	Macroeconomic simulation model. Contains 39 equations, 39 endogenous, 7 exogenous and 82 lag variables with a time lag of up to 5 yers.
SIMUR No. IV (1989 -	Next step of the SIMUR project.	Multisector macroeconomic model. Contains more than 100 equations.

the help of a "competing" procedure between the model variation.

4) APS "SIMUL" - it is used for complex evaluation of the adequacy of the synthesized model using a great number of criteria and for conducting different simulation experiments with the model.

These APSs cover the main tasks during automated design of simulation models.

Application

Together with the design of a procedure for automated synthesis of simulation models, we designed a family of macroeconomic models called SIMUR. The models from this family can be considered as a result and at the same time as a base for this procedure.

Each model from the SIMUR family is designed with a special purpose - experimentation and improvement of a given part of the MSSP for synthesis of SSE. The conclusions obtained from each experiment with a model are used on one hand for designing the next model and on the other for improving MSSP. In this way we have achieved a parallel development and interaction between the family of models SIMUR and the MSSP for synthesis of SSE.

Short information for the models is presented in table 1.

We have to remark, that during the design of new simulation models and with new acquired experience, we continuously enrich the set of algorithms and programs for automated synthesis of SSE.

Conclusions

This paper presents a new and perspective line in research and modelling of economic systems - using a multi-stage selection procedure for synthesis of simulation models. The proposed approach for automated design of simulation models gives us the possibility to shorten the time expenditure and efforts for the design of a simulation model and consequently widens the field of application of simulation modelling. The accumulated experimental data (3,4) shows, that MSSP has great advantages over the existing methods - system dynamics, econometric approach, aggregated modelling etc.

In this connection, the presented approach has a large field of application and needs further development in the following main directions:

I. The design of new simulation macroeconomic models. In this direction, special interest represents the design of the macro models for different countries using the same set of variables. On the basis of this, it will be possible to make valuable comparisons and analysis of the common and specific characteristics reflected in the different models.

II. A further development of the proposed approach, by designing new algorithms and programs for automated construction of simulation models. The aggregated experience and the achieved results show that in the future the models could be constructed on a very high automation level, up to a fully automated interactive procedure which includes: the input of the initial data for the modelled system; preliminary analysis and processing of the real data; synthesis of the separate equations in the process of multi-stage selection procedures; generating and selecting the multitude of competing systems of equations; conclusive adjustment and evaluation of the model as a whole.

References

(1) A.G. Ivachnenko and J.A. Muller, Selbst-organisation von Vorhersage-modellen. Berlin: VEB Verlag Technic. Kiew: Verlag Technika. 1984.

(2) E. Malinvaud, Methodes Statistiques de L'econometrie. Paris: Dunod. 1969, ch.5, pp. 207-306.

(3) A.A Marchev and M.R. Motzev, "Simulation Macroeconomic Models Designed in the Form of Systems of Simultaneous Equations Synthesizing During the Run of Multi-stage Selection Procedure", Scientific Papers of the Economic University "K. Marx", Sofia, Vol.I, pp. 303-349, 1983.

(4) A.A. Marchev and M.R. Motzev, "An Approach for Constructing a Simulation Macroeconomic Models", Scientific Papers of the Economic University "K. Marx", Sofia, Vol.II, pp. 511-547, 1984.