DATA REQUIREMENTS AND EVALUATION

H. KASNAKOGLU

As far as I know, with the exception of very few, most of the agricultural sector models are either developed by academicians, research institutes or by international agencies like the World Bank. Furthermore, due to the reasons raised in Prof. Bauer's overview paper, they have not reached the stage of being employed as tools in the policy making process.

The developing of such models involve on the one hand gathering a highly demanding data set, and conceptual model formulation, solution algorithms and software on the otherhand. Considering the fact the data requirements have to be met from different sources, both published and unpublished, official and non- official, and the fact that a reasonably medium sized model easily riches the memory limitations of most of the hardware and the software, there arises important trade-offs, which almost in most cases work against the data bases of such models.

This is understandable, since as far as the academicians are concerned as Prof.Bauer has pointed out, the benefits of long investments in data bases, are far below their costs with the prevailing academic norms which can be summarized as "publish or perish". Furthermore, the academician does not pay for the consequences of his mostly "temporary-preliminary" results. In the case of the international agencies like the World Bank, it is the trade off between time and the good data base which moves the pendulum against the quality of the data base. The sector models, being employed as tools in agricultural sector missions to LDC's generally leave no more than couple of months for the data base, to be put together by those experts who are foreigners to the agricultural sector modeled.

This is a vicious circle, which results many times in data sets which originated as "guesstimates" but which later becomes facts and their origins are very soon forgotten.

We have experienced this in the case of TASM. For example, the price elasticities employed in the model to incorporate consumer demand functions, were estimated from the income elasticities borrowed from the World Bank and transformed via Frisch method, but were also calibrated during the base solutions. Such calibrated (adjusted) price elasticities, after appearing in World Bank publications became facts and used by many others, including ourselves, by making reference not to their origins, but to the latter World Bank publication.

Everybody who used risk in sector models, knows about the risk coefficient theta employed, which is generally assumed to be 0.5, 1 or 2, because those happened to calibrate some of the well known sector models and not because anybody knows where they come from or what they really mean.

Another very important factor contributing to the poor data base of many of the sector models, is the conventional validation and calibration methods employed. In general, many of the sector models end up with different data bases than the ones they have started with, in the process of calibration which involves adjusting critical data and parameters to bring the model estimates closer to those observed to pass the validation tests.

Since the journal articles, even books don't reserve any space for presentation of the model data employed, and the sources of data employed and adjustments made are not precisely reported, it makes it very difficult if not impossible to realistically evaluate the presented models and their results. I have not seen any published sector model with unsatisfactory results. They all validate, calibrate and are right on in policy simulations. I don't believe, although I would like to, that the picture is that bright. I don't also believe that such an approach of painting model results in pink is helping the sector modeling profession in the

long-run. On the one hand ad-hoc data manipulations hide the real deficiencies of the models, and on the other hand, make the life of developed models very short since it is very difficult to update ad-hoc data and parameters, even by the initiators of the original ad-hoc data, let alone by others to take over.

I therefore suggest that, the agricultural sector modelers set up some norms in the standard outputs reported to accompany model results. In the case of models published in full length, in books or monographs, I suggest that the raw as well as the model data be presented. I think that 10 additional pages will pay far more than their costs. In the case of journal articles presenting model results, I suggest that, the shadow prices for fixed inputs as well as those for "flexibility" constraints are presented as standard output, like the "t" statistics or standard errors in econometric studies.

We have already suggested a procedure in another paper (Bauer and Kasnakoglu; 1988) and employed it for the results presented in our paper earlier in this conference, to at least limit data adjustments to consistency efforts (which is not avoidable given that the data comes from different non-compatible sources), free the original data from ad-hoc adjustments for calibration and validation purposes and finally to provide an explicit account of model cost structures.

I hope that our discussions today could at least form the starting point for setting up some norms in the presentation and evaluation of the agricultural sector models.

DATA-BASE MANAGEMENT IN RELATION TO ECONOMIC MODELLING

M.KEYZER

It is self-evident that empirical models need data-bases to feed them. Less evident is the notion that data bases need models to structure them. This will be the topic of this introductory note.

In the absence of further structure, a data-base is a set of data which are retrievable. Data-base management deals with the way to structure the data so as to make them more easily retrievable, say on the base of some characteristic and with the computer software used to generate and access these data on acomputer. This has seemingly little to do with economics and therefore with this seminar, until one realizes that the structure given to a database should have some theoretical foundation in its field of application, otherwise the database becomes a sheer collection of indicators.

In economics the physical flow of commodities from source to destinations via intermediate nodes is the object under study. Commodities are produced, traded, consumed, carried-over, etc. It is therefore natural to choose such a flow as the guiding principle for database structure. The flows is expressed as physical quantities per time-unit and may accumulate through stocks. The economic conditions under which the flow occurs are to a large extent described through the payments which take place against them. Differences in value per unit of volume at various points in the network reflect costs of moving the commodities through the network. These costs may relate to physical transportation, storage, or processing, i.e. to physical transformation processes, as well as to government intervention through tariffs and subsidies. Having represented flows in value and volume for commodities, one may obtain budget accounts through aggregation, except that for some variables, such as savings and income taxes, there are value flows without a volume counterpart. Thus, the structure of an economic model can be mapped onto the data base. Once the structure of the flows has been specified, the data base must be filled with numbers. It is then that the major problems come up as the numbers are usually not published in the classification required and several sources give different numbers for the same variables. This is where the actual statistical accounting work begins. One has to reclassify, choose between sources, or scale numbersto make alternative sources compatible. Thus, there is a lengthy process of manipulation between data-source and data base. It is imperative for data base maintenance, that this flow of operations be well-documented and computerized. Commercial data base packages tend to put emphasis on easy retrieval rather than on documentation of data-manipulations. In our work at the Centre for World Food Studies we feel that this emphasis should be reversed. As the data needs of a given economic model do not change overnight, data retrieval is only the tail-end of a long process of data manipulation and documentation. Overbosch (cf (1)) has developed a data-management software which is tailored to these needs and which we use in our database work.

Once a consistent data base has been obtained it is time to turn to model simulation. Two sets of data are needed. First there is a consistent set of economic accounts for some given year. This set, called a Social Accounting Matrix (SAM) in value and volume, is mainly used to provide initializing values for endogenous as well as exogenous variables.

Then, there is time series information which is needed, in addition to the SAM, to econometrically estimate behavioural parameters as well as to provide time-series values for exogenous variables.

When initialized properly and with appropriate scaling of behavioural parameters, the economic model should in the base-year immediately reproduce the data base, i.e. without any iterative steps. This feature has great practical value since it permits to detect