SOME NEW METHODOLOGICAL CONCEPTS AND THE USE OF THE COMPUTER IN QUANTITATIVE ECONOMIC HISTORY

H. VAN DER WEE, Hoogleraar K.U.L.
E. VAN CAUWENBERGHE, Aspirant N.F.W.O.
H. DAEMS

1. The archives-technical progress, the manifold publications of important source material, the refinement of historical critique and the enlargement of scientific research make it possible now to propel the socio-economic history to a new phase of analysis. In the past the methodological anxiety of the historians was focused on the improvement of both the research instruments and the criteria of criticism concerning the reliability of the preserved data.
Obviously, this anxiety is maintained, but the technological attainments in the heuristic field have definitely shifted emphasis in the methodological problems to the field of the interpretation. It must be admitted that early specialists in socio-economic history have not only been confined to unearthing and compiling reliable data. These data have also been described, ordened and published. Moreover, on the basis of ordening and compilation of data, explanations have been advanced and presented. In the field of explanation, however, historical methodology has remained rather primitive. Most of the interpretations of historical phenomena mainly emanated from the intuitive apprehending of a relationship between data, resulting from a simple ordening or a comparison. Consequently, relationships between two or three variables at a maximum were looked for. In addition it has always been taken for granted that the relationship was only confined to one dimension, i.e. cause-result, while less concern has been attached to the simultaneous dimension. Hence the refinement of the explanation methodology is the main problem for the historians of the forthcoming generation. Here, the computer will come forward as an essential auxiliary instrument. In order to clarify its importance in this subject, it is necessary, however, to review briefly the various possibilities for an improvement of the interpretation techniques in socio-economic history.

2. Each historical phenomenon is caracterised by a whole range of variables, which affect it either causally or simultaneously. The insight into and the elaboration of the causal or simultaneous relationship of these variables give the explanation for the studied phenomenon. On the other hand, taking the whole of these influencing factors, some can be revealed to be particularly relevant : the detection of the relevant factors can be obtained either by direct empirical evidence, by theoretical considerations, by intuitive apprehending, or by combining these three alternatives. The preliminary selection of these apparently relevant variables and their coherence form the starting-point of the research : it is to be the working-hypothesis, the working model. The framing of a model is scientifically justified as long as sufficient evidence is available for provisionally maintaining the proposed hypothesis, not enough other evidence

Extrait de la Revue (R.E.L.O.)
VII, 1 à 4, 1971. C.I.P.L. - Université de Liège

being present to counter it. Obviously, in the selection of such a model there is always a subjective element. Yet, as the selection ranks as an interim starting-point, subjectivity is minimized.

The historian has to deal with another problem of subjectivity. The data, kept in the archives, are mostly not only fragmentary, but what's more, they are intrinsically incomplete : they are merely remainders of historical reality, accidentally fossilated in the preserved documents. With these few fragments the complexity of the living past has to be reconstructed. This is always a risky matter, for the preserved parts can be joined in various ways, each time resulting into a different image of reality. Moreover, through reasoning or statistical intervention, inter- or extrapolations can be carried through, from which the fragmentary data can be filled up in such a way that tentatively a more complete reality is obtained. Both kinds of reconstructions form an important basis of information for the formulation of historical working hypotheses. Consequently, from this point of view also, working models rely heavily on a subjective basis. Yet, the refinement of the interpretative methodology can sensitively minimize the above mentioned subjectivity factors and substantially increases the objectivity of the link between the model and historical reality.

A last subjective factor can be discovered in the process of the data-collecting itself. Though always intrinsically fragmentary and incomplete, the historical data for many periods and phenomena are nevertheless so numerously available, that a selection becomes inevitable. Therefore scientific objectivity is strictly required here, which, next to serious source criticism, also implies honesty in the choice of the usable material.

3. The character of working models can be either qualitative or quantitative. The *qualitative* models offer a realm of relations, which are neither mathematically nor statistically measurable in view of elucidating the studied phenomenon. Consequently, the point at issue in a qualitative model is the analysis of the relationship between variables of a psychological, philosophical, sociological, aesthetical, religious or any other origin, which can nowhere be brought back to a reliable quantitative expression.

   Mostly the qualitative model is intuitively established. Moreover, the testing of the working model with the compiled data often relies on a rather weak basis : in this field, many social sciences are still underdeveloped, and often remain intuitive. In many social sciences every intelligent researcher can rather easily put forward a new theory, for, next to the building up of the working model, the testing and interpretation of the compiled data greatly remains subjective. It goes without saying that greater objectivity in the interpretation of qualitative models in the social sciences is possible; yet this will require a long process of methodological renewal and refinement of the qualitative measuring and testing techniques. Although not perfect, a close substitute to this new methodology in decreasing subjectivity consists in the critical and comparative analysis of the presented model by co-researchers.

4. The *quantitative* models are hypotheses concerning a well-defined historical phenomenon, which, by way of elucidation, express a quantitative relationship between a set of variables. It should be stressed again that also the construction of a quantitative model is strongly influenced by subjective factors. During the preparatory process of apprehending and conceptualising the relationship between some quantitative variables of the studied phenomenon, even during the next stage of choosing the empirical methods and

techniques of measurement, the personal point of view and the subjective valuebias undoubtedly are extant to a large degree. In that way the quantitative hypothesis does not essentially deviate from the qualitative one. Yet, figures have a special advantage : they are precise and thus, to the formulation of the working model, they offer the possibility to fasten down the preconceived relationship in a simplified and clear-cut way. Such a precision cannot always be attained so easily for the qualitative models. In the testing phase the advantages of figures are even more obvious. Although the starting point, as formulated in the hypothesis, keeps its subjective character, yet the testing of this hypothesis becomes more objective. For to the quantitative relationships, provisionally preconceived by the working model, the refined measuring techniques of the positive sciences can be applied. The introduction of these precision instruments gives a firm backing to the testing and interpretation of the data and consequently contributes largely to the objectivity of the research.

The quantitative model can be worked out in two ways. First, it can be confined to an *indication* of relations by means of statistical operations. In this group statistics are applied to give a quantitative expression to the apprehended relations, and to check this expression in a further stage on the basis of new source material. Moreover, the above mentioned indications can be conceived statically or dynamically. In a *static* model the statistical operations concerned with the building up and further control of the model, will especially become operational in a systematic ordening of the numerous data, compiled with respect to a historical event that occurred at a given moment. A surprisingly new insight in the structure of the complex phenomenon can for example be revealed when the relative share of the different kinds of data are compared to the data-set as a whole. Where such relative distributions are calculated from a great number of different points of view and when, moreover, the data are split up in ever smaller subgroups or categories, a far-reaching analysis of the historical event is possible, and the relationship between the variables, which defined the event, can reliably be reconstructed. The statistical sifting of the data, as presented here, is called *cross-tabulation.* Cross-tabulations allow for such a great number of combinations that the *computer* can be inserted as the ideal operating instrument. The socio-medical study on mental ill care in Geel (1855-1965), undertaken at this very moment by Leuven historians under our guidance (1), now carries out a whole series of cross-tabulations, on the basis of the data available per mental patient (cards of 40.000 patients have been predeved). Analogue cross-tabulations are applied to the data of domanial accounts of medieval Flanders, i.a. concerning the relative importance of the various kinds of income in total income, and of the relative importance of the various forms of lease in total rent-income (for more details, see infra).

Cross-tabulations are not only used in purely-static analyses, they can also be applied to static comparisons throughout the time : in that case cross-tabulations on data from a given period are compared to the results from other periods acquiring by that way a dimension of time. For that purpose the data have to be put together in chronological groups : with regard to the research on the Geel data three groups are made, namely 1855-1914, 1914-1940, 1940-1965; with regard to the research of the Flemish domanial accounts, collections of 50 years are made from 1350 up to 1600. Shifts in the relative importance of the different sets of data, which come to light by mutual comparison, reveal the modified rela-

tionship of the phenomenon studied in its chronological dimension. It is clear that the understanding of the phenomenon as a whole is made much easier and deeper by this kind of chronological analysis.
The *computer,* as we have already said, is an indispensable auxiliary instrument to any intensive use of the cross-tabulation method. It increases the possibility of multiplying the number of combinations and comparisons, and supports, by this way, the building-up and checking of quantitative models, which *indicate* certain relations. Nevertheless, it should be pointed out that, in the case of checking indicated relations, the function of the *computer* remains limited : it is essentially the function of a perfect and automated *sorter,* completed with some simple statistical operations of addition and percentage calculation. To be sure the correlation between the various variables can already be *indicated* in a *quantitative* way; yet its formulation remains vague and indefinite, so that the explanation of the phenomenon, i.e. the building-up of a comprehensive model, is always partly anchored in the intuitive sphere. Besides, attention will always be too firmly rivetted on the relations "cause-effect", while the simultaneous influencing elements can hardly be considered, or are not considered at all.

5. The quantitative model, which confines itself to indicating the relations, can also take a *dynamical* position, next to a statical one. With a similar dynamic approach of the relationship between variables, the analysis of time-series will come to the fore. Here, refined statistical methods are available : seasonal movements, cyclical fluctuations and secular or even intersecular trends can be calculated in a representative way. The techniques used for obtaining homogeneous data, comparable in the train of time (especially the calculation of reliable index figures or the translation into logarithmic), will largely come in handy at this moment.
A next phase in the research is the mutual comparison between the different time-series through time, made possible by the application of the various methods of correlation-computation. In this way not only two, but even more time series can be compared with each other (single and multiple correlation-computation). The obtained correlation-coefficients allow a quantitative *indication* of the relationship between various available time-series. They can form the basis for the *formulation* of a quantitative working model; they can further be used as a *test* for the reliability of a qualitative working hypothesis, or - the supply of new data provided - as a *touchstone* for the reliability of a quantitative working model. Besides, the statistical techniques allow a chronological follow-up of the correlation. In this way the discovery of shifts through time in the indicated relationship can add a dynamic dimension to the explanation of the studied phenomenon. Here, the spectral analysis can come forward as a very important auxiliary method : the correlation comparison between the various time-series is not only carried through here from the view-point of the intensity and periodicity of the fluctuations, but also from the view-point of the simultaneousness or lags in the fluctuations. From this a whole network of indicative relations can be deduced, and consequently, the analysis of the historical phenomenon can be deepened to an large degree.
The statistical methods, that are connected to correlation computation, already presuppose rather complicated operations, especially when a great number of data has to be assimilated. Here once more, the *computer* is the ideal auxiliary instrument. There are standard programs available for the application of most statistical correlation methods so that the processing of the mass of data, inserted in the *computer,*

can pass off relatively smoothly and without difficulties.
Though the *computer* can be used in a sophisticated way for the analysis of time-series and for the correlation computation, in order to plot relationship between variables and to check this relationship, yet some important limitations still are in force here. Now, the function of the computer does not revert to that of a perfected *sorter* as described earlier, but rather to that of a perfected calculating-machine. In essence, the *computer* is used here for the automatic execution of a whole series of complicated calculations concerning a mass of data collected in time-series. From the analysis and comparison of the results, the researcher can only indicatively deduce a relationship or a succession of relationships. Consequently, the intuitive and subjective character of the interpretation is not entirely eliminated. The other criticisms, formulated in connection with the cross-tabulations, are also in evidence here, on the understanding that the stage of measuring is undoubtedly more objectified by the introduction of ever finer techniques.

6. Quantitative models do not only refer to the *indication* of relations between variables to elucidate and explain a historical phenomenon, they can also *specify* these relations, so far as data are available. The point is that the relationship between the variables is expressed in a formal connection. Next to statistics, mathematics also comes forward as an interdiscipline. The variables are placed in a mathematical equation with regard to the studied phenomenon, and their functional relation to the phenomenon is elaborated in specifying formula. The mathematical specification of a functional relation can take on various forms, such depending on the kind of relationship existing between the variables as the linear, the logarithmic and the exponential forms. Next to the relations "cause-effect", the mathematical equations can also reproduce simultaneous relations, and they can further be elaborated both statically and dynamically. A specified quantitative model can be set up on a basis of intuitive experience (i.e. departing from a qualitative working model) or on a basis of an indicating quantitative relationship (i.e. departing from an quantitatively indicated hypothesis). As soon as the working model is specified, it can be tested by way of the method of regression. This method defines the functional linear relation between the quantitative data of the studied phenomenon and the quantitative data of the various groups of variables (e.g. between growth of population of a certain town on one hand, and the differences in income and the differences in infrastructure between town and countryside on the other hand). The parameters, which appear from the regression of equations, define for each group of variables the linear functional relationship with the studied phenomenon. For each of these parameters a set of statistical testing coefficients can be acquired : they indicate to what extent the functional relation is relevant and well fitted, in other words to what extent it can be considered as a strong or weak relation (2).
If irrelevant relationships are obtained from the regression of certain equations, the working model might not be applicable. If this is true, new determinants must be looked for in the residue factor, which normally appear in the formal specification of the function. These new determinants have to be taken up in order to be tested.
Seen in this light, quantitative history is a form of laboratory work. However, in the explanatory study of historical phenomenons a complete exhaustion of the residue factor will mostly appear impossible, not only because certain determinating data have got lost or are not directly and completely quantitatively

measurable; but also because the irrational and non-deterministic factor in human behaviour, even in the perspective of macro-history, cannot be disregarded. On the other hand, the regression of certain equations can result into a set of very relevant relationships. On the basis of these results, an explanatory model for the studied phenomenon can be established. This comprehensive model too remains essentially provisional, i.e. a kind of working model. For a further analysis can bring the co-defining influence of new variables to light, so that each model always remains susceptible to completion or alteration.
Obviously the statistical methods for analysing and indexing the time-series, as they are discussed above, will also be useful for the preparation of data in the specified quantitative models. Finally, for some studies the introduction of simulation techniques can deepen the insight in the relationship. The historian can simulate the appearance of certain variables, and besides he can research what their relationship with the other variables would have been, in order to check what effect the new situation would have had on the historical phenomenon which is to be elucidated. This method is rejected by many historians, because it analyses a phenomenon that historically never has occurred. For some well-defined studies, however, the simulation-method can bring about a broader dimension to the insight.
Once more the *computer* is an ideal auxiliary instrument for the application of the method of regression and for the use of simulation techniques. Standard programs are available for the various elementary regression-methods, so that the elaboration of the data can easily be carried out. Here as well, the *computer* essentially remains a perfected calculating machine. Yet the signification of its calculating intervention goes much farther than in the previous cases. At present, the *computer* searches, in a mathematically precise way, for the specified functional relation between variables and at the same time it reproduces the relevancy of this relation. For a rather extensive calculation, which is often the case for the study of historic data, only the *computer* can efficiently carry out the processing. With thát an imposing methodological progress is made. For the transition from the confrontation between the working model and the compiled data to the explanatory model now goes along a strongly objectified way, in other words, during this important phase of scientific research, the space for intuitive, subjective interpretation is greatly sensibly reduced. Besides, the relationship between various historical variables can both causally and simultaneously be defined in a very precise way, so that the historical interpretation gains in purity and objectivity.

7. *Computer* operations on historic data can only lead to reliable results if the data have been previously well conditioned. Data, as they appear in documents, mostly cannot be fed immediately into the computer. Even if the documents have a framing, which allows a direct feeding in (which is rarely the case), the data would mostly have to be made homogeneous first, in order to be susceptible to statistical and mathematical assimilation. A good example of this problem is provided by the study of the domanial accounts of the County of Flanders, 1300-1600.
We would like to draw attention to the fact that, with this contribution, we only introduce a scheme for the working out (by ordinator) of our quantitative material. Obviously, we cannot give a complete description of the employed analysis methods and programming techniques. Therefore, we stick to a surveyable framework, in which the various phases of the procedure are clearly reflected.

01. *THE CONCEPTION PERIOD*

011. *The research or the inquiry*

* first task : definition of the problem
* importance of the source
* description of the source
* scheme of the accounts : receipts and expenditure
* source-criticism

012. *Architecture :* preparation of the theoretical solution

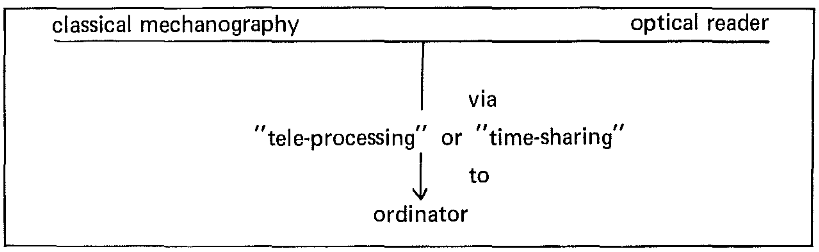
* compiling of data from the sources
* transcription of the non-classified information
* codification
* preparation of the mechanographic questionnaire

013. *Serialization of the information*

* instructions for the C.L.U. (Central and Logical Unit)
* data (input data, intermediate results, output data)
* reference data, necessary for a certain application, such as : tables, code lists, constants.

02. *THE REALIZATION*

021. *Choice of the means of production*



Extrait de la Revue (R.E.L.O.)
VII, 1 à 4, 1971. C.I.P.L. - Université de Liège - Tous droits réservés.

022. *Initial period*

* transformation of the theoretical solution into the real solution
* the ordinogram
* testing of the program

023. *Production*

* cfr. ordinogram : from reading instruction to listings
* process control.

03. *THE RESULTS*

031. *Results from primary statistical elaborations or arithmetical operations*

* trends
* percentages
* calculations of productivity

032. *Results from logical operations*

* presentation of the combinations attained
* cross-tabulations
* correlation and comparison of some series
* other combinations and graphical configuration (diagram) (3).

Yet, from experience we know that, for the historian, who ventures into the field of mechanography, the main obstacle is to be found in the first phase of this scheme, that means, in the conception period and mainly in the architectural phase.
That is why we consider it very important to expatiate on this.

8. *Architecture or the preparation of the theoretical solution* (012)

Here, the very bottleneck, i.e. the transformation of historical data into a mechanographically usable matter has to be conquered.
However, the concrete example, which will be described, is in the first instance applicable to the treatment of the rent-series from the medieval accounts of the royal domains in Flanders.

The collecting of data from the sources

It is obvious that the taking over should be performed in function of the way in which these data will be ultimately utilised and assimilated. The realization of the objectives one wants to attain by means of automation, depends on a long period of preparation, i.e. the period of analysis. The important point here is the definition of the object and of the grounds for existence for each operation in the treatment. In order to check all the consequences of each elementary factor, it is necessary to keep considering the different exploratory data from the beginning up to their ultimate assimilation as results. Thus, the analysis consists of the elaboration of a dynamic treatment system, which alters the available information and which follows, step by step, the evolution of these alterations in order to attain the wanted results. On the basis of this, the program should be made, so that from the first beginning the ordinator should be able to assimilate the complete statistical part of the matter and also be able to execute the wanted operations and print the intended results. For a historical study the procedure, outlined here, remains an almost unattainable ideal. Due to the very nature of quantitative historical research and due to the complex form of appearance of the sources, such a working-method will turn out to be less workable, if not impossible, in most cases. Hence the necessity to conceive the compiling of information as an intermediate phase, in which a favourable basic document can be drawn up by means of an adaptation process.

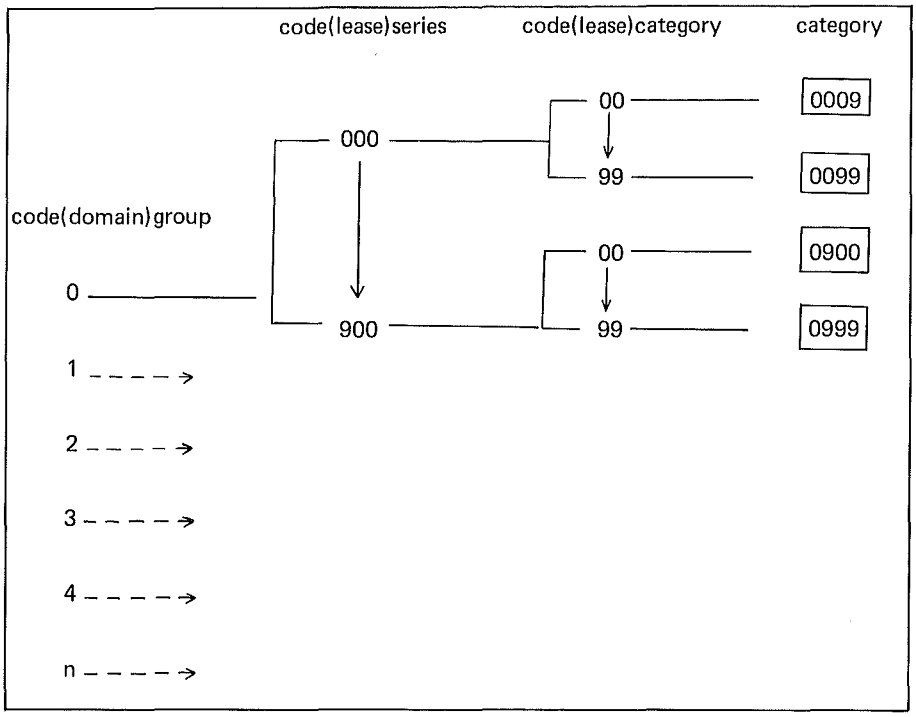
The transcription of the non-classified information

The transcription must be functional, because it will form the starting-point for the further mechanographical treatment. In theory it was sufficient to lay out chronologically arranged forms for all receipts and expenditures.
On these lists all data have to be taken over item by item. Moreover, the possibility of insertion of new items, if necessary, should be allowed for. The building-up of these forms has been kept rather simple. A detailed definition of the item dealt with is given at the top of the form; on the left the calculation years are mentioned and on the right the figure material.
In the middle, between those two data, one finds all the qualitative information, such as : the identity of the tenants or the identity of persons or institutions that made a certain payment, or vice versa, that received a certain amount.
In this way a very extensive collection of items is available some of which had disappeared for years and some of which only turned up after one century (4).
It is obvious that, what we attained along these lines, was only a better ranged version of the contents of the domanial accounts.

Codification

Codification is a necessity for each mechanographic treatment. For one has to accumulate a maximum of information by means of a minimum of data. In order to allow for the working-out, the latter should generally be presented in a numerical or binary form. In the classical punched-card system, codification is mainly an element for sorting or selection.
Though both functions are of great importance to the ordinator, codification in this instance is mainly a basis for certain logical choices by the ordinator. A well appointed codification system reduces the

work of the rent-roll series from the domanial accounts. A codification method can schematically be presented as follows :



Mechanographic questionnaire

As a mechanographic model we used the I.B.M. Fortran Coding Form. It should be stressed that, when filling up this list, the indicative data (here : the code of the leaseholders) come first. For indication always contains a constant factor, on which different variable elements can be made conditional.
The transcription of data is performed by means of either the transformed source or the basic document and with consultation of the codification system. Once all elements have been compiled, one turns to the input and after serial classification of the information one can turn to the realization period, as reproduced under point -02- in the scheme.

CONCLUSION

This set-out procedure, though estimated generally, is only justified for *one* concrete field of application. For, each quantitative historical study has its own analytical problems and there is no standard architecture at all.
That is why we believe that the historian has to go in for *team-work;* he should make an effort in order to get thoroughly acquainted with the techniques of the new auxiliary science, especially that which concerns the methods of analysis. If he does, he will be able to use, in his rational way, better manageable and financially more profitable processes. First we would like to mention the *tele-processing system,* which has extensively reduced the distance between the user and the data-processing centre : there it is possible to bring computer capacity to the exact point where it is needed. Further the time sharing-system permits many users at remote terminals to use the system simultaneously.

(1) This study is part of a Research Project on Mental Ill and Family Care in Geel, organised by Columbia University (New York) and by the University of Louvain, under the direction of Prof. Dr. L. Srole and Dr. J. Schrijvers.

(2) J. JOHNSTON, *Econometric Methods,* New York, Mc Graw-Hill, 1960.

(3) For the review of some of these results we refer to Van der Wee, H. - Van Cauwenberghe, E., *L'utilisation de l'ordinateur pour l'étude des domaines royaux aux Pays-Bas (XIV-XVIIIe siècles).* Communication of the Vth International Congress of Economic History, 10-14-VIII, 1970, Leningrad.

(4) These lists were printed according to this pattern :

Evaluation of the item :		receipts : expenditure :		
Calculation years	Qualitative information :	Amount :		
1329-1330 ↓ 1603-1604		lb.	s.	d.
Source :	Remarks :			