Journal of Econometrics 22 (1983) 1–12. North-Holland Publishing Company

EDITORS' INTRODUCTION

1. Introduction

To an ever increasing extent, econometricians do what psychometricians have been doing since the early days of their science: analyze large data sets. As a result, the developments in psychometrics are becoming more and more of interest to econometricians. The aim of this issue of the *Annals of Applied Econometrics* is to highlight eclectically a number of recent developments in psychometrics that are of actual or potential relevance in econometrics. In selecting the subjects, we have made no attempt at being very precise in defining what does or does not belong to psychometrics; some contributions might better fit under the heading of 'multivariate analysis' or just 'statistics', and at least some of the authors would certainly not call themselves psychometricians — be it as it is, we preferred to be led by the 'relevance' criterion without bothering unduly about deliminational problems.

The idea of looking towards psychometrics is of course not new. In his 1971 *Psychometrika* paper, entitled 'Econometrics and psychometrics: A survey of communalities', Arthur Goldberger described a number of themes shared by both sciences, and in his Schultz lecture in the same year, published in *Econometrica* in 1972, he indicated the relevance to econometricians of path analysis models and, more generally, structural equation models with latent variables. Due, to a large extent, to these efforts, latent variable modelling has received a lot of interest in econometrics over the past decade, reviving and generalizing the classical 'errors-in-variables'-model that econometricians are familiar with and that, due to its inherent identification problem, has until recently been considered to constitute a hopeless problem, as is readily confirmed by a glance in almost any econometrics textbook. Yet, embedding an equation containing error-ridden or essentially unobservable variables in a multiple-equation or simultaneous equations context has now become common practice among econometricians.

Apart from the theme of *latent variables*, which appear in almost all contributions to this issue, another important theme dominates throughout, to wit that of *discrete* or *categorical variables*. The link between both themes is evident: a discrete variable can often be thought of as the manifestation of

an underlying latent, continuous variable. Yet there is an important difference in the econometric and the psychometric tradition: whereas the latent variables 'revival' in econometrics was strongly inspired by psychometrics, the developments in the field of discrete variables have taken place almost independently in both sciences, the distinction roughly being that econometrics has emphasized choice behavior modelling, and psychometrics the analysis of large multi-way frequency tables. As a result, most econometricians will be unfamiliar with techniques like correspondence analysis, loglinear models and multidimensional scaling, methods that aim at detecting the main relationships in such tables, and which are popular among psychometricians.

Although these methods, by their 'exploratory' background, run the risk of being considered by econometricians with suspicion and disdain, it seems nevertheless worthwhile to present them here (they are discussed in several papers), as the practice in econometrics is not always so 'confirmatory' after all, and there is no reason why they can not be fruitfully integrated in economic modelling and estimation. We hope that this issue contributes to such integration.

In the remainder of this Introduction, we briefly sketch the contents of the issue, and the interrelation between the various papers.

2. Developments in linear structural models

In the mid-seventies the basic models of path analysis (from biology and sociology), factor analysis (from psychology), and simultaneous equation theory (from economics) were merged into a single comprehensive model, mainly through the efforts of Jöreskog and Goldberger. And, perhaps even more importantly, computer programs were written which made it possible to fit and test additional parametric specifications within the basic model. In his contribution to this issue Bentler presents an up-to-date review of current developments, starting with Jöreskog's important computer implementation, known as LISREL.

For ease of reference we briefly summarize the basic model here, using the convention of printing random variables and vectors in bold type. The *structural part* of the model is

$$B\boldsymbol{\eta} = \Gamma\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{1}$$

with $\boldsymbol{\eta}$ the *endogenous* variables, $\boldsymbol{\xi}$ the *exogenous* variables, and $\boldsymbol{\zeta}$ the *disturbances* or *shocks* or *errors-in-equations*. None of these three sets of variables need to be observed or even observable; they are linked to the observed variables $\boldsymbol{x}$ and $\boldsymbol{y}$ by the *measurement part* of the model. This is

$$x = \Lambda_x \xi + \delta, \tag{2a}$$

$$y = \Lambda_y \eta + \varepsilon, \tag{2b}$$

with δ and ε the *measurement errors* or *errors-in-variables*. The general model is a long way from being identified. Some steps in the direction of identification are taken by imposing the *general specifications* that ξ, ζ, δ, ε are all uncorrelated and in deviations from the mean. Moreover B is non-singular. These assumptions make it possible to express the covariance matrix of the observed variables in terms of the parameter matrices occurring in (1)–(2) and in terms of the covariance matrices of ξ, ζ, δ, ε, which contain additional parameters. The parameters can be either free, restricted to be equal to given constants, or restricted to be equal to other parameters. If the *additional specifications* are sufficient for identification of the model, then we can proceed and estimate the parameters by fitting the expected covariance matrix to the observed covariance matrix. Fitting is done by some *minimum distance method*, i.e., we define a distance-like measure on the space of covariance matrices and minimize the distance between expected and observed. Distance can be defined by ordinary unweighted least squares, by multinormal weighted least squares, or by multinormal maximum likelihood. The computer algorithm used in LISREL is the Davidon–Fletcher–Powell variable metric method. The appropriateness of the model is tested and the asymptotic dispersion of the estimates is computed by making the additional assumption that the data are a random sample from a multivariate normal population.

This is a brief outline of the LISREL-system of Jöreskog, which consists of a model, an algorithm, and a computer program. Bentler points out that the usual econometric, psychometric, and sociometric models are special cases of the general model (1)–(2), and that they can consequently be fitted with the LISREL-program. Development of ad-hoc estimation methods and ad-hoc large sample theory for specific models is not necessary any more, because it is given directly by the general system. This does not mean, of course, that the system cannot be improved. Indeed Bentler discusses various possibilities for improvement. He reformulates the general model in such a way that it becomes more convenient to apply in some situations. In his EQS-system he also incorporates modelling of the expected values (or the moments around the origin). And, perhaps most importantly, he discusses alternative measures of distance between the observed and expected covariance matrix with superior computational and/or statistical properties. It is clear that the statistical component of Jöreskog's LISREL-system is its weakest part. All statistical statements suppose that the data are a random sample from a multivariate normal distribution, an assumption which is both difficult to test

and almost certainly a poor approximation in most econometric and psychometric situations. Bentler presents results of Browne which make it possible to incorporate asymptotically distribution-free estimation into the linear structural equation context. These estimates, and their properties, are also discussed in this issue by De Leeuw.

More or less independently from the work of Jöreskog another system which incorporates latent variables in simultaneous equation models has been developed by Hermann Wold. His work is discussed in this volume by Dijkstra, who compares Wold's PLS-system in detail with the LISREL-system. It turns out that the differences are mainly on the statistical and computational level. The model Wold uses is a special case of (1)–(2), in which the matrices Λ_x and Λ_y are the direct sums of a number of vectors. [Remember that the direct sum of $n_i \times m_i$ matrices A_i is a $(\sum n_i) \times (\sum m_i)$ matrix with the A_i as diagonal blocks and zeroes elsewhere.] Thus, for example, the m endogeneous variables ξ_j partition the observed variables x into m subsets of $k_1, \ldots, k_m$ variables each. Element (i, j) of Λ_x is non-zero if and only if observed variable x_i is in the subset corresponding to (is an indicator of) latent variable ξ_j.

Statistically the main difference between PLS and LISREL is that Wold does not assume multivariate normality, which is one reason why he calls his models *soft*. The softer assumptions are the linearity of the conditional expectations. In PLS the structural parameters are estimated by *partial least squares*. We start with 'estimates' or *proxies* for the latent variables ξ and η. Given these proxies we can compute estimates of the structural parameters by using simple linear regression. Then, given these estimates, new proxies are constructed. The new proxies are linear combinations of the corresponding indicators, with weights determined by the current estimates of the structural parameters. The two partial least squares steps (improve estimates given proxies and improve proxies given estimates) are repeated until convergence. Dijkstra studies the resulting algorithms and estimates in detail, and compares them with the corresponding LISREL-estimates. His discussion of the consistency of PLS-estimates of the structural parameters is especially interesting, because it shows how the problem of factor score indeterminacy familiar from psychometrics causes inconsistencies also in this context. It would also be interesting in this context to connect the PLS-methods discussed by Dijkstra with the EM-algorithm, briefly discussed in this issue by Muthén, Bartholomew, and De Leeuw, because the EM-algorithm computes maximum likelihood estimates by using least squares methods (and consequently does give consistent estimates).

3. Developments in exploratory multivariate analysis

Factor analysis can be interpreted as a structural errors-in-variables model,

which fits into the general framework of the previous section. Factor analysis, however, is confused continually with principal component analysis, which is basically an exploratory technique. In this context exploratory has several meanings, all of them rather vague. The most important aspects of exploratory data analysis seem to be that no definite statistical model is imposed and tested. The usual statistical optimality considerations consequently cannot be applied. Moreover there is a heavy emphasis on graphical techniques. In a sense the soft models of Wold's PLS-system are closer to exploratory multivariate analysis than the hard, confirmatory models of the LISREL- or EQS-systems.

The most popular exploratory multivariate analysis technique is principal component analysis, which has as its principal aim a graphical representation of the data matrix. Row-objects (often individuals) and column-objects (often variables) are presented as points in a low-dimensional Euclidean space. This is done in such a way that row-objects i and k are close in the representation if they have similar values on all column-objects, and column objects j and l are close if they have similar values on all row-objects. Many variations of this basic idea are possible; some of them go under the name of principal component analysis, others are called multidimensional scaling techniques. In all of them a data matrix is transformed into a picture, which is supposed to portray the most important relationships in the matrix.

A particular form of principal component analysis, which has recently become quite popular, is *correspondence analysis*. It is discussed in this issue by Deville and Saporta, but it also occurs in the contributions of Heiser and Meulman, Keller and Wansbeek, Fienberg and Meyer, and De Leeuw. Deville and Saporta summarize some of the important work done in France on this class of techniques. Originally, correspondence analysis was developed to construct graphical representations of contingency tables. It has been extended from this basically bivariate situation to multivariate situations, and even to time series problems in which there is an infinite number of variables, Deville and Saporta discuss the relationships with classical principal component analysis, using the idea of optimal scaling. In another contribution to this issue Heiser and Meulman present correspondence analysis as a particular form of multidimensional scaling. They also discuss other graphical scaling techniques, such as unfolding and restricted scaling.

At first it may seem as if the graphical exploratory techniques discussed in this section and the confirmatory techniques in the previous section are not related at all. A more precise comparison is possible by using a more formal definition. An important class of multidimensional scaling techniques can be interpreted as fitting the model

$$x_{ij} \sim f(a_i, b_j). \tag{3}$$

Here the x_{ij} are elements of the data matrix, the a_i are p-vectors which

represent the row-objects in p-space, and the b_j are p-vectors which represent the column-objects in p-space. The function f defines the *geometric model*; it is usually distance in Euclidean space, but it can also be cosine of angle or inner product. There is no explicit stochastic structure. The algorithms, which can be very complicated indeed, minimize some distance-type measure between the observed and the reconstructed data matrix.

Both multidimensional scaling and principal component analysis have been generalized in such a way that they can deal with categorical data (with *nominal* and *ordinal* variables). Heiser and Meulman discuss these forms of *non-metric* multidimensional scaling. In addition to (3) it is specified that

$$x_{ij} = g(x_{ij}^*), \tag{4}$$

where the x_{ij}^* are now the observed data, and g is a transformation or quantification of the observed data. Instead of having a single transformation g it is also possible to have a separate transformation for each row or a separate transformation for each column. The transformations are usually chosen *optimally*, i.e., the criterion that was minimized over representations a_i and b_j is now minimized in addition over transformations g, with the restriction that $g \in G$, a class of admissible transformations. The class G can be the class of all monotone transformations or the class of all quadratic polynomials, or whatever.

It is now possible to compare (1) with (3) in the case that the row-objects are individuals and the column-objects are variables. Because (3) introduces parameters for individuals it is a *functional model*. In the usual statistical context in which we increase the number of observations for a fixed number of variables, the b_j are *structural* parameters and the a_i are *incidental* parameters. Specification (4) can be compared with the measurement model (2); instead of an additive stochastic perturbation it provides for a nonlinear deterministic relation of the observed and latent variable. Clearly model (1)–(2) is much more specific than model (3)–(4).

4. Developments in categorical data analysis

We have seen in the discussion of Bentler's contribution to this issue that multivariate normality is the exception rather than the rule in social science situations. In the linear structural model context this has led to the construction of asymptotically distribution-free estimates; it is also part of the motivation for the soft models of Wold. Exploratory multivariate analysis has developed partly as a reaction to the confirmatory forms of multivariate analysis expounded in textbooks, which did not seem applicable in most social science situations. But recently it has become more clear that the assumption of continuous multivariate normality is not only dubious because

of the normality, but also because of the continuity. After all, all variables are categorical, because of the limited precision with which we measure them. In physics, and often in economics, there are plenty of variables which can be approximated very nicely by continuous models. In psychology and sociology, however, and often in econometrics, variables classify individuals into a small number of categories, and they are inherently *discrete* or *caregorical.*

In statistics the study of categorical variables and their association was started by Yule. His theory has developed, through the seminal work of Goodman and Kruskal on measures of association, into general loglinear analysis, which is discussed in this issue by Fienberg and Meyer. Loglinear analysis decomposes a discrete multivariate distribution (a multidimensional contingency table) in the same way as a conventional fixed-effects analysis of variance decomposes a higher-way layout. There are main effects, interactions of various orders, just as in the analysis of variance. The analysis is called loglinear, because we decompose the logarithms of the observed frequencies. Taking logarithms is useful both for statistical purposes and for ease of interpretation. Of course the assumptions of the analysis of variance (normal independent errors in each cell, with equal variance) can no longer be true. The statistical procedures for loglinear analysis use multinomial large-sample theory. Loglinear analysis has turned out to be an extremely useful class of techniques to analyze multivariate contingency tables. It is used mainly in sociology, under the influence of Leo Goodman, and it has been incorporated in computer-systems such as ECTA, BMDP, GLIM. Fienberg and Meyer discuss the basic saturated (just-identified) model for a multivariate table, and the various additional overidentified submodels that are possible. They discuss the elegant maximum likelihood theory for hierarchical loglinear models, and the equally elegant iterative proportional fitting procedure. Their discussion covers the more classical linear logit models for binary dependent variables.

Psychometricians had to deal with categorical variables right from the beginning. In psychometric test theory the answers to items in the test were either wrong or correct, which means that all variables are binary. In the early days this problem was evaded by computing subtest scores, which were then used in ordinary linear structural models such as factor analysis, but this solution is now considered to be unsatisfactory unless preceded by a detailed item analysis of the binary variables themselves. The model used for these purposes is the *latent structure model*, which is discussed in this issue in the contributions of Bartholomew, Andersen, De Leeuw, and Fienberg and Meyer. The basic idea in latent structure theory is that there are latent variables and observed variables, and that the observed variables are conditionally independent given the latent variables. This last assumption is very powerful; it generalizes the measurement model (2) in an elegant way.

Bartholomew shows that various interesting special cases arise by combining continuous/discrete latent/observed variables. Discrete latent together with discrete observed variables constitute Lazarsfeld's latent class model, which has been fitted neatly into the general loglinear model by Goodman. Continuous latent together with continuous observed defines factor analysis, at least if we assume in addition that regressions of observed on latent are linear. This is closely related to the soft model assumptions used by Wold. In psychometrics, however, models with continuous latent variables and discrete observed variables were especially interesting.

In his contribution to this issue Andersen discusses latent trait models, with one single continuous latent variable and binary observed variables. A typical latent trait model is of the form

$$\text{prob}(x_j = 1 \mid \xi = \xi) = \Phi(\xi - \theta_j), \tag{5}$$

with ξ the latent variable and Φ the *item characteristic*, which is usually either the standard normal or the logistic distribution function. By using conditional independence and by integrating over the latent variable we derive a model for the observed 2^m-contingency table. More often, however, the model (5) is used in its functional form. We use

$$\text{prob}(x_{ij} = 1) = \Phi(\xi_i - \theta_j). \tag{6}$$

If Φ is logistic, then (6) defines the Rasch model, which is discussed extensively by Andersen. Fienberg and Meyer also discuss the Rasch model, and the related BTL-model for paired comparisons, and give various ways to integrate it into the general loglinear model. Andersen gives a detailed discussion of the various statistical techniques used for fitting the Rasch model and similar latent trait models.

5. Relationships

In the previous three sections we have distinguished three content areas in which there have been important developments. Developments in linear structural equation modelling are covered in this issue by Bentler and Dijkstra, developments in exploratory multivariate analysis by Deville and Saporta and by Heiser and Meulman, and developments in categorical data analysis by Andersen and by Fienberg and Meyer. The remaining four papers in this issue deal mainly with relationships between these three content areas.

We have seen that the statistical component in the LISREL-system is its weakest part: social science data are not multivariate normal, in fact they often are not even continuous. Muthén presents a drastic modification of the

statistical component, which makes it much more realistic, and which at the same time brings it close to recent developments in categorical data analysis. In addition to the usual specifications (1)–(2) Muthén assumes that

$$x_i^* = \alpha_i(x_i), \tag{7a}$$

$$y_j^* = \beta_j(y_j). \tag{7b}$$

In (7) the x_i, defined by (1)–(2), are not observed any more, they are the *latent response variables*. The same thing is true for the y_j. The latent response variables are related to the observed variables x_i^* and y_j^* by (7a) and (7b), in which the α_i and β_j are non-decreasing step functions if the observed variables are ordered categorical and are identities if the observed variables are numerical, in which case they coincide with the latent response variables. In Muthén's system the latent variables are assumed to be multivariate normal, which means that the observed variables are either normally distributed or are discretizations of normally distributed variables.

This specification makes Muthén's work a far-reaching generalization of Karl Pearson's system of tetrachoric, polychoric, biserial, and polyserial correlation coefficients. Pearson had to abandon his work because of the unsurmountable computational difficulties associated with estimation of the parameters. Muthén reviews a number of techniques which are quite practical if there are not too many variables. The very same model, at least the measurement part (7), is also discussed this issue by De Leeuw. He proposes some alternative statistical methods, which could be more profitable. He also shows that the latent trait model (5), with Φ the cumulative standard normal, is equivalent to the factor analysis model with one common factor in Muthén's system. Bartholomew discusses the factor analysis model with q independent common factors, in which Φ is the logistic. He gives a very useful approximation to the cross-product ratios, which can be used to construct approximate fitting methods that work even on very large data sets. It is clear from the papers of Muthén, De Leeuw, and Bartholomew that the fields of linear structural equation models and categorical data analysis are being integrated rapidly, although it is unlikely that a single all-embracing model will arise in the end.

Keller and Wansbeek's contribution to this issue integrates various forms of exploratory multivariate analysis with linear functional equation models. They start out with

$$x_{ij} = \sum_{s=1}^{q} a_{is}\, b_{js} + \zeta_{ij}, \tag{8}$$

which is clearly a specialization of (3). By introducing stochastic structure in

a specific way the symbol $\sim$ in (3) gets a definite meaning. The function f in (3) is specialized to the inner product. It is assumed in addition that the vectors ζ_i are independent, identically distributed, multinormal, with a known covariance matrix. Additional specifications on the form of this covariance matrix lead to various familiar regression and component type models. But Keller and Wansbeek do not stop here. They extend their approach in such a way that it can deal also with categorical variables. Suppose the observed variables x_{ij}^* are categorical, variable j having k_j possible values. We code the categorical variables by using *dummies*, i.e., binary vectors of length k_j in which exactly one element is equal to one. The location of the one in the vectors indicates the value of the variable. Concatenate the dummies in an $n \times (\sum k_j)$ matrix $\boldsymbol{X}^*$, and add to (8) the specification

$$\boldsymbol{X} = \boldsymbol{X}^* W, \tag{9}$$

with W an $(\sum k_j) \times m$ matrix of *scale values* (or *weights* or *quantifications*). W must be the direct sum of m vectors w_j, of length k_j. Eq. (9) is a specialization of (4). Of course the combination of (8) and (9) must be considered an approximate model for categorical variables, because (8) says that $\boldsymbol{X}$ is continuous multinormal and (9) says that $\boldsymbol{X}$ is discrete. Keller and Wansbeek give an interpretation of the model they are approximating in which the multinormal density is approximated on a discrete grid of points. The same model is briefly treated by Fienberg and Meyer, who relate it to earlier work of Goodman, and by De Leeuw, who calls it the point multinormal model to contrast it with the block multinormal model based on (7). Both Keller and Wansbeek and Fienberg and Meyer point out that the point multinormal model is related to correspondence analysis. In De Leeuw's contribution a final model called the regression multinormal model is discussed and compared with block and point multinormal models. It is based on Lancaster's work in the linear (not loglinear) analysis of multidimensional contingency tables, and it turns out to be even more closely related to correspondence analysis.

In summary it is clear that the relationships discussed in this section are related to properties of the multinormal distribution that are being generalized to categorical data. In the block multinormal model the idea of discretization of latent response variables remains very close to the multinormal tradition. The corresponding 2×2 association measure is Pearson's tetrachoric correlation. For the point multinormal model the idea that is generalized is the simple product structure of the bivariate interactions, and the sufficiency of the bivariate marginals. The corresponding association measure is the cross-product ratio. Bartholomew's logit model also uses the cross-product ratio, but only as an approximation. The regression multinormal model, finally, generalizes the linearity of the

regressions. The corresponding association measure is the maximal correlation, which degenerates to the phi coefficient or point correlation in the 2×2 case. Much research remains to be done on the properties of these models, and on their appropriateness in the applications, but it is clear that a reconciliation of Pearson and Yule and possibly of exploratory and confirmatory multivariate analysis is in the making.

6. Concluding remarks

In this introduction, the ten papers in this issue have all been considered from several points of view. Keeping in line with the spirit of this issue, one may like to have some kind of overall view of the interrelation of the various papers and try to derive some simple representation thereof.

In order to do so, we constructed a matrix of 'distances' between the papers, in the following simple manner: we counted, for each two papers, the number of references that they had in common, and grouped the result in a 10×10 matrix; the diagonal elements were set equal to the number of references in a paper that occurred in at least one other paper. (This choice for the diagonal elements is motivated by the large differences in number of references per paper — if the number *sec* would be taken, undesired distortions in the representation would come up.) This matrix was next subjected to a correspondence analysis. The result is given in fig. 1.[1]

Fig. 1 represents the 10×10 matrix according to its two main axes. According to the first (horizontal) axis, two papers are separated from the rest: Heiser and Meulman, and Deville and Saporta. These papers stand apart due to their purely exploratory character. The second (vertical) axis can be interpreted as giving an ordering from discrete (top) to continuous (bottom); the papers by Andersen and Bartholomew focus on zero–one variables (and are indeed closely related), whereas the papers by Bentler and Dijkstra are all about continuous variables models. The papers by Muthén and Keller and Wansbeek take an intermediate position, in that the former extends the LISREL-model for continuous variables to ordered categorical ones, and the latter gives a unified setup, for two general classes of linear models, for both continuous and discrete variables. With this figure in hand, the reader is invited to choose his own route along the papers in this issue.

Finally, we would like to thank the following people: Dennis Aigner for his enthusiasm when we proposed the project leading to this issue; the authors, who stood up remarkably well to a number of sharp deadlines; our Dutch colleagues Herman Bierens, Henk Don, Abby Israëls, Arie Kapteyn, Teun Kloek, Peter Kooiman, Peter Nijkamp, Franz Palm, Jeroen Pannekoek, Piet Rietveld, Dirk Sikkel, Henk Stronkhorst, Herman van Dijk, Wynand van de

[1]We are grateful to Anco Hundepool, Richard Jager and Dirk Sikkel for performing the counting and programming work for this analysis.

Ven and Albert Verbeek, who provided expert opinion in reviewing the papers, and Wanda Hendriksz and Sandra Ikkersheim for their secretarial assistance.

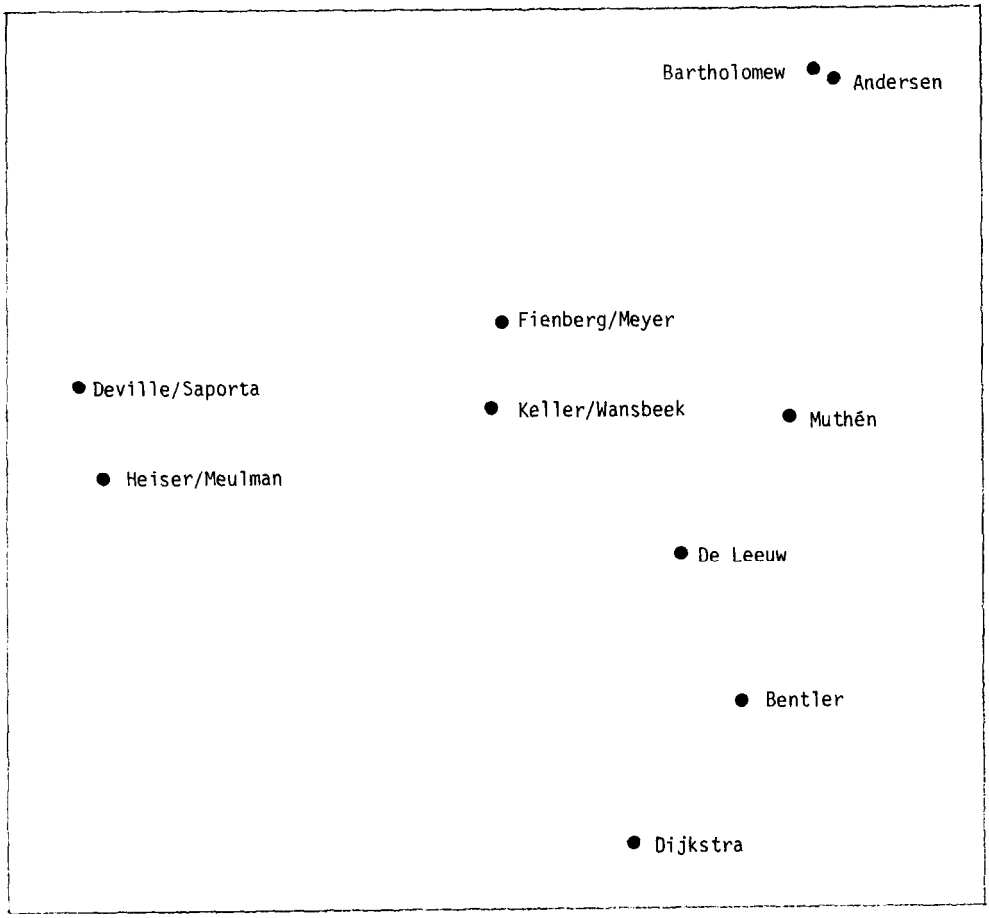

Fig. 1. Two-dimensional representation of distances between papers.

Jan DE LEEUW

University of Leyden, Leyden, The Netherlands

Wouter J. KELLER and Tom WANSBEEK

Netherlands Central Bureau of Statistics, Voorburg, The Netherlands