Psychometrika

CONTENTS

VOLUME SIXTY-EIGHT **JUNE 2003** **NUMBER 2**

MANUSCRIPTS RECENTLY ACCEPTED FOR PUBLICATION

Listed in the Order They will be Published

Building an Identifiable Latent Class Model with Covariate Effects on Underlying and Measured Variables
Guan-Hua Huang and Karen Bandeen-Roche

Reliability as a Function of the Number of Item Options Derived from the "Knowledge or Random Guessing" Model
Robert MacCann

Exact Distributions of Intraclass Correlation and Cronbach's Alpha with Gaussian Data and General Covariance
Emily O. Kistner and Keith E. Muller

Interval Estimation of Gamma for a R $\times$ S Table
Kung-Jong Lui and William G. Cumberland

Generalized Multilevel Structural Equation Modeling
Sophia Rabe-Hesketh, Anders Skrondal, and Andrew Pickles

Remarks on "Equivalent Linear Logistic Test Models" by Bechger, Verstralen, and Verhelst (2002)
Gerhard H. Fischer

Equivalent Linear Logistic Test Models: A Rejoinder
Timo M. Bechger, Huub, H.H.F.M., Verstralen, Norman, D. Verhelst, and Gunter Maris

Locally Dependent Latent Trait Model for Polytomous Responses with Application to Inventory of Hostility
Edward H. Ip, Yuchung J. Wang, Paul de Boeck, and Michel Meulders

Asymptotic Biases in Exploratory Factor Analysis and Structural Equation Modeling
Haruhiko Ogasawara

Global and Local Covert Visual Attention: Evidence From A Bayesian Hidden Markov Model
John Liechty, Rik Pieters, and Michel Wedel

Rotation to Simple Loadings using Component Loss Functions: The Orthogonal Case
Robert I. Jennrich

Analysis of Distractor Difficulty in Multiple-Choice Items
Javier Revuelta

PSYCHOMETRIKA—VOL. 68, NO. 2, 169–191
JUNE 2003

BAYESIAN MODELING OF MEASUREMENT ERROR IN PREDICTOR VARIABLES USING ITEM RESPONSE THEORY

JEAN-PAUL FOX AND CEES A.W. GLAS

UNIVERSITY OF TWENTE

It is shown that measurement error in predictor variables can be modeled using item response theory (IRT). The predictor variables, that may be defined at any level of an hierarchical regression model, are treated as latent variables. The normal ogive model is used to describe the relation between the latent variables and dichotomous observed variables, which may be responses to tests or questionnaires. It will be shown that the multilevel model with measurement error in the observed predictor variables can be estimated in a Bayesian framework using Gibbs sampling. In this article, handling measurement error via the normal ogive model is compared with alternative approaches using the classical true score model. Examples using real data are given.

Key words: classical test theory, Gibbs sampler, item response theory, hierarchical linear models, Markov Chain Monte Carlo, measurement error, multilevel model, multilevel IRT, two-parameter normal ogive model.

Introduction

In many research areas, and especially in social sciences, studies may involve variables that cannot be observed directly or are observed with error. Further, many forms of human response behavior are inherently stochastic in nature. In the sequel, all these types of variation will be categorized under the heading measurement error. In this context, Lord and Novick (1968, chap. 2) adhere to the so-called stochastic subject view in which it is assumed that responses of the subjects depend on small variations in the circumstances in which the response is generated. Accordingly, response variance is the variation in responses to the same question repeatedly administered to the same person. The use of unreliable explanatory variables leads to biased estimation of the regression coefficients and the resulting statistical inference can be very misleading unless careful adjustments are made (see, e.g., Carroll, Ruppert, & Stefanski, 1995; Cook & Campbell, 1979; Fuller, 1987).

Models developed to account for measurement error in regression models are commonly known as measurement error models. The enormous amount of literature on measurement error in linear regression is summarized by Fuller (1987). Generally, measurement error is handled by the classical additive measurement error model. An example is the classical test theory model used in educational measurement. Goldstein (1995) extended some of the techniques to handle measurement errors in the independent variables in linear models to the multilevel model.

In the present paper, attention is focused on an alternative way of handling response error in the independent variables using an item response theory (IRT) model. This has several advantages. First, measurement error is defined conditionally on the value of the latent ability. That is, measurement error can be defined locally, for instance, as the posterior variance of the ability parameter given a response pattern. This local definition of measurement error results in heteroscedasticity. In the Rasch model, for instance, the posterior variance of the ability parameter given an extreme score is greater than the posterior variance of the ability parameter given an intermediate score (see, for instance, Hoijtink & Boomsma, 1995, pp. 59, Table 4.1). Second, IRT

This paper is part of the dissertation by Fox (2001) that won the 2002 Psychometric Society Dissertation Award.
Requests for reprints should be send to Jean-Paul Fox, Department of Educational Measurement and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, THE NETHERLANDS. E-Mail: Fox@edte.utwente.nl

0033-3123/2003-2/2000-0804-A $00.75/0

PSYCHOMETRIKA—VOL. 68, NO. 2, 193–211
JUNE 2003

ASYMPTOTIC STANDARD ERRORS OF IRT OBSERVED-SCORE EQUATING METHODS

HARUHIKO OGASAWARA

OTARU UNIVERSITY OF COMMERCE

A method of the IRT observed-score equating using chain equating through a third test without equating coefficients is presented with the assumption of the three-parameter logistic model. The asymptotic standard errors of the equated scores by this method are obtained using the results given by M. Liou and P.E. Cheng. The asymptotic standard errors of the IRT observed-score equating method using a synthetic examinee group with equating coefficients, which is a currently used method, are also provided. Numerical examples show that the standard errors by these observed-score equating methods are similar to those by the corresponding true score equating methods except in the range of low scores.

Key words: IRT, observed-score equating, equipercentile equating, chain equating, asymptotic standard errors.

In item response theory (IRT) we have two types of equating methods. The first type is concerned with equating of the underlying metrics for two or more tests via the item parameters or the examinee proficiency scores. The reason for equating the metrics stems from the well-known arbitrariness of the location and scale units. The methods using the moments of the estimates of item parameters such as the mean/mean method (Loyd & Hoover, 1980) and the method using estimated test characteristic curves (Stocking & Lord, 1983) have been developed to equate item and ability parameters. The second type of equating methods (the IRT true- and observed-score equating methods) is associated with raw scores of tests such as number-correct scores. Even after we have equated item parameters, we often have the practical necessity of equating raw scores of different tests, since actual decisions are usually made about the raw or scaled scores in e.g., professional certification programs. The IRT true- and observed-score equating methods equate the raw scores with IRT models (see, e.g., Lord, 1980, 1982a; Kolen & Brennan, 1995, chap. 6). Though the true score is different from the observed score, the true-score equating is often used as an approximation to the observed-score equating because of its simplicity.

When a third test (anchor test) is available, the true-score equating can be carried out without equating item parameters. The standard errors of the equated scores by the method using an external anchor test (i.e., the anchor test which is not a subtest of the two tests to be equated) were given by Lord (1982b). The standard errors for the true-score equating methods in the case of an internal anchor test with or without IRT equating coefficients have been given by Ogasawara (2001b). The IRT true-score equating seems to give results similar to those by the IRT observed-score equating (see, e.g., Han, Kolen & Pohlmann, 1997; Lord, 1977; Lord & Wingersky, 1984) while somewhat different results have been given elsewhere (e.g., Kolen, 1981; Tsai, Hanson, Kolen & Forsyth 2001). The similar results reported are concerned with similar equated scores while the different results show different stabilities. It is to be noted that the IRT true- and observed-score equating methods have variations (e.g., with or without equating coefficients, separate or simultaneous estimation of item parameters with respect to examinee groups, different methods of estimating equating coefficients). Tsai et al.'s (2001) recent comparison study covers typical variations of the two methods using the three-parameter logistic model. They showed

The author is indebted to Michael J. Kolen for access to the real data used in this article and anonymous reviewers for their corrections and suggestions on this work.

Requests for reprints should be sent to Haruhiko Ogasawara, Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501 JAPAN. E-Mail: hogasa@res.otaru-uc.ac.jp

0033-3123/2003-2/2000-0744-A $00.75/0

PSYCHOMETRIKA—VOL. 68, NO. 2, 213–228
JUNE 2003

A PERSON-FIT INDEX FOR POLYTOMOUS RASCH MODELS, LATENT CLASS MODELS, AND THEIR MIXTURE GENERALIZATIONS

MATTHIAS VON DAVIER

EDUCATIONAL TESTING SERVICE

IVO W. MOLENAAR

UNIVERSITY OF GRONINGEN

A normally distributed person-fit index is proposed for detecting aberrant response patterns in latent class models and mixture distribution IRT models for dichotomous and polytomous data.

This article extends previous work on the null distribution of person-fit indices for the dichotomous Rasch model to a number of models for categorical data. A comparison of two different approaches to handle the skewness of the person-fit index distribution is included.

Key words: person-fit, polytomous Rasch model, mixed Rasch model, latent class analysis, appropriateness measurement.

Introduction: Assessing Person-Fit in Complex Psychometric Models

The aim of this paper is to propose an extension of the work of Molenaar and Hoijtink (1990) on person-fit indices to mixture distribution IRT models for polytomous categorical data.

Starting out with an extension of Molenaar and Hoijtink (1990) results to the polytomous Rasch model in the next section, this paper will show how to apply the results to increasingly complex models. A person-fit index is developed for the Latent Class Analysis (LCA; see Lazarsfeld & Henry, 1968). This index can also be used for mixture distribution generalizations of the Rasch model (Rost, 1990; von Davier & Rost, 1995) and applies also to Hybrid models (Yamamoto, 1987) and other discrete mixture IRT models.

The final section includes a demonstration of the distributional properties of the proposed index and a comparison of two different approaches to approximate the distribution of person-fit indices as proposed by Molenaar and Hoijtink (1990) and Bedrick (1997).

A Person-Fit Index for Polytomous Models

The person-fit index of Drasgow, Levine and Williams (1985) can be used both for dichotomous and polytomous models. This so-called appropriateness index is based on logarithmized response probabilities. For IRT models, it is defined as

$$\mathrm{L}(\mathbf{x} \mid \theta) = \sum_{i=1}^{k} \ln P_i(X = x_i \mid \theta), \tag{1}$$

which is the log-likelihood of the response vector $\mathbf{x} = (x_1, \ldots, x_k)$ given trait level θ. In order to use the appropriateness index $\mathrm{L}(\mathbf{x} \mid \theta)$ for comparing subjects, information about its distribution

Major parts of this paper were written while the first author worked at the Institute for Science Education, Kiel, Germany. Any opinions expressed in this paper are those of the authors and not necessarily of Educational Testing Service. The results presented in this paper were improved by valuable comments from J. Rost, K. Yamamoto, N.D. Verhelst, E. Bedrick and two anonymous reviewers.

Requests for reprints should be sent to Matthias von Davier, ETS, MS O2-R, Princeton NJ 08541. E-Mail: mvondavier@ets.org

0033-3123/2003-2/1998-0661-A $00.75/0

PSYCHOMETRIKA—VOL. 68, NO. 2, 229–249
JUNE 2003

DISCREPANCY RISK MODEL SELECTION TEST THEORY FOR COMPARING POSSIBLY MISSPECIFIED OR NONNESTED MODELS

R.M. GOLDEN

APPLIED COGNITION AND NEUROSCIENCE PROGRAM
UNIVERSITY OF TEXAS AT DALLAS

A new model selection statistical test is proposed for testing the null hypothesis that two probability models equally effectively fit the underlying data generating process (DGP). The new model selection test, called the Discrepancy Risk Model Selection Test (DRMST), extends previous work (see Vuong, 1989) on this problem in four distinct ways. First, generalized goodness-of-fit measures (which include log-likelihood functions) can be used. Second, unlike the classical likelihood ratio test, the models are not required to be fully nested where the nesting concept is defined for generalized goodness-of-fit measures. The DRMST also differs from the likelihood ratio test by not requiring that either competing model provides a completely accurate representation of the DGP. And, fourth, the DRMST may be used to compare competing time-series models using correlated observations as well as data consisting of independent and identically distributed observations.

Key words: asymptotic statistical theory, model selection, hypothesis-testing, model misspecification, time-series, m-estimation

Considerable research in the field of model selection has focussed upon the Model Selection Criterion (MSC) problem where multiple competing descriptions of some Data Generating Process (DGP) are compared (e.g., Akaike, 1973; Balasubramanian, 1997; Bozdogan, 1987; Clarke & Barron, 1990; Djuric, 1998; Kass & Wasserman, 1995; Linhart & Zuchhini, 1986; Myung, Forster, & Browne, 2000; Qian & Kunsch, 1998; Rissanen, 1996; Schwarz, 1978). In this paradigm, some goodness-of-fit measure is used to estimate the (true) expected goodness-of-fit of each *model* (i.e., a family of probability distributions) to the underlying DGP. The model (or models) which has (have) the smallest *estimated* goodness-of-fit is (are) then selected.

This paper focuses attention upon a closely related and equally important problem called the Model Selection Test (MST) problem which has received relatively less attention in the literature. In this paradigm, the standard error of the difference in the relative fits of the two models is estimated and used to test the null hypothesis that both models provide equally effective fits to the underlying DGP at a chosen significance level.

An important early approach to the MST problem is the Generalized Likelihood Ratio Test (GLRT) which is now widely used (Wilks, 1938). The essential strategy for applying the GRLT is to compute maximum likelihood estimates of the model's parameters using the observed data. The goodness-of-fit of the model is then assessed by using the computed maximum likelihood estimates to calculate the observed likelihood of the data given the "full" model. It is also assumed that the DGP is a probability distribution contained in the full model. Some subset of the model's parameters are then usually set equal to a constant (such as zero) and another maximum likelihood estimation procedure is then used to estimate the free parameters of the "restricted"

This research was supported in part by a University of Texas at Dallas Special Faculty Development Award and the NSF Information Technology Research Initiative through the Research on Learning and Education Program Award 0113369 to the author at the University of Texas at Dallas. This research was also supported by the National Institute on Alcohol Abuse and Alcoholism grant to Martingale Research Corporation (Grant No. R44AA11607). The author thanks Halbert White, Steven Henley, and the UTD A^2N^2 research group for discussions regarding issues related to this manuscript. The author also thanks three anonymous reviewers for feedback and comments regarding a previous version of this paper.

Requests for reprints should be sent to Richard M. Golden, Applied Cognition and Neuroscience Program (GR4.1), University of Texas at Dallas, Box 830688, Richardson, TX 75083-0688. E-Mail: golden@utdallas.edu

0033-3123/2003-2/2000-0822-A $00.75/0

PSYCHOMETRIKA—VOL. 68, NO. 2, 251–265
JUNE 2003

USING RESPONSE TIMES TO DETECT ABERRANT RESPONSES IN COMPUTERIZED ADAPTIVE TESTING

WIM J. VAN DER LINDEN AND EDITH M.L.A. VAN KRIMPEN-STOOP

UNIVERSITY OF TWENTE

A lognormal model for response times is used to check response times for aberrances in examinee behavior on computerized adaptive tests. Both classical procedures and Bayesian posterior predictive checks are presented. For a fixed examinee, responses and response times are independent; checks based on response times offer thus information independent of the results of checks on response patterns. Empirical examples of the use of classical and Bayesian checks for detecting two different types of aberrances in response times are presented. The detection rates for the Bayesian checks outperformed those for the classical checks, but at the cost of higher false-alarm rates. A guideline for the choice between the two types of checks is offered.

Key words: aberrant response patterns, computerized adaptive testing, posterior predictive checks, person misfit, residual analysis, response times.

Though the primary use of response vectors in testing is to construct accurate ability estimates, they also contain useful information to detect possible aberrances in examinee behavior. Most statistical analyses to identify such behavior belong to the class of statistical procedures known as residual analysis. That is, they are based on the residuals of an examinee's response vector left after a model known to explain the responses of a population of regular examinees has been fitted. The first step in this residual analysis is to check for examinees with unexpected behavior. A more challenging second step is to diagnose their response vector for specific types of aberrances. Ideally, the analysis would support the hypothesis of one type of aberrance and exclude competing hypotheses. Papers with seminal techniques for this type of analysis are Bradlow, Weiss, and Cho (1998), Drasgow, Levine and Williams (1985), Levine and Rubin (1979), Molenaar and Hoijtink (1990), and Trabin and Weiss (1983). A review of the literature is given in Meijer and Sijtsma (1995).

The introduction of computerized adaptive testing (CAT) has increased the necessity of checks on examinee behavior. For example, a new type of aberrance is response behavior due to preknowledge of some of the items in the pool. To return investments, item pools in CAT have to remain operational for some time. Particularly in high-stakes testing programs, examinees may try to use this time to memorize and share items in the pool. Another potential new source of aberrant behavior is due to differential speededness of the test. Adaptive tests are selected to have optimal information at the ability level of the examinee. However, items differ not only in their information but also in the amount of time they require. As a consequence, some CAT examinees operate under higher time pressure than others. A recent study revealed that high-ability examinees may suffer especially from this type of speededness. For those examinees, item selection CAT results in more difficult items, and more difficult items generally require more time (van der Linden, Scrams & Schnipke, 1999).

Also, residual analysis of response vectors has difficulty maintaining its power when applied to adaptive tests. One reason is that adaptive tests are typically much shorter than paper-and-

This study received funding from the Law School Admission Council (LSAC). The opinions and conclusions contained in this paper are those of the authors and do not necessarily reflect the policy and position of LSAC. The authors are most indebted to Wim M. M. Tielen for his computational assistance and to the US Defense Manpower Data Center for the permission to use the ASVAB data set in the empirical examples.

Requests for reprints should be sent to W.J. van der Linden, Department of Measurement Research Methodology and Data Analysis, University of Twente, P.O. Box 217, 7500 AE Enschede, THE NETHERLANDS. E-Mail: w.j.vanderlinden@utwente.nl

0033-3123/2003-2/2001-0874-A $00.75/0

PSYCHOMETRIKA—VOL. 68, NO. 2, 267–287
JUNE 2003

MULTILEVEL LOGISTIC REGRESSION FOR POLYTOMOUS DATA AND RANKINGS

ANDERS SKRONDAL

DIVISION OF EPIDEMIOLOGY
NORWEGIAN INSTITUTE OF PUBLIC HEALTH, OSLO

SOPHIA RABE-HESKETH

DEPARTMENT OF BIOSTATISTICS AND COMPUTING
INSTITUTE OF PSYCHIATRY, LONDON

We propose a unifying framework for multilevel modeling of polytomous data and rankings, accommodating dependence induced by factor and/or random coefficient structures at different levels. The framework subsumes a wide range of models proposed in disparate methodological literatures. Partial and tied rankings, alternative specific explanatory variables and alternative sets varying across units are handled. The problem of identification is addressed. We develop an estimation and prediction methodology for the model framework which is implemented in the generally available gllamm software. The methodology is applied to party choice and rankings from the 1987–1992 panel of the British Election Study. Three levels are considered: elections, voters and constituencies.

Key words: multilevel models, generalized linear latent and mixed models, factor models, random coefficient models, polytomous data, rankings, first choice, discrete choice, permutations, nominal data, gllamm.

Introduction

Two kinds of nominal outcome variable are considered in this article; unordered polytomous variables and permutations. The outcome for a unit in the unordered polytomous case is one among several objects, whereas the outcome in the permutation case is a particular ordering of objects. The objects are nominal in the sense that they do not possess an inherent ordering shared by all units as is assumed for ordinal variables. The development in this paper is coaxed in the terminology of decision theory. Hence, objects are henceforth denoted *alternatives*, unordered polytomous variables denoted *first choices* and permutations denoted *rankings*. For instance, in election studies a central outcome variable is the first choice of a voter (say Conservatives) among a set of alternatives (say Labour, Conservatives and Liberals). Sometimes additional information is obtained in the form of rankings of the alternatives (say Liberals preferred to Labour preferred to Conservatives).

The standard statistical model for first choices and rankings is logistic regression. It has been pointed out in the econometric and psychometric literature that these models involve a questionable independence assumption known as "Independence from Irrelevant Alternatives" (IIA). Discussions of the ramifications of IIA have been confined to one-level designs. However, first choice and ranking data are often of a multilevel nature where units are nested within clusters. In the context of the election example, multilevel data could for instance arise from two-level designs where voters are nested within constituencies or election occasions nested within vot-

Parts of this work were completed while Anders Skrondal visited the Biostatistics Group at The University of Manchester, UK. gllamm and the script for the analyses in this article can be downloaded from: http://www.iop.kcl.ac.uk/IoP/Departments/BioComp/programs/gllamm.html. Requests for reprints should be sent to Anders Skrondal, Division of Epidemiology, Norwegian Institute of Public Health, P.O. Box 4404 Nydalen, N-0403 Oslo, NORWAY. E-Mail: anders.skrondal@fhi.no

0033-3123/2003-2/2001-0913-A $00.75/0

PSYCHOMETRIKA—VOL. 68, NO. 2, 289–298
JUNE 2003

A TWO-STAGE LOGISTIC REGRESSION MODEL FOR ANALYZING INTER-RATER AGREEMENT

STUART R. LIPSITZ

DEPARTMENT OF BIOMETRY AND EPIDEMIOLOGY
MEDICAL UNIVERSITY OF SOUTH CAROLINA, CHARLESTON

MICHAEL PARZEN

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO

GARRETT M. FITZMAURICE

DEPARTMENT OF BIOSTATISTICS
HARVARD SCHOOL OF PUBLIC HEALTH

NEIL KLAR

DIVISION OF PREVENTIVE ONCOLOGY
CANCER CARE ONTARIO

Studies of agreement commonly occur in psychiatric research. For example, researchers are often interested in the agreement among radiologists in their review of brain scans of elderly patients with dementia or in the agreement among multiple informant reports of psychopathology in children. In this paper, we consider the agreement between two raters when rating a dichotomous outcome (e.g., presence or absence of psychopathology). In particular, we consider logistic regression models that allow agreement to depend on both rater- and subject-level covariates. Logistic regression has been proposed as a simple method for identifying covariates that are predictive of agreement (Coughlin et al., 1992). However, this approach is problematic since it does not take account of agreement due to chance alone. As a result, a spurious association between the probability (or odds) of agreement and a covariate could arise due entirely to chance agreement. That is, if the prevalence of the dichotomous outcome varies among subgroups of the population, then covariates that identify the subgroups may appear to be predictive of agreement. In this paper we propose a modification to the standard logistic regression model in order to take proper account of chance agreement. An attractive feature of the proposed method is that it can be easily implemented using existing statistical software for logistic regression. The proposed method is motivated by data from the Connecticut Child Study (Zahner et al., 1992) on the agreement among parent and teacher reports of psychopathology in children. In this study, parents and teachers provide dichotomous assessments of a child's psychopathology and it is of interest to examine whether agreement among the parent and teacher reports is related to the age and gender of the child and to the time elapsed between parent and teacher assessments of the child.

Key words: binary data, generalized estimating equations, kappa, odds ratio.

Introduction

Agreement studies are increasingly common in psychiatric research. For example, researchers are often interested in the agreement among radiologists in their review of brain scans

The authors thank the Associate Editor and the referees for their helpful comments and suggestions. We also thank Gwen Zahner for use of data from the Connecticut Child Study, which was conducted under contract to the Connecticut Department of Children and Youth Services. This research was supported by grants HL 69800, AHRQ 10871, HL52329, HL61769, GM 29745, MH 54693 and MH 17119 from the National Institutes of Health.

Requests for reprints should be sent to Garrett Fitzmaurice, Department of Biostatistics, Building II, 655 Huntington Ave., Rm. 423, Boston MA 02115. E-Mail: fitzmaur@hsph.harvard.edu

0033-3123/2003-2/2000-0846-A $00.75/0

PSYCHOMETRIKA—VOL. 68, NO. 2, 299–321
JUNE 2003

OBLIQUE FACTORS AND COMPONENTS WITH INDEPENDENT CLUSTERS

HARUHIKO OGASAWARA

OTARU UNIVERSITY OF COMMERCE

Relationships between the results of factor analysis and component analysis are derived when oblique factors have independent clusters with equal variances of unique factors. The factor loadings are analytically shown to be smaller than the corresponding component loadings while the factor correlations are shown to be greater than the corresponding component correlations. The condition for the inequality of the factor/component contributions is derived in the case with different variances for unique factors. Further, the asymptotic standard errors of parameter estimates are obtained for a simplified model with the assumption of multivariate normality, which shows that the component loading estimate is more stable than the corresponding factor loading estimate.

Key words: oblique factors, component loadings, orthoblique rotation, standard errors, independent clusters, sphericity.

Introduction

Comparative studies on factor analysis (FA) and component analysis (CA) have been made since the 1970s by several researchers. Among others Velicer and his colleagues gave various works on the topic (Fava & Velicer, 1992; Velicer, 1977; Velicer & Fava, 1987, 1998; Velicer, Peacock, & Jackson, 1982). Velicer and Jackson (1990) reviewed associated problems extensively and were relatively in favor of CA over FA (see also the comments and reply following the article). Borgatta, Kercher and Stull (1986), Hubbard and Allen (1987), and Snook and Gorsuch (1989) emphasized the differences between the results given by FA and CA while Wilkinson (1989a, 1989b) had opposite views (see also Borgatta, 1989; Hubbard & Allen, 1989). These discussions are mainly based on empirical and simulated data.

On the other hand we have analytical results for the similarities and differences of FA and CA. Bentler and Kano (1990) showed the equivalence of a common factor in a one-factor model and the corresponding principal component when the number of observed variables becomes infinite. Sato (1990, 1992) gave analytical relationships between the factor and component loadings in the one-factor model with a finite number of observed variables. Sato (1992) also derived the similar result as given by Bentler and Kano. Schneeweiss and Mathes (1995) gave various relationships for FA and CA when we have multiple common factors in situations similar to those in Bentler and Kano. Schneeweiss (1997) treated the case with near sphericity (i.e., near equal variances of unique factors; see also Hayashi & Bentler, 2000). Ogasawara (2000a, 2000b) gave the asymptotic correlations between the parameter estimates in FA and CA.

The purpose of this study is to derive some relationships between oblique factors and corresponding components. It is to be noted that in practice factors and components are frequently rotated by oblique methods such as the promax method. So far, the comparative studies of FA and CA with analytical results have not sufficiently dealt with the cases of oblique factors and components though Schneeweiss and his colleague's works implicitly involve oblique factors and components because they treated the canonical correlations between factors and components. Ogasawara (2000a, 2000b) treated the asymptotic correlations between the parameter estimates for obliquely rotated factors and components. Rozeboom (1990, p. 64) emphasized the importance of the comparison of factor correlations and component correlations in oblique factors and components. Widaman (1993) derived the reduction of component correlations in comparison

Request for reprints should be sent to Haruhiko Ogasawara, Department of Information and Management Science, Otaru University of Commerce, 3-5-21, Midori, Otaru 047-8501 Japan. E-Mail: hogasa@res.otaru-uc.ac.jp

0033-3123/2003-2/2000-0868-A $00.75/0

PSYCHOMETRIKA—VOL. 68, NO. 2, 323–326
JUNE 2003

REVIEW

K. Sijtsma & I.W. Molenaar. *Introduction to nonparametric item response modeling* (Measurement methods for the social sciences, Vol. 5). Thousand Oaks, CA: Sage Publications, 168 pp., 2002, Cloth $69.95 ISBN 0–7619–0812–9, Paperback $31.95 ISBN 0–7619–0813–7

General Remarks

Item response theory (IRT) deals with the modeling of response probabilities for person-item combinations. In Sage's *Measurement Methods for the Social Sciences* (MMSS) series, Hambleton, Swaminathan and Rogers (1991) wrote about the fundamentals of IRT, mainly focusing on parametric models. As an overview of the theory and practice of nonparametric IRT modeling (NIRT), the volume of Sijtsma and Molenaar is a welcome supplement to this series. Parametric models, like the well-known Rasch model, imply assumptions about the form of item response curves (probability functions about item responses conditional on respondents' value of some latent trait θ). Nonparametric models are less demanding on the specific form of those response curves, but not quite so on the efforts needed to make a reliable item scale.

The treated models in the book basically originate from a scale analysis model for dichotomous items, described by Mokken (1971) in his doctoral thesis. The authors cover the development of nonparametric item response modeling since then, with a major emphasis on the monotone homogeneity model (MHM) and the double monotonicity model (DMM). Conditional on their statistical adequacy, those models enable an ordering of persons, and persons as well as items, respectively. Based on such (reliable and valid) orderings, practical decisions can be made regarding personnel selection, social policy evaluation, assessment of deviant behavior, school admission, significance of group differences, and so on. The two key models use the unweighted total item score for rank ordering the persons on a latent trait θ. The main objective of using a specific NIRT probability model is to select those items that satisfy the assumptions of the model, that is, to define a proper item scale with the ultimate goal to order the persons on some latent trait of interest. Compared to parametric IRT modeling no numeric, interval-scaled estimates of the θ's can be obtained; the best available proxy would be the sum of the item scores, which scale is *ordinal* only.

The authors' intention is to teach researchers how to use nonparametric item response theory—how to analyze available sample data, and how to build substantive tests and questionnaires, that is. Although the authors claim that the dominant angle is practical, the contents of the chapters constitute a fine balance between theory and application: in principle, chapters handling item response theory are followed by chapters in which the foregoing theory is illustrated by real data analysis. Throughout the book, the software being used in the applications is Mokken Scale Analysis for Polytomous Items (MSP), developed by Molenaar and Sijtsma (2000). It is beyond this review to evaluate this program (under update now), which manual harbors numerous examples.

Besides serving as a reference book for researchers, the book is well suited as a textbook for a course on NIRT with practical exercises using the MSP software. In a one-semester course the full contents of the book could be covered—an ideal prospect. As a prerequisite, students should have a basic course in statistics, they should also be familiar with and have working knowledge of (conditional) probability, discrete distributions like the binomial, expectation and covariance operators, and the concept of stochastic independence. In addition, for a full grasp of the theory, it

0033-3123/2003-2/2003-0001-R $00.75/0

Behaviormetrika: An official journal of The Behaviormetric Society of Japan

Chief Editor: Yutaka Kano, kano@hus.osaka-u.ac.jp
Volume 28 Number 2 July 2001

CONTENTS

PDF files can be downloaded at
http://wwwsoc.nii.ac.jp/bsj/behaviormetrika_index.html

Psychometrika
A JOURNAL OF QUANTITATIVE PSYCHOLOGY

PSYCHOMETRIC SOCIETY MEMBERSHIP APPLICATION

☐ New Member ☐ Renewal

Name: ____________________

Affiliation: ____________________

E-Mail: ____________________ Fax: ____________________

Address where journal should be sent: ____________________

____

____

____

Payment Method (✓one): ☐ Check enclosed♣ ☐ Visa™ or ☐ Mastercard™

Card#: ____________________

Expiration Date: ________/________

Signature: ____________________

♣Make check payable to the Psychometric Society.

Type of Membership (✓one): ☐ Regular Member ($60 USD)

☐ Family Member ($75 USD) ☐ Student Member* ($25 USD)

*Advisor's name: ____________________ *Advisor's Signature: ____________________

Return form with payment to: PSYCHOMETRIC SOCIETY
Carol Earey
Center for Educational Research and Evaluation
University of North Carolina at Greensboro
210 Curry Building
P.O. Box 26171
Greensboro, NC 27402-6171 USA

E-Mail: cearey@aol.com
Society website: http://www.psychometricsociety.org
Journal website: http://www.psychometrika.org

Join online at http://www.psychometricsociety.org

Psychometrika

Psychometrika, the official journal of the Psychometric Society, is devoted to the development of psychology as a quantitative rational science. Issued four times a year, on March 15, June 15, September 15, and December 15.

MARCH 2003, VOLUME 68, NUMBER 1

(ISSN 0033-3123) is published by the Psychometric Society, 326 Cox Statistics Building, Research Triangle Institute, P.O. Box 12194, Research Triangle Park, NC 27709-2194 U.S.A. and at additional mailing offices. Editorial office: *Psychometrika*, University of North Carolina at Greensboro, P.O. Box 26171, Greensboro, NC 27402-6171 U.S.A (e-mail: Pmetrika@s.psych.uiuc.edu). Managing Editor: P.O. Box 7104, San Jose, CA 95150-7104 U.S.A. (e-mail: Pmetrika@nullset.org).

Subscription Price: The regular subscription rate is $125.00 per volume. All annual subscriptions start with the March issue and cover the calendar year. All bills and orders are payable in advance. Checks should be made payable to the *Psychometric Society*. All checks, notices of change of address, and business communications (including notification of the nonreceipt of journals) should be sent to the Treasurer. Membership in the Psychometric Society includes a subscription to *Psychometrika*. Regular members pay annual dues of $60.00, except that family members pay annual dues of $75.00. The annual dues for a student in a recognized college or university is $25.00.

Application for membership and student membership in the Psychometric Society, together with a check for dues for the calendar year in which application is made, should be sent to the Treasurer.

Back Issues: Back issues for the previous 6 volumes only are available from the Treasurer at the address below. Persons claiming replacement for issues lost in the mail must do so with the Treasurer within one year. Microfilm copies of all volumes are available from Johnson Associates, Inc., P.O. Box 1017, Greenwich, CT, 06830, and Bell & Howell Information and Learning, P.O. Box 1346, Ann Arbor, MI 48106-1346 (http://wwwlib.umi.com/sim/). *Psychometrika* is abstracted in Statistical Theory and Method Abstracts (STMA).

LORI D. MCLEOD, Treasurer
Psychometric Society
326 Cox Statistics Building
Research Triangle Institute
P.O. Box 12194
Research Triangle Park, NC 27709-2194 USA

(Continued on the back inside cover page)

Psychometrika contains articles on the development of quantitative models for psychological phenomena, and on quantitative methodology in the social and behavioral sciences, including new mathematical and statistical techniques for the evaluation of psychological data, and the application of such techniques. Empirical studies will be considered only if they involve new or particularly interesting uses of quantitative techniques.

Submission of a paper to this journal will be taken to imply that it represents original work not previously published, and that it is not being considered elsewhere for publication. Authors submitting a manuscript do so on the understanding that if it is accepted for publication, copyright in the article, including the right to reproduce the article in all forms and media, shall be assigned exclusively to the Publisher. The Publisher will not refuse any reasonable request by the author for permission to reproduce any of his or her contributions to the journal.

The editor removes, as far as possible, the author's identity from manuscripts before sending them for review.

Paper submissions must include five copies of a manuscript. (Authors wishing to make an electronic submission should refer to the submission information posted on the Journal's web site.) Prospective authors are referred to the "Instructions to *Psychometrika* Authors" contained in the September 1994 issue. Reprints are available from the Editor upon request. A manuscript which fails to comply with these requirements will be returned to the author for revision. Authors will receive 100 reprints without covers, free of charge.

Manuscripts for publication in *Psychometrika* should be sent to:

Ulf Böckenholt, Editor
Psychometrika
Univ. of N.C. at Greensboro
210 Curry Building
P.O. Box 26171
Greensboro, NC 27402

Books for review in *Psychometrika* should be sent to:

Laura M. Koehly, Review Editor
Psychometrika
Department of Psychology
Texas A&M University
College Station, TX 77843-4235

Electronic Addresses:

Ulf Böckenholt, Editor: Pmetrika@uncg.edu
Cynthia H. Null, Managing Editor: Pmetrika@nullset.org
Tim Null, Technical Editor: Pmetrika@nullset.org
Laura M. Koehly, Review Editor: lmk@psyc.tamu.edu
24 Hour Fax/Voice Mail: 1 (888) 467-4101
Journal web site: http://www.psychometrika.org
Psychometric Society web site: http://www.psychometricsociety.org

The Psychometric Society

PRINTED BY CADMUS PROFESSIONAL COMMUNICATION—BYRD PRESS DIVISION
COMPOSITION BY INTEGRE TECHNICAL PUBLISHING CO., INC.

VOLUME 68
NUMBER 2
JUNE 2003

ISSN 0033-3123

Psychometrika

A JOURNAL OF QUANTITATIVE PSYCHOLOGY

THE PSYCHOMETRIC SOCIETY

JUNE 2003
PSYCHOMETRIKA
VOLUME 68 NO. 2 165–332