

# Nonlinear Factor Analysis as a Statistical Method

Ilker Yalcin and Yasuo Amemiya

*Abstract.* Factor analysis and its extensions are widely used in the social and behavioral sciences, and can be considered useful tools for exploration and model fitting in multivariate analysis. Despite its popularity in applications, factor analysis has attracted rather limited attention from statisticians. Three issues, identification ambiguity, heavy reliance on normality, and limitation to linearity, may have contributed to statisticians' lack of interest in factor analysis. In this paper, the statistical contributions to the first two issues are reviewed, and the third issue is addressed in detail. Linear models can be unrealistic even as an approximation in many applications, and often do not fit the data well without increasing the number of factors beyond the level explainable by the subject-matter theory. As an exploratory model, the conventional factor analysis model fails to address nonlinear structure underlying multivariate data. It is argued here that factor analysis does not need to be restricted to linearity and that nonlinear factor analysis can be formulated and carried out as a useful statistical method. In particular, for a general parametric nonlinear factor analysis model, the errors-in-variables parameterization is suggested as a sensible way to formulate the model, and two procedures for model fitting are introduced and described. Tests for the goodness-of-fit of the model are also proposed. The procedures are studied through a simulation study. An example from personality testing is used to illustrate the issues and the methods.

*Key words and phrases:* Latent variable modeling, multivariate data analysis, identification, test of fit.

## 1. LINEAR AND NONLINEAR FACTOR ANALYSIS

Factor analysis and more general structural equation modeling are used very widely in the applied sciences. Gnanadesikan and Kettenring (1984) conducted a comprehensive survey of literature in a variety of fields to assess the role and use of multivariate techniques in applications. They reported that, in five of eight surveyed fields, factor analysis

was the most widely used multivariate method, and that, in education, psychology, and sociology, factor analysis was applied six to eight times more often than the second most popular multivariate method, discriminant analysis. Note that this survey was conducted more than 15 years ago, before the explosion in the use of structural equation modeling in social science research. Factor and structural equation analyses utilize latent variable models which express observed measurements or indicators in terms of underlying unobservable characteristics or traits. Such models seem to correspond very well with the subject-matter theory in the social and behavioral sciences, where unobservable but well-conceived characteristics such as attitude, personality, and opinion are to be studied, and where items or questions designed to measure or relate to such characteristics can be constructed. In those areas, the wide use of latent variable statistical techniques has in turn had some impact on the

---

*Ilker Yalcin is with Eli Lilly and Company, Lilly Corporate Center, Bldg. 58/2, Drop Code 0721, Indianapolis, Indiana 46285. Yasuo Amemiya is Professor, Department of Statistics, Iowa State University, Ames, Iowa 50011, and is Research Staff Member, IBM T. J. Watson Research Center, Box 218, Yorktown Heights, New York 10598 (e-mail: yasuo@us.ibm.com).*

way the subject-matter concepts and questions are formulated quantitatively.

The idea behind latent variable modeling is useful also in multivariate exploratory data analysis, where some lower-dimensional structure of high-dimensional data may be of interest, but may not be directly observable due to errors. Nevertheless, only a handful of statistical researchers have worked on the development of factor analysis and related methods. Some statisticians seem to be skeptical about the usefulness and appropriateness of factor analysis in applications. As a result, the coverage of factor analysis and latent variable modeling in graduate statistics education tends to be minimal, and often does not reflect the current practice.

For a  $p \times 1$  observation  $\mathbf{Z}_t$ , from the  $t$ th individual, the standard factor analytic structure can be expressed by a model,

$$(1) \quad \mathbf{Z}_t = \mathbf{g}^*(\mathbf{f}_t^*) + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, n,$$

where  $\mathbf{f}_t^*$  is a  $k \times 1$  underlying unobservable factor vector, and  $\boldsymbol{\varepsilon}_t$  is a  $p \times 1$  unobservable error vector with mean zero. Here,  $\mathbf{g}^*(\mathbf{f}_t^*)$  is a  $p$ -variate function of  $\mathbf{f}_t^*$ , representing the fact that the true value (or systematic part) of the  $p$ -dimensional observation  $\mathbf{Z}_t$  lies on some  $k$ -dimensional surface ( $k < p$ ). The linear factor analysis model is a special case of (1) with

$$(2) \quad \mathbf{g}^*(\mathbf{f}_t^*) = \boldsymbol{\mu}^* + \boldsymbol{\Lambda}^* \mathbf{f}_t^*,$$

where  $p \times 1$   $\boldsymbol{\mu}^*$  and  $p \times k$   $\boldsymbol{\Lambda}^*$  are unknown parameters. A basic factor analysis assumption is that all inter-relationships among observed variables are explained by the underlying common factors but not by the errors. Thus, the  $\mathbf{f}_t^*$ 's and  $\boldsymbol{\varepsilon}_t$ 's are assumed to be independent, and the  $p$  components of  $\boldsymbol{\varepsilon}_t$  are also assumed to be independent. Hence,  $\text{Cov}\{\mathbf{f}_t^*, \boldsymbol{\varepsilon}_t\} = \mathbf{0}$ , and the covariance matrix  $\boldsymbol{\Psi}$  of  $\boldsymbol{\varepsilon}_t$  is diagonal. It is also assumed that no relationship exists among the distributions of the components of  $\boldsymbol{\varepsilon}_t$ , so that, for example, the  $p$  variances in  $\boldsymbol{\Psi}$  are generally unrelated. This assumption guarantees that the structure specified by model (1) is equivariant under a scale or measurement unit change of any of the  $p$  components of the observation  $\mathbf{Z}_t$ . Some practitioners may think that some of the models used in practice do not satisfy the independence assumptions for  $\mathbf{f}_t^*$  and the components of  $\boldsymbol{\varepsilon}_t$ . For example, the linear model (2) with a nondiagonal error covariance matrix  $\boldsymbol{\Psi}$  (and with sufficient conditions on  $\boldsymbol{\Lambda}$  to identify parameters) is occasionally used. However, such correlated error structure can also be written in the form (1) or (2) with a diagonal  $\boldsymbol{\Psi}$  by taking a view that any dependency is due to a common

underlying factor and by introducing additional factors accounting for the correlation between errors. The general factor analysis model (1) provides a very unrestrictive and natural way to represent a multivariate structure suggested by subject-matter theory or conjectured by exploratory data analysis.

From a technical perspective, there are several aspects of factor analysis that statisticians have traditionally considered as deficiencies, and that may have contributed to their skepticism and lack of interest. Three of these aspects are identification problems, reliance on normality, and limitation to linearity.

The identification problem in factor analysis refers to the fact that a particular lower-dimensional structure in model (1) may be uniquely specified but can be expressed using many equivalent parameterizations with different  $\mathbf{f}_t^*$  and  $\mathbf{g}_t^*$  of a similar form. The linear factor analysis model (2) can be written

$$(3) \quad \begin{aligned} \mathbf{Z} &= \boldsymbol{\mu}^* - \boldsymbol{\Lambda}^* \mathbf{C}^{-1} \mathbf{d} + \boldsymbol{\Lambda}^* \mathbf{C}^{-1} (\mathbf{C} \mathbf{f}_t^* + \mathbf{d}) + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\mu}^{**} + \boldsymbol{\Lambda}^{**} \mathbf{f}_t^{**} + \boldsymbol{\varepsilon}_t, \end{aligned}$$

and is not identified due to the possibility of a nonsingular linear transformation of factors  $\mathbf{f}_t^{**} = \mathbf{C} \mathbf{f}_t^* + \mathbf{d}$ . For the general nonlinear model (1),  $\mathbf{g}^*(\mathbf{f}_t^*)$  defines a unique  $k$ -dimensional nonlinear surface in the  $p$ -dimensional space, but the same surface may be expressed using other  $\mathbf{g}^*$  and  $\mathbf{f}^*$ . The identification issue for nonlinear models is less straightforward than the linear case. For the linear model (2), the most widely used model fitting procedure is the maximum likelihood approach based on the assumption of normal  $\mathbf{Z}_t$ , that is, normal  $\mathbf{f}_t^*$  and normal  $\boldsymbol{\varepsilon}_t$ . It is a common practice to apply the normality based method to nonnormal or discrete data. This is the problem described as the heavy reliance on normality. For the nonlinear model (1), the normality issue is more complex in the sense that normal observations  $\mathbf{Z}_t$  and nonlinear  $\mathbf{g}^*$  imply nonnormal factors  $\mathbf{f}_t^*$ , and that normal factors  $\mathbf{f}_t^*$  and nonlinear  $\mathbf{g}^*$  imply nonnormal observations  $\mathbf{Z}_t$ . Although there is no inherent reason to restrict  $\mathbf{g}^*$  in (1) to be linear, the models used in practice have been almost exclusively limited to the linear model (2). This limitation has invited criticism from statisticians in view of the fact that most traditional statistical methods have been extended to accommodate some form of nonlinearity.

In this paper, Section 2 reviews statistical development regarding the identification issue and the heavy use of normality-based methods. Sections 3–8 address the problem concerning nonlinearity, and deal with the general parametric nonlinear factor analysis model. After a general discussion in Section 3, our approach to the identification of

the nonlinear model is covered in Section 4. Two new statistical methods for nonlinear model fitting and checking are introduced in Sections 5 and 6. A simulation study is reported in Section 7. An example where nonlinear factor analysis can be usefully applied is described in the remainder of this section and in Section 8. The Appendix presents the computing algorithms for the two methods described in Sections 5 and 6.

For our example, we use a part of a data set obtained in a personality testing study assessing the vocational interest of college students. The *Self-Directed Search* test was designed to measure six vocational personality aspects proposed by theory: investigative, artistic, realistic, social, enterprising and conventional. Dumenci (1993) applied the test to 370 female and 305 male college students with the eventual goal of explaining the vocational interest aspects by their relationships to the personality factors obtained through another personality test based on the so-called big-five theory of personality. For the vocational interest data, he reports that a linear model with six overall factors fits the data very poorly for each gender group, and concludes that the observed test scores are not consistent with the six-aspect theory used in designing the test. As a result, he abandoned the analysis relating the theory-suggested vocational interest aspects to the big-five personality factors. However, it should be pointed out that the *Self-Directed Search* vocational test items were carefully constructed based on the six-aspect theory. In such construction, the developers may have known which items are supposed to be related to each of the six aspects, but they may not have been particularly thinking the relationships to be linear. Thus, by expanding the scope of factor analysis to include nonlinear models, it may still be possible to fit a model consistent with the theory, to use the fitted model to construct estimates of the interpretable six aspects, and address the subject-matter question regarding possible relationships between the well-defined vocational interest aspects and the personality factors.

To simplify and focus our discussion, we consider only the scale-level analysis of a small part of the female data on items which were designed by the test developer to correspond to the investigative type factor. This factor is supposed to be measured by four constructed scales: Activities (e.g., fixing electrical equipment), Competencies (e.g., using a microscope), Occupations (e.g., becoming a biologist), and a Self Estimate (e.g., rating themselves on a particular investigative skill). We will use  $Z_1$ ,  $Z_2$ ,  $Z_3$ , and  $Z_4$ , respectively, to denote the four scales. Each scale score was obtained based on a

large number of items. By design, these four scales are supposed to be four constructs of a common factor representing the investigative aspect. Thus, the subject-matter theory suggests a one-factor factor analysis model as a measurement model for this part of the data. Figure 1 is a jittered scatter-plot matrix of  $Z_1, \dots, Z_4$  which shows a reasonably strong relationship between every pair of the scales along with nontrivial variation in every direction. The linear exploratory (unrestricted) factor analysis model (2) with one factor ( $k = 1$ ) fails to fit the data well. The likelihood ratio goodness-of-fit test statistic based on the assumption of normal observations gives  $\chi^2 = 17.45$  with 2 degrees of freedom ( $p = 0.0002$ ), and the one-factor linear measurement model is rejected. The observed variables  $Z_1, \dots, Z_4$  are discrete and skewed, and the normal assumption is violated. But, as reviewed in the next section, this goodness-of-fit test is valid in large samples for most nonnormal data.

The poor fit of the one-factor linear model for this part of the data may have contributed to the failure of the overall six-factor linear model in explaining the entire data. Note that, for  $p = 4$  observed variables, the (unrestricted) linear factor analysis model (2) can be fitted only with  $k = 1$ , and that two or more factors cannot be fitted. Thus, no linear model fits this data well, because the linear model with the maximum possible number of factors provides a very poor fit. Also, the scatter-plot matrix in Figure 1 suggests that the relationships among  $Z_1, \dots, Z_4$  may not be linear. The nonlinear relationships among observed variables partially explain why the model with linear relationships between the observed variables and a single latent factor is rejected, but does not necessarily mean that the one-factor theory is rejected in general. Recall that these four scores were constructed to be related in some manner to a common "investigative" factor, and that the observed variables show moderately strong monotone positive relationships among themselves. Thus, it may be natural to consider fitting a one-factor model with nonlinear relationships between the observed variables and the factor. This example will be revisited in Section 8 using the nonlinear factor analysis approach developed in Sections 4–7.

## 2. IDENTIFICATION AND NORMALITY IN LINEAR MODELS

In this section, the statistical development and relevant literature for the identification problem and the reliance on normality based methods for linear factor analysis are reviewed. The review

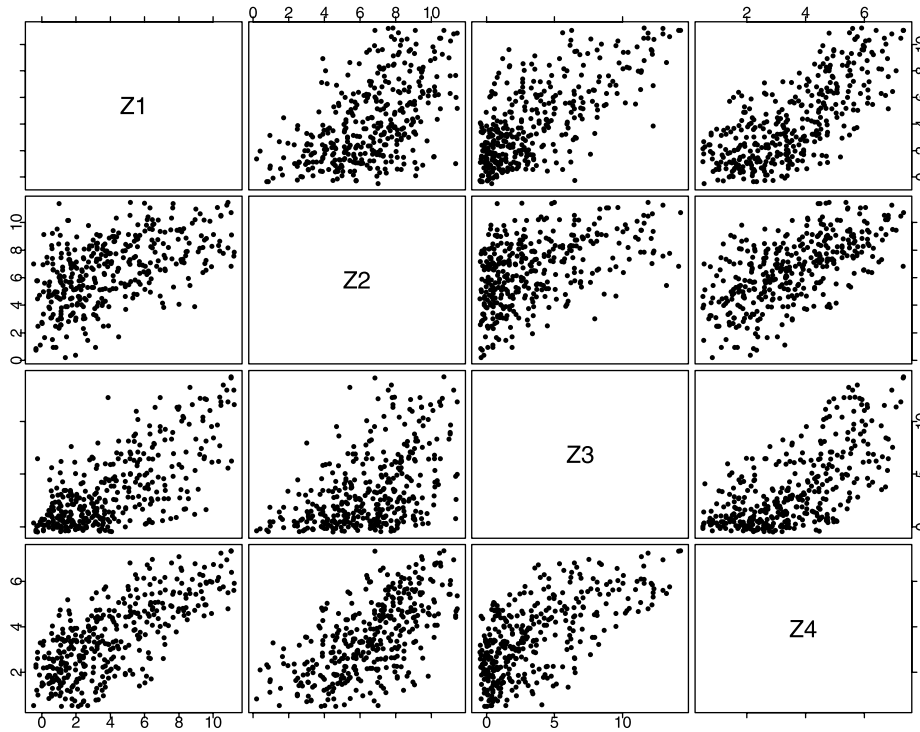


FIG. 1. Scatter-plot matrix of the four scales designed to measure the investigative type personality.

here focuses only on those references directly relevant for our approach to nonlinear factor analysis development, and is not meant to be exhaustive.

The linear model (2) uniquely defines the  $k$ -dimensional linear factor space in the  $p$ -dimensional observation space. But, as indicated in (3), the factor space can be parametrized in various ways using different definitions of  $\mathbf{f}_t^*$ . The traditional parameterization used by psychologists assumes a standardized factor vector  $\mathbf{f}_t^*$  with zero mean and identity covariance matrix. This does not remove the indeterminacy associated with rotation or orthogonal transformation of  $\mathbf{f}_t^*$  using an orthogonal  $\mathbf{C}$  in (3). Although this parameterization was originally introduced for computational ease, the practice of rotation spread widely. The casual use of factor rotation especially in conjunction with interpretation and inference has become a source of statisticians' skepticism and criticism concerning factor analysis. However, in recent years, the use of this parameterization allowing rotation has been largely replaced by another type of parameterization of the linear model (3). Starting with Jöreskog (1970, 1973) and culminating in Jöreskog and Sörbom (1989), structural equation modeling based on the so-called errors-in-variables parameterization has been developed. As the popularity of the structural equation modeling increased, so did the use of the same parameterization for the linear factor analysis model (considered a special case of the

structural equation model). Additional support for the parameterization was provided by those working in the area of errors-in-variables. See, for example, Fuller (1987, Section 4.3). The errors-in-variables parameterization places the identification conditions only on the coefficient matrices, and leaves the distribution of the factor vector unrestricted. The factor vector  $\mathbf{f}_t^*$  is identified as the "true values," and is measured with error by the  $k$  corresponding components of  $\mathbf{Z}_t$ . By placing, without loss of generality, such  $k$  components as the last  $k$  components, we write  $\mathbf{Z}_t = (\mathbf{Y}_t', \mathbf{X}_t')$  for  $(p-k) \times 1$   $\mathbf{Y}_t$  and  $k \times 1$   $\mathbf{X}_t$ , and write (3) in the form

$$(4) \quad \begin{aligned} \mathbf{Y}_t &= \boldsymbol{\lambda} + \boldsymbol{\Lambda} \mathbf{f}_t + \mathbf{e}_t, \\ \mathbf{X}_t &= \mathbf{f}_t + \mathbf{u}_t, \end{aligned} \quad t = 1, 2, \dots, n,$$

where  $\boldsymbol{\varepsilon}_t = (\mathbf{e}_t', \mathbf{u}_t')$ ,  $\mathbf{f}_t$  is the  $k \times 1$  factor identified as the true value of  $\mathbf{X}_t$ , the  $(p-k) \times 1$   $\boldsymbol{\lambda}$  and  $(p-k) \times k$   $\boldsymbol{\Lambda}$  are unknown factor-loading or relationship parameters. In terms of model (3), the identifying restrictions are

$$\boldsymbol{\mu}^* = \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{0} \end{pmatrix}, \quad \boldsymbol{\Lambda}^* = \begin{pmatrix} \boldsymbol{\Lambda} \\ \mathbf{I}_k \end{pmatrix}.$$

The errors-in-variables linear model (4) defines the factor vector unambiguously without possibility of rotation, and allows interpretation based on the equivalence to the errors-in-variables regression

of  $\mathbf{Y}_t$  on  $\mathbf{X}_t$ . The use of this parameterization has largely removed statisticians' criticisms of linear factor analysis associated with factor indeterminacy and rotation, as well as with their effect on interpretation.

For many years, factor analysis methods based on the assumption of normally distributed observations have been applied to various obviously nonnormal data. This practice, along with the limited development of methods designed for nonnormal data, was another source of the criticism of factor analysis. As in many other statistical methods, the initial statistical interest in factor analysis was the development of efficient estimation methods under normality assumptions. However, unlike other traditional statistical methods, this development was conducted over a very long period of time, because the computation of the maximum normal likelihood estimator was not straightforward. Initiated by Lawley (1940, 1941, 1943), the development of a practical algorithm for such an estimator was not completed until Jöreskog (1967, 1969) and Jennrich and Robinson (1969). With the development of computer packages, the normality-based factor analysis became the standard method in applications, regardless of the actual distributional form. As a result, the attention to nonnormality was neglected for some time. Browne (1984) considered an asymptotically distribution free estimation procedure. In the late 1980s and early 1990s, statistical research was conducted to assess the appropriateness of the use of normality-based packages in the factor analysis of nonnormal data without specifying a particular distributional form. Note that if model (3) holds for the nonnormal observation  $\mathbf{Z}_t$ , then at least one of  $\mathbf{f}_t^*$  and  $\boldsymbol{\varepsilon}_t$  must be nonnormal. Amemiya, Pantula, and Fuller (1987) and Browne (1987) showed that, in the errors-in-variables formulation (4), the maximum normal likelihood estimators of unrestricted  $\boldsymbol{\lambda}$ ,  $\boldsymbol{\Lambda}$ , and the error variances have a common limiting covariance matrix for a broad class of random or fixed  $\mathbf{f}_t$  as long as  $\boldsymbol{\varepsilon}_t$  is normal. Partially based on the foundational work of Anderson and Rubin (1956), Anderson and Amemiya (1988) proved that the limiting covariance matrix of the maximum normal likelihood estimator of possibly restricted  $\boldsymbol{\Lambda}$  in (4) is the same for virtually any normal or nonnormal  $\boldsymbol{\varepsilon}_t$  and any normal, nonnormal, or fixed  $\mathbf{f}_t$ , provided  $\mathbf{f}_t$  and the  $p$ -components of  $\boldsymbol{\varepsilon}_t$  are all independent. Amemiya and Anderson (1990) showed that various goodness-of-fit test statistics have a common chi-square limiting distribution under the general linear factor analysis model (3) with any  $\boldsymbol{\varepsilon}_t$  and any random or fixed  $\mathbf{f}_t^*$  given the independence

of  $\mathbf{f}_t^*$  and the  $p$ -components of  $\boldsymbol{\varepsilon}_t$ . Using a different approach Browne and Shapiro (1988) derived related results for a class of linear latent variable models. Subsequently, a series of psychometric papers on the same topic have followed. See, for example, Hu, Bentler and Kano (1992) and Satorra (1992). These results imply that the blind use of the normality-based packages for factor analysis of nonnormal data can produce useful information. In particular, the standard errors and goodness-of-fit test results printed out by the existing packages based on normality are in fact appropriate for a large class of nonnormal data, provided that the errors-in-variables parameterization (4) is used, and that the basic independence assumption given following model (1) is taken as a part of the factor analysis model formulation. In addition to providing further statistical support for the use of the errors-in-variables parameterization, these results have partially dispelled the previously held view that heavy reliance on the normality-based packages tends to produce inappropriate inferences and limits the practical usefulness of factor analysis.

### 3. NONLINEAR FACTOR ANALYSIS

In applications of factor analysis, linear models (2) are used almost exclusively. Statisticians have considered this practice as a limitation of factor analysis in view of the fact that most other traditional statistical methods have been extended to accommodate some form of nonlinearity. The wide and long use of the normality-based linear factor analysis packages contributed to the practice in social and behavioral sciences of trying to explain the subject-matter relationships only through covariances and correlations. The popularity of this practice has in turn resulted in neglect of the possibility of nonlinear relationships and in lack of motivation for development of a nonlinear factor analysis. However, when applied scientists try to represent some subject-matter theory in a latent variable model, they may have a reasonably good idea about the number of underlying factors and their possible monotone relationships to the observations, but they often do not have a definite idea about the actual form of the relationships. The appeal and appropriateness of factor analysis in practice rarely come from a particular model form, but usually lie in the basic principle of distinguishing observed and latent variables and of explaining relationships through a certain number of underlying factors. There are situations where subject knowledge indicates the inappropriateness of linear models and suggests a particular form of

nonlinear model. For example, in aptitude or ability testing, a test with low difficulty tends to discriminate poorly between subjects with high ability, but to discriminate well between those with low ability, implying possible nonlinear relationships between the test scores and abilities. Also, the scores of constructed scales or tests fall in certain ranges, while the underlying abilities, attitudes or personality traits can be interpreted more naturally as unbounded and continuous random variables. Then, the relationship between the observed scores and underlying factors cannot be linear. In these situations, a model allowing nonlinear structures is much more realistic, and corresponds more closely to the subject-matter theory. When the subject-matter theory suggests nonlinear relationships, fitting such nonlinear models is much more desirable than the use of ad hoc linear models.

Additional needs for considering nonlinear factor analysis comes from the practical difficulty, often encountered and pointed out in the behavioral and social sciences, that a linear model with factors suggested by the subject-matter theory frequently does not fit the data well and a good fit can be obtained only with a larger number of factors than originally suggested. This phenomenon may not necessarily mean the inadequacy of the theory, but may merely suggest the possible existence of nonlinear structure. Further motivation for considering the general model of the form (1) is that applied scientists, for example, psychologists, prefer to work with models directly expressed for the observed variables of their interest and are rarely willing to transform the measurements familiar to them, for example, the scores of constructed scales.

In the popular structural equation modeling, the overall model consists of two parts: a structural model representing relationships among latent variables and a measurement model expressing relationships between observed and latent variables. Linear factor analysis models have been exclusively used as measurement models. However, the inadequacy of linear measurement models can affect the fit of the overall system of equations, and may possibly lead to an erroneous rejection of the proposed structural model for latent variables. Capabilities to consider nonlinear factor analysis models as measurement models can potentially expand the scope of the structural equation modeling.

Considering nonlinear models is also important when a factor analytic model is used in exploratory multivariate data analysis. The ability to model possibly nonlinear lower-dimensional structure hidden in high-dimensional data would enrich the set of

available statistical tools. Examination of data plots with the errors-in-variables view that errors occur in directions of all axes can be informative in revealing the inadequacy of linear models and in suggesting nonlinear models. If any possible underlying nonlinear relationships among observed variables are detected, then linear models are inappropriate. With the understanding that errors appear only in the directions of axes, some nonlinear models for underlying relationships can be suggested by scatter plots.

The concept of nonlinear factor analysis was initiated by Bartlett (1953). McDonald (1962) discussed the basic notions of nonlinear factor analysis, and introduced a model which is nonlinear in factors but linear in factor loadings, for example, a model with observed variables being polynomials in factors except for errors. He proposed a two-step estimation procedure. First, a linear orthogonal factor model is fitted and the scatter plots of estimated factor scores are used to detect possible nonlinearity. If a "significant" nonlinear relationship is seen on the plots, the function representing the relationship is estimated. The paper states that this procedure has some potential problems and limitations in practice. One severe restriction is that the number of variables need to be very large compared to the number of considered factors so that a linear model with a large number of terms possibly representing powers of underlying true factors can be fitted. An empirical application of this method is given in McDonald (1965). McDonald (1967a, b) focused his discussion on models which are simple polynomials in factors, and applied the procedure introduced in his first paper. McDonald (1979) suggested a model fitting method for a parametric version of the general nonlinear model

$$(5) \quad \mathbf{Z}_t = \mathbf{g}^*(\mathbf{f}_t^*; \boldsymbol{\beta}^*) + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, n,$$

where  $\mathbf{g}^*$  is a  $p$ -variate function of  $\mathbf{f}_t^*$  indexed by an unknown parameter vector  $\boldsymbol{\beta}^*$ . His method estimates  $\boldsymbol{\beta}^*$  and  $\mathbf{f}_t^*$  by minimizing

$$(6) \quad l_r = \frac{1}{2} \log |\text{diag } \mathbf{Q}| - \log |\mathbf{Q}|,$$

where

$$\mathbf{Q} = \frac{1}{n} \sum_{t=1}^n (\mathbf{Z}_t - \mathbf{g}^*(\mathbf{f}_t^*; \boldsymbol{\beta}^*)) (\mathbf{Z}_t - \mathbf{g}^*(\mathbf{f}_t^*; \boldsymbol{\beta}^*))',$$

and estimates the error variances by the diagonal elements of  $\mathbf{Q}$  evaluated at the estimates of  $\boldsymbol{\beta}^*$  and  $\mathbf{f}_t^*$ . His method reduces to the maximum normal likelihood estimation, if the model is the unrestricted linear model and if the maximum likelihood estimates of the error variances are all

positive. Little discussion was given to the identification issue. Etezadi-Amoli and McDonald (1983) applied the method suggested by McDonald (1979) for the polynomial factor analysis model (5) with all  $p$  components of  $\mathbf{g}^*$  being polynomials in the elements of  $\mathbf{f}_t^*$ . They suggested the use of  $n$  times the minimum of the criterion  $l_r$  in (6) as a chi-square test statistic for the fit of the model with some unspecified degrees of freedom. They provided some discussion of the identification issue from the rotation point of view and mentioned that the identification problem for the general polynomial model is not expected to be simple.

Kenny and Judd (1984) suggested an approach to structural equation model fitting with a single equation involving a cross product or a quadratic factor term. Their approach can be used for fitting a quadratic factor analysis model (with multiple quadratic relationships) under strong distributional assumptions, although the implementation can be tedious. Jöreskog and Yang (1994, 1997) gave further development of the Kenny–Judd method. A similar moment-based method was suggested by Mooijaart and Bentler (1986) for quadratic factor analysis. These methods rely heavily on the normality of the underlying factor and error vectors. Wall and Amemiya (2001) presented a modified version of the Kenny–Judd type estimator that does not require any distributional assumption. The basic approach used in these papers is applicable only for low-order polynomial (essentially second-order) models, and does not extend to more general nonlinear models. Also in the context of polynomial structural equation analysis, Arminger and Muthén (1998) considered a Bayes approach assuming the normality of factors and errors. Treating the factors as unknown constants, Wall and Amemiya (2000) introduced a new method for fitting polynomial structural equation models. Their procedure can be used for any single polynomial model in a system of equations, but is not directly applicable for the general nonlinear factor analysis model being considered here.

For an identifiable version of the general parametric nonlinear model (5), Amemiya (1993a, b) introduced a single-equation model fitting procedure based on the instrumental variable approach. His estimation procedure contains an adjustment for parameter estimate bias due to nonlinearity. Theoretical properties of the estimator based on the so-called small- $\sigma$  asymptotics were also presented in support of the procedure. Yalcin and Amemiya (1995) considered a class of additive nonlinear factor analysis models that are nonlinear in factors but linear in coefficients. They proposed an approximate

conditional likelihood model-fitting procedure, and used the small- $\sigma$  asymptotics to derive the properties of parameter estimators and a fit statistic. Compared to these previous papers by the current authors, this paper treats more general nonlinear models, and presents more comprehensive model fitting procedures.

Some topics in structural equation analysis may be considered to be tangentially related to nonlinear factor analysis. Latent growth curve analysis fits a model for an underlying trend over the observed variables, typically measured over time, in terms of a linear combination of known or unknown basis functions. See, for example, McArdle and Epstein (1987) and Meredith and Tisak (1990). In this type of analysis, the interest is in estimating a single trend over the observed measurements, and the relationships between the observed variables and factors are still linear. Another related topic is structural equation modeling of dichotomous and polytomous observed variables, as discussed by, for example, Muthén (1984) and Jöreskog (1994). The models used in such analysis represent nonlinear relationships between the observed measurements and factors, but the relationships are direct consequences of the distributional assumptions on the observed variables. This differs from the topic of interest in this paper, that is, to explore and express underlying nonlinear structure in continuous or scale-level multivariate data.

In this paper, we consider the general nonlinear factor analysis model (5) without imposing particular distributional forms for unobservable factors and errors. Some examples of a typical component, say the  $i$ th component, of  $\mathbf{g}^*$  in (5) are

$$\begin{aligned} g_i^*(\mathbf{f}_t^*, \boldsymbol{\beta}^*) &= \beta_0^* + \frac{\beta_1^*}{1 + e^{(\beta_2^* - \beta_3^* f_{1t}^* - \beta_4^* f_{2t}^*)}}, \\ g_i^*(\mathbf{f}_t^*, \boldsymbol{\beta}^*) &= \beta_0^* + \beta_1^* e^{\beta_2^* f_t^*}, \\ g_i^*(\mathbf{f}_t^*, \boldsymbol{\beta}^*) &= \beta_0^* + \beta_1^* f_{1t}^* + \beta_2^* f_{1t}^{*2} + \beta_3^* f_{2t}^* \\ &\quad + \beta_4^* f_{2t}^{*2} + \beta_5^* f_{1t}^* f_{2t}^*. \end{aligned}$$

Note that the model does not have to be polynomial in factors nor linear in the parameter  $\boldsymbol{\beta}^*$ . We suggest a nonlinear version of the errors-in-variable parameterization as a means to deal with the identification issue in model (5), as discussed in the next section. This parametrization also allows the systematic development of two statistical model fitting procedures with a minimal condition on the number of factors to be fitted. Both of the proposed procedures, described in Sections 5 and 6, incorporate some adjustments for bias due to nonlinearity. Also, the procedures produce reasonable parameter estimators without specific distributional assumptions on the factor and error terms.

#### 4. IDENTIFICATION OF NONLINEAR MODELS

As shown in (3), the linear model (2) contains the indeterminacy due to possible nonsingular linear transformation of  $\mathbf{f}_t^*$ . The parameter and factor vector in the nonlinear model (5) can have a more complex identification problem. Consider a one-to-one nonlinear transformation  $\mathbf{f}_t^{**} = \mathbf{h}(\mathbf{f}_t^*; \boldsymbol{\beta}^*)$ , possibly depending on  $\boldsymbol{\beta}^*$ . Then (5) can be written in an equivalent form

$$(7) \quad \mathbf{Z}_t = \mathbf{g}^*(\mathbf{h}^{-1}(\mathbf{f}_t^{**}; \boldsymbol{\beta}^*); \boldsymbol{\beta}^*) + \boldsymbol{\varepsilon}_t = \mathbf{g}^{**}(\mathbf{f}_t^{**}; \boldsymbol{\beta}^*) + \boldsymbol{\varepsilon}_t,$$

with  $\mathbf{f}_t^{**}$  as the factor vector. If  $\mathbf{g}^{**}(\mathbf{f}_t^{**}; \boldsymbol{\beta}^*)$  can be written in the form  $\mathbf{g}^*(\mathbf{f}_t^{**}; \boldsymbol{\beta}^{**})$  for some  $\boldsymbol{\beta}^{**}$ , then models (5) and (7) cannot be distinguished based on data, and the factor vector as well as the parameter are nonidentifiable. That is, the model is not identified in general, because a given nonlinear structure can be expressed using different sets of  $\mathbf{f}_t^*$  and  $\boldsymbol{\beta}^*$ . With the possibility for nonlinear transformation of  $\mathbf{f}_t^*$ , the identification issue for the nonlinear model (5) is not straightforward. Assuming a specific distributional form of  $\mathbf{f}_t^*$  may remove the possibility of nonlinear transformation. However, such an assumption on the unobservable  $\mathbf{f}_t^*$  for the purpose of identification is hard to justify and cannot be used when, as in linear factor analysis, the practical interest is in fitting models without specifying a particular distributional form for  $\mathbf{f}_t^*$ . Finding a generic identification condition in terms of the parameterization of  $\mathbf{g}^*$  in the general model (5) seems difficult, and may not be particularly useful. Instead, we suggest here a simple and practical solution of considering only those models that are written in a known identifiable form. As reviewed in Section 2, for linear factor analysis, the errors-in-variables parameterization provided an identifiable model and played a key role in the theoretical development for nonnormal data. The errors-in-variables version of the nonlinear model (5) can also be defined. We say model (5) is in the errors-in-variables parameterization, if the model can be written as

$$(8) \quad \begin{aligned} \mathbf{Y}_t &= \mathbf{g}(\mathbf{f}_t; \boldsymbol{\beta}) + \mathbf{e}_t, \\ \mathbf{X}_t &= \mathbf{f}_t + \mathbf{u}_t, \end{aligned} \quad t = 1, 2, \dots, n,$$

where, with possible reordering of the components of  $\mathbf{Z}_t$ ,  $\mathbf{Z}_t = (\mathbf{Y}_t', \mathbf{X}_t')$ ,  $\boldsymbol{\varepsilon}_t = (\mathbf{e}_t', \mathbf{u}_t')$ ,  $\mathbf{Y}_t$  and  $\mathbf{e}_t$  are  $(p - k) \times 1$ ,  $\mathbf{X}_t$  and  $\mathbf{u}_t$  are  $k \times 1$ ,  $\mathbf{f}_t$  is the  $k \times 1$  factor vector,  $\mathbf{g}$  is a  $(p - k)$ -valued nonlinear function of  $\mathbf{f}_t$  and  $\boldsymbol{\beta}$  is the parameter for such a function  $\mathbf{g}$ . In this form, the factor  $\mathbf{f}_t$  is identified as the "true value" of the observed  $\mathbf{X}_t$ , and there is no factor indeterminacy. With no possibility for transformation of  $\mathbf{f}_t$ , the relationship parameter  $\boldsymbol{\beta}$  is unambiguously determined

provided that there is no redundancy among the elements of  $\boldsymbol{\beta}$ . Thus, this errors-in-variables nonlinear factor analysis model is identifiable. In addition, the interpretation of model (8) is straightforward based on that of multivariate nonlinear regression of  $\mathbf{Y}_t$  with explanatory variable  $\mathbf{f}_t$  measured by  $\mathbf{X}_t$  with error. If some theory suggests particular nonlinear models, we try to transform the models to those given in (8) by possible reordering of the  $p$  components of  $\mathbf{Z}_t$ , by an appropriate choice of one-to-one  $\mathbf{h}$  in (7), and by possible transformation of some elements of  $\mathbf{Z}_t$ . In our approach, the models that cannot be transformed to the errors-in-variables form (7) are not considered in the analysis, because their identification status is unclear. This approach somewhat restricts the form of the general nonlinear models (5) that can be considered. What we are suggesting for practice is to try to write down models directly in the errors-in-variables form (8) without going through the general form (5). That is, the form (8) can be used as the starting point in specifying possible models for a given problem. In this way, all models under consideration are identified and can be fitted. This approach is in fact natural and practical in searching for plausible nonlinear models in exploratory data analysis. Nonlinear models in the errors-in-variables parameterization can be easily suggested based on observed variables through the measurement error regression interpretation of scatter plots. For example, it would be difficult to suggest, based only on observations, a model where each observed variable is a logistic function of an unobservable factor. On the other hand, a simple scatter-plot matrix can easily suggest a model where each of  $p - 1$  observed variables is a logistic function of the true value of the remaining variable. Also, we do not consider a model where all observed variables are polynomials in, say, two factors. Such a model may not be identified without additional conditions. On the other hand, a model where  $p - 2$  observed variables are polynomials in the true values of two remaining variables is identified and can be suggested by data based on plots. In some sense, the errors-in-variables parameterization focuses the model building on the relationships among the true values of observed variables, avoiding the difficulty of searching for an arbitrarily defined underlying factor and its relationship to observed variables at the same time.

In the next two sections, we will propose procedures for fitting and testing nonlinear factor analysis models given in the errors-in-variables form (8). Our procedures take advantage of the errors-in-variables form and of recent ideas in statistical model fitting. Some of the advantages of the errors-in-variables parameterization in statistical analysis



of the linear models are expected to carry over to the nonlinear models. Throughout our development, we also assume that

$$(9) \quad 2^{-1}(p-k)(p-k+1) \geq p.$$

This condition assumes that the dimension  $k$  of the factor vector cannot be very large compared to the number of observed variables  $p$ . However, this is a very unrestrictive condition, as it is identical to that required for the linear unrestricted (exploratory) factor analysis. We assume this condition, because the general nonlinear factor model (5) includes such a linear model as a special case, and because it plays an important role in our model fitting and checking procedures. Thus, our approach to the general nonlinear model does not require any more restrictive condition on the number of factors than the linear case. Note that there is no restriction on the dimension of the parameter vector  $\beta$  in  $\mathbf{g}$ . Thus, for example, even with a small  $k$ , a polynomial model of any order with any number of mixed power terms can be fitted and estimated with parameterization (8).

With the use of the errors-in-variables parameterization (8), some discussion of the literature on nonlinear measurement error model analysis may be relevant. Most existing methods for nonlinear errors-in-variables models are summarized in three books. Fuller (1987) presented estimation methods for a single equation (with a single dependent variable) when the measurement error variances are either known or estimated. Carroll, Ruppert, and Stefanski (1995) considered various nonlinear measurement error situations with a single equation, including those with a dichotomous dependent variable and with availability of instrumental variables. Cheng and Van Ness (1999) treated a single polynomial model with either known error variances or replicate observations. In the factor analysis problem addressed in this paper, replicate observations are not available, error variances are unknown, and multiple equations are involved. Hence, the existing methods for nonlinear errors-in-variables analysis summarized in these three books are not directly applicable for our nonlinear factor analysis model (8). In developing the model fitting procedures given in the next two sections, we have favored methods that are as closely related to traditional fitting methods for linear factor analysis models as possible, and that are accompanied by convenient and interpretable tests for goodness-of-fit for the model.

## 5. EXTENDED LINEAR MAXIMUM LIKELIHOOD METHOD

As described in Section 1, for the linear factor analysis model, maximum likelihood estimation

under the normality assumption is the most widely used model fitting method, and produces estimators and test statistics with good properties even for nonnormal data. For the nonlinear models, the normal likelihood itself may not be simple to handle. In the nonlinear model (5), the normality of  $\mathbf{f}_t^*$  may not lead to a simple distributional form for  $\mathbf{Z}_t$ , and the assumption of normal  $\mathbf{Z}_t$  may not be meaningful without the reasonableness of the corresponding distribution for  $\mathbf{f}_t^*$ . Thus, the direct extension of the normal likelihood approach to the nonlinear model is not straightforward. However, we show in this section that a maximum likelihood computational algorithm for the linear model can be extended naturally to the nonlinear model. The extended procedure for the nonlinear model reduces to the maximum normal likelihood method when the model happens to be linear. The linear model algorithm proposed by Pantula and Fuller (1986) takes advantage of the factor analysis structure that the observation is the sum of a lower-dimensional factor part and an error with a diagonal covariance matrix, the structure applicable also for nonlinear models. Their iterative algorithm for the linear model alternates between the relationship parameter and the error variances. Given error variance estimates, a relationship parameter estimate can be expressed explicitly as a solution to a minimization problem involving a ratio of quadratic forms. The error variances are estimated by generalized least squares applied to the elements of a residual covariance matrix obtained using a given relationship parameter estimate. Jöreskog (1967, 1977) also suggests the use of some alternation between the error variances and the relationship parameters. Such an alternation can be directly extended to the nonlinear cases, although each part becomes more complex and the use of factor score estimates is required.

The iterative maximum likelihood algorithm for the linear model (4) in the errors-in-variables parameterization consists of the following two alternating steps. Given an estimate of the error covariance matrix  $\hat{\Psi}$ , the updated estimate of the relationship parameters  $\lambda$  and  $\Lambda$  are obtained by minimizing

$$(10) \quad \sum_{t=1}^n \mathbf{v}_t'(\lambda, \Lambda) \Sigma^{-1}(\Lambda, \hat{\Psi}) \mathbf{v}_t(\lambda, \Lambda),$$

where

$$(11) \quad \begin{aligned} \mathbf{v}_t(\lambda, \Lambda) &= \mathbf{Y}_t - \lambda - \Lambda \mathbf{X}_t, \\ \Sigma(\Lambda, \Psi) &= (\mathbf{I}_{p-k}, -\Lambda) \Psi (\mathbf{I}_{p-k}, -\Lambda)'. \end{aligned}$$

Given  $\hat{\lambda}$  and  $\hat{\Lambda}$ , the updated estimate of  $\Psi$  is obtained by applying linear generalized least

squares to the unique elements of

$$(12) \quad \frac{1}{n} \sum_{t=1}^n \mathbf{v}_t(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\Lambda}}) \mathbf{v}_t'(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\Lambda}}),$$

where the weight matrix is estimated using  $\hat{\boldsymbol{\Lambda}}$  and the previous estimate of  $\boldsymbol{\Psi}$ .

For the nonlinear model (8), we attempt to extend the essential idea underlying this maximum likelihood algorithm for the linear model. It is necessary to use a certain approximation to the nonlinear function. The linear approximation is inappropriate, leading to nonlinearity bias. Thus, a certain form of quadratic approximation to the nonlinear function is utilized. This quadratic approximation to the function  $\mathbf{g}$  evaluated at the observed  $\mathbf{X}_t$  is given by an expansion around the underlying  $\mathbf{f}_t$ ,

$$(13) \quad \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}) \doteq \mathbf{g}(\mathbf{f}_t; \boldsymbol{\beta}) + \mathbf{G}(\mathbf{f}_t; \boldsymbol{\beta}) \mathbf{u}_t + \frac{1}{2} \mathbf{H}(\mathbf{f}_t; \boldsymbol{\beta}) \text{vec } \boldsymbol{\Psi}_{uu},$$

where

$$(14) \quad \mathbf{G}(\mathbf{f}; \boldsymbol{\beta}) = \frac{\partial}{\partial \mathbf{f}'} \mathbf{g}(\mathbf{f}; \boldsymbol{\beta}),$$

$$(15) \quad \mathbf{H}(\mathbf{f}; \boldsymbol{\beta}) = [\mathbf{h}_1(\mathbf{f}; \boldsymbol{\beta}), \mathbf{h}_2(\mathbf{f}; \boldsymbol{\beta}), \dots, \mathbf{h}_{p-k}(\mathbf{f}; \boldsymbol{\beta})]',$$

$$\mathbf{h}_i(\mathbf{f}; \boldsymbol{\beta}) = \text{vec} \left[ \frac{\partial^2 g_i(\mathbf{f}; \boldsymbol{\beta})}{\partial \mathbf{f} \partial \mathbf{f}'} \right],$$

$$\boldsymbol{\Psi}_{uu} = \text{Var} \{ \mathbf{u}_t \},$$

and  $g_i$  is the  $i$ th component of  $\mathbf{g}$ . We also used the vec operator that lists the elements of a matrix in a column vector by stacking the columns of the matrix. The approximation in (13) is based on ignoring the terms of order higher than two and on replacing  $\mathbf{u}_t \mathbf{u}_t'$  with its expectation  $\boldsymbol{\Psi}_{uu}$ . This type of expansion was used by Amemiya and Fuller (1988) for the bias adjustment in the nonlinear errors-in-variables problem. By the basic assumption of factor analysis,  $\mathbf{u}_t$  is independent of  $\mathbf{f}_t$  whether  $\mathbf{f}_t$  is treated as random or fixed. Thus, given  $\mathbf{f}_t$ , the conditional moments of the right-hand side of (13) can be obtained easily. As yet, (13) is not a simple linear approximation and makes an adjustment for the nonlinearity of  $\mathbf{g}$  through the second derivative term. Using (13), we obtain

$$(16) \quad \mathbf{Y}_t - \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}) \doteq \mathbf{e}_t - \mathbf{G}(\mathbf{f}_t; \boldsymbol{\beta}) \mathbf{u}_t - \frac{1}{2} \mathbf{H}(\mathbf{f}_t; \boldsymbol{\beta}) \text{vec } \boldsymbol{\Psi}_{uu}.$$

Based on this approximation, a quantity corresponding to  $\mathbf{v}_t(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$  used in (10) and (12) can be defined as

$$(17) \quad \mathbf{v}_t(\boldsymbol{\beta}, \boldsymbol{\Psi}, \mathbf{f}_t) = \mathbf{Y}_t - \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}) + \frac{1}{2} \mathbf{H}(\mathbf{f}_t; \boldsymbol{\beta}) \text{vec } \boldsymbol{\Psi}_{uu}.$$

The  $\boldsymbol{\Sigma}(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$  in (11) for the linear model is the covariance matrix of  $\mathbf{v}_t(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$ , which is free of  $\mathbf{f}_t$  and  $t$ . For the nonlinear model, the conditional covariance matrix of (17) given  $\mathbf{f}_t$  depends on  $t$ , and is given by

$$(18) \quad \boldsymbol{\Sigma}_{tt}(\boldsymbol{\beta}, \boldsymbol{\Psi}, \mathbf{f}_t) = [\mathbf{I}_{p-k}, -\mathbf{G}(\mathbf{f}_t; \boldsymbol{\beta})] \times \boldsymbol{\Psi} [\mathbf{I}_{p-k}, -\mathbf{G}(\mathbf{f}_t; \boldsymbol{\beta})]'$$

Hence, an extension of the linear model iterative procedure can be obtained by replacing  $\mathbf{v}_t(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$  and  $\boldsymbol{\Sigma}(\boldsymbol{\Lambda}, \boldsymbol{\Psi})$  in (10) and (12) by (17) and (18). For a nonlinear model, the minimization of the form (10) is nontrivial, requiring some iterative procedure. However, the generalized least squares step corresponding to (12) can still be given explicitly, because the conditional expectation of  $\mathbf{v}_t(\boldsymbol{\lambda}, \boldsymbol{\Lambda}) \mathbf{v}_t'(\boldsymbol{\lambda}, \boldsymbol{\Lambda})$  given  $\mathbf{f}_t$  is linear in  $\boldsymbol{\Psi}$ . Since (17) and (18) depend on  $\mathbf{f}_t$ , some estimate of  $\mathbf{f}_t$  is needed at each iteration. Thus, unlike the linear case, the nonlinear case requires factor score estimation at each step. For this, we can extend the linear errors-in-variables true-value estimator based on a residual vector as given in Fuller (1987, page 364). Using (17) as an approximate residual, a simple factor score estimate given a previous estimate  $\hat{\mathbf{f}}_t$  of  $\mathbf{f}_t$  is

$$(19) \quad \hat{\mathbf{f}}_t = \mathbf{X}_t + \boldsymbol{\Psi}_{uu} \mathbf{G}(\hat{\mathbf{f}}_t; \boldsymbol{\beta}) \boldsymbol{\Sigma}_{tt}^{-1}(\boldsymbol{\beta}, \boldsymbol{\Psi}, \hat{\mathbf{f}}_t) \mathbf{v}_t(\boldsymbol{\beta}, \boldsymbol{\Psi}, \hat{\mathbf{f}}_t).$$

Thus, the extended linear maximum likelihood (ELM) method is an iterative process over three sets of parameters,  $\mathbf{f}_t$ ,  $\boldsymbol{\beta}$  and  $\boldsymbol{\Psi}$ .

The Appendix gives explicit formulas needed for the algorithm that incorporates modification for the parameter space restriction and for numerical stability. The Appendix also gives an initial estimate that can be used to start the iterative process. In practice, a few iterations of this procedure need to be performed. [In the context of a different problem, Carroll and Ruppert (1988) gave the same recommendation for a similar iterative method.] In our experience, convergence of the parameter estimates and the related goodness-of-fit statistic usually occurred within four to five iterations. The ELM estimates  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\boldsymbol{\Psi}}$  and  $\hat{\mathbf{f}}_t$  are those values obtained at the final step of the iterative algorithm. The generalized least squares residual sum of squares in the step corresponding to (12) serves as a test statistic for the fit of the model and can be compared to the upper percentage points of the chi-square distribution with  $d = \frac{1}{2}(p-k)(p-k+1) - p$  degrees of freedom. For the linear model, this test statistic reduces to the normal likelihood ratio test statistic for the model fit. As in the linear case, the factor covariance matrix estimation is not required during the iteration. If desired, an estimate of the factor covariance matrix can be obtained easily

once the estimates of other intrinsic parameters are obtained. The Appendix gives a formula.

For the linear model, the maximum normal likelihood estimation has been recommended also for fixed or nonnormal  $\mathbf{f}_t$  and nonnormal  $\boldsymbol{\varepsilon}_t$ . In fact, the estimation of  $\boldsymbol{\lambda}$  and  $\boldsymbol{\Lambda}$  minimizing (10) for a given  $\boldsymbol{\Psi}$  is also the maximum likelihood estimators of  $\boldsymbol{\lambda}$  and  $\boldsymbol{\Lambda}$  for the linear model with fixed  $\mathbf{f}_t$  and known  $\boldsymbol{\Psi}$ . Also, generalized least squares applied to (12) makes sense without specifying the distributional form of  $\boldsymbol{\varepsilon}_t$ . Hence, we expect the ELM method to provide a reasonable model fitting procedure for a nonlinear model regardless of distributional assumptions on  $\mathbf{f}_t$  and  $\boldsymbol{\varepsilon}_t$ .

### 6. APPROXIMATE CONDITIONAL LIKELIHOOD METHOD

The ELM model fitting procedure developed in the previous section has some good properties especially when it is considered as an extension of the linear model case. However, the algorithm requires the computation of factor score estimates for all individuals at each iteration step. Recall that the factor score estimate for the  $k \times 1 \mathbf{f}_t$  is essentially based on the  $p \times 1 \mathbf{Z}_t$  for each fixed  $t$ . Such individual factor score computation and the evaluation of various quantities at the estimated factor scores may potentially introduce some numerical instability and increased variability. In fact, our experience and simulation study have shown that the ELM procedure tends to produce some outlying values of parameter estimates in small samples. This led us to consider a model fitting procedure without involving factor score estimation. To achieve this, an expansion of  $\mathbf{g}(\mathbf{f}_t; \boldsymbol{\beta})$  around the observed  $\mathbf{X}_t$  is used instead of (16), and all necessary quantities are evaluated at  $\mathbf{X}_t$  instead of  $\mathbf{f}_t$  to avoid the factor score estimation. Then, the approximate conditional distribution of  $\mathbf{Y}_t$  given  $\mathbf{X}_t$  is utilized in developing a proper model fitting procedure. For the linear model, this conditional approach is not as efficient as the ELM approach which is, for this case, the normal maximum likelihood estimation. In addition, some restriction on the distributional form of  $\mathbf{f}_t$  and  $\mathbf{u}_t$  is required for the conditional approach to be fully justified. On the other hand, for many practical situations, the computational simplicity of the conditional approach is expected to produce more numerically reliable parameter estimates in small samples than the ELM approach.

To introduce this conditional approach, we first assume hypothetically in model (8) that  $\mathbf{f}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Phi})$  and  $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Psi})$ . Since  $\mathbf{Y}_t$  is non-normal for any  $\mathbf{g}$  nonlinear in  $\mathbf{f}_t$ , the likelihood

based on  $\mathbf{Z}_t$  cannot be written down explicitly. To obtain a workable form of the likelihood, we use the quadratic approximation to the function  $\mathbf{g}$  around the observable  $\mathbf{X}_t$ , that is,

$$(20) \quad \mathbf{g}(\mathbf{f}_t; \boldsymbol{\beta}) \doteq \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}) - \mathbf{G}(\mathbf{X}_t; \boldsymbol{\beta})\mathbf{u}_t + \frac{1}{2}\mathbf{H}(\mathbf{X}_t; \boldsymbol{\beta}) \text{vec } \boldsymbol{\Psi}_{uu},$$

where  $\mathbf{G}(\mathbf{f}; \boldsymbol{\beta})$  and  $\mathbf{H}(\mathbf{f}; \boldsymbol{\beta})$  are defined in (14) and (15), respectively. With this approximation, the equation for  $\mathbf{Y}_t$  in (8) becomes

$$(21) \quad \mathbf{Y}_t \doteq \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}) + \frac{1}{2}\mathbf{H}(\mathbf{X}_t; \boldsymbol{\beta}) \text{vec } \boldsymbol{\Psi}_{uu} + \mathbf{e}_t - \mathbf{G}(\mathbf{X}_t; \boldsymbol{\beta})\mathbf{u}_t.$$

The conditional distribution of  $\mathbf{e}_t - \mathbf{G}(\mathbf{X}_t; \boldsymbol{\beta})\mathbf{u}_t$  given  $\mathbf{X}_t$  is

$$(22) \quad N(\boldsymbol{\Sigma}_{vutt}(\boldsymbol{\Phi} + \boldsymbol{\Psi}_{uu})^{-1}(\mathbf{X}_t - \boldsymbol{\mu}), \boldsymbol{\Gamma}_{tt}),$$

where

$$(23) \quad \begin{aligned} \boldsymbol{\Gamma}_{tt} &= \boldsymbol{\Sigma}_{vtt} - \boldsymbol{\Sigma}_{vutt}(\boldsymbol{\Phi} + \boldsymbol{\Psi}_{uu})^{-1}\boldsymbol{\Sigma}'_{vutt}, \\ \boldsymbol{\Sigma}_{vutt} &= -\mathbf{G}(\mathbf{X}_t; \boldsymbol{\beta})\boldsymbol{\Psi}_{uu}, \\ \boldsymbol{\Sigma}_{vtt} &= \boldsymbol{\Psi}_{ee} + \mathbf{G}(\mathbf{X}_t; \boldsymbol{\beta})\boldsymbol{\Psi}_{uu}\mathbf{G}(\mathbf{X}_t; \boldsymbol{\beta})', \\ \boldsymbol{\Psi}_{ee} &= \text{Var}\{\mathbf{e}_t\}. \end{aligned}$$

Using (21) and (22), the approximate conditional distribution of  $\mathbf{Y}_t$  given  $\mathbf{X}_t$  is

$$N(\mathbf{v}_t, \boldsymbol{\Gamma}_{tt}),$$

where

$$\mathbf{v}_t = \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}) + \frac{1}{2}\mathbf{H}(\mathbf{X}_t; \boldsymbol{\beta}) \text{vec } \boldsymbol{\Psi}_{uu} + \boldsymbol{\Sigma}_{vutt}(\boldsymbol{\Phi} + \boldsymbol{\Psi}_{uu})^{-1}(\mathbf{X}_t - \boldsymbol{\mu}).$$

The approximate likelihood function is a product of the densities of  $\mathbf{Z}_t$ 's, and each density is the product of the marginal density of  $\mathbf{X}_t$  and the conditional density of  $\mathbf{Y}_t$  given  $\mathbf{X}_t$ . Thus, the approximate log likelihood is, except for a multiplier  $(-1/2)$  and an additive constant,

$$l = l_1 + l_2,$$

where

$$\begin{aligned} l_1 &= \sum_{t=1}^n [\log |\boldsymbol{\Gamma}_{tt}| + (\mathbf{Y}_t - \mathbf{v}_t)' \boldsymbol{\Gamma}_{tt}^{-1} (\mathbf{Y}_t - \mathbf{v}_t)], \\ l_2 &= \sum_{t=1}^n [\log |\boldsymbol{\Phi}^*| + (\mathbf{X}_t - \boldsymbol{\mu})' \boldsymbol{\Phi}^{*-1} (\mathbf{X}_t - \boldsymbol{\mu})], \\ \boldsymbol{\Phi}^* &= \boldsymbol{\Phi} + \boldsymbol{\Psi}_{uu}. \end{aligned}$$

We observe that the normal marginal likelihood  $l_2$  is valid only for normal  $\mathbf{X}_t = \mathbf{f}_t + \mathbf{u}_t$ . However, simple

unbiased estimators,

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}} = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t,$$

$$\hat{\boldsymbol{\Phi}}^* = \mathbf{m}_{XX} = \frac{1}{n-1} \sum_{t=1}^n (\mathbf{X}_t - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})',$$

are reasonable estimators of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Phi}^*$  regardless of the distributional forms of random  $\mathbf{f}_t$  and  $\mathbf{u}_t$ . Thus, to estimate  $\boldsymbol{\beta}$  and  $\boldsymbol{\Psi}$ , we consider  $l_1$  with  $\hat{\boldsymbol{\mu}}$  and  $\hat{\boldsymbol{\Phi}}^*$  substituted. This is the so-called pseudolikelihood approach. The  $l_1$ , which is also a normal likelihood, does not directly depend on the normality of  $\mathbf{f}_t$ , but can be justified if the distribution of  $\mathbf{e}_t$  and the conditional distribution of  $\mathbf{u}_t$  given  $\mathbf{X}_t$  are approximately normal. Note that  $l_1$  is only an approximate likelihood, and that  $\Gamma_{tt}$  is not the exact conditional covariance matrix. The use of such an approximate expression for a covariance matrix in a role other than a (possibly incorrect) weight for a quadratic form (to be minimized) is known to produce estimators with potentially poor properties. (See, e.g., Carroll and Ruppert, 1982, and van Houwelingen 1988.) Hence, instead of minimizing  $l_1$  with respect to  $\boldsymbol{\beta}$  and  $\boldsymbol{\Psi}$ , we separate the two parameters, and iterate between the two. For the estimation of  $\boldsymbol{\beta}$  for a given value of  $\boldsymbol{\Psi}$ , we consider minimizing the exponent part of the conditional likelihood  $l_1$  using the approximate  $\Gamma_{tt}$  only as a weight. Given a value of  $\boldsymbol{\beta}$ , we take the conditional residual of the form  $\mathbf{Y}_t - \hat{\mathbf{v}}_t$ , and apply the conditional generalized least squares to the residual sum of squares matrix. An explicit expression for the iterative process is given in the Appendix. This iterative procedure should be continued for a few iterations only. In our experience of the actual implementation, the estimates in most cases converged within three to four iterations. The final step produces the approximate conditional likelihood (ACL) estimates  $\hat{\boldsymbol{\beta}}$  and  $\hat{\boldsymbol{\Psi}}$ . The generalized least squares residual sum of squares at the final iteration can be compared to the chi-square distribution with  $d = \frac{1}{2}(p-k)(p-k+1) - p$  degrees of freedom for testing the fit of the model. This test is based on the conditional distribution, and differs from that for the ELM.

The minimization of the exponent part of  $l_1$  to obtain an estimate of  $\boldsymbol{\beta}$  may look somewhat similar to that used for the ELM procedure, but there are some important differences. The quantity to be minimized for the ACL is based on the approximate conditional distribution and uses the expansion around  $\mathbf{X}_t$  instead of around  $\mathbf{f}_t$  for the ELM. Thus, for example, the second derivative bias adjustment, based on similar ideas, has different signs in the two approaches. Also, as given explicitly in

the Appendix, the function being minimized in the ACL step is a quadratic form in some deviation term with a given value of the weight. On the other hand, within the ELM function,  $\boldsymbol{\beta}$  appears in both the deviation and the weight. Hence, the ACL minimization is considerably simpler than that for the ELM.

For a special case of the linear model with normal assumptions on  $\mathbf{f}_t$  and  $\boldsymbol{\varepsilon}_t$ , the ACL procedure iterated to convergence does not reduce to maximum normal likelihood estimation. This is because the information on  $\boldsymbol{\Psi}_{ee}$  is not used to obtain an estimate of  $\boldsymbol{\beta}$ , and because some information about  $\boldsymbol{\Phi}$  contained in  $l_1$  is ignored. However, for the linear model, the resulting estimator possesses reasonable efficiency and is simple to compute. Thus, we might expect similar properties for the nonlinear model. The convenient form of the approximate conditional likelihood is appealing for handling a complex likelihood based on a nonlinear function of a random factor. The ACL method with a simpler minimization and without function evaluation at new factor estimates in every iteration has a computational advantage over the ELM method and is expected to be more numerically stable.

For both ELM and ACL estimators and fit statistics, the small- $\sigma$  asymptotic theory could be applied to derive approximate distributional properties and to provide justification, although the details can be extremely tedious. Some theoretical results show that the small- $\sigma$  asymptotic justification for the ACL approach requires some distributional assumption on  $\mathbf{f}_t$  and  $\mathbf{u}_t$ . In short, how much the conditional mean of  $\mathbf{f}_t$  given  $\mathbf{X}_t$  is allowed to depart from the linearity depends on how small the error variances are. For the practical use of the ACL, it is necessary to choose the  $k$  components in the reference variable  $\mathbf{X}_t$  in the errors-in-variables parameterization so that  $\mathbf{X}_t$  is roughly normal or the error  $\mathbf{u}_t$  is small. See Yalcin and Amemiya (1995). In the asymptotics, the ELM approach may be preferred. But, as we shall see in the next section, the finite sample comparison is not that simple. To assess the accuracy of the parameter estimates in either approach, approximate standard error estimates could be obtained using the small- $\sigma$  asymptotics, or some resampling procedures could be useful in taking advantage of the multivariate random sample structure of the data.

## 7. A SIMULATION STUDY

A Monte Carlo simulation study was conducted to assess the finite sample properties of the parameter estimates obtained by the ELM and ACL procedures

developed in Sections 5 and 6. Model (8) with  $p = 4$  observed variables and  $k = 1$  factor was considered. For  $t = 1, 2, \dots, n$ , the model is

$$\begin{aligned} Y_{1t} &= \beta_1 + \frac{\beta_2}{1 + e^{\beta_3 - \beta_4 f_t}} + e_{1t}, \\ Y_{2t} &= \beta_5 + \beta_6 f_t + \beta_7 f_t^2 + e_{2t}, \\ Y_{3t} &= \beta_8 + \beta_9 f_t + \beta_{10} f_t^2 + e_{3t}, \\ X_t &= f_t + u_t, \end{aligned}$$

where the true parameter values for the coefficients were chosen to be

$$\begin{aligned} \boldsymbol{\beta}' &= (\beta_1, \beta_2, \dots, \beta_{10}) \\ &= (0, 7, 10, 0.5, 50, -5, 0.1, 65, -7, 0.2). \end{aligned}$$

We generated random samples, assuming

$$\begin{aligned} f_t &\sim N(20, 36), \\ \begin{pmatrix} \mathbf{e}_t \\ u_t \end{pmatrix} &\sim N(\mathbf{0}, \text{diag}(\psi_{ee11}, \psi_{ee22}, \psi_{ee33}, \psi_{uu})). \end{aligned}$$

To study the effect of the error variance size relative to the total variance, we considered two sets of the error variances,

$$\begin{aligned} \psi_{eei} &= \delta \text{Var}\{Y_{it}\}, \quad i = 1, 2, 3, \\ \psi_{uu} &= \delta \text{Var}\{X_t\}, \end{aligned}$$

with  $\delta = 0.1$  and  $\delta = 0.2$  corresponding to cases with error variances approximately 10% and 20% of the total variances. (The numerical approximation was used for  $\text{Var}\{Y_{1t}\}$ ). Two sample sizes,  $n = 300$  and  $500$ , were considered. For each of the four combinations of  $\delta$  and  $n$ , 1000 Monte Carlo samples were generated. From each sample, initial estimates for  $\boldsymbol{\beta}$  and  $\psi$  described in the Appendix were obtained. Using these initial estimates as starting values, a step consisting of (E1)–(E3) in the Appendix was iterated twice to obtain the extended linear maximum likelihood (ELM) estimates of  $\boldsymbol{\beta}$  and  $\psi$ . Also using the same initial values, a step consisting of (A1) and (A2) in the Appendix was iterated twice to obtain the approximate conditional likelihood (ACL) estimates of  $\boldsymbol{\beta}$  and  $\psi$ .

Table 1 presents the square root of the mean squared error (RMSE) and the relative bias (RB = bias/true value) for some selected parameters for the four combinations of the sample sizes and variance ratios. Although the estimators of  $\boldsymbol{\beta}$  may not possess moments, these tables provide useful summaries and comparisons of the estimation procedures. In addition, some boxplots are given in Figure 2 and Figure 3 representing general patterns of empirical distributions of the estimates. In the boxplots, whiskers are drawn to the nearest

value not beyond  $1.5 \times$  (*interquartile range*) from the quartiles. There were a few outliers hidden in the boxplots. Such values were somewhat away from the remaining values, but were not extreme outliers.

For estimation of  $\boldsymbol{\beta}$ , the initial estimator is the ordinary nonlinear least squares estimator (as specified in the Appendix) which ignores the error in  $\mathbf{X}_t$ . As can be seen in the table and boxplots, the initial estimator of  $\boldsymbol{\beta}$  has a large bias which increases with  $\delta$  and does not decrease with  $n$ . In fact, the initial estimator is so biased that all 1000 samples in some cases give the values on one side (either larger or smaller) of the true value. Thus, the ordinary nonlinear least squares estimator of  $\boldsymbol{\beta}$  is useless unless the error variances are very small. On the other hand, the ELM and ACL estimators of  $\boldsymbol{\beta}$  have small biases, and the ACL is nearly median unbiased. The small bias and the nearly symmetric empirical distribution around the true value are two common features of the ELM and ACL estimators appealing for practical use. The ELM estimator of  $\boldsymbol{\beta}$  takes some outlying values when the relative size  $\delta$  of the error variance is large or the sample size  $n$  is small. This also affects the RMSE and RB, and may possibly indicate the nonexistence of moments in some cases. If we exclude a few outlying values, the ELM estimator of  $\boldsymbol{\beta}$  generally tends to have slightly smaller variability than the ACL.

For estimation of  $\psi$ , the initial estimator given in the Appendix is an unweighted estimator, while the ELM and ACL estimators are weighted estimators (with bias adjustment). As a result, the ELM and ACL estimators improve largely over the initial estimator in terms of variability. The performances of the ELM and ACL estimators are similar to each other, except that the ELM tends to have a smaller bias but a slightly larger number of outlying values. As in estimation of  $\boldsymbol{\beta}$ , the outlying values are not extreme, and the number of samples containing them is very small. The estimation of  $\psi$  becomes increasingly more difficult when the error variances become large relative to the total variances.

Yalcin (1995) examines histograms of the parameter estimates, excluding the outliers. He reports that the empirical distributions of the estimators are approximately normal, that the ELM estimator of a  $\boldsymbol{\beta}$  parameter generally tends to be less variable (excluding outliers) but more biased than the ACL, and that, for estimating  $\psi_{uu}$ , ELM is less biased but more variable than the ACL. In general, both bias and variability decrease for both estimators as the error variance decreases or as the sample size increases. For estimating  $\psi_{eei}$ , the two estimators are generally similar.

TABLE 1  
 Root mean squared error and relative bias for three estimators

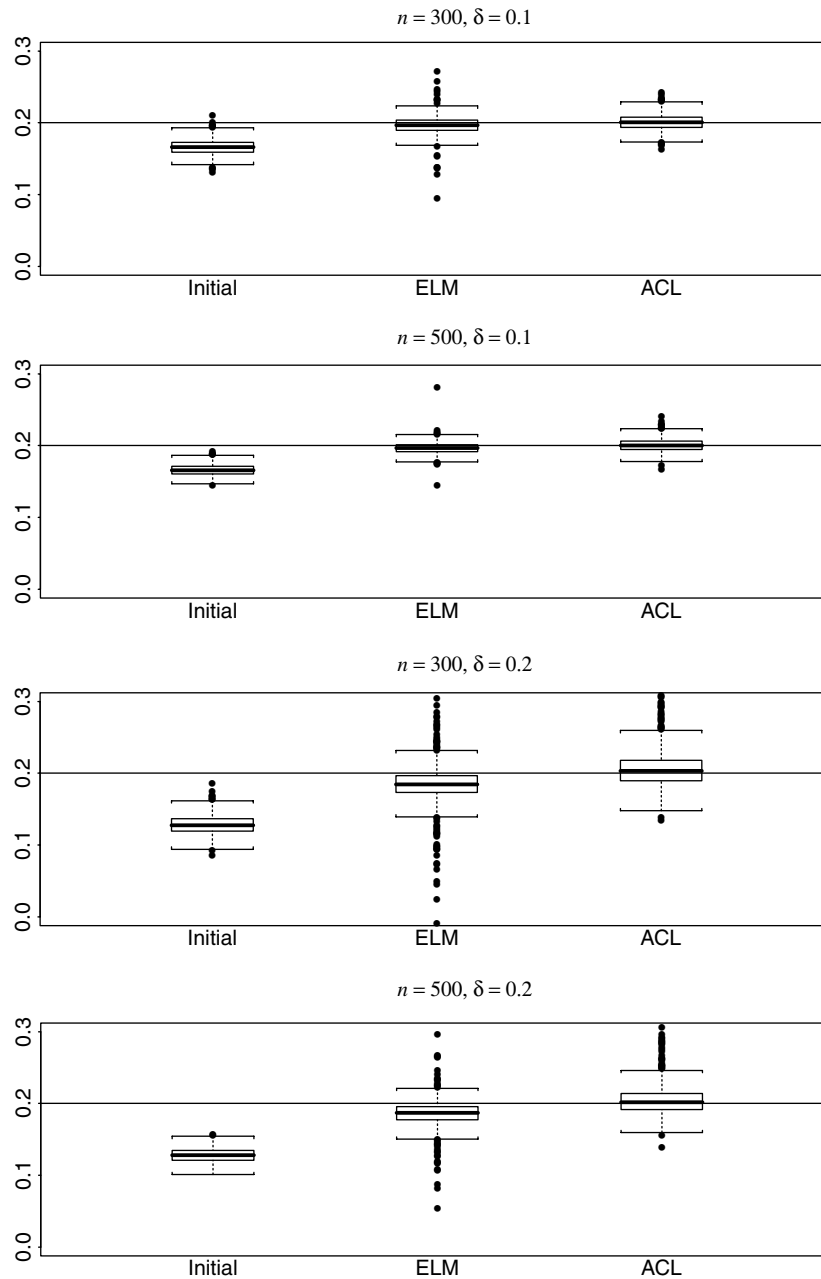
Parameter	Initial		ELM		ACL	
	RMSE	RB	RMSE	RB	RMSE	RB
<b><math>n = 300 \delta = 0.1</math></b>						
$\beta_4$	0.1132	-0.2093	0.0701	-0.0848	0.0073	0.0331
$\beta_5$	8.7747	-0.1667	4.1434	-0.0176	2.7802	-0.0006
$\beta_{10}$	0.0352	-0.1683	0.0155	-0.0140	0.0110	0.0036
$\psi_{ee11}$	0.2792	-0.0853	0.1474	-0.0873	0.1534	-0.0555
$\psi_{ee22}$	2.0480	-0.0900	1.3914	-0.0212	1.2358	0.0054
$\psi_{uu}$	0.9302	0.0475	0.5045	-0.0196	0.4576	-0.0449
<b><math>n = 500 \delta = 0.1</math></b>						
$\beta_4$	0.1115	-0.2120	0.0546	-0.0789	0.0563	0.0251
$\beta_5$	8.5938	-0.1666	2.4003	-0.0193	2.3242	-0.0012
$\beta_{10}$	0.0348	-0.1693	0.0087	-0.0158	0.0088	0.0033
$\psi_{ee11}$	0.2182	-0.0799	0.1114	-0.0747	0.1165	-0.0416
$\psi_{ee22}$	1.6591	-0.0903	0.9675	-0.0092	0.9386	0.0073
$\psi_{uu}$	0.7485	0.0781	0.3742	-0.0185	0.3724	-0.0437
<b><math>n = 300 \delta = 0.2</math></b>						
$\beta_4$	0.2015	-0.3868	0.0546	-0.2424	0.0563	-0.0826
$\beta_5$	18.0742	-0.3544	9.9081	-0.0993	6.0836	0.0110
$\beta_{10}$	0.0730	-0.3590	0.0343	-0.0407	0.0269	0.0306
$\psi_{ee11}$	0.7565	-0.2245	0.3564	-0.0611	0.3487	0.0645
$\psi_{ee22}$	6.8998	-0.2275	3.8092	-0.0165	3.1516	0.0400
$\psi_{uu}$	2.5832	0.1325	1.3115	-0.0257	1.4179	-0.1119
<b><math>n = 500 \delta = 0.2</math></b>						
$\beta_4$	0.2018	-0.3941	0.1461	-0.2303	0.1001	-0.0908
$\beta_5$	18.0186	-0.3562	7.7465	-0.0878	4.7030	0.0029
$\beta_{10}$	0.0725	-0.3593	0.0246	-0.0683	0.0208	0.0227
$\psi_{ee11}$	0.6835	-0.2342	0.2608	-0.0452	0.2836	0.0711
$\psi_{ee22}$	6.3766	-0.2361	2.6675	0.0013	2.4453	0.0397
$\psi_{uu}$	2.3628	0.1784	1.0265	-0.0256	1.2965	-0.1128

Overall, the computationally simpler ACL fitting procedure is generally more stable than the ELM and seems to produce estimators with at least similar efficiency. Because the normality of the factor and error vectors used in this simulation corresponds to the ideal case for the ACL, and because the ELM estimator is more efficient for linear models, the relative efficiency of the ACL against the ELM may decrease for other distributional situations. But, for the studied situation, the factor score estimation and the repeated evaluation of functions at the score estimates in the ELM procedure seem to result in finite sample instability, producing more outliers than the ACL. It may be possible to develop further modification of the ELM so that the occurrence of outlying estimates decrease and the finite sample properties improve. These and a number of other methodological and theo-

retical issues regarding nonlinear factor analysis require further investigation. However, the results of this study are promising for the ELM and ACL approaches as practical model fitting procedures.

## 8. THE EXAMPLE REVISITED

In this section, the *Self-Directed Search* vocational interest data of Section 1 is further analyzed. Recall that the fit of the linear one-factor model was very poor ( $p = 0.0002$ ) and that no linear factor analysis model fits  $Z_1, \dots, Z_4$  well. Based on the scatter plots in Figure 1, it may be reasonable to suggest that, except for measurement errors,  $Z_1$  and  $Z_3$  are quadratic functions of  $Z_4$ , and  $Z_2$  is either linearly or quadratically related to  $Z_4$ . Thus, as an exploratory nonlinear model, we might consider a general quadratic model with  $Z_4$  as the reference variable corresponding to the underlying

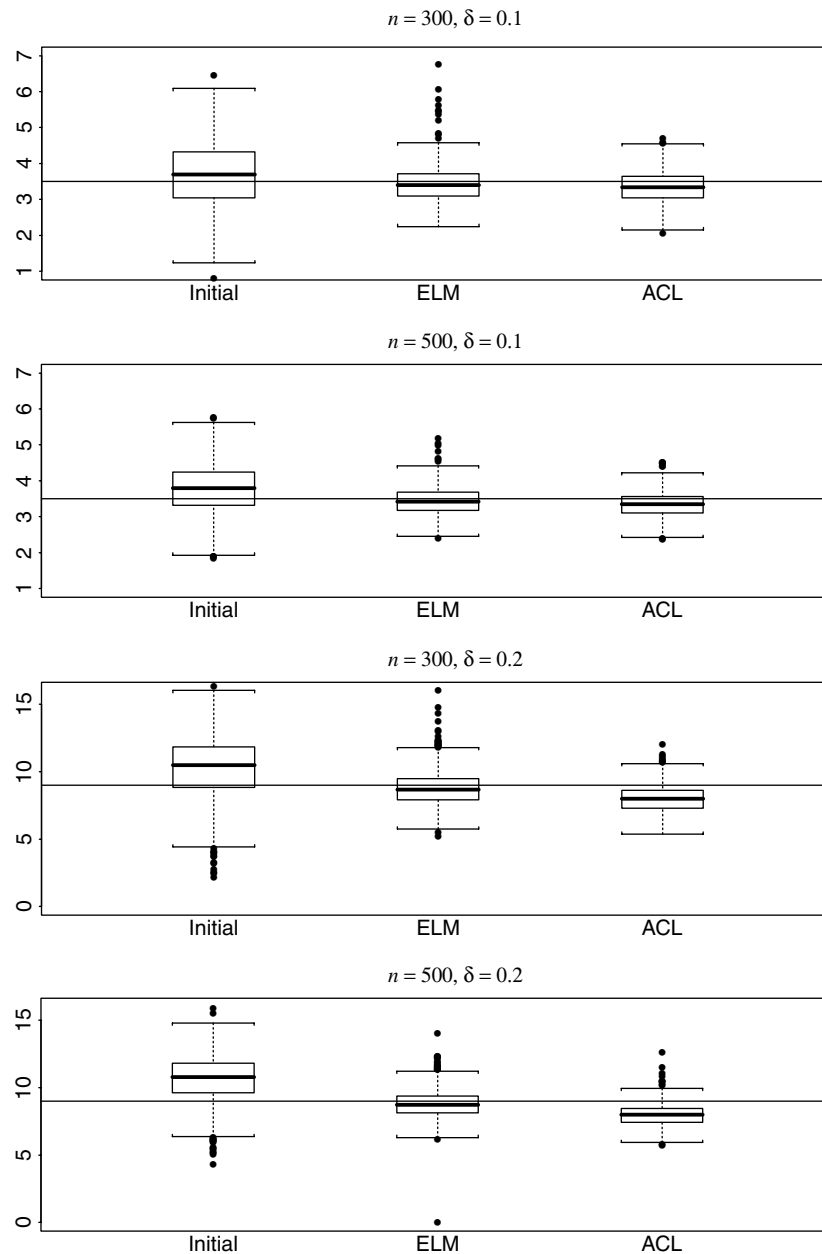
FIG. 2. Boxplots for estimates of  $\beta_{10}$ .

ing investigative factor in the errors-in-variables parameterization (8) as given by

$$(24) \quad \begin{aligned} Z_{it} &= \beta_{0i} + \beta_{1i}f_t + \beta_{2i}f_t^2 + \varepsilon_{it}, \quad i = 1, 2, 3, \\ Z_{4t} &= f_t + \varepsilon_{4t}, \quad t = 1, 2, \dots, n. \end{aligned}$$

This model is fitted to  $Z_1, \dots, Z_4$  here using the ELM and ACL procedures described in Sections 5 and 6. For both procedures, the estimates converged after five iterations. The goodness-of-fit test statistics for the whole system of equations introduced in Sections 5 and 6 were computed. For the ELM procedure, the statistic is  $\chi^2 = 4.37$  with 2 degrees

of freedom ( $p = 0.11$ ). For the ACL, the statistic is  $\chi^2 = 3.78$  also with 2 degrees of freedom ( $p = 0.15$ ). Thus, according to these goodness-of-fit tests, the quadratic model (24) with one underlying factor for the data  $Z_1, \dots, Z_4$  cannot be rejected. Figure 4 is the scatter plots of  $Z_1, Z_2$ , and  $Z_3$  versus  $Z_4$  with three fitted lines; linear factor analysis fit and two quadratic factor analysis fits by the ELM and ACL. By examining the first and the third plots, the inadequacy of the linear model in describing the data is clear. The two quadratic fitted lines capture the apparent curvature in the data. For  $Z_1$  and  $Z_3$ , the

FIG. 3. Boxplots for estimates of  $\psi_{uu}$ .

ELM fit with near monotonicity may make more sense from the subject-matter point of view. The good fit of the quadratic model with one factor does not deny the subject-matter theory that these four scales measure a common underlying trait, and indicates that the data for this particular set of students are not necessarily inconsistent with the theory used to design the investigative aspect part of the vocational interest test. On the other hand, if the analysis was restricted to linear models, then the only possible conclusion would be that this data set and the one-factor theory are not compatible. By expanding the analysis to include the possible

nonlinear models, a well-fitting exploratory model based on the one-factor theory was found. If we trust the theory used in designing the test, then we can use the fitted quadratic model to represent approximate relationships between one underlying factor and the four observed scales  $Z_1, \dots, Z_4$ , and to define an “investigative” factor corresponding to the errors-in-variables parameterization used in model (24). A similar analysis can be applied to the remaining scale scores in the data set designed to measure each of the five other vocational interest aspects. Alternatively, for all scale variables, a reasonable nonlinear factor analysis model with



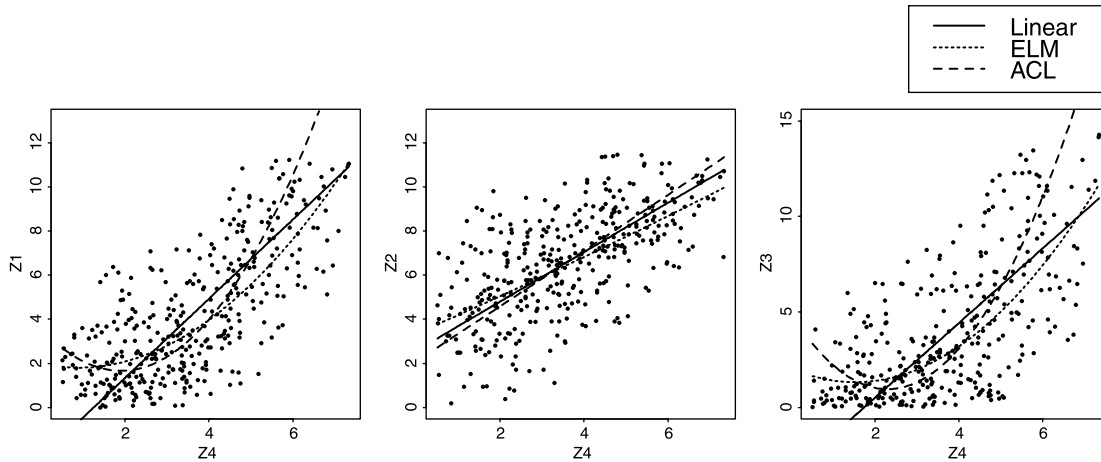


FIG. 4. Vocational interest data with the linear and quadratic (ELM and ACL) factor analysis model fits.

six factors matching the theory can be searched using the techniques described in this paper. Either way, if models with good fit are obtained, then the six-factor score estimates corresponding to the theoretical six vocational interest aspects can be obtained. In the ELM procedure, factor score estimates are computed at each iteration using the formula (19). For the ACL method, the final fitted model and a formula similar to (19) can be used to obtain factor score estimates. For either procedure, more complicated estimates without using a linear approximation are possible, once the final fit of a model is available. Then, such theoretically meaningful factor score estimates and possibly their covariance matrix estimate can be used to study, through structural equation analysis, the relationships between the vocational interest factors and the personality factors, for example, based on the so-called big-five theory, as originally intended by Dumenci (1993).

This type of study using factor score estimates, sketched above but not discussed here in detail, enables the separate analysis and model-checking of the measurement and structural parts of the overall model and allows the incorporation of nonlinear measurement models. This approach also makes the nonlinear exploratory model building simpler by concentrating on a rather small number of variables at a time. When, as in this example, the theory suggests a particular number of factors but not a particular form of relationship, the ability to consider a broad class of nonlinear measurement models allows the extraction of much more information about the underlying factors from the observed variables than the analysis restricted to linearity. For applied problems such as this example, the nonlinear factor analysis can provide alternative

procedures for fitting, building and checking models supported by the theory and can potentially lead to new discovery. In other problems where the theory suggests a particular nonlinear model, the ELM or ACL procedure described in this paper can be used to fit the model without resorting to an approximate linear model nor an unwelcome transformation. The nonlinear factor analysis also increases the available tools for analyzing, exploring, and describing multivariate data and can possibly change the interpretation and the conceptualization of the latent variable modeling from those mostly based on correlations and covariances to those based on regression-like equations with measurement error. This paper has shown that factor analysis can be a sound statistical method, that nonlinear factor analysis and nonlinear structural equation analysis are possible and useful, that the ELM and ACL approaches can form bases for developing statistical procedures for such analyses, and that more statistical investigation in this area is desired.

## APPENDIX

The ELM and ACL algorithms are presented here using explicit computational formulas. Some vector-matrix operation notation is used. For a  $p \times p$  symmetric  $\mathbf{A}$ , let  $\text{vech } \mathbf{A}$  be the  $2^{-1}p(p+1) \times 1$  vector listing the elements of  $\mathbf{A}$  on or below the diagonal beginning with the first column. Let  $\mathbf{K}_p$  be the  $p^2 \times 2^{-1}p(p+1)$  matrix such that  $\text{vec } \mathbf{A} = \mathbf{K}_p \text{vech } \mathbf{A}$ . Also, define  $\mathbf{K}_p^+ = (\mathbf{K}_p' \mathbf{K}_p)^{-1} \mathbf{K}_p'$ , so that  $\text{vech } \mathbf{A} = \mathbf{K}_p^+ \text{vec } \mathbf{A}$ . For a  $p \times p$  diagonal  $\mathbf{A}$  with  $\mathbf{a} = \text{diag } \mathbf{A}$  being the  $p \times 1$  vector of the diagonal elements, let the  $p^2 \times p$  matrix  $\mathbf{L}_p$  be such that  $\text{vec } \mathbf{A} = \mathbf{L}_p \mathbf{a}$ . Define  $\boldsymbol{\psi}$  to be the  $p \times 1$  vector consisting of the diagonal elements of  $\boldsymbol{\Psi}_{ee}$  and  $\boldsymbol{\Psi}_{uu}$ . Then,  $\text{vech } \boldsymbol{\Psi} = \mathbf{K}_p^+ \text{vec } \boldsymbol{\Psi} = \mathbf{K}_p^+ \mathbf{L}_p \boldsymbol{\psi}$ .

### Initial Estimate

Both ELM and ACL procedures require some initial values. An initial estimate  $\boldsymbol{\beta}^{(0)}$  of  $\boldsymbol{\beta}$  can be obtained by combining Amemiya's (1993a, b) single-equation instrumental variable estimates or by minimizing

$$\sum_{t=1}^n [\mathbf{Y}_t - \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta})]' [\mathbf{Y}_t - \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta})].$$

An initial estimate  $\boldsymbol{\Psi}^{(0)}$  can be obtained by applying unweighted least squares to

$$\mathbf{m}^{(0)} = \frac{1}{n} \sum_{t=1}^n \mathbf{r}_t^{(0)} \mathbf{r}_t^{(0)'},$$

where

$$\mathbf{r}_t^{(0)} = \mathbf{Y}_t - \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}^{(0)}).$$

That is, for the  $p \times 1$   $\boldsymbol{\psi}^{(0)}$  satisfying  $\text{vec } \boldsymbol{\Psi}^{(0)} = \mathbf{L}_p \boldsymbol{\psi}^{(0)}$ ,

$$\boldsymbol{\psi}^{(0)} = (\mathbf{C}^{(0)' \mathbf{C}^{(0)}})^{-1} \mathbf{C}^{(0)' } \text{vech } \mathbf{m}^{(0)},$$

where

$$\mathbf{C}^{(0)} = \mathbf{K}_{p-k}^+ \{ \mathbf{L}_{p-k}, \frac{1}{n} \sum_{t=1}^n (\mathbf{G}(\mathbf{X}_t; \boldsymbol{\beta}^{(0)}) \otimes \mathbf{G}(\mathbf{X}_t; \boldsymbol{\beta}^{(0)})) \mathbf{L}_k \}.$$

For the ELM method,  $\mathbf{f}_t^{(0)} = \mathbf{X}_t$  serves as the initial value for the factor scores.

### ELM Algorithm

The  $i$ th step of the iterative ELM procedure consists of the following:

(E1)  $\mathbf{f}_t^{(i)}$  is obtained by

$$\mathbf{f}_t^{(i)} = \mathbf{X}_t + \boldsymbol{\Psi}_{uu}^{(i-1)} \widehat{\mathbf{G}}_t^{(i-1)} \widehat{\boldsymbol{\Sigma}}_{tt}^{(i-1)-1} \mathbf{v}_t^{(i-1)},$$

where

$$\begin{aligned} \widehat{\mathbf{G}}_t^{(i-1)} &= \mathbf{G}(\mathbf{f}_t^{(i-1)}; \boldsymbol{\beta}^{(i-1)}), \\ \widehat{\boldsymbol{\Sigma}}_{tt}^{(i-1)} &= \left[ \mathbf{I}_{p-k}, -\widehat{\mathbf{G}}_t^{(i-1)} \right] \left( \frac{n-1}{n} \boldsymbol{\Psi}^{(i-1)} + \frac{1}{n} \mathbf{m}_{ZZ} \right) \\ &\quad \times \left[ \mathbf{I}_{p-k}, -\widehat{\mathbf{G}}_t^{(i-1)} \right]', \\ \mathbf{v}_t^{(i-1)} &= \mathbf{Y}_t - \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}^{(i-1)}) + \frac{1}{2} \mathbf{H}(\mathbf{f}_t^{(i-1)}; \boldsymbol{\beta}^{(i-1)}) \\ &\quad \times \text{vec } \boldsymbol{\Psi}_{uu}^{(i-1)}, \\ \mathbf{m}_{ZZ} &= \frac{1}{n-1} \sum_{t=1}^n (\mathbf{Z}_t - \bar{\mathbf{Z}})(\mathbf{Z}_t - \bar{\mathbf{Z}})', \\ \bar{\mathbf{Z}} &= \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t = (\bar{\mathbf{Y}}', \bar{\mathbf{X}}_t)', \end{aligned}$$

and  $\mathbf{G}$  and  $\mathbf{H}$  are defined in (14) and (15).

(E2)  $\boldsymbol{\beta}^{(i)}$  is obtained by minimizing

$$\sum_{t=1}^n \hat{\mathbf{v}}_t'(\boldsymbol{\beta}) \widehat{\boldsymbol{\Sigma}}_{tt}^{-1}(\boldsymbol{\beta}) \hat{\mathbf{v}}_t(\boldsymbol{\beta}),$$

where

$$\hat{\mathbf{v}}_t(\boldsymbol{\beta}) = \mathbf{Y}_t - \mathbf{g}(\mathbf{X}_t; \boldsymbol{\beta}) + \frac{1}{2} \mathbf{H}(\mathbf{f}_t^{(i)}; \boldsymbol{\beta}) \text{vec } \boldsymbol{\Psi}_{uu}^{(i-1)},$$

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}_{tt}(\boldsymbol{\beta}) &= \left[ \mathbf{I}_{p-k}, -\mathbf{G}(\mathbf{f}_t^{(i)}; \boldsymbol{\beta}) \right] \left( \frac{n-1}{n} \boldsymbol{\Psi}^{(i-1)} + \frac{1}{n} \mathbf{m}_{ZZ} \right) \\ &\quad \times \left[ \mathbf{I}_{p-k}, -\mathbf{G}(\mathbf{f}_t^{(i)}; \boldsymbol{\beta}) \right]'. \end{aligned}$$

(E3)  $\boldsymbol{\Psi}^{(i)}$  is obtained by applying generalized least squares to

$$\boldsymbol{\gamma}^{(i)} = \text{vech } \frac{1}{n} \sum_{t=1}^n \hat{\mathbf{v}}_t(\boldsymbol{\beta}^{(i)}) \hat{\mathbf{v}}_t'(\boldsymbol{\beta}^{(i)}).$$

For the  $p \times 1$   $\boldsymbol{\psi}^{(i)}$  satisfying  $\text{vec } \boldsymbol{\Psi}^{(i)} = \mathbf{L}_p \boldsymbol{\psi}^{(i)}$ ,

$$\boldsymbol{\psi}^{(i)} = (\widehat{\mathbf{C}}^{(i)' \widehat{\mathbf{V}}^{-1} \widehat{\mathbf{C}}^{(i)}})^{-1} \widehat{\mathbf{C}}^{(i)' \widehat{\mathbf{V}}^{-1} \boldsymbol{\gamma}^{(i)},$$

where

$$\widehat{\mathbf{C}}^{(i)} = \mathbf{K}_{p-k}^+ \{ \mathbf{L}_{p-k}, \frac{1}{n} \sum_{t=1}^n (\widehat{\mathbf{G}}_t^{(i)} \otimes \widehat{\mathbf{G}}_t^{(i)}) \mathbf{L}_k \},$$

$$\widehat{\mathbf{G}}_t^{(i)} = \mathbf{G}(\mathbf{f}_t^{(i)}; \boldsymbol{\beta}^{(i)}),$$

$$\widehat{\mathbf{V}} = \frac{2}{n^2} \mathbf{K}_{p-k}^+ \sum_{t=1}^n (\ddot{\boldsymbol{\Sigma}}_{tt}^{(i-1)} \otimes \ddot{\boldsymbol{\Sigma}}_{tt}^{(i-1)}) \mathbf{K}_{p-k}^+,$$

$$\begin{aligned} \ddot{\boldsymbol{\Sigma}}_{tt}^{(i-1)} &= \left[ \mathbf{I}_{p-k}, -\widehat{\mathbf{G}}_t^{(i)} \right] \left( \frac{n-1}{n} \boldsymbol{\Psi}^{(i-1)} + \frac{1}{n} \mathbf{m}_{ZZ} \right) \\ &\quad \times \left[ \mathbf{I}_{p-k}, -\widehat{\mathbf{G}}_t^{(i)} \right]'. \end{aligned}$$

The  $n^{-1} \mathbf{m}_{ZZ}$  modification in (E1), (E2), and (E3) is for numerical stability when taking inverse matrices in small samples. Step (E3) requires some modifications related to the parameter space for  $\boldsymbol{\Psi}$ . Because of the nonnegativity of variances, any negative element of  $\boldsymbol{\psi}^{(i)}$  in step (E3) should be replaced by zero, and the remaining elements should be reestimated by reduced generalized least squares. In addition, a practically useful upper bound modification can be developed based on the observation that, for the general nonlinear model with no linear relationship, the sample covariance matrix  $\mathbf{m}_{ZZ}$  estimates the sum of  $\boldsymbol{\Psi}$  and a positive definite matrix. First, the largest root  $\hat{\lambda}$  of  $|\boldsymbol{\Psi}^{(i)} - \lambda \mathbf{m}_{ZZ}| = 0$  is obtained. If  $\hat{\lambda} < 1 + \frac{1}{n}$ , then  $\boldsymbol{\Psi}^{(i)}$  is unchanged. If  $\hat{\lambda} \geq 1 + \frac{1}{n}$ , then we set  $\boldsymbol{\Psi}^{(i)} = \frac{1}{\hat{\lambda} - n^{-1}} \boldsymbol{\Psi}^{(i)}$ . This modification prevents  $\boldsymbol{\Psi}^{(i)}$  from becoming "too large" and retains the variability due to the factor part in  $\mathbf{m}_{ZZ} - \boldsymbol{\Psi}^{(i)}$ . The same upper bound modification and reestimation due to the negative estimates should also be incorporated in obtaining the initial estimate  $\boldsymbol{\Psi}^{(0)}$ .

If an estimate of the factor covariance matrix  $\Phi = V\{\mathbf{f}_t\}$  is needed after obtaining the ELM estimates  $\hat{\beta}$ ,  $\hat{\Psi}$  and  $\hat{\mathbf{f}}_t$ , then a simple unweighted estimator

$$\hat{\Phi} = \frac{1}{n} \sum_{t=1}^n \left[ (\hat{\mathbf{f}}_t - \bar{\mathbf{f}})(\hat{\mathbf{f}}_t - \bar{\mathbf{f}})' + \hat{\Psi}_{uu} \hat{\mathbf{G}}_t' \hat{\Sigma}_{tt}^{-1} \hat{\mathbf{G}}_t \hat{\Psi}_{uu} \right] - \hat{\Psi}_{uu}$$

can be used, where

$$\bar{\mathbf{f}} = \frac{1}{n} \sum_{t=1}^n \hat{\mathbf{f}}_t,$$

$$\hat{\mathbf{G}}_t = \mathbf{G}(\hat{\mathbf{f}}_t; \hat{\beta}),$$

$$\hat{\Sigma}_{tt} = \left[ \mathbf{I}_{p-k}, -\hat{\mathbf{G}}_t \right] \left( \frac{n-1}{n} \hat{\Psi} + \frac{1}{n} \mathbf{m}_{ZZ} \right) \left[ \mathbf{I}_{p-k}, -\hat{\mathbf{G}}_t \right]'$$

To guarantee the nonnegative definiteness of  $\hat{\Phi}$ , the modification of Amemiya (1985) should be incorporated.

**ACL Algorithm**

At the  $i$ th iteration of the ACL procedure, estimates of  $\beta$  and  $\Psi$  are obtained as follows:

(A1)  $\beta^{(i)}$  is obtained by minimizing

$$\sum_{t=1}^n \left[ \mathbf{Y}_t - \mathbf{v}_t^{(i-1)}(\beta) \right]' \Gamma_{tt}^{(i-1)-1} \left[ \mathbf{Y}_t - \mathbf{v}_t^{(i-1)}(\beta) \right],$$

where

$$\begin{aligned} \mathbf{v}_t^{(i-1)}(\beta) &= g(\mathbf{X}_t; \beta) + \frac{1}{2} \mathbf{H}(\mathbf{X}_t; \beta) \text{vec} \Psi_{uu}^{(i-1)} \\ &\quad - \mathbf{G}(\mathbf{X}_t; \beta) \Psi_{uu}^{(i-1)} \mathbf{m}_{XX}^{-1} (\bar{\mathbf{X}}_t - \mathbf{X}), \\ \Gamma_{tt}^{(i-1)} &= \Psi_{ee}^{(i-1)} + \tilde{\mathbf{G}}_t^{(i-1)} \left[ \Psi_{uu}^{(i-1)} - \Psi_{uu}^{(i-1)} \mathbf{m}_{XX}^{-1} \Psi_{uu}^{(i-1)} \right] \\ &\quad \times \tilde{\mathbf{G}}_t^{(i-1)'} + \frac{1}{n} \left[ \mathbf{I}_{p-k}, -\tilde{\mathbf{G}}_t^{(i-1)} \right] \mathbf{m}_{ZZ} \\ &\quad \times \left[ \mathbf{I}_{p-k}, -\tilde{\mathbf{G}}_t^{(i-1)} \right]', \\ \tilde{\mathbf{G}}_t^{(i-1)} &= \mathbf{G}(\mathbf{X}_t; \beta^{(i-1)}). \end{aligned}$$

(A2)  $\Psi^{(i)}$  is obtained by generalized least squares applied to

$$\mathbf{m}^{(i)} = \frac{1}{n} \sum_{t=1}^n \left[ \mathbf{r}_t^{(i)} \mathbf{r}_t^{(i)'} + \mathbf{A}_t^{(i)} \right],$$

where

$$\begin{aligned} \mathbf{r}_t^{(i)} &= \mathbf{Y}_t - \mathbf{v}_t^{(i-1)}(\beta^{(i)}), \\ \mathbf{A}_t^{(i)} &= \tilde{\mathbf{G}}_t^{(i)} \Psi_{uu}^{(i-1)} \mathbf{m}_{XX}^{-1} \Psi_{uu}^{(i-1)} \tilde{\mathbf{G}}_t^{(i)'} \\ &\quad + \mathbf{D}_t^{(i)} \left[ \sum_{s=1}^n \mathbf{D}_s^{(i)'} \Gamma_{ss}^{(i-1)-1} \mathbf{D}_s^{(i)} \right]^{-1} \mathbf{D}_t^{(i)'}, \\ \mathbf{D}_t^{(i)} &= \frac{\partial \mathbf{g}(\mathbf{X}_t; \beta^{(i)})}{\partial \beta'}. \end{aligned}$$

For the  $p \times 1 \psi^{(i)}$  satisfying  $\text{vec} \Psi^{(i)} = \mathbf{L}_p \psi^{(i)}$ ,

$$\psi^{(i)} = \left( \tilde{\mathbf{C}}^{(i)'} \tilde{\mathbf{V}}^{-1} \tilde{\mathbf{C}}^{(i)} \right)^{-1} \tilde{\mathbf{C}}^{(i)'} \tilde{\mathbf{V}}^{-1} \text{vech} \mathbf{m}^{(i)},$$

where

$$\tilde{\mathbf{C}}^{(i)} = \mathbf{K}_{p-k}^+ \left( \mathbf{L}_{p-k}, \frac{1}{n} \sum_{t=1}^n (\tilde{\mathbf{G}}_t^{(i)} \otimes \tilde{\mathbf{G}}_t^{(i)}) \mathbf{L}_k \right),$$

$$\tilde{\mathbf{V}} = \frac{2}{n^2} \mathbf{K}_{p-k}^+ \sum_{t=1}^n \left( \tilde{\Gamma}_{tt}^{(i-1)} \otimes \tilde{\Gamma}_{tt}^{(i-1)} \right) \mathbf{K}_{p-k}^+$$

$$\begin{aligned} \tilde{\Gamma}_{tt}^{(i-1)} &= \Psi_{ee}^{(i-1)} + \tilde{\mathbf{G}}_t^{(i)} \left[ \Psi_{uu}^{(i-1)} - \Psi_{uu}^{(i-1)} \mathbf{m}_{XX}^{-1} \Psi_{uu}^{(i-1)} \right] \tilde{\mathbf{G}}_t^{(i)'} \\ &\quad + \frac{1}{n} \left[ \mathbf{I}_{p-k}, -\tilde{\mathbf{G}}_t^{(i)} \right] \mathbf{m}_{ZZ} \left[ \mathbf{I}_{p-k}, -\tilde{\mathbf{G}}_t^{(i)} \right]'. \end{aligned}$$

An estimator of the factor covariance matrix  $\Phi$  can be obtained by  $\mathbf{m}_{XX} - \tilde{\Psi}_{uu}$  with modification suggested in Amemiya (1985). The terms involving  $n^{-1} \mathbf{m}_{ZZ}$  in (A1) and (A2) are included as a small order adjustment for numerical stability. The lower and upper bounds modifications used in (E3) should also be incorporated in (A2).

REFERENCES

AMEMIYA, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite? *Amer. Statist.* **39** 112–117.

AMEMIYA, Y. (1993a). Instrumental variable estimation for nonlinear factor analysis. In *Multivariate Analysis: Future Directions 2* (C. M. Cuadras and C. R. Rao, eds.) 113–129. North-Holland, Amsterdam.

AMEMIYA, Y. (1993b). On nonlinear factor analysis. *Proceedings of the ASA Social Statistics Section* 290–294.

AMEMIYA, Y. and ANDERSON, T. W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *Ann. Statist.* **18** 1453–1463.

AMEMIYA, Y. and FULLER, W. A. (1988). Estimation for the nonlinear functional relationship. *Ann. Statist.* **16** 147–160.

AMEMIYA, Y., PANTULA, S. G. and FULLER, W. A. (1987). The asymptotic distribution of some estimators for a factor analysis model. *J. Multivariate Anal.* **22** 51–64.

ANDERSON, T. W. and AMEMIYA, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *Ann. Statist.* **16** 759–771.

ANDERSON, T. W. and RUBIN, H. (1956). Statistical inference in factor analysis. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **5** 111–150. Univ. California Press, Berkeley.

ARMINGER, G. and MUTHÉN, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63** 271–300.

BARTLETT, M. S. (1953). Factor analysis in psychology as a statistician sees it. In *Uppsala Symposium on Psychological Factor Analysis* 23–24. Almqvist and Wiksell, Stockholm.

BROWNE, M. W. (1984). Asymptotically distribution-free methods for the analysis covariance structures. *British J. Math. Statist. Psych.* **37** 62–83.

BROWNE, M. W. (1987). Robustness of statistical inference in factor analysis and related models. *Biometrika* **74** 375–384.

- BROWNE, M. W. and SHAPIRO, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British J. Math. Statist. Psych.* **41** 193–208.
- CARROLL, R. J. and RUPPERT, D. (1982). A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear model. *Amer. Statist. Assoc.* **77** 878–882.
- CARROLL, R. J. and RUPPERT, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, New York.
- CARROLL, R. J., RUPPERT, D. and STEFANSKI, L. A. (1995). *Measurement Error in Nonlinear Models*. Chapman and Hall, New York.
- CHENG, C. and VAN NESS, J. W. (1999). *Statistical Regression with Measurement Error*. Arnold, London.
- DUMENCI, L. (1993). Personality from a broad perspective: Hierarchical item-level factor analysis of personality, temperament, and vocational interests. Ph.D. dissertation, Iowa State Univ. Ames, IA.
- ETEZADI-AMOLI, J. and MC DONALD, R. P. (1983). A second generation nonlinear factor analysis. *Psychometrika* **48** 315–342.
- FULLER, W. A. (1987). *Measurement Error Models*. Wiley, New York.
- GNANADESIKAN, R. and KETTENRING, J. R. (1984). A pragmatic review of multivariate methods in applications. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds.) 309–337. Iowa State Univ. Press, Ames, IA.
- HU, L., BENTLER, P. and KANO, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin* **112** 351–362.
- JENNRICH, R. I. and ROBINSON, S. M. (1969). A Newton–Raphson algorithm for maximum likelihood factor analysis. *Psychometrika* **34** 111–123.
- JÖRESKOG, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika* **32** 443–482.
- JÖRESKOG, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34** 183–202.
- JÖRESKOG, K. G. (1970). A general method for analysis of covariance structure. *Biometrika* **57** 239–252.
- JÖRESKOG, K. G. (1973). A general method for estimating a linear structural equation system. In *Structural Equation Models in the Social Sciences* (A. S. Goldberger and O. D. Duncan, eds.) 85–112. Seminar Press, New York.
- JÖRESKOG, K. G. (1977). Factor analysis by least squares and maximum likelihood methods. In *Statistical Methods for Digital Computers III* (K. Enslein, A. Ralston and H. S. Wilf, eds.) **13** 125–153. Wiley, New York.
- JÖRESKOG, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika* **59** 381–389.
- JÖRESKOG, K. G. and SÖRBOM, D. (1989). *LISREL 7: User's Reference Guide*. Scientific Software, Mooresville, IN.
- JÖRESKOG, K. G. and YANG, F. (1994). Non-linear structural equation models: The Kenny-Judd model with interaction effects. In *Advanced Structural Equation Modeling: Issues and Techniques* (G. A. Marcoulides and R. E. Schumacker, eds.) 57–88. Erlbaum, Mahwah, NJ.
- JÖRESKOG, K. G. and YANG, F. (1997). Estimation of interaction models using the augmented moment matrix: comparison of asymptotic standard errors. In *SoftStat '97: Advances in Statistical Software* (W. Bandilla and F. Faulbaum, eds.) **6** 467–478. Lucius, Stuttgart.
- KENNEY, D. A. and PJUDD, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin* **96** 201–210.
- LAWLEY, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edinburgh Sec. A* **60** 64–82.
- LAWLEY, D. N. (1941). Further investigations in factor estimation. *Proc. Roy. Soc. Edinburgh Sec. A* **61** 176–185.
- LAWLEY, D. N. (1943). The application of the maximum likelihood method to factor analysis. *British J. Psychology* **33** 172–175.
- MC ARDLE, J. J. and EPSTEIN, D. (1987). Latent growth curves within developmental structural equation models. *Child Development* **58** 110–133.
- MC DONALD, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika* **27** 397–415.
- MC DONALD, R. P. (1965). Difficulty factors and nonlinear factor analysis. *British J. Math. Statist. Psych.* **18** 11–23.
- MC DONALD, R. P. (1967a). Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika* **32** 77–112.
- MC DONALD, R. P. (1967b). Factor interaction in nonlinear factor analysis. *British J. Math. Statist. Psych.* **20** 205–215.
- MC DONALD, R. P. (1979). The simultaneous estimation of factor loadings and scores. *British J. Math. Statist. Psych.* **32** 212–228.
- MEREDITH, W. and TISAK, J. (1990). Latent curve analysis. *Psychometrika* **55** 107–122.
- MOOLJAART, A. and BENTLER, P. (1986). Random polynomial factor analysis. In *Data Analysis and Informatics* (E. Diday, Y. Escoufier, L. Lebart, J. Pages, Y. Schektman and R. Tomasone, eds.) **4** 241–250. North-Holland, Amsterdam.
- MUTHÉN, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika* **49** 115–132.
- PANTULA, S. G. and FULLER, W. A. (1986). Computational algorithms for the factor model. *Comm. Statist. Ser. B* **15** 227–259.
- SATORRA, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology* **22** 249–278.
- VAN HOUWELINGEN, J. C. (1988). Use and abuse of variance models in regression. *Biometrics* **44** 1073–1081.
- WALL, M. M. and AMEMIYA, Y. (2000). Estimation for polynomial structural equation models. *J. Amer. Statist. Assoc.* **95** 929–940.
- WALL, M. M. and AMEMIYA, Y. (2001). Generalized appended product indicator procedure for fitting polynomial structural models. *J. Educational and Behavioral Statist.* **26** 1–29.
- YALCIN, I. (1995). Nonlinear factor analysis. Ph.D. dissertation, Iowa State Univ. Ames, IA.
- YALCIN, I. and AMEMIYA, Y. (1995). Additive nonlinear factor analysis. Preprint 95-16, Statistical Laboratory, Iowa State Univ. Ames, IA.