T. W. Anderson's Contributions to the Study of Linear Statistical Relationship Models

by

Leon Jay Gleser
Purdue University

Technical Report # 88-6

Department of Statistics
Purdue University

March, 1988

T. W. ANDERSON'S CONTRIBUTIONS TO THE STUDY OF LINEAR STATISTICAL RELATIONSHIP MODELS

by

Leon Jay Gleser
Purdue University

ABSTRACT

This article discusses some of the major contributions of T. W. Anderson to the statistical formulation and analysis of linear statistical relationship (LSR) models. A secondary goal is to briefly introduce some LSR models, including errors-in-variables regression models, linear functional relationship models, linear structural relationship models and factor analysis models, and to describe theoretical and methodological problems that arise in analyzing such models.

T. W. Anderson's Contributions to the Study of Linear Statistical Relationship Models

Leon Jay Gleser

Department of Statistics

Purdue University

A large class of statistical models share the assumption that certain latent variables are linearly related, and that these variables can only be observed subject to random errors of measurement. In his 1982 Wald Lectures [106]*, T. W. Anderson called such models *linear statistical relationship models.*

The most basic linear statistical relationship (LSR) model is the *linear functional relationship model.* A collection of p-variate random vectors $X_1, X_2, \ldots, X_n$ obeys a linear functional relationship model if the mean vectors $u_1, u_2, \ldots, u_n$ of these observations lie in an r-dimensional hyperplane $\mathcal{H}$ in p-dimensional space, $r < p$. One can regard the components of the u_i's as being values of the latent variables, and the elements of

$$e_i = X_i - u_i$$

as being random errors of measurement. The random vectors e_i are usually assumed to be independent and identically distributed with mean vector $\mathbf{0}$ and covariance matrix Σ. The dimension r of the hyperplane $\mathcal{H}$ containing the u_i's is called the *rank* of the model.

The fact that the vectors u_i lie in a hyperplane of dimension r can be expressed parametrically in several equivalent ways. For example, many of Anderson's papers make use of the implicit form:

$$A\, u_i = \mathbf{a}, \quad i = 1, 2, \ldots, n, \tag{1}$$

*Numbers in square brackets refer to the bibliography of Anderson's publications included in *The Collected Papers of T.W. Anderson: 1943-1985* (G.P.H. Styan, ed.), to be published by John Wiley & Sons. An excerpt from this bibliography is given as an appendix to the present article.

where A is a $(p-r)\times p$ matrix of rank $p-r$, and $\mathbf{a}$ is a $(p-r)$-dimensional column vector. Alternatively, one can assume that

$$u_i = \Lambda f_i + \delta, \quad i = 1, 2, \ldots, n, \tag{2}$$

where Λ is a $p \times r$ matrix of rank r. This is the familiar model of factor analysis. Finally, $p-r$ elements of u_i can be expressed as a linear transformation of the remaining r elements; that is,

$$v_i = Bw_i + c, \quad i = 1, 2, \ldots, n, \tag{3}$$

where B is a $(p-r) \times r$ matrix of slopes, and v_i, w_i are subvectors of u_i. This is the *errors-in-variables linear regression model* widely studied in the statistical literature (see Fuller, 1987.). Note that model (3) is a special case of model (2), in which certain of the coordinates of Λ have specified values (0 or 1).

In each of the parametric forms (1), (2), (3) of the linear functional relationship model, the vectors u_i (or f_i or w_i) are unknown parameters of the model. Since these parameters are not usually of primary interest, they are called *incidental parameters.* The parameters A and $\mathbf{a}, \Lambda$ and δ, B and c which serve to identify the hyperplane $\mathcal{H}$ are called *structural parameters*, as is also the error covariance matrix Σ. When the rank r of the model is unspecified, it also is regarded as a structural parameter. The structural parameters are typically the primary focus of statistical inference.

Linear structural relationship models are related to linear functional relationship models in the same way that random factor (Model II) ANOVA models are related to fixed factor (Model I) ANVOA. That is, the vectors u_i, f_i, w_i in (1), (2), (3), respectively, are now regarded as being randomly sampled from some common population (distribution). The linear relationships (1), (2), or (3) consequently determine a parametric structure for the covariance matrix of the observations X_i. Psychometric factor analysis models are usually of the linear structural relationship type.

Both the implicit form (1) and the factor analysis form (2) of linear functional or structural relationships models are not identifiable. For example, the mathematical form of (1) is not changed by replacing A by AT and u_i by $T^{-1}u_i$ for any p-dimensional non-singular matrix T. Similarly, the transformations $\Lambda \rightarrow \Lambda M, f_i \rightarrow M^{-1}f_i$ do not change the form of (2). The errors-in-variables linear regression model (3) appears identifiable in this sense, but there is an implicit restriction (noted by Anderson and Sawa [72]) resulting from specifying which elements of u_i form the vector w_i. Any identifying restrictions imposed to remove the parametric indeterminacies in (1) or (2), or to choose which elements of u_i appear in w_i in (3), are not inherent in the model, but rather are imposed externally by the investigator. (In addition a structure must often be imposed upon the error covariance matrix Σ to prevent the linear relationships among the elements of u_i from being confounded with correlations among the elements of the measurement error vectors e_i.) Different contexts of application have generated different identifying restrictions, thus making communcations among specialists in the applications of LSR models difficult and obscuring the essential similarity of the models used.

In bivariate ($p = 2$) LSR models of rank $r = 1$, Anderson [69] suggests use of the angle θ made by the hyperplane $\mathcal{H}$ (here, a line) with one of the axes as a structural parameter of the model. (The choice of the axis to be used is immaterial.) This angle is clearly an intrinsic property of the model; generalizations of this suggestion to higher-dimensional ($p > 2, r \geq 1$) LSR models are straightforward. When the error covariance matrix Σ is spherical ($\Sigma = \sigma^2 I$) and normality assumptions hold, Anderson [69] shows that the exact distribution of the maximum likelihood estimator $\hat{\theta}$ of θ can be obtained from the distribution of $\hat{\theta}$ when $\theta = 0$. The distribution of the maximum likelihood estimator $\hat{B}$ of the slope B in the errors-in-variables linear regression model (3) is then easily obtained from that of $\hat{\theta}$, and confidence sets for both θ and B can be constructed. (See also [72].) Obtaining the exact distribution of $\hat{B}$ from that of $\hat{\theta}$ offers a considerable simplification when compared to previous direct derivations of the distribution of $\hat{B}$. However, these exact distributions do not have a simple closed form. This lack of mathematical simplicity is common to the exact distributions of virtually all estimators proposed for the structural parameters of LSR models. Consequently, evaluation and comparison of the properties of such estimators require some type of distributional approximation or expansion, even in the simple bivariate case studied in [69]. Anderson, with various collaborators, has developed several types of distributional expansion for estimators in bivariate ($p = 2$) linear functional and structural relationship models. See [66], [67], [69], [72], [73], [83], [90], [97], [100], [101], [104] and [105]. The papers [69] and [90] provide useful summaries of this topic.

In LSR models, ordinary least squares estimators of the structural parameters in any of the forms (1), (2) or (3) of such models are biased (and also inconsistent). However, the familiarity of the least squares estimator (and the existence of plentiful software for computing it) leads many investigators to want to use this estimator. On the other hand, theoretical considerations favor use of the maximum likelihood estimator. To settle the question of which estimator to use, the usual approach would be to compare expected mean square errors of these estimators. Unfortunately, the maximum likelihood estimators fail to even have a well-defined mean, and have infinite mean squared error. Even so, large-sample ($n \to \infty$) comparisons favor the MLE, but need not be relevant to finite-sample situations. In [69], Anderson makes an interesting and important contribution to this question by comparing the distributional concentrations of the least squares and maximum likelihood estimators of the slope B in the functional linear errors-in-variables regression model (3) when $p = 2, r = 1$, and the errors are normally distributed with covariance matrix $\Sigma = \sigma^2 I$. Subject to the accuracy of the distributional expansions used for the least squares estimator $\tilde{B}$ and the maximum likelihood estimator $\hat{B}$, it is shown that

$$P\{|\hat{B} - B| \leq x\} \leq P\{|\tilde{B} - B| \leq x\}$$

for all $x > 0$, except when the magnitude of the slope B is small. This supports the intuitive belief (based also in part on practical experience) that the maximum likelihood estimator is the superior estimator.

As already noted, in the bivariate case ($p = 2, r = 1$) of LSR models the angle θ made by the line (hyperplane) $\mathcal{H}$ with an axis is more intrinsically meaningful as a structural parameter than the slope B in (3). In a joint paper [107] with C. Stein and A. Zaman, Anderson shows that under the usual normality assumptions (and with $\Sigma = \sigma^2 I$) the maximum likelihood estimator $\hat{\theta}$ of θ is the best equivariant (under rotation) estimator of this parameter under a certain invariant loss function. From this result, it follows that $\hat{\theta}$ is admissible. This is the only exact decision-theoretic optimality result for estimation of a structural parameter in LSR models obtained to date.

In linear functional relationship models, maximum likelihood estimators of the structural parameters do not necessarily exist (see [23], and also Solari, 1969; Willassen, 1979). Even if they exist, these estimators need not be consistent. These difficulties arise because the number of incidental parameters (which must be accounted for in maximizing the likelihood) increases proportionally to the sample size n. A variety of theoretical approaches have been suggested for producing estimators in such contexts. Two such approaches were suggested by Anderson and Herman Rubin in [23]: (a) to maximize the likelihood of a certain non-sufficient reduction of the data (the sample covariance matrix) whose likelihood depends on the incidental parameters only through a finite-dimensional vector function of these parameters, and (b) to use the maximum likelihood estimator of the structural parameters in the corresponding linear *structural* relationship model. Proposal (a) is complicated by the fact that the sample covariance matrix has a non-central Wishart distribution (see [3], [4]), so that the likelihood does not possess a convenient mathematical form. Proposal (b), which is one of the earliest examples of a parametric empirical Bayes approach to estimation, has strongly influenced my own approach to constructing estimators in linear functional relationship models (see Gleser, 1985, for example). Recent research (Gleser, 1983; Nussbaum, 1984; Bickel and Ritov, 1987) shows that, under normality assumptions, the structural-model maximum likelihood estimators of the structural parameters of linear functional relationship models are best asymptotic normal and asymptotically minimax. The estimators of Anderson and Rubin's proposal (a) have similar large-sample properties.

We have noted that even for LSR models of the form (3) identification of the structural parameters requires knowledge of the error covariance matrix Σ, or the specification of a restricted structure for this matrix (e.g., that $\Sigma = \sigma^2 I$). Alternatively, one can try to estimate Σ. Two widely used approaches for providing estimators of Σ are to use replications (several X_i's with the same value of u_i), or to use instrumental variables (covariates Z_i which are "correlated" with the latent vectors u_i, but not with the error vectors e_i). Both approaches yield a classical multivariate linear regression model in which a matrix $\triangle$ of slopes satisfies unknown linear constraints. Consequently, the columns X_i of the least squares estimators $\hat{\triangle}$ of $\triangle$ obey a linear functional relationship model, and the residual covariance matrix provides an independent consistent estimator of the error covariance matrix Σ of these X_i's. Statistical inference (maximum likelihood estimators, likelihood ratio tests of hypotheses) for such a model under normality assumptions is thoroughly treated by Anderson in [14]. Unfortunately, this important paper has often been overlooked by researchers, with the consequence that its main results have been repeatedly rediscovered. (For a comment on one such case, see Gleser, 1983.)

The results in [14] apply only to linear functional relationship models. Similar results are obtained for linear structural relationship models by Anderson in [108], who indicates that his results had been obtained in 1946, but left unpublished. (See also Anderson, Anderson, and Olkin, 1986). The procedures obtained in [14] and [108] have important applications in psychometric mental test theory (see Healy, 1979) and in statistical genetics. In [14], and many of his later papers, Anderson explicitly points out the application of the results in [14] to estimation and fitting problems arising in studies of systems of linear stochastic equations by econometricians. (See also [8], [11], written with Herman Rubin.)

Because estimators and test statistics in LSR models have complex distributions, large-sample distributions are usually used to construct confidence regions for structural parameters and rejection regions for tests of hypothesis. These tests and estimators (particularly likelihood ratio tests and maximum likelihood estimators) are derived from the principal components and corresponding eigenvectors of sample covariance matrices, or more generally (see [14], [108]) from roots and corresponding eigenvectors of determinantal equations in sample covariance matrices. Anderson has made basic contributions to the determination of the asymptotic distributions of these roots and eigenvectors ([7], [11], [13], [23], [37]).

Earlier in this article it was noted that LSR models can be expressed in various mathematical forms, and that statistical identifiability of the parameters of such models require imposing restrictions on these parameters that are not intrinsic to the model. Different contexts of application of LSR models have led to different identifying restrictions. Consequently, specialists in the use of LSR models in various fields of applications have not always been aware that they are working on common problems, and considerable duplication of effort has occured. In his 1982 Wald lectures [106], Anderson set himself the task of expounding the relationships among the various linear statistical relationship models used by statisticians and by specialists in econometrics, psychometrics and other scientific fields. This paper succeeds brilliantly in providing an elegant unifying framework for such models, and in summarizing several decades of research. The debt owed by the present article to [106] is apparent.

For reasons of space, it has not been possible mention all of Ted Anderson's contributions to the statistical formulation and analysis of linear statistical relationship models. Enough has been said, hopefully, to indicate his seminal influence on this area of research, and his continuing leadership in its development.

REFERENCES

Anderson, B.M., Anderson, T.W. and Olkin, I. (1986). Maximum likelihood estimators and likelihood ratio criteria in multivariate components of variance. *Annal of Statistics*, **14**, 405-417.

Bickel, P.J. and Ritov, Y. (1987). Efficient estimators in the errors in variables model. *Annals of Statistics*, **15**, 513-540.

Fuller, W.A. (1987). *Measurement Error Models*, John Wiley, New York.

Gleser, L.J. (1983). Functional, structural and ultrastructural errors-in-variables models. *Proceedings Business and Economic Section American Statistical Association*, 57-66.

Gleser, L.J. (1985). A note on G.R. Dolby's ultrastructural model. *Biometrika*, **72**, 117-124.

Healy, J.D. (1979). On Kristoff's test for a linear relation between true scores of two measures. *Psychometrika*, **44**, 235-238.

Ibragimov, I.A. and Has'minskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory*, Springer, New York.

Nussbaum, M. (1984). An asymptotic minimax risk bound for estimation of a linear functional relationship. *Journal of Multivariate Analysis*, **14**, 300-314.

Solari, M.E. (1969). The 'maximum likelihood solution' to the problem of estimating a linear functional relationship. *Journal of the Royal Statistical Society, Series B*, **31**, 372-375.

Willassen, Yngve (1979). Two clarifications on the likelihood surface in functional models. *Journal of Multivariate Analysis*, **9**, 138-149.

APPENDIX

The following is a list of publications by T.W. Anderson cited in this report (see footnote on page 1).

[1] Some significance tests for normal bivariate distributions (with D.S. Villars), *Annals of Mathematical Statistics*, **14** (1943), 141-148.

[3] Some extensions of the Wishart distribution (with M.A. Girshick), *Annals of Mathematical Statistics*, **15** (1944), 345-357. [Correction: **35** (1964), 923-924.]

[4] The non-central Wishart distribution and certain problems of multivariate statistics, *Annals of Mathematical Statistics*, **17** (1946), 409-431. [Correction: **35** (1964), 923-924.]

[7] The asymptotic distributions of the roots of certain determinantal equations, *Journal of the Royal Statistical Society, Series B*, **10** (1948), 132-139.

[8] Estimation of the parameters of a single equation in a complete system of stochastic equations (with Herman Rubin), *Annals of Mathematical Statistics*, **20** (1949), 46-63. [Reprinted in *Readings in Econometric Theory* (J. Malcolm Dowling and Fred R. Glahe, *eds*), Colorado Associated University Press, 1970, 358-375.]

[11] The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equation (with Herman Rubin), *Annals of Mathematical Statistics*, **21** (1950), 570-582. [Reprinted in *Readings in Econometric Theory* (J. Malcolm Dowling and Fred R. Glahe, *eds.*), Colorado Associated University Press, 1970, 376-388.]

[13] The asymptotic distribution of certain characteristic roots and vectors, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, *ed.*), University of California Press, Berkeley, 1951, 103-130.

[14] Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics*, **22** (1951), 327-351. [Correction: *Annals of Statistics*, **8** (1980), p. 1400.]

[23] Statistical inference in factor analysis (with Herman Rubin), *Proceedings of the*

Third Berkeley Symposium on Mathematical Statistics and Probability (Jerzy Neyman, ed.), University of California Press, 1956, **5**, 111-150.

[37] Asymptotic theory for principal component analysis, *Annals of Mathematical Statistics*, **34** (1963), 122-148.

[66] Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions (with Takamitsu Sawa), *Econometrica*, **41** (1973), 683-714.

[67] An asymptotic expansion of the distribution of the limited information maximum likelihood estimate of a coefficient in a simultaneous equation system, *Journal of the American Statistical Association*, **69** (1974), 565-573. [Correction: **71** (1976), 1010.]

[69] Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics (with discussion), *Journal of the Royal Statistical Society, Series B*, **38** (1976), 1-36.

[72] Two-stage least squares: In which direction should the residuals be minimized? (with Takamitsu Sawa), *Journal of the American Statistical Association*, **72** (1977), 187-191.

[73] Asymptotic expansions of the distributions of estimates in simultaneous equations for alternative parameter sequences, *Econometrica*, **45** (1977), 509-518.

[83] Evaluation of the distribution function of the two-stage least squares estimate (with Takamitsu Sawa), *Econometrica*, **47** (1979), 163-182.

[90] Recent results on the estimation of a linear functional relationship, *Multivariate Analysis-V* (P.R. Krishnaiah, *ed.*), North-Holland Publishing Co., Amsterdam, 1980, 23-34.

[97] Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient (with Takamitsu Sawa), *Journal of the Royal Statistical Society, Series B*, **44** (1982), 52-62.

[100] Evaluation of the distribution function of the limited information maximum like-

lihood estimator (with Naoto Kunitomo and Takamitsu Sawa), *Econometrica*, **50** (1982), 1009-1027.

[101] Some recent developments on the distributions of single-equation estimators, *Advances in Econometrics* (Werner Hildenbrand, *ed.*), Cambridge University Press, Cambridge, 1982, 109-122.

[104] The numerical values of some key parameters in econometric models (with Kimio Morimune and Takamitsu Sawa), *Journal of Econometrics*, **21** (1983), 229-243.

[105] Comparison of the densities of the TSLS and LIMLK estimators (with Naoto Kunitomo and Takamitsu Sawa), *Global Econometrics, Essays in Honor of Lawrence R. Klein* (F. Gerard Adams and Bert G. Hickman, *eds.*), MIT Press, Cambridge, Massachusetts, 1983, 103-124.

[106] Estimating linear statistical relationships (1982 Abraham Wald Memorial Lectures), *Annals of Statistics*, **12** (1984), 1-45.

[107] Best invariant estimation of a direction parameter (with Charles Stein and Asad Zaman), *Annals of Statistics*, **13** (1985), 526-533.

[108] Components of variance in MANOVA, *Multivariate Analysis-VI* (P.R. Krishnaiah, *ed.*), North-Holland, Amsterdam, 1985, 1-8.