KM 16(1984)
pag 97-108

MAXIMUM LIKELIHOOD ESTIMATION OF SUM-CONSTRAINED LINEAR MODELS WHEN SAMPLES ARE SMALL*

by

P.M.C. de Boer and R. Harkema**

Abstract

Maximum likelihood procedures for estimating sum-constrained models like demand systems, brand choice models and so on, break down or produce very unstable estimates when the number of categories is large as compared with the number of observations available. In empirical studies this difficulty is mostly resolved by postulating the contemporaneous covariance matrix of the dependent variables at time t to equal $\sigma^2(I_n - n^{-1}\iota_n\iota_n')$. In this paper we develop a maximum likelihood procedure based on a contemporaneous covariance matrix which allows that the variances per category may be different, while the number of observations required is substantially less than the number that would be required in the case of a completely unrestricted contemporaneous covariance matrix.

* This article is a highly condensed version of de Boer and Harkema (1983). For proofs and some special cases the reader is referred to the original publication, which may be requested from the authors.

** Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, Tel. 010-525511, ext. 3033/3093.

1. INTRODUCTION

Sum-constrained models, i.e., models in which subsets of the dependent variables sum to a fixed number, occur in almost every field of applied econometric research. In demand analysis the amounts spent on the categories of consumer goods and services that are distinguished add up to total expenditure, in production theory the cost shares of the various factors of production add up to unity, in marketing analysis the probabilities that a specific brand will be chosen add up to unity, in international trade the flows of exports from a specific country to different destinations add up to total exports, and so on. Sum-constrained linear models may generally be represented by means of the following system of seemingly unrelated regression equations

$$(1.1) \qquad y_i = Z_i\beta_i + u_i \qquad i = 1, \ldots, n$$

where y_i denotes a $T\times 1$ vector of observations on the i-th dependent variable, Z_i denotes a $T\times k_i$ matrix of observations on a set of k_i explanatory variables which are specific for the i-th dependent variable, β_i is a $k_i\times 1$ vector of unknown parameters to be estimated, u_i is a $T\times 1$ vector of zero-mean disturbances and n, the number of categories that are distinguished, is supposed to be larger than 2. The adding-up restrictions imply that the vectors of dependent variables y_i add up to a vector of fixed numbers m. Hence

$$(1.2) \qquad \sum_{i=1}^{n} y_i = m$$

Summing (1.1) over i and taking expectations it follows that

$$(1.3) \qquad \sum_{i=1}^{n} u_i = 0$$

and

$$(1.4) \qquad \sum_{i=1}^{n} Z_i\beta_i = m$$

Evidently, (1.3) reflects the wellknown fact that the vectors of disturbances in sum-constrained linear models are linearly dependent. The restrictions (1.4) are usually accomodated by imposing linear constraints on the vectors of parameters β_i. To give but a few examples, in the Rotterdam model (see e.g. Theil (1975)) and in the simplified version of the Almost Ideal Demand System of Deaton and Muellbauer (1980), the vectors of parameters β_i add up to the unit vector, while in the Multiplicative Competitive Interaction Model of Nakanishi and Cooper (1974) the vectors of parameters β_i are supposed to be the same for all brands.

A major difficulty in estimating sum-constrained linear models is caused by the fact that the method of maximum likelihood is very demanding with respect to the number of observations that is required. Maximum likelihood procedures frequently break down or produce very unstable estimates because of lack of data even when only a moderate number of categories is considered. Laitinen (1978), for example, has shown that the minimum number of observations required for maximum likelihood estimation of the Rotterdam model equals 2n. In applied research this problem is usually resolved by imposing far-reaching restrictions on the contemporaneous covariance matrix of the disturbances. Denoting this matrix by Ω, McGuire et al. (1968), Solari (1971), Deaton (1975), and Deaton and Muellbauer (1980), for example, impose $\Omega = \sigma^2(I_n - n^{-1}\iota_n\iota_n')$. In developing his theory of rational random behavior Theil (1971, 1974, 1980) proposes to impose $\Omega = -\sigma^2 S$, where S denotes the matrix of Slutsky-coefficients. Both approaches have in common that, apart from a constant of proportionality, the structure of the contemporaneous covariance matrix is completely specified beforehand.

The purpose of the present paper is to introduce a more flexible specification of the contemporaneous covariance matrix which allows for n parameters to be estimated freely and possesses the attractive property that the number of observations that is required need not be larger than $\max_i\{k_i + 1\}$. For the case considered by Laitinen this means that the minimum number of observations required is only n+2 instead of 2n.

The plan of the paper is as follows. In Section 2 we introduce the specification of the contemporaneous covariance matrix and derive the corresponding maximum likelihood estimators and their asymptotic distribution. In Section 3 we present an outline of the estimation procedure for the covariance matrix and in Section 4 we summarize our findings and discuss some extensions.

2. MAXIMUM LIKELIHOOD ANALYSIS

We start by rewriting (1.1) according to

$$y_i = X_i\beta + u_i \qquad i = 1, \ldots, n \tag{2.1}$$

where y_i and u_i are as before, X_i denotes a $T\times k$ matrix containing the columns of Z_i and $k-k_i$ zero columns and β is the $k\times 1$ vector of parameters that is obtained by writing the vectors β_i in stacked form. From (1.2)-(1.4) it follows that (2.1) is subject to the following constraints

$$\sum_{i=1}^{n} y_i = m \qquad \sum_{i=1}^{n} X_i\beta = m \qquad \sum_{i=1}^{n} u_i = 0 \tag{2.2}$$

As said before, the constraints $\Sigma_{i=1}^{n} X_i\beta = m$ are usually accomodated by imposing linear constraints on the vector of parameters β. Therefore we impose[1]

$$R\beta = r \tag{2.3}$$

where R denotes a $q\times k$ matrix of full row rank and r represents a $q\times 1$ vector. Of course, (2.3) may also represent other linear constraints like those resulting from homogeneity and symmetry conditions in linear demand systems.

As regards the vectors of disturbances u_i, we assume that the vector $u' = [u_1' \ldots u_n']$ is distributed according to a nT-variate normal distribution with zero mean and variance-covariance matrix $[\Omega_n \otimes I_T]$, Ω_n being a positive semi-definite symmetric matrix of rank (n-1). More specifically, in the present paper Ω_n will be specified as follows

$$\Omega_n = D_n - d^{-1}\delta_n\delta_n' \tag{2.4}$$

with

1. For expository reasons all restrictions that may exist with respect to the vector of parameters β have been collected in (2.3). From a computational viewpoint, however, it may be advantageous to eliminate all restrictions right away from the start.

$$D_n = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{bmatrix} \qquad \delta_n' = [d_1 \ldots d_n] \qquad d = \sum_{i=1}^{n} d_i$$

In scalar notation we may write (2.4) as

$$\omega_{ii} = \mathrm{var}(u_{it}) = d_i - \frac{d_i^2}{d} \qquad i = 1, \ldots, n$$

$$\omega_{ij} = \mathrm{cov}(u_{it}, u_{jt}) = -\frac{d_i d_j}{d} \qquad i,j = 1, \ldots, n;\ i \neq j$$

The specification (2.4) originally arose from a straightforward generalization of the specification $\Omega_n = \sigma^2(I_n - n^{-1}\iota_n\iota_n')$. Recently, however, Don (1984) has shown that the specification (2.4) corresponds to the least informative error distribution in the sense of having maximum entropy within the class of all error distributions with finite variances.

From (2.4) one easily verifies that $\Omega_n \iota = 0$. As a consequence the density function of the vector u will be degenerate as it should be. Barten (1969), however, has shown that this problem may be handled by simply deleting one category. Choosing without any loss of generality the last one, we delete the last row and column of Ω_n. Denoting the resulting matrix by Ω_{n-1}, straightforward matrix calculation shows that

$$\Omega_{n-1}^{-1} = \begin{bmatrix} d_1^{-1} + d_n^{-1} & d_n^{-1} & \cdots & d_n^{-1} \\ d_n^{-1} & d_2^{-1} + d_n^{-1} & & d_n^{-1} \\ \vdots & \vdots & & \vdots \\ d_n^{-1} & d_n^{-1} & \cdots\ d_{n-1}^{-1} & + d_n^{-1} \end{bmatrix} \tag{2.5}$$

In addition, it can be shown that

$$|\Omega_{n-1}| = d^{-1} \prod_{i=1}^{n} d_i \tag{2.6}$$

and that Ω_{n-1} will be positive-definite in the following two mutually exclusive and exhaustive cases: (i) all d_i's are positive and (ii) at most one d_i is negative with d being negative as well.

From our assumptions about the distribution of the vector u and the restrictions (2.2) one easily verifies that the loglikelihood function may be written as

$$\text{(2.7)}\qquad \begin{aligned}\log \ell(y_1 \ldots y_{n-1}) = &- \frac{T(n-1)}{2} \log 2\pi - \frac{T}{2} \log\left[d^{-1} \prod_{i=1}^{n} d_i\right] + \\ &- \tfrac{1}{2} \sum_{i=1}^{n} \left[d_i^{-1}(y_i - X_i\beta)'(y_i - X_i\beta)\right]\end{aligned}$$

From (2.7) it is clear that the loglikelihood function and hence the resulting maximum-likelihood estimators are invariant with respect to the category that is deleted as it should be. The restricted maximum-likelihood estimators may be obtained from the following system of equations

$$\text{(2.8)}\qquad \hat{\beta} = \bar{\beta} - CR'(RCR')^{-1}(R\bar{\beta} - r)$$

$$\text{(2.9)}\qquad \hat{d}_i - \frac{\hat{d}_i^2}{\hat{d}} = \frac{\hat{u}_i'\hat{u}_i}{T} \qquad i = 1, \ldots, n$$

where[2]

$$\bar{\beta} = \left[\sum_{i=1}^{n} \hat{d}_i^{-1} X_i'X_i\right]^{-1}\left[\sum_{i=1}^{n} \hat{d}_i^{-1} X_i'y_i\right] = \left[\sum_{i=1}^{n} X_i'X_i\right]^{-1}\left[\sum_{i=1}^{n} X_i'y_i\right]$$

$$C = \left[\sum_{i=1}^{n} \hat{d}_i^{-1} X_i'X_i\right]^{-1}$$

and

$$\hat{u}_i = y_i - X_i\hat{\beta}$$

Evidently, (2.8)-(2.9) constitutes a system of highly nonlinear equations. It can be shown, however, that, conditional upon $\hat{u}_i'\hat{u}_i$, solving (2.9) can be accomplished by means of a numerical search procedure for the unique real root of an equation in only one variable. The full system (2.8)-(2.9) may therefore be solved by applying the following iterative procedure:

(i) Choose initial values $\hat{d}_i^0$ for d_i, for example $\hat{d}_i^0 = 1$ $(i = 1,\ldots,n)$;

(ii) Calculate $\hat{\beta}^0$ according to (2.8) and $\hat{u}_i^0$ according to $\hat{u}_i^0 = y_i - X_i\hat{\beta}^0$;

(iii) Obtain first-round estimates $\hat{d}_i^1$ by solving (2.9) conditional upon

2. Note that $\bar{\beta}$ is the estimator that is obtained by writing the ordinary least-squares estimators for each separate equation in stacked form.

$\hat{u}_i'\hat{u}_i = (\hat{u}_i^0)'\hat{u}_i^0$;

(iv) Calculate the first-round estimates $\hat{\beta}^1$ according to (2.8) and $\hat{u}_i^1$ according to $\hat{u}_i^1 = y_i - X_i\hat{\beta}^1$ and so forth, until convergence.

Under very mild conditions, Sargan (1964) and Oberhofer and Kmenta (1974) have proved that the above procedure actually converges to a solution of the system (2.8)-(2.9). Under suitable regularity conditions, it follows from standard arguments that $\hat{\beta}$ and $\hat{d}_i$ are consistent estimators for β and d_i and that $\sqrt{T}(\hat{\beta} - \beta)$ is asymptotically distributed according to a k-variate normal distribution with zero mean and variance-covariance matrix PQP where

$$P = \operatorname*{plim}_{T\to\infty} T[C - CR'(RCR')^{-1}RC]$$

$$Q = \operatorname*{plim}_{T\to\infty} T^{-1}[C^{-1} - \hat{d}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} X_i'X_j] \tag{2.10}$$

Finally, it should be noted that the assumption of normal disturbances is not crucial for establishing the above asymptotic results. Under appropriate conditions $\hat{\beta}$ and $\hat{d}_i$ may also be interpreted as quasi-maximum likelihood estimators without affecting these results.

3. ESTIMATING THE COVARIANCE MATRIX: AN OUTLINE

In this section we summarize how to obtain the estimates $\hat{d}_i$ for the covariance parameters in stage (iii) of the iterative procedure as described in Section 2. Without any loss of generality we assume that

$$T^{-1}\hat{u}_n'\hat{u}_n \overset{\text{def}}{\equiv} \hat{\alpha}_n \geq T^{-1}\hat{u}_i'\hat{u}_i \overset{\text{def}}{\equiv} \hat{\alpha}_i \qquad i = 1, \ldots, n-1$$

Apart from border-cases, which will not be treated here, it appears that three cases must be distinguished. Below, we summarize all three cases. First, we define

$$f_1(d) = \sum_{i=1}^{n} \left(1 - \frac{4\hat{\alpha}_i}{d}\right)^{\frac{1}{2}} - (n-2) \qquad \text{for } d \geq 4\hat{\alpha}_n \tag{3.2}$$

$$f_2(d) = \sum_{i=1}^{n-1} \left(1 - \frac{4\hat{\alpha}_i}{d}\right)^{\frac{1}{2}} - \left(1 - \frac{4\hat{\alpha}_n}{d}\right)^{\frac{1}{2}} - (n-2) \quad \text{for } d \geq 4\hat{\alpha}_n \text{ or } d < 0 \tag{3.3}$$

and

$$\hat{\gamma} = f_1(4\hat{\alpha}_n) = f_2(4\hat{\alpha}_n) = \sum_{i=1}^{n-1} \left(1 - \frac{\hat{\alpha}_i}{\hat{\alpha}_n}\right)^{\frac{1}{2}} - (n-2). \tag{3.4}$$

Then, we have

Case 1: $\hat{\alpha}_n < \sum_{i=1}^{n-1} \hat{\alpha}_i$ and $\hat{\gamma} \leq 0$

We have to solve on $[4\hat{\alpha}_n, \infty)$

$$f_1(\hat{d}) = 0$$

It can be shown that the graph of $f_1(d)$ looks as follows

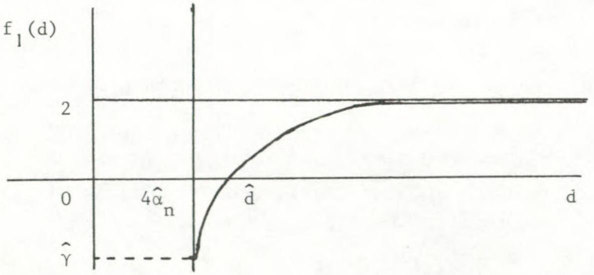

Figure 1

The solution $\hat{d}$ has to be substituted into

$$\hat{d}_i = \frac{\hat{d}}{2} - \frac{\hat{d}}{2}\left(1 - \frac{4\hat{\alpha}_i}{\hat{d}}\right)^{\frac{1}{2}} \qquad i = 1, \ldots, n$$

in order to obtain the solution in terms of the covariance parameters $\hat{d}_i$, which are all positive.

<u>Case 2</u>: $\hat{\alpha}_n < \sum_{i=1}^{n-1} \hat{\alpha}_i$ and $\hat{\gamma} > 0$

We have to solve on $(4\hat{\alpha}_n, \infty)$

$$f_2(\hat{d}) = 0$$

Graphically, the shape of the function $f_2(d)$ looks as follows

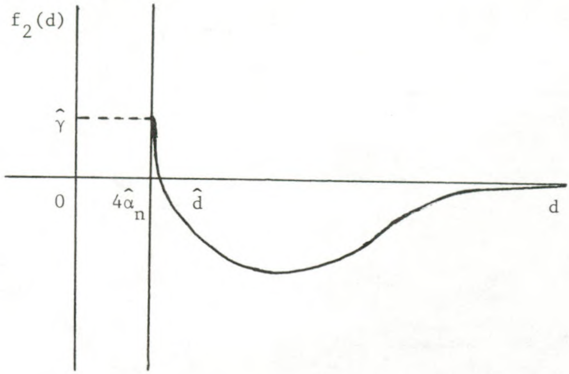

Figure 2

The solution $\hat{d}$ should be substituted into

$$\hat{d}_i = \frac{\hat{d}}{2} - \frac{\hat{d}}{2}\left(1 - \frac{4\hat{\alpha}_i}{\hat{d}}\right)^{\frac{1}{2}} \qquad i = 1, \ldots, n-1$$

$$\hat{d}_n = \frac{\hat{d}}{2} + \frac{\hat{d}}{2}\left(1 - \frac{4\hat{\alpha}_n}{\hat{d}}\right)^{\frac{1}{2}}$$

in order to find the solution in terms of the parameters $\hat{d}_i$, which are all positive again.

[3]Case 3: $\sum_{i=1}^{n-1} \hat{\alpha}_i < \hat{\alpha}_n < \left(\sum_{i=1}^{n-1} \hat{\alpha}_i^{\frac{1}{2}} \right)^2$

We have to solve on $(-\infty, 0)$

$$f_2(\hat{d}) = 0$$

In this case the graph of the function $f_2(d)$ looks as follows

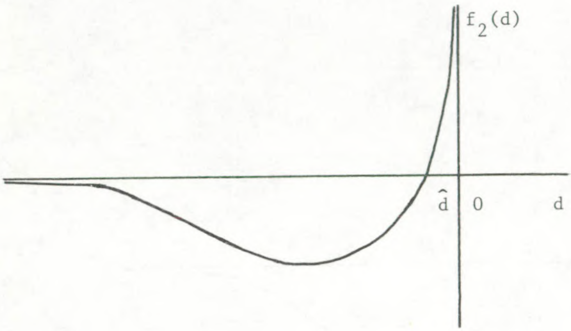

Figure 3

The solution $\hat{d}$ should be substituted into

$$\hat{d}_i = \frac{\hat{d}}{2} - \frac{\hat{d}}{2} \left(1 - \frac{4\hat{\alpha}_i}{\hat{d}}\right)^{\frac{1}{2}} \qquad i = 1, \ldots, n-1$$

$$\hat{d}_n = \frac{\hat{d}}{2} + \frac{\hat{d}}{2} \left(1 - \frac{4\hat{\alpha}_n}{\hat{d}}\right)^{\frac{1}{2}}$$

in order to obtain the solution in terms of the parameters $\hat{d}_i$. Obviously, $\hat{d}_i$ $(i = 1, \ldots, n-1)$ is positive, $\hat{d}_n$ is negative, with $\hat{d}$ being negative as well.

4. CONCLUSION

In the present paper a new specification is introduced for the

3. Because of the adding-up restriction on the vectors u_i $(i = 1, \ldots, n)$, it can be shown that $\hat{\alpha}_n$ cannot become larger than $(\sum_{i=1}^{n-1} \hat{\alpha}_i^{\frac{1}{2}})^2$.

contemporaneous covariance matrix of the error structure of sum-constrained linear models. In applied research one usually specifies this covariance matrix either to be restricted only by logical constraints or to be equal to $\sigma^2(I_n - n^{-1}\iota_n\iota_n')$, n denoting the number of categories. The former specification is the most flexible one but suffers from the drawback that it is very demanding with respect to the number of observations when the number of categories is large. The latter specification is the most rigid one that can be thought of but requires only a small number of observations even when the number of categories is large. The specification that we propose is intermediate in the sense that it allows for n covariance parameters to be estimated freely, while it possesses the attractive property of not requiring too many observations. The estimates involved may be obtained by a simple iterative scheme. In each stage of this scheme one numerical search procedure has to be carried out in order to determine the unique real root of an equation in only one variable. Therefore, it may be expected that the costs associated with the newly proposed estimation procedure will not be much higher than those associated with the estimation procedure under the most rigid specification.

In this paper we have restricted ourselves to linear models. However, the estimation procedure can easily be extended so as to include nonlinear models like the linear expenditure system or the Almost Ideal Demand System in its extensive form. Actually, nonlinearities will not affect the estimation procedure for the covariance parameters, but only the estimation procedure for the parameters of the deterministic part of the model. In a similar way, the present approach can easily be extended so as to apply to sum-constrained simultaneous equations models as well. Finally, it would be useful to generalize the estimation procedure to models with serially correlated errors in order to be able to test for dynamic misspecifications. We hope to address this question as well as the problem of testing the present specification against alternative ones in the near future.

REFERENCES

Barten, A.P. (1969), "Maximum likelihood estimation of a complete system of demand equations", European Economic Review, Vol. 1, Fall 1969.

de Boer, P.M.C. and R. Harkema (1983), "Undersized samples and maximum likelihood estimation of sum-constrained linear models", Report 8331/E, Econometric Institute, Erasmus University Rotterdam.

Deaton, A. (1975), Models and projections of demand in post-war Britain, Chapman and Hall, London.

Deaton, A. and J. Muellbauer (1980), "An almost ideal demand system", American Economic Review, Vol. LXX, June 1980.

Don, F.J.H. (1984), "The specification of least informative error distributions", Research Memorandum, Central Planning Bureau, The Hague.

Laitinen, K. (1978), "Why is demand homogeneity so often rejected?", Economics Letters, Vol. 1., No. 3.

McGuire, T.W., J.U. Farley, R.E. Lucas and L.W. Ring (1968), "Estimation and inference for linear models in which subsets of the dependent variable are constrained", Journal of the American Statistical Association, Vol. 63, December 1968.

Nakanishi, M. and L.G. Cooper (1974), "Parameter estimation for a multiplicative competitive interaction model-least squares approach", Journal of Marketing Research, Vol. XI, August 1974.

Oberhofer, W. and J. Kmenta (1974), "A general procedure for obtaining maximum likelihood estimates in generalized regression models", Econometrica, Vol. 42, May 1974.

Sargan, J.D. (1964), "Wages and prices in the United Kingdom: a study in econometric methodology", in P.E. Hart, G. Mills and J.K. Whitaker (eds.): Econometric analysis for national economic planning, Buttonworth.

Solari, L. (1971), Théorie des choix et fonctions de consommation semi-agrégées: modèles statiques, Librairie Droz, Genève.

Theil, H. (1971), "An economic theory of the second moments of disturbances of behavioral equations", American Economic Review, Vol. LXI, March 1971.

Theil, H. (1974), "A theory of rational random behavior", Journal of the American Statistical Association, Vol. 69, June 1974.

Theil, H. (1975), Theory and measurement of consumer demand, Vol. 1, North-Holland, Amsterdam.

Theil, H. (1980), The system-wide approach to microeconomics, Basil Blackwell, Oxford.