IDENTIFIABLE REPARAMETERIZATION OF AN OVER-PARAMETERIZED LIKELIHOOD FUNCTION

D. S. Robson

BU-576-M

January, 1976

Abstract

A likelihood function $e^{L(\underset{\sim}{x};\underset{\sim}{\theta})}$ of an observable variable $\underset{\sim}{x}$ and an unobservable parameter $\underset{\sim}{\theta} = (\theta_1, \cdots, \theta_p)$ is overparameterized if the rank of the information matrix is $k < p$. If the $p - k$ linear dependencies among the p partials $\partial L/\partial\theta_j$ are expressed in the form

$$\frac{\partial L}{\partial\theta_i} = \sum_{j=1}^{k} \beta_{ij}(\underset{\sim}{\theta}) \frac{\partial L}{\partial\theta_{p-k+j}}, \qquad i = 1, \cdots, p - k$$

then the solution to the system of partial differential equations

$$\frac{\partial\theta_{p-k+j}}{\partial\theta_i} = -\beta_{ij}(\underset{\sim}{\theta}) \qquad \begin{array}{l} i = 1, \cdots, p - k \\ j = 1, \cdots, k \end{array}$$

defines an identifiable, k-dimensional reparameterization.

IDENTIFIABLE REPARAMETERIZATION OF AN OVER-PARAMETERIZED LIKELIHOOD FUNCTION

D. S. Robson

BU-576-M January, 1976

Over-parameterization of a statistical model is a common though sometimes inadvertent practice in applied statistics. Examples of the common practice are "main effect plus interaction" linear models of cell means in two- and higher-way cross-classifications, where linear constraints such as "main effects sum to zero" must then be imposed to reduce the dimensionality of the parameter space to at most the dimensionality of the space of observed cell means. Examples of inadvertent over-parameterizations are rare in the literature because nonidentifiability is ordinarily detected before the publication stage, and either an identifiable reparameterization of the original statistical model appears in published form or the model is not published at all. Perhaps the most common such correction is the change from absolute values to relative values of certain parameters in the model, typically reducing the dimensionality of the original parameter space by one in order to achieve identifiability.

A problem in applied statistics may usually be partitioned longitudinally into several component problems, one of which is to devise a stochastic model which approximately characterizes the data generating process. Typically, the stochastic model takes the form of a specific parametric family $F_{\underset{\sim}{X}}(\underset{\sim}{x};\underset{\sim}{\theta})$ of possible probability distributions of the observable random variable $\underset{\sim}{X}$, where at least some of the components of the parameter $\underset{\sim}{\theta}$ are unobservable and unknown. For present purposes all components of $\underset{\sim}{\theta}$ are assumed to be in this category, with any observable parameters being subsumed in the functional form of $F_{\underset{\sim}{X}}$.

A stochastic model of an observable random variable $\underset{\sim}{X}$ is called a statistical model, or at least becomes so in the next component of the problem which is to devise formal methods of making inferences about the unknown $\underset{\sim}{\theta}$ on the basis of the observed value $\underset{\sim}{x}$ of the random variable $\underset{\sim}{X}$. Identifiability, which is not an issue in stochastic modeling, becomes an issue when addressing questions concerning which values of $\underset{\sim}{\theta}$ governed the process which generated the observation $\underset{\sim}{x}$. If more than one value of $\underset{\sim}{\theta}$ can produce the same probability distribution of the observable random variable $\underset{\sim}{X}$ then the data $\underset{\sim}{x}$ provide us with zero capability for distinguishing among such values of $\underset{\sim}{\theta}$. In this circumstance $\underset{\sim}{\theta}$ $\left(\text{or } F_{\underset{\sim}{X}}(\underset{\sim}{x};\underset{\sim}{\theta})\right)$ is said to be nonidentifiable or nonestimable, though in some statistical circles "estimable" is used in the more restricted sense of the existence of an unbiased estimator.

Inadvertent nonidentifiability is probably not an infrequent occurrence in the practice of an applied statistician who is called upon to develop statistical models in unfamiliar subject matter areas. Its surprise occurrence almost always causes some distress, particularly when discovered after the data $\underset{\sim}{x}$ have been collected, and more particularly when the statistican had a voice in the critical matter of determining which variables were to be observed. Distress may be only temporary, however, if an identifiable reparameterization of the statistical model $F_{\underset{\sim}{X}}(\underset{\sim}{x};\underset{\sim}{\theta})$ adequately serves the purposes of the investigator as, for example, when identifiable relative values of some parameters serve his purpose as well as nonidentifiable absolute values.

Nonidentifiability, planned or inadvertent, introduces additional components into the statistician's problem; namely, the problem of fitting the data to a nonidentifiable model and usually also the problem of constructing an identifiable reparameterization which the client finds comprehensible and suitable for his needs. The order in which these two component problems are attacked need not be critical, as in the case of over-parameterized linear models where linear constraints may be imposed first in order to reduce the "design matrix" to full column rank for computational convenience in fitting the model, or generalized inverse matrices may be employed to fit the over-parameterized model with no constraints imposed on these excessively abundant parameters. The latter tactic, however, usually amounts to only a postponement of the question of identifiable (estimable) reparameterizations; i.e., the needs of the client are usually not fully met simply by producing a fit to the nonidentifiable model. In an estimation problem, identifiable parameters need to be specified, point and/or interval estimates computed and interpreted; in an hypothesis testing problem nonidentifiable alternative hypotheses need to be characterized in a comprehensible manner so that the investigator will understand what hypothesis is being rejected or accepted.

Confronted with a nonidentifiable statistical model, the statistician's first task is to analyze the nonidentifiability to determine which if any of the components of $\underset{\sim}{\theta}$ are identifiable and to determine the dimension of the identifiable parameter space. In the linear model context the latter is simply the (column) rank of the design matrix, and in the likelihood context is the rank of the information matrix. A design matrix is observable, however, while entries in an information matrix are more generally functions of the unobservable, unknown, nonidentifiable parameters. The tasks facing the statistician are therefore somewhat different in these two circumstances, and considerations here will be focused upon the more general case of an over-parameterized likelihood function in which none of the parameters are identifiable.

The following "fitness" example from the field of population genetics is given to illustrate and motivate one approach to the analysis of a nonidentifiable likelihood function. In the absence of perturbing forces the three genotypes AA, Aa, and aa should occur with relative frequencies $p_{AA} = p^2$, $p_{Aa} = 2pq$, and $p_{aa} = q^2$ in a randomly mating (infinite) population, where

$$p = p_{AA} + \tfrac{1}{2}p_{Aa} \qquad q = p_{aa} + \tfrac{1}{2}p_{Aa} \qquad p + q = 1 \quad .$$

If rates of survival to adulthood differ among the three types, however, then relative frequencies in the next mating generation become

$$\left(p_{AA}\frac{S_{AA}}{S},\ p_{Aa}\frac{S_{Aa}}{S},\ p_{aa}\frac{S_{aa}}{S}\right) = \left(p^2\frac{S_{AA}}{S},\ 2pq\frac{S_{AA}}{S},\ q^2\frac{S_{AA}}{S}\right)$$

where S_{AA}, S_{Aa}, S_{aa} are survival probabilities conditional on genotype and $S = p_{AA}S_{AA} + p_{Aa}S_{Aa} + p_{aa}S_{aa}$ is the marginal probability of survival to adulthood. Parameters called "fitness values" are defined as relative survival rates, thus $W_1 = S_{AA}/S_{Aa}$, $W_2 = S_{aa}/S_{Aa}$ are the fitness values of the two homozygotes relative to the fitness of the heterozygous genotype, and $\bar{W} = W_1p^2 + 2pq + W_2q^2$ is called the average (relative) fitness.

The parameters (W_1, W_2, p) obtained by this reduction from $(S_{AA}, S_{Aa}, S_{aa}, p)$ are still not identifiable, however, with respect to the counts X_{AA}, X_{Aa}, X_{aa} of genotypes in a random sample of fixed size $n = X_{AA} + X_{Aa} + X_{aa}$ from the (infinite) population of adult survivors. The linear dependency among the observations automatically precludes the possibility of estimating three linearly independent parameters (W_1, W_2, p) from such data obtained at a single point in time. This nonidentifiability would become immediately apparent if an attempt were made to construct maximum likelihood estimators of W_1, W_2, and p from the trinomial likelihood

$$e^L = P(\underset{\sim}{X} = \underset{\sim}{x}) = \frac{n!}{x_{AA}!x_{Aa}!x_{aa}!}\,\frac{\left(W_1p^2\right)^{x_{AA}}\left(2pq\right)^{x_{Aa}}\left(W_2q^2\right)^{x_{aa}}}{\left(W_1p^2 + 2pq + W_2q^2\right)^n} .$$

A linear dependence is seen to exist among the three partial derivatives of $L(\underset{\sim}{x};\, W_1,\, W_2,\, p)$, namely,

$$\frac{\partial L}{\partial p} = \frac{W_1}{pq}\frac{\partial L}{\partial W_1} - \frac{W_2}{pq}\frac{\partial L}{\partial W_2} ,$$

and the 3 X 3 matrix of expected values of the second partials of L would be found to have rank 2.

A similar result obtains if independent samples are selected from k different local populations having different "gene frequencies" p_i but subjected to common selection forces which produce common fitness values W_1, W_2; thus, if

$$L\left(\underset{\sim}{x};W_1,W_2,p_1,\cdots,p_k\right) = \sum_1^k L\left(\underset{\sim}{x}_i;W_1,W_2,p_i\right)$$

then

$$\sum_1^k p_iq_i \frac{\partial L}{\partial p_i} = W_1 \frac{\partial L}{\partial W_1} - W_2 \frac{\partial L}{\partial W_2} \quad . \tag{1}$$

The observation variable $\underset{\sim}{x}$ now consists of 2k linearly independent variables while the parameter space has dimension only $k+2$; but Neyman-factorization of the likelihood would helpfully reveal a sufficient statistic of dimension only $k+1$, obviating any need in this case to use numerical methods in determining the dimension of an identifiable parameter space. With this information in hand, knowing that a linear dependency lies hidden among the likelihood equations, the statistician could presumably stare at the partial derivatives of L until (1) is revealed unto him.

Knowing a specific form (1) of the linear dependency is not essential, however, in solving the likelihood equations, for when all of the parameters are nonidentifiable then one possible tactic is simply to assign an arbitrary numerical value to one of the parameters and eliminate the corresponding likelihood equation. The only precaution required is to avoid boundary values in this assignment; i.e., avoid the value 0 or 1 for p_i or the value 0 for W_1 or W_2. Letting $W_2 = 1$, for example, reduces the space of unknown parameters to a $k+1$ dimensional linear subspace of the original parameter space; this tactic thus carries over from the familiar methods of analysis of over-parameterized linear models.

An identifiable parameter space is not necessarily a linear subspace of the original parameter space, but linear subspaces of this particular form, say W_2 = a constant, are necessarily identifiable. More generally, an identifiable parameterization is any $k+1$-vector $\underset{\sim}{\eta} = (\eta_1, \cdots, \eta_{k+1})$ of (nonlinear) functions of the original parameters,

$$\eta_i = \eta_i\left(W_1, W_2, p_1, \cdots, p_k\right) \qquad i = 1, \cdots, k+1 \tag{2}$$

satisfying an identity of the form

$$L\left(\underset{\sim}{x}; W_1, W_2, p_1, \cdots, p_k\right) \underset{\underset{\sim}{x}}{\equiv} G\left(\underset{\sim}{x}; \eta_1, \cdots, \eta_{k+1}\right) \quad . \tag{3}$$

For any fixed value of W_2, say, the $k+1$ equations (2) could be solved for $W_1, p_1, \cdots, p_k$ in terms of W_2 and $\underset{\sim}{\eta}$,

$$W_1 = W_1\left(W_2; \underset{\sim}{\eta}\right), \; p_1 = p_1\left(W_2; \underset{\sim}{\eta}\right), \; \cdots, \; p_k = p_k\left(W_2; \underset{\sim}{\eta}\right) \; ; \tag{4}$$

a unique solution must exist because $\underset{\sim}{\eta}$ is identifiable, and hence this solution is also an identifiable parameterization.

When the solution (4) is substituted into the likelihood function L it follows from (3) that

$$L\left(\underset{\sim}{x};W_1(W_2;\underset{\sim}{\eta}),W_2,p_1(W_2;\underset{\sim}{\eta}),\cdots,p_k(W_2;\underset{\sim}{\eta})\right) \underset{\underset{\sim}{x}}{\equiv} G\left(\underset{\sim}{x};\underset{\sim}{\eta}\right) \quad .$$

Taking derivatives of both sides with respect to W_2 thus gives

$$\frac{\partial L}{\partial W_1}\frac{\partial W_1}{\partial W_2} + \frac{\partial L}{\partial W_2} + \frac{\partial L}{\partial p_1}\frac{\partial p_1}{\partial W_2} + \cdots + \frac{\partial L}{\partial p_k}\frac{\partial p_k}{\partial W_2} \underset{\underset{\sim}{x}}{\equiv} \frac{\partial G}{\partial W_2} \underset{\underset{\sim}{x}}{\equiv} 0 \qquad (5)$$

which must therefore be an expression of the linear dependency (1) which exists among the first partials of L with respect to the original, nonidentifiable parameters $(W_1,W_2,p_1, \cdots, p_k)$. Relating (5) to (1) we thus obtain the $k+1$ differential equations

$$\frac{\partial W_1}{\partial W_2} = -\frac{W_1}{W_2}, \quad \frac{\partial p_i}{\partial W_2} = \frac{p_i(1-p_i)}{W_2}, \qquad i = 1, \cdots, k \quad . \qquad (6)$$

Letting $\underset{\sim}{\eta} = (\eta_1, \cdots, \eta_{k+1})$ denote the "constants of integration" we obtain the solutions

$$W_1 = \frac{\eta_1}{W_2}, \qquad p_i = \frac{W_2}{W_2+\eta_{i+1}}, \qquad i = 1, \cdots, k \qquad (7)$$

or

$$\eta_1 = W_1W_2, \qquad \eta_{i+1} = W_2\,\frac{q_i}{p_i}, \qquad i = 1, \cdots, k \quad . \qquad (8)$$

Provided $W_2 > 0$ then substitution of (7) into the likelihood L will indeed produce a function of only $\underset{\sim}{x}$ and $\underset{\sim}{\eta}$, since

$$\frac{\left[W_1p_i^2, 2p_iq_i, W_2q_i^2\right]}{W_1p_i^2+2p_iq_i+W_2q_i^2} = \frac{\left[\frac{\eta_1}{W_2}\left(\frac{W_2}{W_2+\eta_{i+1}}\right)^2, \frac{2W_2\eta_{i+1}}{\left(W_2+\eta_{i+1}\right)^2}, W_2\left(\frac{\eta_{i+1}}{W_2+\eta_{i+1}}\right)^2\right]}{\frac{\eta_1}{W_2}\left(\frac{W_2}{W_2+\eta_{i+1}}\right)^2 + \frac{2W_2\eta_{i+1}}{\left(W_2+\eta_{i+1}\right)^2} + W_2\left(\frac{\eta_{i+1}}{W_2+\eta_{i+1}}\right)^2}$$

$$= \frac{\left[\eta_1, 2\eta_{i+1}, \eta_{i+1}^2\right]}{\eta_1+2\eta_{i+1}+\eta_{i+1}^2} .$$

The identifiable reparameterization $\underset{\sim}{\eta}$ in (8) clearly occupies a nonlinear subspace of the original, nonidentifiable parameter space, but the equations (7) with $W_2 = 1$ transform this nonlinear subspace into the $k+1$-dimensional linear subspace described earlier. Any other identifiable reparameterization is likewise simply a transformation of $\underset{\sim}{\eta}$; for example, let

$$p_i^* = \frac{\sqrt{\eta_1}}{\sqrt{\eta_1}+\eta_{i+1}} , \quad i = 1, \cdots, k \qquad W = \frac{1}{\sqrt{\eta_1}}$$

so that

$$\frac{\left[W_1p_i^2, 2p_iq_i, W_2q_i^2\right]}{W_1p_i^2+2p_iq_i+W_2q_i^2} = \frac{\left[p_i^{*2}, 2p_i^*q_i^*W, q_i^{*2}\right]}{p_i^{*2}+2p_i^*q_i^*W+q^{*2}}$$

$$= \frac{\left[\frac{1}{W}\, p_i^{*2}, 2p_i^*q_i^*, \frac{1}{W}\, q_i^{*2}\right]}{\frac{1}{W}\, p_i^{*2}+2p_i^*q_i^*+\frac{1}{W}\, q_i^{*2}}$$

thus revealing to the population geneticist that unequal fitness of the two homozygotes is observationally indistinguishable from equal fitness in this context.

The system (8) thus effectively characterizes identifiable reparameterizations, analogous to a characterization of "estimable linear functions" in the case of linear models, and manipulation of (8) through transformations into contextually meaningful parameters can be helpful in comprehending the implications of the nonidentifiability. One disconcerting implication is that if the population geneticist had originally postulated the p_i^*, W model of equal fitness then identifiability would have obtained from the start, the above investigation would not have been pursued, and the observational equivalence of equal and unequal fitness of homozygotes might well have gone unnoticed.

The general form of the tactic illustrated by (1) -- (8) may be described in terms of an overparameterized log-likelihood function $L(\underset{\sim}{x};\underset{\sim}{\theta})$ where none of the p components of $\underset{\sim}{\theta} = (\theta_1, \cdots, \theta_p)$ are identifiable. Let $\underset{\sim}{\eta} = (\eta_1, \cdots, \eta_k)$ denote an identifiable reparameterization of dimension $k < p$, so that for all $\underset{\sim}{\theta}$ satisfying $\eta_i(\underset{\sim}{\theta}) = \eta_i$, $i = 1, \cdots, k$, the likelihood depends only upon $\underset{\sim}{\eta}$, say

$$L(\underset{\sim}{x};\underset{\sim}{\theta}) \underset{\underset{\sim}{x}}{\equiv} G(\underset{\sim}{x};\underset{\sim}{\eta}) \quad .$$

If the functions $\eta_i(\underset{\sim}{\theta})$ were specified then the (row vector of) first partials of L with respect to $\theta_1, \cdots, \theta_p$ could be expressed as

$$\underset{(1\times p)}{\frac{\partial L(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}} \underset{\underset{\sim}{x}}{\equiv} \underset{(1\times k)}{\frac{\partial G(\underset{\sim}{x};\underset{\sim}{\eta})}{\partial \underset{\sim}{\eta}}} \underset{(k\times p)}{\left[\frac{\partial \eta_i(\underset{\sim}{\theta})}{\partial \theta_j}\right]}$$

and the matrix $\mathcal{L}$ of expected values of second partials of L with respect to θ would have the corresponding form

$$\underset{(p\times p)}{\mathcal{L}} = \underset{(p\times k)}{\left[\frac{\partial\eta_i(\underset{\sim}{\theta})}{\partial\theta_j}\right]} \; \underset{(k\times k)}{\mathcal{G}} \; \underset{(k\times p)}{\left[\frac{\partial\eta_i(\underset{\sim}{\theta})}{\partial\theta_j}\right]'}$$

where $\mathcal{G}$ is the k × k (full rank) matrix of expected second partials of G with respect to $\underset{\sim}{\eta}$.

The p - k linear dependencies which exist among the p partials $\partial L/\partial\theta_i$, expressed in the form

$$\frac{\partial L(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial\theta_i} = \sum_{j=1}^{k} \beta_{ij}(\underset{\sim}{\theta}) \frac{\partial L(\underset{\sim}{x};\underset{\sim}{\theta})}{\partial\theta_{p-k+j}}, \qquad i = 1, \cdots, p - k \tag{9}$$

may then be related to the k solutions $\theta_{p-k+j}(\theta_1, \cdots, \theta_{p-k};\underset{\sim}{\eta})$ of the k equations

$$\eta_\nu(\underset{\sim}{\theta}) = \eta_\nu, \qquad \nu = 1, \cdots, k$$

solved for $\theta_{p-k+1}, \cdots, \theta_p$ in terms of $\theta_1, \cdots, \theta_{p-k}$ and $\underset{\sim}{\eta}$. Evaluated at this solution, $L(\underset{\sim}{x};\underset{\sim}{\theta})$ becomes the function $G(\underset{\sim}{x};\underset{\sim}{\eta})$ which is independent of $\theta_1, \cdots, \theta_{p-k}$,

$$L(\underset{\sim}{x};\underset{\sim}{\theta}) \underset{\underset{\sim}{x}}{\equiv} L\Big(\underset{\sim}{x};\theta_1,\cdots,\theta_{p-k},\theta_{p-k+1}(\theta_1,\cdots,\theta_{p-k};\underset{\sim}{\eta}),\cdots,\theta_p(\theta_1,\cdots,\theta_{p-k};\underset{\sim}{\eta})\Big)$$

$$\underset{\underset{\sim}{x},\theta_1,\cdots,\theta_{p-k}}{\equiv} G(\underset{\sim}{x};\underset{\sim}{\eta}) \;,$$

and hence

$$\frac{\partial L}{\partial\theta_i} + \sum_{j=1}^{k} \frac{\partial\theta_{p-k+j}}{\partial\theta_i} \frac{\partial L}{\partial\theta_{p-k+j}} \equiv \frac{\partial G}{\partial\theta_i} \equiv 0, \qquad i = 1, \cdots, p - k \;. \tag{10}$$

Relating (9) to (10) thus produces the system of partial differential equations

$$\frac{\partial\theta_{p-k+j}}{\partial\theta_i} = -\beta_{ij}(\underset{\sim}{\theta}), \qquad \begin{array}{l} i = 1, \cdots, p - k \\ j = 1, \cdots, k \end{array} \quad . \tag{11}$$

An identifiable reparameterization $\underset{\sim}{\eta} = (\eta_1, \cdots, \eta_k)$ is then obtained as the k "constants of integration" in the solution of this system of differential equations:

$$\theta_{p-k+j} = \theta_{p-k+j}(\theta_1, \cdots, \theta_p; \underset{\sim}{\eta}),\ j=1,\cdots,k \Leftrightarrow \eta_j = \eta_j(\underset{\sim}{\theta}),\ j=1,\cdots,k\ ,$$

just as (7) ⇔ (8).

The advantage of this tactic over the blind tactic of simply assigning arbitrary numerical values to, say, $\theta_1, \cdots, \theta_{p-k}$, is the functional characterization $\underset{\sim}{\eta}(\underset{\sim}{\theta})$ of identifiable reparameterizations. In simple cases the latter might be constructed simply by inspection of $L(\underset{\sim}{x};\underset{\sim}{\theta})$, but (9) -- (11) provide an analytical construction when insight fails. Insight is required, of course, in determining the specific form (9) of the p - k linear dependencies, as well as in determining their number, p - k, and herein lies the disadvantage as compared to the blind tactic which requires only the determination of k.

As a last resort, numerical and graphical aids to insight might be employed in determining the functional form of the coefficients (11) appearing in the linear dependencies (9). At any specified numerical value of $\underset{\sim}{\theta}$ the numerical values of $\beta_{ij}(\underset{\sim}{\theta})$ could be calculated by conventional linear methods; e.g., by calculating the vector $\partial L(\underset{\sim}{x};\underset{\sim}{\theta})/\partial\underset{\sim}{\theta}$ at a number of points $\underset{\sim}{x}$ in the sample space and fitting the multiple linear regression equations (9) to obtain $\beta_{ij}(\underset{\sim}{\theta})$. Repeating this operation at several strategically chosen values of $\underset{\sim}{\theta}$ and plotting $\beta_{ij}(\underset{\sim}{\theta})$ against components of $\underset{\sim}{\theta}$ might then provide the necessary clues leading to an analytical expression of (9). Note that at any fixed $\underset{\sim}{\theta}$ the "sample" covariance matrix of the vector $\partial L(\underset{\sim}{x};\underset{\sim}{\theta})/\partial\underset{\sim}{\theta}$ obtained by this operation has rank k, thus allowing numerical determination of k without calculation of the information matrix.

In the special case of linear models (with known error variance) the coefficients $\beta_{ij}(\underset{\sim}{\theta})$ are independent of the (unobservable) parameter $\underset{\sim}{\theta}$, so in this case the solution to (11) is simply

$$\theta_{p-k+j} = \eta_j - \sum_{i=1}^{p-k} \beta_{ij}\theta_i \qquad j = 1, \cdots, k$$

or the identifiable parameterization

$$\eta_j = \theta_{p-k+j} + \sum_{i=1}^{p-k} \beta_{ij}\theta_i \qquad j = 1, \cdots, k$$

characterizes the "estimable linear functions"; i.e., any linear function $\Sigma C_j\eta_j$ is estimable, and any estimable linear function of $\underset{\sim}{\theta}$, is expressible in the form $\Sigma C_j\eta_j$. Identifiability in this case is achieved by imposing any set of p - k linearly independent nonestimable linear constraints on $\underset{\sim}{\theta}$. Note that this tactic is not applicable in the nonlinear case where $\beta_{ij}(\underset{\sim}{\theta})$ does depend upon $\underset{\sim}{\theta}$.

The present development treats only the case where all components of $\underset{\sim}{\theta}$ are nonidentifiable, and hence excludes the case of overparameterized linear models with unknown but identifiable error variance. Tactical modifications required to accommodate such cases have not been explored by this author, but clearly require partitioning the vector $\underset{\sim}{\theta}$ into component vectors such that linear dependencies among the first partials of L exist only within and not between these components.