Constrained Maximum Likelihood *

Ronald Schoenberg
Aptech Systems, Inc.
and
The University of Washington

11 November 1996

*Comments or suggestions may be addressed to Ronald Schoenberg, Aptech Systems, Inc., 23804 SE Kent-Kangley Rd, Maple Valley, WA 98038, or by e-mail to rons@aptech.com.

Abstract

Constrained Maximum Likelihood (**CML**) is a new software module developed at Aptech Systems for the generation of maximum likelihood estimates of statistical models with general constraints on parameters. These constraints can be linear or nonlinear, equality or inequality. The software uses the Sequential Quadratic Programming method with various descent algorithms to iterate from a given starting point to the maximum likelihood estimates.

Standard asymptotic theory asserts that statistical inference regarding inequality constrained parameters does not require special techniques because for a large enough sample there will always be a confidence region at the selected level of confidence that avoids the constraint boundaries. Sufficiently large, however, can be quite large, in the millions of cases when the true parameter values are very close to these boundaries. In practice, our finite samples may not be large enough for confidence regions to avoid constraint boundaries, and this has implications for all parameters in models with inequality constraints, even for those that are not themselves constrained.

The usual method for statistical inference, comprising the calculation of the covariance matrix of the parameters and constructing t-statistics from the standard errors of the parameters, fails in the context of inequality constrained parameters because confidence regions will not generally be symmetric about the estimates. When the confidence region impinges on the constraint boundary, it becomes truncated, possibly in a way that affects the confidence limit. It is therefore necessary to compute confidence intervals rather than t-statistics.

CML computes two classes of confidence intervals, by inversion of the Wald and likelihood ratio statistics, and by simulation of the distributions of the parameters by weighted and unweighted bootstrap. Confidence intervals computed by inverting the likelihood ratio statistic are also called profile likelihood confidence intervals.

While the inversion methods of computing confidence intervals produce regions of the correct shape, they do not take into account the affect the truncation has on the **size**, i.e., on the extent to which the observed proportion of the regions which fail to include the true value of the parameter matches the nominal level of confidence. The correction for the effects of the truncated regions on the size takes the form of an adjustment to the statistic used in the inversion. Thus, in circumstances to be described, the statistics used in the inversion of the Wald or likelihood ratio statistics are modified when the confidence region is within a specified distance of a constraint boundary.

For the construction of a confidence limits for a parameter, the following three cases must be considered: (1) when the parameter has a constraint placed on it and the limit is within a specified distance of the constraint boundary, (2) when the confidence region of a nuisance parameter correlated with the parameter which is on or near its constraint boundary, and (3) when the confidence regions of both the parameter of interest and the nuisance parameter are near constraint boundaries.

A method is proposed for solving the first size problem. Monte Carlo evidence is presented which shows that the solution for (1) is satisfactory for computing confidence limits by inversion of likelihood ratio and Wald statistics. There is no known solution for (2) and (3).

KEY WORDS: Maximum Likelihood, Inequality Constraints, Profile Likelihood Statistical Inference

1 Introduction

Nearly all statistical models contain constrained parameters. Even the simplest models contain them; for example, in ordinary least squares the estimate of the residual variance is constrained to be positive.

Many methods have been devised to enforce these restrictions; for example, the use of concentrated log-likelihoods, or standard deviations are estimated rather than variances. Other techniques for positivity include estimating the square root or log of a parameter. The hyperbolic cosine function can be used for correlations, and the logistic function for intervals.

Often, however, constraints are ignored, or tested for post hoc, for example, stationarity constraints in time dependent processes. Coefficient matrices in simultaneous equation models with lagged variables require specific constraints to ensure stationarity of the system (Greene, 1990:644), but this constraint is enforced by rejection, i.e., estimates where the constraint fails are rejected and the model is re-specified.

Over the last twenty years or so, statistical models have become more complicated, and the trend seems to be models with more constraints in them. For example, the GARCH model requires a complex set of equality and inequality constraints to ensure stationarity. Models of categorical data require normalizations which amount to constraints on parameters. Limited dependent variables contain variance and correlation restrictions.

In psychology, the covariance structure model (Browne and Arminger, 1995) estimates covariance matrices as well as coefficient matrices. The covariance matrices must be positive definite in addition to the the coefficient matrices being constrained by stationarity requirements.

Transformations of parameters and penalty methods have been customarily used to enforce constraints in statistical models. Convergence to a solution with these methods, however, was not always reliable.

Han (1977) proposed the Sequential Quadratic Programming (SQP) method for the optimization of functions with general equality and inequality constraints. This method was not initially exploited for estimating constrained statistical models because it was mostly known in the Operations Research field where the requirement of the method that the Hessian be positive definite was a serious drawback.

Such a requirement is not a hindrance in statistical estimation problems and finally in 1993 it was applied to a statistical problem (Jamshidian, et al., 1993). Then software began to appear – Matlab's optimization toolbox, SAS's Proc NLP, and Aptech System's **CML**. **CML** is the first implementation of the SQP method explicitly for the maximum likelihood estimation of constrained statistical models.

2 CML

CML) is a set of procedures written in the GAUSS programming language

(Schoenberg, 1995) for the estimation of the parameters of models via the maximum likelihood method with general constraints on the parameters

CML solves the general weighted maximum likelihood problem

$$L = \sum_{i=1}^{N} \log P(Y_i; \theta)^{w_i},$$

where N is the number of observations, w_i is a weight. $P(Y_i, \theta)$ is the probability of Y_i given θ, a vector of parameters, subject to the linear constraints,

$$A\theta = B,$$
$$C\theta \geq D,$$

the nonlinear constraints

$$G(\theta) = 0,$$
$$H(\theta) \geq 0,$$

and bounds

$$\theta_l \leq \theta \leq \theta_u.$$

$G(\theta)$ and $H(\theta)$ are functions provided by the user and must be differentiable at least once with respect to θ.

CML finds values for the parameters in θ such that L is maximized using the Sequential Quadratic Programming method. In this method the parameters are updated in a series of iterations beginning with a vector of starting values. Let θ_t be the current parameter values. Then the succeeding values are

$$\theta_{t+1} = \theta_t + \rho\delta,$$

where δ is a $K \times 1$ *direction* vector, and ρ a scalar *step length.*

Define

$$\begin{aligned} \Sigma(\theta) &= \frac{\partial^2 L}{\partial\theta\partial\theta'}, \\ \Psi(\theta) &= \frac{\partial L}{\partial\theta}, \end{aligned}$$

and the Jacobians

$$\dot{G}(\theta) = \frac{\partial G(\theta)}{\partial\theta},$$
$$\dot{H}(\theta) = \frac{\partial H(\theta)}{\partial\theta}.$$

For the purposes of this exposition, and without loss of generality, we may assume that the linear constraints and bounds have been incorporated into G and H.

The direction, δ is the solution to the quadratic program

$$minimize\ \frac{1}{2}\delta'\Sigma(\theta_t)\delta + \Psi(\theta_t)\delta,$$

$$\begin{aligned} subject\ to\quad & \dot{G}(\theta_t)\delta + G(\theta_t) = 0, \\ & \dot{H}(\theta_t)\delta + H(\theta_t) \geq 0. \end{aligned}$$

This solution requires that Σ be non-negative definite (In practice, CML incorporates a slight modification in the quadratic programming solution which relaxes this requirement to positive semi-definite).

In practice, linear constraints are specified separately from the G and H because their Jacobians are known and easy to compute. And the bounds are more easily handled separately from the linear inequality constraints.

The SQP method requires the calculation of a Hessian, Σ, and various gradients and Jacobians, Ψ, $\dot{G}(\theta)$, and $\dot{H}(\theta)$. **CML** computes these numerically if procedures to compute them are not supplied.

Descent Algorithms. The Hessian may be very expensive to compute at every iteration, and poor start values may produce an ill-conditioned Hessian. For these reasons alternative algorithms are provided in **CML** for updating the Hessian rather than computing it directly at each iteration. These algorithms, as well as step length methods, may be modified during the execution of **CML**.

Beginning with an initial estimate of the Hessian, or a conformable identity matrix, an update is calculated. The update at each iteration adds more "information" to the estimate of the Hessian, improving its ability to project the direction of the descent. Thus after several iterations the secant algorithm should do nearly as well as Newton iteration with much less computation.

There are two basic types of secant methods used in **CML**, the BFGS (Broyden, Fletcher, Goldfarb, and Shanno), and the DFP (Davidon, Fletcher, and Powell). They are both rank two updates, that is, they are analogous to adding two rows of new data to a previously computed moment matrix. The Cholesky factorization of the estimate of the Hessian is updated using the functions **CHOLUP** and **CHOLDN**.

BFGS is the method of Broyden, Fletcher, Goldfarb, and Shanno, and DFP is the method of Davidon, Fletcher, and Powell. These methods are complementary (Luenberger 1984, page 268). BFGS and DFP are like the Gauss-Newton method in that they use both first and second derivative information. However, in DFP and BFGS the Hessian is approximated, reducing considerably the computational requirements. Because they do not explicitly calculate the second derivatives, they are sometimes called "quasi-Newton" methods. While typically needing more iterations than the Gauss-Newton method, these approximations can be expected to converge in less overall time (though Gauss-Newton with analytical second derivatives can be quite competitive).

The secant methods are commonly implemented as updates of the inverse of the Hessian. This is not the best method numerically for the BFGS algorithm (Gill and Murray, 1972). This version of **CML**, following Gill and Murray (1972), updates the Cholesky factorization of the Hessian instead. The new direction is computed using a Cholesky solve, as applied to the updated Cholesky factorization of the Hessian and the gradient.

Line Search Methods. Define the merit function

$$m(\theta) = L + \max \mid \kappa \mid \sum_j \mid g_j(\theta) \mid - \max \mid \lambda \mid \sum_\ell \min(0, h_\ell(\theta)),$$

where g_j is the j-th row of G, h_ℓ is the ℓ-th row of H, κ is the vector of Lagrangean coefficients of the equality constraints, and λ the Lagrangean coefficients of the inequality constraints. The line search finds a value of ρ that minimizes or decreases

$$m(\theta_t + \rho\delta),$$

where ρ is a constant. Given θ and d, this is function of a single variable ρ. Line search methods attempt to find a value for ρ that decreases m. Several methods are available in **CML**, STEPBT, a polynomial fitting method, BRENT and BHHHSTEP, golden section methods, and HALF, a step-halving method.

Constraints. There are two general types of constraints, nonlinear equality constraints and nonlinear inequality constraints. However, for computational convenience they are divided into five types: linear equality, linear inequality, nonlinear equality, nonlinear inequality, and bounds.

Linear constraints are of the form

$$A\theta = B,$$

where A is an $m_1 \times k$ matrix of known constants, and B an $m_1 \times 1$ vector of known constants, and θ the vector of parameters.

Linear constraints are of the form

$$C\theta \geq D,$$

where C is an $m_2 \times k$ matrix of known constants, and D an $m_2 \times 1$ vector of known constants, and θ the vector of parameters.

Nonlinear equality constraints are of the form

$$G(\theta) = 0,$$

where θ is the vector of parameters, and $G(\theta)$ is an arbitrary, user-supplied function.

Nonlinear inequality constraints are of the form:

$$H(\theta) \geq 0,$$

where θ is the vector of parameters, and $H(\theta)$ is an arbitrary, user-supplied function.

Bounds are a type of linear inequality constraint. For computational convenience they are specified separately from the other inequality constraints.

Covariance Matrix of the Parameters. **CML** computes a covariance matrix of the parameters that is an approximate estimate when there are constrained parameters in the model (Gallant, 1987, Wolfgang and Hartwig, 1995). When the model includes inequality constraints, however, confidence limits computed from the usual t-statistics – dividing the parameter estimates by their standard errors – are incorrect because they do not account for boundaries placed on the distributions of the parameters by the inequality constraints. Confidence limits will have to be calculated therefore. Methods using inversion of the the Wald and likelihood ratio statistics are presented in the next section. The Wald method, however, requires an estimate of the covariance matrix of the parameters.

An argument based on a Taylor-series approximation to the likelihood function (e.g., Amemiya, 1985, page 111) shows that

$$\hat{\theta} \rightarrow N(\theta, A^{-1}BA^{-1}),$$

where

$$\begin{aligned} A &= E\left[\frac{\partial^2 L}{\partial\theta\partial\theta'}\right], \\ B &= E\left[\left(\frac{\partial L}{\partial\theta}\right)'\left(\frac{\partial L}{\partial\theta}\right)\right]. \end{aligned}$$

Estimates of A and B are

$$\hat{A} = \frac{1}{N}\sum_{i}^{N}\frac{\partial^2 L_i}{\partial\theta\partial\theta'},$$
$$\hat{B} = \frac{1}{N}\sum_{i}^{N}\left(\frac{\partial L_i}{\partial\theta}\right)'\left(\frac{\partial L_i}{\partial\theta}\right).$$

Assuming the correct specification of the model plim(A) = plim(B), and thus

$$\hat{\theta} \rightarrow N(\theta, \hat{A}^{-1}).$$

Without loss of generality we may consider two types of constraints, the nonlinear equality and the nonlinear inequality constraints (the linear constraints are included in nonlinear, and the bounds are regarded as a type of linear inequality). Furthermore, the inequality constraints may be treated as equality constraints with the introduction of "slack" parameters into the model:

$$H(\theta) \geq 0$$

is changed to

$$H(\theta) = \zeta^2,$$

where ζ is a conformable vector of slack parameters.

Further we distinguish *active* from *inactive* inequality constraints. Active inequality constraints have nonzero Lagrangeans, γ_j, and zero slack parameters, ζ_j, while the reverse is true for inactive inequality constraints. Keeping this in mind, define the diagonal matrix, Z, containing the slack parameters, ζ_j, for the inactive constraints, and another diagonal matrix, Γ, containing the Lagrangean coefficients. Also, define $H_{\oplus}(\theta)$ representing the active constraints, and $H_{\ominus}(\theta)$ the inactive.

The likelihood function augmented by constraints is then

$$\begin{aligned} L_A \;=\; & L + \lambda_1 g(\theta)_1 + \cdots + \lambda_I g(\theta)^I + \gamma_1 h_{\oplus 1}(\theta) + \cdots + \gamma_J h_{\oplus J}(\theta) + \\ & h_{\ominus 1}(\theta)_i - \zeta_1^2 + \cdots + h_{\ominus K}(\theta) - \zeta_K^2, \end{aligned}$$

and the Hessian of the augmented likelihood is

$$E\left(\frac{\partial^2 L_A}{\partial\theta\partial\theta'}\right) = \begin{bmatrix} \Sigma & 0 & 0 & \dot{G}' & \dot{H}'_{\oplus} & \dot{H}'_{\ominus} \\ 0 & 2\Gamma & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2Z \\ \dot{G} & 0 & 0 & 0 & 0 & 0 \\ \dot{H}_{\oplus} & 0 & 0 & 0 & 0 & 0 \\ \dot{H}_{\ominus} & 0 & 2Z & 0 & 0 & 0 \end{bmatrix},$$

where the dot represents the Jacobian with respect to θ, $L = \sum_{i=1}^{N}\log P(Y_i;\theta)$, and $\Sigma = \partial^2 L/\partial\theta\partial\theta'$. The covariance matrix of the parameters, Lagrangeans, and slack parameters is the Moore-Penrose inverse of this matrix.

Usually we are interested only in the covariance matrix of the parameters, i.e., the upper left portion of the inverse associated with the parameters. This portion of the inverse can be produced using methods described in Hartmann and Hartwig (1995). For example, construct the partitioned array

$$\tilde{B} == \begin{bmatrix} \dot{G} \\ \dot{H}_{\oplus} \\ \dot{H}_{\ominus} \end{bmatrix}.$$

Let Ξ be the orthonormal basis for the null space of $\tilde{B}$, then the covariance matrix of the parameters is

$$\Xi(\Xi'\Sigma\Xi)^{-1}\Xi'.$$

Rows of this matrix associated with active inequality constraints may not be available, i.e., the rows and columns of Ω associated with those parameters may be all zeros.

Heteroskedastic-consistent Covariance Matrix. **CML** computes a heteroskedastic-consistent covariance matrix of the parameters when requested. Define $B = (\partial L_A/\partial\theta)'(\partial L_A/\partial\theta)$ evaluated at the estimates. Then the covariance matrix of the parameters is $\Omega B \Omega$.

3 Inference

CML includes two classes of methods for analyzing the distributions of the estimated parameters:

- confidence limits by inversion of the two types of chi-squared statistics - Wald and likelihood ratio
- simulation of the distribution of the parameters by weighted and unweighted bootstrap

The **CML** procedure for the weighted bootstrap implements a Bayesian simulation of the posterior distribution of the parameters based on Newton and Raftery's (1994) weighted likelihood bootstrap method.

3.1 Confidence Limits by Inversion

Proper confidence regions can be computed for constrained models through the inversion of the chi-squared Wald or likelihood ratio statistics. The inversion of the Wald statistic is discussed in the context of the restricted least squares model by Rust and Burrus (1972) and O'Leary and Rust (1986). Discussion of the inversion of the likelihood ratio statistic in computing confidence regions is found in Cox (1974), Cook and Weisberg (1990), Meeker and Escobar (1995).

Inversion of the Likelihood Ratio Statistic. Partition a k-vector of parameters, $\theta = (\theta_1 \; \theta_2)$, and let $\tilde{\theta}$ be a maximum likelihood estimate of θ, where θ_1 is fixed to some value. A $100(1-\alpha)\%$ confidence region for the parameters in θ_1 is defined by

$$-2 * log(L(\tilde{\theta})/L(\hat{\theta})) \leq \chi^2_{(1-\alpha,k)}.$$

Let

$$F_{lr}(\phi) = \min(-2 * log(L(\tilde{\theta})/L(\hat{\theta})) \mid \eta_i'\theta = \phi, H(\theta) \geq 0)$$

where η is a vector with a one in the i-th position and zeros elsewhere, and $H(\theta)$ is a function describing the constraints. The lower limit of the $(1-\alpha)$ interval for θ_i is the value of ϕ such that

$$F_{lr}(\phi) = \chi^2_{(1-\alpha,k)}. \qquad (1)$$

A modified secant method is used to find the value of ϕ that satisfies (1). The upper limit is found by defining F_{lr} as a maximum.

Inversion of the Wald statistic. A $(1-\alpha)$ joint Wald-type confidence region for θ is the hyper-ellipsoid

$$JF(J, N-K; \alpha) = (\theta - \hat{\theta})'V^{-1}(\theta - \hat{\theta}),$$

where V is the covariance matrix of the parameters.

The lower limit of the confidence limit is the solution to

$$\min\left\{\eta_k'\theta \mid (\theta-\hat{\theta})'V^{-1}(\theta-\hat{\theta}) \geq JF(J,N-K;\alpha), G(\theta)=0, H(\theta)\geq 0)\right\},$$

where η can be an arbitrary vector of constants and $J = \sum \eta_k \neq 0$, and where again we have assumed that the linear constraints and bounds have been folded in among nonlinear constraints. The upper limit is the maximum of this same function.

In this form, the function to be minimized is not convex and cannot be solved by the usual methods. However, the problem can be re-stated as a parametric nonlinear programming problem (Rust and Burrus, 1972). Let

$$F_{wald}(\phi) = \min((\theta-\hat{\theta})'V^{-1}(\theta-\hat{\theta}) \mid \eta_k'\theta = \phi, G(\theta)=0, H(\theta)\geq 0).$$

The upper and lower limits of the $1-\alpha$ interval are the values of ϕ such that

$$F_{wald}(\phi) = JF(J,N-K;\alpha). \tag{2}$$

A modified secant method is used to find the value of ϕ that satisfies (2) (O'Leary and Rust, 1986). The upper limit is found by defining F_{wald} as a maximum.

3.2 Bootstrap

The bootstrap method is used to generate empirical distributions of the parameters, thus avoiding the difficulties with the usual methods of statistical inference described above.

CML provides two types of bootstrap methods, a simple bootstrap that generates a simulated distribution of parameters by random re-sampling with replacement, and a weighted re-sampling that generates a simulation of the Bayesian posterior distribution of the parameters.

Additional procedures are available in **CML** for generating kernel density plots, histograms, contour plots of bivariate distributions, and confidence limits, from the data sets containing the simulated distributions of parameters.

Bayesian Inference. CML procedure can generate a simulation of the posterior distribution of the parameters using a weighted bootstrap method described by Newton and Raftery (1994). Here weighted Dirichlet random variates are used for weights. After generating the weighted bootstrapped parameters, "importance" weights are computed:

$$r(\hat{\theta}) = \pi(\hat{\theta})e^{L(\hat{\theta})}/\hat{g}(\hat{\theta}),$$

where $\pi(\hat{\theta})$ is the prior distribution of the parameters, and $\hat{g}(\hat{\theta})$ is a normal kernel density estimate of the joint density of the parameters using Terrell's (1990) method of maximum smoothing. The SIR algorithm, described in Rubin (1988), is applied to the bootstrapped parameters using these importance weights.

4 Inference When True Values On or Near Constraint Boundaries

It is well known that the distributions of the Wald and likelihood ratio statistics are modified when the true value of a constrained parameter being estimated is on a constraint boundary (Gourieroux, et al., 1982, Self and Liang, 1987, Wolak, 1991). In finite samples these effects occur in the *region* of the constraint boundary, specifically when the true value is within $\epsilon = \sqrt{(\sigma_e^2/N)\chi^2_{(1-\alpha,k)}}$ of the constraint boundary. This has consequences for the calculation of the confidence limits described in the previous sections.

We are concerned here with the unidimensional problem; the determining the confidence limits of one parameter in the model, leaving all other parameters as "nuisance" parameters. This problem can be divided into three cases,

(1) parameter constrained, no nuisance parameters constrained,

(2) parameter unconstrained, one or more nuisance parameters constrained,

(3) parameter constrained, one or more nuisance parameters constrained.

For case 1, when the true value is on the boundary, the statistics are distributed as a simple mixture of two chi-squares. Monte Carlo evidence presented below will show that this holds as well in finite samples for true values within ϵ of the constraint boundary.

For case 2, the statistics are distributed as weighted mixtures of chi-squares when the correlation of the constrained nuisance parameter with the unconstrained parameter of interest is greater than about .8. A correction for these effects is feasible. However, for finite samples, the effects on the statistics due to a true value of a constrained nuisance parameter being within ϵ of the boundary are greater and more complicated than the effects of actually being on the constraint boundary. There is no systematic strategy available for correcting for these effects.

For case 3, the references disagree. Gourieroux, et al. (1982) and Wolak (1991) state that the statistics are distributed as a mixture of chi-squares. However, Self and Liang (1987) show that when the distributions of the parameter of interest and the nuisance parameter are correlated, the distributions of the statistics are not chi-square mixtures.

4.1 Case 1: Confidence Limits of Constrained Parameter of Interest

A Monte Carlo analysis was conducted to explore the effects of a constraint boundary on the true size of the confidence limits computed by two methods, (1) inversion of the Wald statistic, and (2) inversion of likelihood ratio statistic.

The 95 percent likelihood ratio and Wald confidence limits for means constrained to be greater than zero were estimated for 40 models: a Normal with unit variance and 20 different true values for the means ranging from 0 to .18 for each of two sample sizes, 300 and 500.

The proportion of the confidence limits that failed to include the true value is plotted in Figure 1 against the true value. We observe that this proportion, or size, is about one half the correct size up to a threshold where it becomes the full correct size. The theoretical thresholds for confidence limits for a mean with Normal density with N = 300 and 500, at $\alpha = .05$, are .1131, .0876, and for $\alpha = .10$ are .0950, .0736, respectively. These threshold values are quite close to the Monte Carlo results in Figure 1.

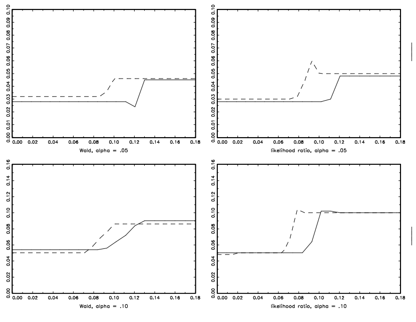

Figure 1: Size of constrained parameter of interest at different distances from the constraint boundary

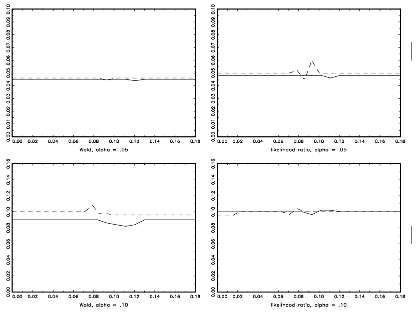

Figure 2: Size of constrained parameter of interest corrected

4.1.1 Correction for Effects of True Value in Region of Constraint Boundary

The effects of having a true value near a constraint boundary may be corrected for by modifying the method of inversion of the chi-square statistic described in Section 3.1. For the likelihood ratio statistic, for example, a ϕ is found that satisfies

$$F_{lr}(\phi) = \begin{cases} \chi^2_{(1-2\alpha,1)}, & H(\tilde{\theta}) < \sqrt{(\sigma_e^2/N)\chi^2_{(1-\alpha,1)}} \\ \chi^2_{(1-\alpha,1)}, & H(\tilde{\theta}) \geq \sqrt{(\sigma_e^2/N)\chi^2_{(1-\alpha,1)}} \end{cases}$$

Upon request, CML applies this correction to the calculation of the confidence limits by inverting the chi square statistics. The Monte Carlo analysis reported in the previous section was repeated with this correction. Results displayed in Figure 2, where the success of the correction for the effects of the proximity of the constraint boundary may be observed.

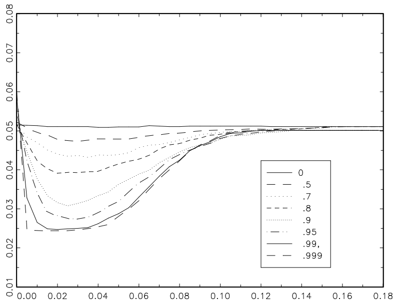

Figure 3: Size of confidence region of unconstrained parameter in the presence of a constrained nuisance parameter

4.2 Case 2: Confidence Limits of Unconstrained Parameter in Presence of Constrained Nuisance Parameters

When a constrained nuisance parameter is in the region of a constraint boundary, the confidence limits of a model parameter are affected even when the parameter is itself unconstrained. This fact is established when the true value of the nuisance parameter is on the boundary (Gourieroux, et al. (1982), Self and Liang (1987), and Wolak (1991)). When the parameter of interest and the nuisance parameter are uncorrelated, however, the effect vanishes.

A Monte Carlo study was conducted to determine possible effects of a constrained nuisance parameter in the region of a constraint boundary on the size of the confidence limits of an unconstrained parameter of interest. The true value of the constrained nuisance parameter varied from 0 to .18 in 37 intervals, and the correlation between the nuisance parameter and the constrained parameter of interest varied in 8 intervals from 0 to .999. 10,000 samples of size 500 were drawn under each of the conditions for a total of 2,960,000 samples. The results are presented in Figure 3.

The abscissa represents the true value of the constrained nuisance parameter, and the ordinate represents the observed size of the distribution of confidence intervals for the parameter of interest, that is, the proportion of intervals that fail to contain zero, the true value of the parameter of interest. The different curves show this relation at different correlations between the nuisance parameter and the parameter of interest.

We see from Figure 3 that any correction for size for a true value of the nuisance parameter on the boundary is not generalizable to true values in the region of the boundary, as it was for Case 1. In fact, effects on size in the region of the boundary are nonlinear and much larger than the effects on the boundary. This means that the corrections to the inversion of the chi-square statistics discussed in Gourieroux, et al. (1982), Self and Liang (1987), and Wolak (1991) are not applicable to cases where the true value of the correlated nuisance parameter falls within ϵ of the constraint boundary.

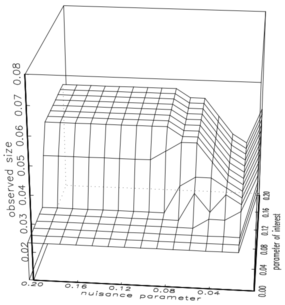

Figure 4: Size of confidence region of constrained parameter of interest in the presence of a constrained nuisance parameter

4.3 Case 3: Confidence Limits of Constrained Parameter in Presence of Constrained Nuisance Parameters

As described in Self and Liang (1987) when both the parameter of interest and a nuisance parameter are on boundaries, the likelihood ratio statistic is not a mixture of chi-squares. One should expect the behavior of the likelihood ratio statistic to be quite complex when both parameters are in the region of their boundaries, and this is confirmed in a Monte Carlo analysis.

The true values of the two means of a bivariate unit Normal distribution with correlation .9 were varied from 0 to .18 in 9 intervals. 500 samples of size 300 were drawn for each of these 225 sets of means. The 95 percent confidence intervals were computed by inversion of the likelihood ratio statistic, and the observed proportion of the intervals that failed to contain the true value are plotted in Figure 4.

Departures from nominal size are indicated by deviations from a flat plane set to .05. It is easily observed that relationship to nominal size is quite complex. There is no known method to compensate for this deviation. As with Case 2, however, departures from nominal size are trivial when the correlations between the parameter of interest and the nuisance parameter are less than about .7.

5 Summary

Constraints are a common feature of statistical models. CML is a computer program designed to produce estimates efficiently for these kinds of models. Methods of statistical inference for these models is more difficult, however. When confidence limits of all parameters n a model are more than $\epsilon = \sqrt{(\sigma_e^2/N)\chi^2_{(1-\alpha,k)}}$, standard methods will apply and the constraints will have no implications for inference. However, when this does not pertain, the constraint boundaries may affect the inference.

When the only parameter in the model within ϵ of a constraint boundary is the parameter of interest, inversion of a mixture of chi-square statistics will be required. A correction for this case is incorporated into CML.

When one (or more) nuisance parameter is within ϵ of a constraint boundary is correlated by more than about .7 with the parameter of interest, whether or not the parameter of interest is itself near a constraint boundary –

the chi-square statistics have complex distributions. While methods exist for determining the properties of their distributions when the true values are on the boundaries, there is no known method for determining them when the true values are only near the boundaries.

It is clear from the Monte Carlo evidence that effects near boundaries overwhelm these effects on the boundary, rendering any standard inference in these circumstances quite hazardous.

6 References

Amemiya, Takeshi, 1985. *Advanced Econometrics.* Cambridge, MA: Harvard University Press.

Bates, Douglas M. and Watts, Donald G., 1988. *Nonlinear Regression Analysis and Its Applications.* New York: John Wiley & Sons.

Berndt, E., Hall, B., Hall, R., and Hausman, J. 1974. "Estimation and inference in nonlinear structural models". *Annals of Economic and Social Measurement* 3:653-665.

Brent, R.P., 1972. *Algorithms for Minimization Without Derivatives.* Englewood Cliffs, NJ: Prentice-Hall.

Browne, Michael W. and Arminger, Gerhard, 1995. "Specification and estimation of mean- and covariance-structure models", in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Gerhard Arminger, Clifford C. Clogg, and Michael E. Sobel (eds.), New York: Plenum.

Cox, D.R., and Hinkley, D.V., 1974. *Theoretical Statistics.* London: Chapman and Hall.

Cook, R.D., and Weisberg, S., 1990. "Confidence Curves in Nonlinear Regression", *Journal of the American Statistical Association*, 85:544-551.

Meeker, W.Q., and L. A. Escobar, 1995. "Teaching about approximate confidence regions based on maximum likelihood estimation", *The American Statistician*, 49:48-53.

Dennis, Jr., J.E., and Schnabel, R.B., 1983. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Englewood Cliffs, NJ: Prentice-Hall.

Efron, Gradley, Robert J. Tibshirani, 1993. *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Fletcher, R., 1987. *Practical Methods of Optimization.* New York: Wiley.

Gallant, A.R., 1987. *Nonlinear Statistical Models.* New York: Wiley.

Gill, P. E. and Murray, W. 1972. "Quasi-Newton methods for unconstrained optimization." *J. Inst. Math. Appl.*, 9, 91-108.

Gourieroux, Christian, Holly, Alberto, and Monfort, Alain, 1982. "Likelihood ratio test, Wald Test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters", *Econometrica*, 50:63-80.

Greene, William H., 1990. *Econometric Analysis.* New York: Macmillan.

Han, S.P., 1977. "A globally convergent method for nonlinear programming." *Journal of Optimization Theory and Applications*, 22:297-309.

Hartmann, Wolfgang M. and Hartwig, Robert E., 1995. "Computing the Moore-Penrose inverse for the covariance matrix in constrained nonlinear estimation", SAS Institute, Inc., Cary, NC.

Hock, Willi and Schittkowski, Klaus, 1981. *Lecture Notes in Economics and Mathematical Systems.* New York: Springer-Verlag.

Jamshidian, Mortaza and Bentler, P.M., 1993. "A modified Newton method for constrained estimation in covariance structure analysis." *Computational Statistics & Data Analysis*, 15:133-146.

Newton, M.A. and Raftery, A.E., 1994. "Approximate Bayesian inference with the weighted likelihood bootstrap", *J.R. Statist. Soc. B*, 56:3-48.

O'Leary, Dianne P., and Rust, Bert W., 1986. "Confidence intervals for inequality-constrained least squares problems, with applications to ill-posed problems". *American Journal for Scientific and Statistical Computing*, 7(2):473-489.

Rubin, D.B., 1988. "Using the SIR algorithm to simulate posterior distributions", in *Bayesian Statistics 3*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith (eds.), pp. 395-402.

Rust, Bert W., and Burrus, Walter R., 1972. *Mathematical Programming and the Numerical Solution of Linear Equations.* New York: American Elsevier.

Schoenberg, Ronald, 1995. *CML Users Guide.* Maple Valley, WA, USA: Aptech Systems, Inc.

Self, Steven G. and Liang, Kung-Yee, 1987. "Asymptotic properties of maximum lieklihood estimators and likelihood ratio tests under nonstandard conditions", *Journal of the American Statistical Association*, 82:605-610.

Terrell, G.R., 1990. "The maximal smoothing principle in density estimation", *Journal of the American Statistical Association*, 85: 470-477.

White, H. 1981. "Consequences and detection of misspecified nonlinear regression models." *Journal of the American Statistical Association* 76:419-433.

White, H. 1982. "Maximum likelihood estimation of misspecified models." *Econometrica* 50:1-25.

Wolak, Frank, 1991. "The local nature of hypothesis tests involving inequality constraints in nonlinear models", *Econometrica*, 59:981-995.