CONSTRAINED COVARIANCE COMPONENT MODELS

By

Frank H. Shaw

and

Charles J. Geyer

IMA Preprint Series # 1189

November 1993
CONSTRUCTED COVARIANCE COMPONENT MODELS

FRANK H. SHAH AND CHARLES J. GEYER

IMA, University of Minnesota

Introduction.

A thorny problem in the use of maximum likelihood for variance component analysis has been that of maximizing the likelihood function in the feasible region of the parameter space. More generally, concerning negative estimates of variance components, Searle (1971) remarks: "It is clearly embarrassing to estimate a variance component as negative, since interpretation of a negative estimate of a non-negative parameter is obviously a problem." A number of methods have been proposed and implemented (Harville, 1977, Callanan and Harville, 1989) to constrain the variance components to non-negative values in univariate analysis. In multivariate analysis, however, the constraint problem takes on a more complicated aspect since the matrix of variance (and now covariance) components is no longer diagonal. We must require, for example, that the correlations between variance components be restricted to the interval [-1,1].

This paper presents an algorithm for enforcing feasibility in analysis using maximum likelihood with multivariate data. The method uses linear inequality constraints corresponding to eigenvectors of the variance–covariance component matrix to insure that all eigenvalues are greater than or equal to some predetermined $\epsilon < 0$. In univariate analysis, this method reduces to the active set method used by (Callanan and Harville, 1989). When the MLE is represented by a constrained maximum, the standard technique of comparing the likelihood ratio between the full and null hypotheses to a $\chi^2$ distribution to calculate $P$–values is no longer valid. We present here an "asymptotic" parametric bootstrap for calculating the distribution of likelihood ratios in which the functions representing the alternate and/or the null hypotheses might be constrained to satisfy feasibility.

Key words and phrases. Variance Components, Cutting Plane Algorithms, Asymptotic Parametric Bootstrap.

This research was supported in part by NSF DIR–9112842

Typeset by AMSTeX
An Algorithm to Enforce Feasibility

**Notation.** The model with which we are concerned is the mixed model

\begin{equation}
\mathbf{y} = \mathbf{X}\alpha + \mathbf{Z}\mathbf{u} + \mathbf{e}.
\end{equation}

wherein \(\mathbf{y}\) is an \(n \times 1\) vector of observations, \(\alpha\) is a \(p \times 1\) vector of unobservable fixed effects, \(\mathbf{u}\) is a \(q \times 1\) vector of unobservable random effects, \(\mathbf{e}\) is an \(n \times 1\) vector of unobservable random errors, \(\mathbf{X}\) is an \(n \times p\) known matrix, and \(\mathbf{Z}\) is an \(n \times q\) known matrix. We have \(\mathbb{E}(\mathbf{u}) = \mathbf{0}\), \(\mathbb{E}(\mathbf{e}) = \mathbf{0}\), \(\text{cov}(\mathbf{u}, \mathbf{e}) = \mathbf{0}\), and

\[\text{var}(\mathbf{y}) = \mathbf{V} = \mathbf{R} + \mathbf{Z}\mathbf{D}\mathbf{Z}'\]

where \(\mathbf{R}\) and \(\mathbf{D}\) are the the variance covariance matrices of \(\mathbf{e}\) and \(\mathbf{u}\) respectively. The elements of \(\mathbf{R}\) and \(\mathbf{D}\) are functions of an unobservable parameter vector \(\theta\). We can rewrite the above in the case of univariate data as

\begin{equation}
\mathbf{V} = \sum_{i=1}^{m} \theta_i \mathbf{G}_i
\end{equation}

where the \(n \times n\) matrices \(\mathbf{G}_i\) are known. In the case of \(r\)-variate data,

\begin{equation}
\mathbf{V} = \sum_{i=1}^{m} \mathbf{T}_i \otimes \mathbf{G}_i
\end{equation}

where \(\otimes\) denotes the outer product and the symmetric \(r \times r\) matrices \(\mathbf{T}_i\) are the "blocks" of the block diagonal matrix \(\mathbf{T}\) in which \(\mathbf{T}_{ij,k}\) is the element of the vector \(\theta\) corresponding to the variance or covariance component of the \(j\)-th and \(k\)-th variates associated with the \(i\)-th random factor. The matrix \(\mathbf{T} = \text{diag}\{\mathbf{T}_i\}_{i=1}^{m}\) contains \(mc = mr(r+1)/2\) distinct parameters. Define \(\mathbf{T}(\theta)\) for \(\theta \in \mathbb{R}^{mc}\) to be the matrix \(\mathbf{T}\) containing the parameter values \(\theta\).

**The Feasible Region.** In the univariate case (2) we define the feasible region of the parameter space \(\Omega_1\) to be \(\mathbb{R}^{m^+}\), the region of \(\mathbb{R}^m\) where all coordinates are positive. In the case of (3) we define the feasible region \(\Omega_r\) to be that in which each \(\mathbf{T}_i\) is positive semi-definite. We can then formally write the expression for the feasible region as

\[\Omega_r = \{\theta \in \mathbb{R}^{cm} | x^T\mathbf{T}(\theta)x \geq 0 \ \forall x \in \mathbb{R}^k}\].

The boundary of the feasible region is then

\[\partial\Omega_r = \{\theta \in \Omega_r | \exists x | \mathbf{T}(\theta)x = 0\}\].
i.e. $T(\theta)$ has a null eigenvector $x$.

Following directly from the definition of $\Omega_r$ we note that each $\theta$ for which there exists $x$ such that $x^T T(\theta) x < 0$ is unfeasible. Then the hyperplane

$$C_x = \{ \theta \in \mathbb{R}^c \mid x^T T(\theta) x = 0 \}$$

separates $\mathbb{R}^c$ into two half-spaces one of which contains the interior of $\Omega_r$ in its entirety. If $\theta'$ is an unfeasible point and $x'$ is the eigenvector corresponding to a negative eigenvalue $\lambda$ of $T(\theta')$. Then $C_{x'}$ separates $\theta'$ from the feasible region. This observation leads us to our algorithm.

**Example 1.** Consider the univariate case with two random factors. Then $r = 1$, $c = 1$ and $m = 2$. The feasible region of $\mathbb{R}^2$ is then the positive-positive quadrant. The matrix $T$ is a diagonal $2 \times 2$ matrix. For $\theta = \{ \theta_1, \theta_2 \}$,

$$T(\theta) = \begin{pmatrix} \theta_1 & 0 \\
0 & \theta_2 \end{pmatrix}$$

and $x^T T(\theta) x = 0$ can be written as $x_1^2 \theta_1 + x_2^2 \theta_2 = 0$. Then $C_x$ is a line with negative slope passing through the origin. Clearly, for $x_1^2 \theta_1 + x_2^2 \theta_2 < 0$ either $\theta_1$ or $\theta_2$ or both must be negative, and thus unfeasible.

If $\theta' \in \mathbb{R}^2$ is such that $T(\theta')$ has a null eigenvector $x'$ then $x'$ represents a coordinate axis.

**Example 2.** Consider the bivariate case with only one random factor. Here $r = 2$, $c = 3$, and $m = 1$ and we have

$$T(\theta) = \begin{pmatrix} \theta_1 & \theta_2 \\
\theta_2 & \theta_3 \end{pmatrix}.$$ 

The plane $C_x$ is defined by the equation $x_1^2 \theta_1 + 2x_1 x_2 \theta_3 + x_2^2 \theta_3 = 0$. If $x_1^2 \theta_1 + 2x_1 x_2 \theta_3 + x_2^2 \theta_3 < 0$, then one of $\theta_1$, $\theta_2$ is negative or $\theta_2^2 > \sqrt{\theta_1 \theta_3}$, the latter violating the requirement that the correlation be between -1 and 1.

Let $\theta'$ and $x' \neq 0$ be such that $T(\theta')x' = 0$, implying that $\theta'$ lies on the boundary of the feasible region. Then $C_{x'}$ intersects the feasible region along the ray $\theta = a\theta'$ for $a \in \mathbb{R}^+$, and we note that every $\theta$ along this ray satisfies $T(\theta)x' = 0$. The feasible region is thus a cone.

**The Algorithm.**

The above characterization of the feasible region is perfectly suited to the use of outer approximation cutting plane techniques long known in the fields of linear, convex, and semi-infinite programming (Horst and Tuy, 1990). The technique is
simple: if the maximum is reached at an unfeasible point, find a hyperplane separating the point from the feasible region and then maximize the function on the half-space containing the feasible region.

1. Set $\epsilon > 0$. Maximize the Log Likelihood without constraint: $\text{argmax } l(\theta) = \theta_0$.

2. For $k = 1, 2, \ldots$. Denote the smallest eigenvalue of $T(\theta_k)$ as $\lambda_k^-$. If $\lambda_k^- > -\epsilon$ we are finished. Otherwise construct the constraint function $c_k(\theta_k) = x^T T(\theta) x$ where $x$ is the eigenvector corresponding to $\lambda_k^-$. Note that $c_k(\theta_k) = \lambda_k^-$. Maximize the Log Likelihood subject to the constraints $c_i \geq 0, i = 1 \ldots k$. Put $\theta_{k+1}$ to be the maximizer subject to these constraints.

Horst and Tuy (Horst and Tuy, 1990, p55) have a theorem which guarantees that this algorithm will converge (as $\epsilon$ tends to zero) to a feasible constrained maximum under the following conditions.

1. The sets $D_k = \{ \theta \in \Omega_0 | c_i(\theta) \geq 0, i = 1, 2, \ldots, k \}$ are closed and global optimization of the Log Likelihood function is possible on the $D_k$.

2. For each convergent subsequence $\{ \theta_{i_q} \} \subset \{ \theta_i \}$ we have

$$c_{i_q}(\theta_{i_q}) \rightarrow 0$$

The $D_k$ are clearly closed since each $D_k$ comes about from the deletion of the intersection of an open halfspace with $D_{k-1}$ from $D_{k-1}$. To establish (2) we suppose that the smallest eigenvalue $\tilde{\lambda}$ of $T(\tilde{\theta})$ is strictly less than zero. Then we will be able to find an $i_q$ such that $c_{i_q}(\tilde{\theta}) < 0$. Let $\epsilon = \frac{|\tilde{\lambda}|}{2r(r+1)}$ where $r$ is the order of $T(\theta)$. We may then find $q$ such that

$$T(\tilde{\theta}) = T(\theta_{i_q}) + \epsilon B$$

where the elements of $B$ have magnitude less than one. Application of the $i_q$–th constraint equation to $\tilde{\theta}$ gives

$$c_{i_q}(\tilde{\theta}) = x_{i_q}^T T(\tilde{\theta}) x_{i_q} = x_{i_q}^T T(\theta_{i_q}) x_{i_q} + \epsilon x_{i_q}^T B x_{i_q} = \lambda_{i_q}^- + \epsilon x_{i_q}^T B x_{i_q}$$

and we note that

$$|\epsilon x_{i_q}^T B x_{i_q}| \leq \epsilon \|B\|_2 \|x_{i_q}\|^2 \leq \epsilon r.$$ 

Furthermore, Gershgorin's theorems (Wilkinson, 1963, pp 72ff) show that the eigenvalues of $T(\tilde{\theta})$ lie in disks centered near the eigenvalues of $T(\theta_{i_q})$ with know radius. From this we can write $|\tilde{\lambda} - \lambda_{i_q}^-| \leq \epsilon r^2$. But then $c_{i_q}(\tilde{\theta}) \leq \epsilon r(r + 1) < 0$. This contradicts the requirement that the inequality constraints $c_i(\tilde{\theta}) \geq 0$. 
**Bootstrap Likelihood Ratio Tests**

According to a theorem of Wilks (1938) the asymptotic distribution of twice the log likelihood ratio is chi-square for a test of nested hypotheses when the hypothesis sets are flat in a neighborhood of the true parameter value. The same result obtains even when smooth equality constraints are imposed (Aitchison and Silvey, 1958). We leave this familiar territory when inequality constraints enter the problem. Then it is sometimes true that the asymptotic distribution of twice the log likelihood ratio is a mixture of chi-square distributions with different degrees of freedom, but not always true (Self and Liang, 1987). The general form of the asymptotic distribution was given by Chernoff (1954), and the general form of the asymptotic distribution of the maximum likelihood estimate is given by Geyer (1994), but these asymptotic distributions generally cannot be calculated analytically. Moreover, the distributions depend on the unknown true value of the parameter and hence cannot be exactly calculated even in principle. This is very different from the unconstrained or equality-constrained case where the asymptotic distribution (chi-square) does not depend on the true parameter value.

The natural procedure for attempting to approximate the sampling distribution of the likelihood ratio test statistic is a parametric bootstrap (Efron, 1979). Given an observed data set \( x \) we calculate the maximum likelihood (or restricted maximum likelihood) estimates \( \hat{\theta}_0(x) \) and \( \hat{\theta}_1(x) \) in the null and alternative hypotheses and form the log likelihood ratio statistic

\[
r(x) = l(\hat{\theta}_1(x)) - l(\hat{\theta}_0(x))
\]

where \( l \) is the log likelihood (or restricted log likelihood for REML). We then simulate new data sets \( x_1^*, \ldots, x_m^* \) from the model indexed by the parameter value \( \hat{\theta}_0(x) \). For each of these simulated data sets we calculate the estimates \( \hat{\theta}_0(x_i^*) \) and \( \hat{\theta}_1(x_i^*) \). The simulation parameter value \( \hat{\theta}_0(x) \) is our best estimate of the true parameter value \( \theta_0 \) under the null hypothesis. So the simulated data sets are our best estimate of the true distribution of the data (under the null hypothesis and assuming the model is correct), the simulation distribution estimates \( \hat{\theta}_0(x_i^*) \) and \( \hat{\theta}_1(x_i^*) \) are our best estimates of the sampling distributions of the estimates, and the simulation distribution of the log likelihood ratio statistic

\[
r(x_i^*) = l(\hat{\theta}_1(x_i^*)) - l(\hat{\theta}_0(x_i^*))
\]

is our best estimate of the sampling distribution of the log likelihood ratio. Our best estimate of the \( P \)-value for the likelihood ratio test is the fraction of the \( r(x_i^*) \) that exceed \( r(x) \). More precisely, if the number \( m \) of bootstrap samples is small, we should count the observed data set among the simulations and use

\[
P(x) = \frac{\#\{ r(x_i^*) \geq r(x) : i = 1, \ldots, m \} + 1}{m + 1}
\]
as our estimate of the $P$-value (Geyer, 1991, equation 6).

In order to justify this procedure, we need a demonstration that the bootstrap is valid. Ever since Efron (1979) called attention to the issue, statisticians have generally agreed that the bootstrap can be said to “work” only if it is consistent, that is if the bootstrap estimate of the sampling distribution of statistic in question (here the likelihood ratio test statistic) and actual sampling distribution of the statistic converge to the same limit as the size of the data set goes to infinity. One is allowed to bootstrap a problem only after one has proved two central limit theorems and shown that they have the same limiting distribution.

It is an unfortunate fact that the bootstrap is generally not consistent in inequality-constrained inference, even in the simplest problems. Consider the following problem, which is the simplest case of inequality-constrained statistical inference. We observe a sample $x_1, \ldots, x_n$ that are i. i. d. normal with unknown mean $\theta$ and known variance 1. The null hypothesis assumes $\theta \geq 0$, and the alternative hypothesis leaves $\theta$ unconstrained. Then it is clear that the maximum likelihood (and least squares) estimate of $\theta$ subject to the constraint $\theta \geq 0$ is

$$\hat{\theta}_0(\bar{x}) = \begin{cases} \bar{x}, & \text{if } \bar{x} \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

where $\bar{x}$ is the sample mean of the $x_i$, and $\hat{\theta}_1(\bar{x}) = \bar{x}$ is the maximum likelihood estimate in the alternative hypothesis. Then twice the log likelihood ratio is

$$2r(x) = n(\bar{x} - \hat{\theta}_0(\bar{x}))^2 - n(\bar{x} - \hat{\theta}_1(\bar{x}))^2 = \begin{cases} n\bar{x}^2, & \text{if } \bar{x} \leq 0 \\ 0, & \text{otherwise} \end{cases}$$

Because of the assumed normality of the data, $\sqrt{n}\bar{x}$ is exactly normal for all $n$, and $n\bar{x}^2$ is exactly $\chi^2$ with one degree of freedom. So the exact sampling distribution of twice the log likelihood ratio is an equal mixture of a $\chi^2$ distribution with one d. f. and an atom at zero. Now let us consider the parametric bootstrap estimate of this distribution when the true parameter value is $\theta_0 = 0$. We simulate $x_i^*$ from a normal distribution with mean $\hat{\theta}_0(\bar{x})$ and variance 1 and calculate their mean $\bar{x}^*$ and twice the log likelihood ratio

$$2r(x^*) = n(\bar{x}^* - \hat{\theta}_0(\bar{x}^*))^2 - n(\bar{x}^* - \hat{\theta}_1(\bar{x}^*))^2 = \begin{cases} n(\bar{x}^*)^2, & \text{if } \bar{x}^* \leq 0 \\ 0, & \text{otherwise} \end{cases}$$

The simulation distribution of $2r(x^*)$ is the bootstrap estimate of the sampling distribution of (1). Despite the similar appearance, (1) and (2) have different distributions because the sampling distribution of $\sqrt{n}\bar{x}$ is centered at the true parameter value $\theta_0 = 0$, and the simulation distribution of $\sqrt{n}\bar{x}^*$ is centered at $\sqrt{n}\hat{\theta}_0(\bar{x})$. 
Since the distribution of $\sqrt{n}\tilde{\theta}_0(\bar{x})$ does not depend on the sample size $n$, the asymptotic distributions are the same as the finite sample distributions, and both the asymptotic and finite-sample-size distributions disagree. The bootstrap is not consistent.

If we think about why the bootstrap doesn’t work here, it is clear that the problem is the maximum likelihood estimate does not always get the right active set. At the true parameter value $\theta_0 = 0$, the constraint $\theta \geq 0$ is active, but at the maximum likelihood estimate $\hat{\theta}_0(\bar{x})$, the constraint is active only when the estimate is zero, which occurs only half the time (when $\bar{x} < 0$). Furthermore, the constraint is not even close to being active, when we measure distance on the proper scale for asymptotics: $\sqrt{n}\tilde{\theta}_0(\bar{x})$ is about one unit away from zero for all $n$ and does not converge in probability to zero.

This observation suggests a way to fix the bootstrap. Let $\tilde{\theta}_0(x)$ be a superefficient estimator of $\theta_0$ such as the original example of Hodges (Le Cam, 1953)

$$
\tilde{\theta}_0(\bar{x}) = \begin{cases} 
\bar{x}, & \text{if } \bar{x} \geq n^{-1/4} \\
0, & \text{otherwise}
\end{cases}
$$

Now $\sqrt{n}\tilde{\theta}_0(\bar{x})$ converges in probability to zero when $\theta_0 = 0$ and converges in probability to $\sqrt{n}\hat{\theta}_0(\bar{x})$ when $\theta_0 > 0$. If we simulate our bootstrap samples from the normal distribution centered at $\tilde{\theta}_0(\bar{x})$, the bootstrap will be consistent. Although we are then entitled to say the bootstrap “works” according to general usage, our procedure has several undesirable properties. The first is that $\tilde{\theta}_0(\bar{x})$ is not generally a better estimator of $\theta$ than $\hat{\theta}_0(\bar{x})$. In fact it performs very badly when $\theta_0$ is near $n^{-1/4}$ (Le Cam, 1953). An even more undesirable property of this or any other superefficient estimator is that it is inherently asymptotic and provides no guidance whatsoever as to how one should analyze a particular data set. After all, in real life we never have a sequence of data sets of sample sizes tending to infinity, and usually have just one data set of fixed size. The $n^{-1/4}$ in (3) can be replaced by $An^{-1/4}$ for any $A > 0$ or by $n^{-\alpha}$ for any $\alpha$ strictly between 1/2 and 0 and the estimator would still be superefficient at zero. For a fixed sample size $n$ and $\bar{x} > 0$, whether we set $\tilde{\theta}_0(\bar{x})$ to zero or $\bar{x}$ depends on which superefficient estimator we choose. Although we have repaired the asymptotics, we have done so in a way which is absolutely useless as a guide to practical data analysis.

It might be thought that there should be some data-dependent way to fix the asymptotics, perhaps in the spirit of cross-validation, but our simple example dashes such hopes. Once we reduce to the sufficient statistic $\bar{x}$, we have essentially the same problem for all sample sizes. There seems to be no way to produce a consistent bootstrap without introducing arbitrariness in the procedure.

We have spent a great deal of discussion on this particular example, but it illustrates the general situation. All inequality-constrained problems will have inconsistency of the bootstrap for exactly the same reasons. Most problems with a
finite number of inequality constraints can have the inconsistency repaired by using super-efficient estimators of the active constraint sets, but while repairing the asymptotics, this provides no guide in data analysis. Even problems with an infinite set of inequality constraints, such as our covariance components problems can have the inconsistency of the bootstrap repaired by a different device suggested to us by R. T. Rockafellar (personal communication). Here we simulate from the maximum likelihood estimate $\hat{\theta}_0(x)$, but change the constraint set. Rather than moving the estimate closer to the constraint boundaries, we move the boundaries toward the estimate, achieving much the same thing. This solution, however, suffers from the same practical problems as the other. The rate of shrinkage must be chosen arbitrarily and gives no guidance as to what to do in a practical problem.

Having gone through this analysis, we see that perhaps the consistency criterion for deciding whether bootstrapping is appropriate is misguided in this case, and perhaps we should just use the parametric bootstrap in these problems without worrying about the inconsistency. What practical experience there is with the parametric bootstrap in inequality-constrained maximum likelihood (Geyer, 1991) seems to suggest that it works well in practice, despite inconsistency. If one is seriously worried about inconsistency, Geyer (1991) presents a diagnostic procedure to allay fears. Essentially this is a simulation study to check the performance of the bootstrap in the problem at hand, using a so-called "double bootstrap" or "bootstrap after bootstrap." Although a great deal of work, the double bootstrap can provide evidence that the bootstrap does work well in the problem at hand even though it would not perform consistently on an infinite sequence of problems. The amount of calculation involved in a double bootstrap can be greatly reduced by importance sampling (Newton and Geyer, 1994).

**The Asymptotic Parametric Bootstrap.**

Leaving aside a double bootstrap, even an ordinary parametric bootstrap can be very expensive in a large variance-covariance components problem. When a single fit may take hours of computer time, a bootstrap may take weeks. To simplify the procedure, we propose doing what we call the "asymptotic" parametric bootstrap in which the actual log likelihood for the problem is replaced by its asymptotic quadratic approximation based on the well-known identities

$$E_\theta[\nabla l_X(\theta)] = 0$$

and

$$\text{var}_\theta[\nabla l_X(\theta)] = -E_\theta[\nabla^2 l_X(\theta)] = I(\theta)$$  \hspace{1cm} (4)

where $l_X(\theta)$ denotes the log likelihood for the parameter value $\theta$ and observation $X$ considered as a random function of $X$ and $I(\theta)$, the (expected) Fisher information, is defined by the second term in (4).
Assuming that the amount of data is large enough so this asymptotic approximation is not too bad, we may replace the exact log likelihood \( l_X(\theta) \) by its usual asymptotic approximation

\[
\tilde{l}_Z(\theta) = (\theta - \theta_0)^T Z - \frac{1}{2} (\theta - \theta_0)^T I(\theta_0)(\theta - \theta_0)
\]

where \( Z \) is a normal random vector with mean zero and variance-covariance matrix \( I(\theta_0) \). As usual in bootstrapping we replace the unknown true parameter value by its estimate \( \hat{\theta}_0(x) \). Now we are replacing the exact log likelihood \( l_X \) by the random quadratic function \( \tilde{l}_Z \) whose gradient is normal with the same mean and variance-covariance matrix as the gradient of \( l_X \) and whose Hessian is the expected value of the Hessian of \( l_X \).

To perform the asymptotic bootstrap we simulate normal random vectors \( z_i^* \) with mean zero and variance-covariance matrix \( I(\hat{\theta}_0(x)) \) and calculate “asymptotic” maximum likelihood estimates in the null and alternative hypotheses \( \hat{\theta}_0(z_i^*) \) and \( \hat{\theta}_1(z_i^*) \) by maximizing (5) with \( \theta_0 \) set to \( \hat{\theta}_0(x) \) over the constrained parameter sets for the two hypotheses. We then form the “asymptotic” log likelihood ratio

\[
r(z_i^*) = \tilde{l}_Z(\hat{\theta}_1(z_i^*)) - \tilde{l}_Z(\hat{\theta}_0(z_i^*))
\]

and the \( P \)-value of the test is the fraction of \( r(z_i^*) \) that exceed the observed log likelihood ratio \( r(x) \).

This replacement of \( l_X \) by \( \tilde{l}_Z \) is exactly what is done in the familiar asymptotics of maximum likelihood, although the description is a bit unusual to accommodate the constrained parameter sets. If we have no constraints, this “asymptotic” bootstrap will just calculate the appropriate \( \chi^2 \) asymptotic distribution of twice the log likelihood ratio when the number of simulated \( z_i^* \) is large. Hence it should be no less acceptable to use this “asymptotic” bootstrap than it would be to use the \( \chi^2 \) asymptotics in the unconstrained case.

The asymptotic bootstrap, of course, does nothing about the maximum likelihood estimate underestimating the number of active constraints. Hence the asymptotic bootstrap, like the ordinary bootstrap is inconsistent. As with the ordinary bootstrap, the inconsistency can be repaired by using superefficient estimators of the active constraint set or by shrinking the parameter set toward the maximum likelihood estimate, but such estimators provide no guidance for finite sample sizes.

**Examples.** Data here are from an experiment to investigate the genetic correlation
between performance in differing environments (Shaw and Platenkamp, MS).

![Asymptotic Bootstrap Log Likelihood Ratio * 2](image)

**Figure 1.**

When feasibility constraints are not imposed, the MLE for the additive genetic variance–covariance matrix is

\[
\begin{pmatrix}
0.05529213 & -0.11960102 \\
-0.11960102 & 0.02106031
\end{pmatrix}
\]

in which the genetic correlation is equal to -3.5. With feasibility constraints applied, the additive genetic variance–covariance matrix becomes

\[
\begin{pmatrix}
0.08605613 & -0.07972882 \\
-0.07972882 & 0.07356314
\end{pmatrix}
\]

where the genetic correlation is now nearly -1. We wish to test the difference of this correlation from zero. Figure 1 is a histogram of 1000 likelihood ratios simulated as detailed above from the information matrix under the null hypothesis of no correlation:

\[
\begin{pmatrix}
0.04977596 & 0.00000000 \\
0.00000000 & 0.02732595
\end{pmatrix}
\]
The line plotted on Figure 1 is the $\chi^2$ density for one degree of freedom.

In Figure 2 we test the null hypothesis that the genetic correlation is greater than zero. Here, the sizable atom at zero (64%) has been removed and the remainder is seen to closely follow the $\chi^2$ density for one degree of freedom. If the atom at zero is ignored, the $\chi^2$ density fits the histogram very poorly. Since the size of the atom at zero depends upon the proximity of the alternative model MLE to the boundary of the feasible region, the asymptotic parametric bootstrap is necessary to determine
the $P$-value.

![Asymptotic Bootstrap Log Likelihood Ratio * 2](image)

**Figure 3.**

In Figure 3 the test is whether the genetic variance–covariance matrix

$$
\begin{pmatrix}
0.07320575 & 0.03277579 \\
0.03277579 & 0.01446851
\end{pmatrix}
$$

sports a genetic correlation different from zero. Because the variance component for the second trait is close to the boundary of the feasible region, the number of constraints imposed on the full and null models in the simulation is variable with an average difference of only 0.6. The superimposed line is the $\chi^2$ with one degree of freedom. Though the $\chi^2$ distribution might appear to be an appropriate model of the distribution of the likelihood ratio in this case, it fails to model the distribution correctly where it counts most: in the tail. The asymptotic bootstrap gives a $P$-value of .05 at the (doubled) ratio 2.50. This number does not even satisfy $P < 0.1$ if the $\chi^2$ comparison is used.

In a reanalysis of data obtained by Wilkinson et al (1990), the additive genetic variance–covariance component matrices from two populations were constrained to
be the same for five traits.

![Asymptotic Bootstrap Log Likelihood Ratio * 2](image)

**Figure 4.**

Thus 15 equality constraints are imposed upon the full model to generate the null hypothesis. In both the full and null model one extra constraint was necessary to force feasibility. This was usually the same constraint and it involved a different random factor from that constrained in the null hypothesis. The average constraint difference between full and null models was then 14.978. In this case, the distribution of the likelihood ratio generated by the asymptotic parametric bootstrap (Figure 4) follows that of the $\chi^2$ with 15 degrees of freedom quite closely.
Bibliography


<table>
<thead>
<tr>
<th>#</th>
<th>Author/s</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1108</td>
<td>Lawrence Markus, A brief history of control</td>
<td>Recent IMA Preprints</td>
</tr>
<tr>
<td>1109</td>
<td>Richard A. Brualdi, Keith L. Chavey &amp; Bryan L. Shader</td>
<td>Bipartite graphs and inverse sign patterns of strong sign-nonsingular matrices</td>
</tr>
<tr>
<td>1110</td>
<td>A. Kersch, W. Morokoff &amp; A. Schuster</td>
<td>Radiative heat transfer with quasi-monte carlo methods</td>
</tr>
<tr>
<td>1111</td>
<td>Jianhua Zhang</td>
<td>A free boundary problem arising from swelling-controlled release processes</td>
</tr>
<tr>
<td>1112</td>
<td>Walter Littman &amp; Stephen Taylor</td>
<td>Local smoothing and energy decay for a semi-infinite beam pinned at several points and applications to boundary control</td>
</tr>
<tr>
<td>1113</td>
<td>Srdjan Stojanovic &amp; Thomas Svobodny</td>
<td>A free boundary problem for the Stokes equation via nonsmooth analysis</td>
</tr>
<tr>
<td>1114</td>
<td>Bronislaw Jakubczyk</td>
<td>Filtered differential algebras are complete invariants of static feedback</td>
</tr>
<tr>
<td>1115</td>
<td>Boris Mordukhovich</td>
<td>Discrete approximations and refined Euler-Lagrange conditions for nonconvex differential inclusions</td>
</tr>
<tr>
<td>1116</td>
<td>Bei Hu &amp; Hong-Ming Yin</td>
<td>The profile near blowup time for solution of the heat equation with a nonlinear boundary condition</td>
</tr>
<tr>
<td>1117</td>
<td>Jin Ma &amp; Jiongmin Yong</td>
<td>Solvability of forward-backward SDEs and the nodal set of Hamilton-Jacobi-Bellman Equations</td>
</tr>
<tr>
<td>1118</td>
<td>Chaocheng Huang &amp; Jiongmin Yong</td>
<td>Coupled parabolic and hyperbolic equations modeling age-dependent epidemic dynamics with nonlinear diffusion</td>
</tr>
<tr>
<td>1119</td>
<td>Jiongmin Yong</td>
<td>Necessary conditions for minimax control problems of second order elliptic partial differential equations related to ecology</td>
</tr>
<tr>
<td>1120</td>
<td>Eitan Altman &amp; Nahum Shimkin</td>
<td>Worst-case and Nash routing policies in parallel queues with uncertain service allocations</td>
</tr>
<tr>
<td>1121</td>
<td>Nahum Shimkin &amp; Adam Shwartz</td>
<td>Asymptotically efficient adaptive strategies in repeated games, part II: Asymptotic optimality</td>
</tr>
<tr>
<td>1122</td>
<td>M.E. Bradley</td>
<td>Well-posedness and regularity results for a dynamic Von Kármán plate</td>
</tr>
<tr>
<td>1123</td>
<td>Zhangxin Chen</td>
<td>Finite element analysis of the 1D full drift diffusion semiconductor model</td>
</tr>
<tr>
<td>1124</td>
<td>Gang Bao &amp; David C. Dobson</td>
<td>Diffractive optics in nonlinear media with periodic structure</td>
</tr>
<tr>
<td>1125</td>
<td>Steven Cox &amp; Enrique Zuazua</td>
<td>The rate at which energy decays in a damped string</td>
</tr>
<tr>
<td>1126</td>
<td>Anthony W. Leung</td>
<td>Optimal control for nonlinear systems of partial differential equations related to ecology</td>
</tr>
<tr>
<td>1127</td>
<td>H.J. Sussmann</td>
<td>A continuation method for nonholonomic path-finding problems</td>
</tr>
<tr>
<td>1128</td>
<td>Yung-Jen Guo &amp; Walter Littman</td>
<td>The null boundary controllability for semilinear heat equations</td>
</tr>
<tr>
<td>1129</td>
<td>Q. Zhang &amp; G. Yin</td>
<td>Turnpike sets in stochastic manufacturing systems with finite time horizon</td>
</tr>
<tr>
<td>1130</td>
<td>I. Györi, F. Hartung &amp; J. Turi</td>
<td>Approximation of functional differential equations with time- and state-dependent delays by equations with piecewise constant arguments</td>
</tr>
<tr>
<td>1131</td>
<td>I. Györi, F. Hartung &amp; J. Turi</td>
<td>Stability in delay equations with perturbed time lags</td>
</tr>
<tr>
<td>1132</td>
<td>F. Hartung &amp; J. Turi</td>
<td>On the asymptotic behavior of the solutions of a state-dependent delay equation</td>
</tr>
<tr>
<td>1133</td>
<td>Pierre-Alain Gremaud</td>
<td>Numerical optimization and quasiconvexity</td>
</tr>
<tr>
<td>1134</td>
<td>Jie Tai Yu</td>
<td>Resultants and inversion formula for N polynomials in N variables</td>
</tr>
<tr>
<td>1135</td>
<td>Avner Friedman &amp; J.L. Velázquez</td>
<td>The analysis of coating flows in a strip</td>
</tr>
<tr>
<td>1136</td>
<td>Eduardo D. Sontag</td>
<td>Control of systems without drift via generic loops</td>
</tr>
<tr>
<td>1137</td>
<td>Yuan Wang &amp; Eduardo D. Sontag</td>
<td>Orders of input/output differential equations and state space dimensions</td>
</tr>
<tr>
<td>1138</td>
<td>Scott W. Hansen</td>
<td>Boundary control of a one-dimensional, linear, thermoelastic rod</td>
</tr>
<tr>
<td>1139</td>
<td>Robert Lipton &amp; Bogdan Vernescu</td>
<td>Homogenization of two phase emulsions with surface tension effects</td>
</tr>
<tr>
<td>1140</td>
<td>Scott Hansen &amp; Enrique Zuazua</td>
<td>Exact controllability and stabilization of a vibrating string with an interior point mass</td>
</tr>
<tr>
<td>1141</td>
<td>Bei Hu &amp; Jiongmin Yong</td>
<td>Pontryagin Maximum principle for semilinear and quasilinear parabolic equations with pointwise state constraints</td>
</tr>
<tr>
<td>1142</td>
<td>Mark H.A. Davis</td>
<td>A deterministic approach to optimal stopping with application to a prophet inequality</td>
</tr>
<tr>
<td>1143</td>
<td>M.H.A. Davis &amp; M. Zervos</td>
<td>A problem of singular stochastic control with discretionary stopping</td>
</tr>
<tr>
<td>1144</td>
<td>Bernardo Cockburn &amp; Pierre-Alain Gremaud</td>
<td>An error estimate for finite element methods for scalar conservation laws</td>
</tr>
<tr>
<td>1145</td>
<td>David C. Dobson &amp; Fadil Santosa</td>
<td>An image enhancement technique for electrical impedance tomography</td>
</tr>
<tr>
<td>1146</td>
<td>Jin Ma, Philip Protter, &amp; Jiongmin Yong</td>
<td>Solving forward-backward stochastic differential equations explicitly — a four step scheme</td>
</tr>
<tr>
<td>1147</td>
<td>Yong Liu</td>
<td>The equilibrium plasma subject to skin effect</td>
</tr>
<tr>
<td>1148</td>
<td>Ulrich Hornung</td>
<td>Models for flow and transport through porous media derived by homogenization</td>
</tr>
<tr>
<td>1149</td>
<td>Avner Friedman, Chaocheng Huang, &amp; Jiongmin Yong</td>
<td>Effective permeability of the boundary of a domain</td>
</tr>
<tr>
<td>1150</td>
<td>Gang Bao</td>
<td>A uniqueness theorem for an inverse problem in periodic diffractive optics</td>
</tr>
<tr>
<td>1151</td>
<td>Angelo Favini, Mary Ann Horn, Irena Lasiecka</td>
<td>Global existence and uniqueness of regular solutions to the dynamic von Kármán system with nonlinear boundary dissipation</td>
</tr>
</tbody>
</table>
E.G. Kalnin & Willard Miller, Jr., Models of q-algebra representations: q-integral transforms and “addition theorems”

E.G. Kalnin, V.B. Kuznetsov & Willard Miller, Jr., Quadrics on complex Riemannian spaces of constant curvature, separation of variables and the Gaudin magnet

A. Kersch, W. Morokoff & Chr. Werner, Selfconsistent simulation of sputtering with the DSMC method

Bing-Yu Zhang, A remark on the Cauchy problem for the Korteweg-de Vries equation on a periodic domain

Gang Biao, Finite element approximation of time harmonic waves in periodic structures

Tao Lin & Hong Wang, Recovering the gradients of the solutions of second-order hyperbolic equations by interpolating the finite element solutions

Zhangxin Chen, Lp-posteriori error analysis of mixed methods for linear and quasilinear elliptic problems

Todd Arbogast & Zhangxin Chen, Homogenization of compositional flow in fractured porous media

L. Qiu, B. Bernhardsson, A. Rantzner, E.J. Davison, P.M. Young & J.C. Doyle, A formula for computation of the real stability radius

Maria Inés Troparevsky, Adaptive control of linear discrete time systems with external disturbances under inaccurate modelling: A case study

Petr Klouček & Franz S. Rys, Stability of the fractional step θ-scheme for the nonstationary Navier-Stokes equations

Eduardo Casas, Luis A. Fernández & Jiongmin Yong, Optimal control of quasilinear parabolic equations

Darrell Duffie, Jin Ma & Jiongmin Yong, Black’s consol rate conjecture

D.G. Aronson & J.L. Vazquez, Anomalous exponents in nonlinear diffusion

Ruben D. Spies, Local existence and regularity of solutions for a mathematical model of thermomechanical phase transitions in shape memory materials with Landau-Ginzburg free energy

Pu Sun, On circular pipe Poiseuille flow instabilities

Angelo Favini, Mary Ann Horn, Irena Lasiecka & Daniel Tataru, Global existence, uniqueness and regularity of solutions to a Von Kármán system with nonlinear boundary dissipation

A. Dončheva, Tz. Donchev & I. Slavov, On the upper semicontinuity of the set of solutions of differential inclusions with a small parameter in the derivative

Jin Ma & Jiongmin Yong, Regular-singular stochastic controls for higher dimensional diffusions — dynamic programming approach

Alex Solomonoff, Bayes finite difference schemes

Todd Arbogast, On the implementation of mixed methods as nonconforming methods for second order elliptic problems

Zhangxin Chen & Bernardo Cockburn, Convergence of a finite element method for the drift-diffusion semiconductor device equations: The multidimensional case

Boris Mordukhovich, Optimization and finite difference approximations of nonconvex differential inclusions with free time

Avner Friedman, David S. Ross, and Jianhua Zhang, A Stefan problem for reaction-diffusion system

Alex Solomonoff, Fast algorithms for micromagnetic computations

Nikan B. Firoozye, Homogenization on lattices: Small parameter limits, H-measures, and discrete Wigner measures

G. Yin, Adaptive filtering with averaging

Wlodzimierz Byrc and Amir Dembo, Large deviations for quadratic functionals of Gaussian processes

Ilja Schmelzer, 3D anisotropic grid generation with intersection-based geometry interface

Alex Solomonoff, Application of multiple methods to two matrix eigenproblems

A.M. Latypov, Numerical solution of steady euler equations in streamline-aligned orthogonal coordinates

Beih Hu & Hong-Ming Yin, Semilinear parabolic equations with prescribed energy

Beih Hu & Jianhua Zhang, Global existence for a class of Non-Fickian polymer-penetrant systems

Rongze Zhao & Thomas A. Posbergh, Robust stabilization of a uniformly rotating rigid body

Mary Ann Horn & Irena Lasiecka, Uniform decay of weak solutions to a von Kármán plate with nonlinear boundary dissipation

Mary Ann Horn, Irena Lasiecka & Daniel Tataru, Well-posedness and uniform decay rates for weak solutions to a von Kármán system with nonlinear dissipative boundary conditions

MaryAnn Horn, Nonlinear boundary stabilization of a von Kármán plate via bending moments only

Frank H. Shaw & Charles J. Geyer, Constrained covariance component models

Tomasz Łuczaka, A greedy algorithm estimating the height of random trees

Timo Seppäläinen, Maximum entropy principles for disordered spins

Yuandian Lin, Eduardo D. Sontag & Yuan Wang, Recent results on Lyapunov-theoretic techniques for nonlinear stability

Svante Janson, Random regular graphs: Asymptotic distributions and contiguity

Rachid Ababou, Random porous media flow on large 3-D grids: Numerics, performance, & application to homogenization