The Annals of Statistics
1996, Vol. 24, No. 3, 1316–1326

ASYMPTOTICS OF LEAST-SQUARES ESTIMATORS FOR CONSTRAINED NONLINEAR REGRESSION[1]

BY JINDE WANG

Nanjing University

This paper is devoted to studying the asymptotic behavior of LS-estimators in constrained nonlinear regression problems. Here the constraints are given by nonlinear equalities and inequalities. Thus this is a very general setting. Essentially this kind of estimation problem is a stochastic optimization problem. So we make use of methods in optimization to overcome the difficulty caused by nonlinearity in the regression model and given constraints.

1. Introduction. Recently, constrained regression problems have been studied by many authors. Most of their papers are restricted to linear regression problems. For example, Liew (1976) considered linear regression with linear constraints. Nagaraj and Fuller (1991) studied linear time series subject to nonlinear equality constraints. Only a few papers are devoted to nonlinear regression problems. Dupacova and Wets (1988) studied very general constrained nonlinear estimation problems. They embedded the statistical estimation problem in the framework of stochastic optimization. However, their results can only be applied to those nonlinear regression models, in which the control variable x can be considered random. This may not always be the case in practical problems.

In this paper we consider nonlinear regression problems with nonlinear equality and inequality constraints and investigate the asymptotic behavior of the LS-estimators in these problems. Thus the problem we are facing is of the following form:

$$
\begin{aligned}
&\text{Min} && \sum_{i=1}^{n} \left(y_i - f(x_i, \theta) \right)^2 \\
&\text{s.t.} && g_i(\theta) \le 0, \qquad i = 1, \dots, p, \\
& && h_j(\theta) = 0, \qquad j = p+1, \dots, q,
\end{aligned}
\tag{1}
$$

where $\theta \in R^m$ is the unknown parameter to be estimated and (x_i, y_i), $i = 1, \dots, n$, are observed data from the nonlinear regression model

$$y_i = f(X_i, \theta) + e_i, \qquad i = 1, \dots, n. \tag{2}$$

Clearly, (1) is a very general problem.

Received July 1993; revised July 1995.

[1] This project was supported by the National Natural Science Foundation of China.

AMS 1991 *subject classifications*. Primary 62J02; secondary 62J12.

Key words and phrases. Nonlinear regression, nonlinear constraints, LS-estimator, asymptotics, stochastic optimization.

Let $\hat{\theta}_n$ be the optimal solution to problem (1) and θ_0 be the true value of θ in model (2). We are interested in the asymptotic behavior of $n^{1/2}(\hat{\theta}_n - \theta_0)$. Because of the appearance of the constraints, especially the inequality constraints, in general, one cannot expect to get normality of $n^{1/2}(\hat{\theta}_n - \theta_0)$ as in the unconstrained regression problems. Moreover, one cannot even expect to get an explicit formula of the asymptotic distribution of the estimators. What we will do in this paper is to show that under some mild conditions $n^{1/2}(\hat{\theta}_n - \theta_0)$ converges in distribution to the optimal solution of a comparatively simpler program, a quadratic stochastic program. This result is given in Theorem 6. In order to get the limit distribution of $n^{1/2}(\hat{\theta}_n - \theta_0)$, we will determine a limit form of problem (1). This will be done in the next section. Then we prove $n^{1/2}(\hat{\theta}_n - \theta_0)$ will converge in distribution to the optimal solution of the limit problem. Mathematically speaking, problem (1) is a stochastic optimization problem. It may be a proper way to use the epigraph convergence theory in optimization to get our desired results.

2. The limit problem. It can be easily seen that problem (1) is equivalent to

$$
\text{(3)}\qquad
\begin{aligned}
\text{Min}\quad & \sum_{i=1}^{n} \left(e_i + f(x_i, \theta_0) - f(x_i, \theta)\right)^2 - \sum_{i=1}^{n} e_i^2 \\
\text{s.t.}\quad & g_i(\theta) \le 0, \qquad i = 1, \dots, p, \\
& h_j(\theta) = 0, \qquad j = p+1, \dots, q.
\end{aligned}
$$

Since we are interested in the asymptotic behavior of $n^{1/2}(\hat{\theta}_n - \theta_0)$, we use $z = n^{1/2}(\theta - \theta_0)$ as the optimization variable. This variable is often used in the statistical literature, for example, in Prakasa Rao (1987). Substituting z into problem (3), we get

$$
\text{(4)}\qquad
\begin{aligned}
\text{Min}\quad & \sum_{i=1}^{n} \left[\left(e_i + f(x_i, \theta_0) - f\left(x_i, \theta_0 + n^{-1/2}z\right)\right)^2 - e_i^2\right] \\
\text{s.t.}\quad & g_i\left(\theta_0 + n^{-1/2}z\right) \le 0, \qquad i = 1, \dots, p, \\
& h_j\left(\theta_0 + n^{-1/2}z\right) = 0, \qquad j = p+1, \dots, q.
\end{aligned}
$$

Denote by $F_n(e, z)$ and S_n the objective function and the feasible solution set of problem (4), respectively. Assume the optimal solution of (4) exists and denote it by $\hat{z}_n$. Then $\hat{z}_n = n^{1/2}(\hat{\theta}_n - \theta_0)$. Thus we need to find the asymptotic distribution of $\hat{z}_n$. Let us first find the limit form of problem (4).

For the limit form of the objective function $F_n(e, z)$, we have the following result.

THEOREM 1. *Suppose*:

(i) $e_1, \dots, e_n$ *are i.i.d. with* $Ee_i = 0$ *and* $\operatorname{Var} e_i = 1$;

(ii) $f(x_i, \theta)$, $i = 1, \dots, n$, *are differentiable in* θ *and there is a neighborhood* W *of* θ_0 *such that for* θ *in* W *it holds that*

$$
f(x_i, \theta) = f(x_i, \theta_0) + \left(\nabla_\theta f(x_i, \theta_0)\right)'(\theta - \theta_0) + r_i(\theta)\left(\|\theta - \theta_0\|^2\right),
$$

where $\nabla_\theta f(x_i|\theta_0)$ *is the gradient vector of* $f(x_i, \theta)$ *with respect to* θ *at* $\theta = \theta_0$. $\|\cdot\|$ *denotes the Euclidean norm in* R^m *and* $r_i(\theta)$ *satisfies*

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} r^2(\theta) < \infty,$$

uniformly on W;

(iii) $$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \left[\nabla_\theta f(x_i, \theta_0)(\nabla_\theta f(x_i, \theta_0))'\right] = K$$

exists and K *is a positive-definite matrix.*

Then for each fixed $z \in R^m$, $F_n(e, z)$ *converges in distribution to*

$$G(\xi, z) = z'Kz - 2z'\xi,$$

where ξ *is an* $N(0, K)$ *random vector.*

PROOF. For any θ in W we have

$$\begin{aligned}
&\sum_{i=1}^{n}\left[\left(e_i + f(x_i, \theta_0) - f(x_i, \theta)\right)^2 - e_i^2\right] \\
&\quad = \sum_{i=1}^{n}\left(f(x_i, \theta) - f(x_i, \theta_0)\right)^2 - 2\sum_{i=1}^{n}\left(f(x_i, \theta) - f(x_i, \theta_0)\right)e_i \\
&\quad = (\theta - \theta_0)' \sum_{i=1}^{n}\left[\nabla_\theta f(x_i, \theta_0)(\nabla_\theta f(x_i, \theta_0))'\right](\theta - \theta_0) + o\left(\|\theta - \theta_0\|^2\right) \\
&\qquad - 2(\theta - \theta_0)' \sum_{i=1}^{n} \nabla_\theta f(x_i, \theta_0)e_i - 2\left[\sum_{i=1}^{n} r_i(\theta)e_i\right]\|\theta - \theta_0\|^2 \\
&\quad = z'\left[n^{-1}\sum_{i=1}^{n} \nabla_\theta f(x_i, \theta_0)(\nabla_\theta f(x_i, \theta_0))'\right]z + o\left(n^{-1}\|z\|^2\right) \\
&\qquad - 2z'\left[n^{-1/2}\sum_{i=1}^{n} \nabla_\theta f(x_i, \theta_0)e_i\right] - 2\|z\|^2\left[n^{-1}\sum_{i=1}^{n} r_i(\theta)e_i\right].
\end{aligned}$$

By Theorems 4 and 5 in Jennrich (1969),

$$n^{-1/2}\sum_{i=1}^{n} \nabla f(x_i, \theta_0)e_i \to_D N(0, K),$$

$$n^{-1}\sum_{i=1}^{n} r_i(\theta)e_i \to 0 \quad \text{a.s.}$$

Hence for any fixed z we have

$$F_n(e, z) \to_D z'Kz - 2z'\xi. \qquad \square$$

Next, we study the limit of the feasible solution set S_n. Here we use the concept of convergence of sets in Kuratowski's sense, because this kind of convergence of sets will lead to convergence of optimal solutions of the related programming problems, as shown later. We write $S = (K)\lim S_n$, if

$$\limsup S_n \subset S \subset \liminf S_n, \tag{5}$$

where

$$\liminf S_n = \{z: \exists\, \{z_n\} \text{ such that } z_n \in S_n \text{ and } z_n \to z\},$$

$$\limsup S_n = \left\{z: \exists\, \{z_{n_j}\} \text{ such that } z_{n_j} \in S_{n_j} \text{ and } z_{n_j} \to z\right\},$$

or equivalently,

for any $z \in S$ there is a sequence $\{z_n\}$ such that $z_n \in S_n$ and $z_n \to z$ and for any sequence $\{z_n\}$ with $z_n \in S_n$ any accumulation point of $\{z_n\}$ must belong to S.

Let $I = \{i\colon g_i(\theta_0) = 0,\ i = 1,\dots,p\}$. We then have the following result.

THEOREM 2. *Suppose*:

(i) g_i, $i = 1,\dots,p$; h_j, $j = p+1,\dots,q$, *are continuously differentiable in* W;

(ii) *vectors* $\nabla g_i(\theta_0)$, $i \in I$; $\nabla h_j(\theta_0)$, $j = p+1,\dots,q$, *are linearly independent*.

Then $(K)\lim S_n = S$, *where S is defined by*

$$S = \left\{z: \nabla g_i(\theta_0)'z \le 0,\ i \in I;\ \nabla h_j(\theta_0)'z = 0,\ j = p+1,\dots,q\right\}.$$

PROOF. We show inclusion (5). Expand $g_i(\theta)$ and $h_j(\theta)$ as follows:

$$g_i\left(\theta_0 + n^{-1/2}z\right) = g_i(\theta_0) + n^{-1/2}\,\nabla g_i(\theta_0)'z + o(n^{-1/2}\|z\|),$$

$$h_j\left(\theta_0 + n^{-1/2}z\right) = h_j(\theta_0) + n^{-1/2}\,\nabla h_j(\theta_0)'z + o(n^{-1/2}\|z\|).$$

Suppose $\{z_n\}$ is a sequence such that $z_n \in S_n$. We are going to show that any accumulation point of $\{z_n\}$ must belong to S. Without loss of generality, we assume $\{z_n\}$ itself is a convergent sequence and $z_n \to z$. Then

$$g_i(\theta_0) + n^{-1/2}\,\nabla g_i(\theta_0)'z_n + o\left(n^{-1/2}\|z_n\|\right) \le 0, \qquad i \in I,$$

$$h_j(\theta_0) + n^{-1/2}\,\nabla h_j(\theta_0)'z_n + o\left(n^{-1/2}\|z_n\|\right) = 0, \qquad j = p+1,\dots,q.$$

Multiplying both sides of these expressions by $n^{1/2}$ and taking limits for $n \to \infty$, we obtain

$$\begin{aligned} &\nabla g_i(\theta_0)'z \le 0, && i \in I,\\ &\nabla h_j(\theta_0)'z = 0, && j = p+1,\dots,q. \end{aligned} \tag{6}$$

Thus $z \in S$. This implies $\limsup S_n \subset S$.

Now we show the second inclusion in (5). Let $\bar{z}$ be a point in S. First we assume it holds that

$$\begin{aligned} &\nabla g_i(\theta_0)'\bar{z} < 0, && i \in I,\\ &\nabla h_j(\theta_0)'\bar{z} = 0, && j = p+1,\dots,q. \end{aligned} \tag{7}$$

By the linear independence condition in assumption (ii) and the theory of mathematical programming [see, e.g., Section 4.3 Bazaraa and Shetty (1979)], for this $\bar{z}$ there exists, $\{\theta_n\}$ such that

$$\begin{aligned} & g_i(\theta_n) \le 0, \qquad i \in I, \\ & h_j(\theta_n) = 0, \qquad j = p+1, \dots, q, \\ & \theta_n \to \theta_0, \qquad n^{1/2}(\theta_n - \theta_0) \to \bar{z}. \end{aligned} \tag{8}$$

Let $z_n = n^{1/2}(\theta_n - \theta_0)$. Then

$$\begin{aligned} & g_i\big(\theta_0 + z_n/n^{1/2}\big) \le 0, \qquad i \in I, \\ & h_j\big(\theta_0 + z_n/n^{1/2}\big) = 0, \qquad j = p+1, \dots, q. \end{aligned} \tag{9}$$

For $i \in \{1, \dots, p\} \setminus I$, because $g_i(\theta_0) < 0$, we certainly have

$$g_i\big(\theta_0 + z_n/n^{1/2}\big) \le 0.$$

Thus, $z_n \in S_n$ and $z_n \to \bar{z}$. So for $\bar{z} \in S$ satisfying (7) there is a sequence $\{z_n\}$ such that $z_n \in S_n$ and $z_n \to \bar{z}$.

Now assume $\bar{z} \in S$ and $\nabla g_i(\theta_0)'\bar{z} = 0$ for at least one $i \in I$. Since vectors $\nabla g_i(\theta_0)$, $i \in I$; $\nabla h_j(\theta_0)$, $j = p+1, \dots, q$, are linearly independent, the set S is the intersection of the set $\{z: \nabla h_j(\theta_0)'z = 0,\ j = p+1, \dots, q\}$ with the convex polyhedral cone $\{z: \nabla g_i(\theta_0)'z \le 0,\ i \in I\}$. The latter set has a nonempty interior. Then for this $\bar{z}$ there must be a sequence z_k such that $z_k \to z$ and

$$\begin{aligned} & \nabla g_i(\theta_0)' z_k < 0, \qquad i \in I, \\ & \nabla h_j(\theta_0)' z_k = 0, \qquad j = p+1, \dots, q. \end{aligned}$$

For each z_k by the argument given above, we can find a sequence z_{nk} such that $z_{nk} \in S_n$ and $z_{nk} \to z_k$ as $n \to \infty$. By the method of diagonalization, one can get a subsequence $\{z_{nk}(n)\}$ of the double-indexed sequence $\{z_{nk}\}$ such that $z_{nk}(n) \to z$ and $z_{nk}(n) \in S_n$.

Thus in any case for a point $\bar{z} \in S$ we can find a sequence z_n such that $z_n \in S_n$ and $z_n \to \bar{z}$. This means $\liminf S_n \subset S$. Then the proof is complete. □

With Theorems 1 and 2 we can formulate a limit problem of problem (4):

$$\begin{aligned} \min \quad & z'Kz - 2z'\xi \\ \text{s.t.} \quad & \nabla g_i(\theta_0) \le 0, \qquad i \in I, \\ & \nabla h_j(\theta_0) = 0, \qquad j = p+1, \dots, q. \end{aligned} \tag{10}$$

We call (10) a formal limit problem, because it has not been shown that the optimal solutions of (10) are in distribution limits of the optimal solutions of problem (4). This is what we are going to do in the last section. Before proving this convergence result, we have to show that the optimal solutions of problem (4) are bounded in probability. This will be done in the next section.

3. Boundedness of $\hat{z}_n$. For the boundedness of $\hat{z}_n$, we have the following result.

THEOREM 3. *Under the assumptions made in Theorem* 1, $\hat{z}_n$ *is bounded in probability.*

PROOF. Since $\hat{z}_n$ is the optimal solution of (4) and $z_0 = n^{1/2}(\theta_0 - \theta_0) = 0$ is a feasible solution of (4), we have

$$\begin{aligned} 0 &\geq F_n(e, \hat{z}_n) - F_n(e, z_0) \\ &= \hat{z}_n'\left[n^{-1}\sum_{i=1}^{n}\nabla_\theta f(x_i, \theta_0)(\nabla_\theta f(x_i, \theta_0))'\right]\hat{z}_n + o\left(n^{-1}\|\hat{z}_n\|^2\right) \\ &\quad - 2\hat{z}_n'\left[n^{-1/2}\sum_{i=1}^{n}\nabla_\theta f(x_i, \theta_0)e_i\right] - 2\|\hat{z}_n\|^2\left[n^{-1}\sum_{i=1}^{n} r_i(\theta)e_i\right]. \end{aligned}$$

Since

$$n^{-1}\sum_{i=1}^{n}\nabla f(x_i, \theta_0)(\nabla f(x_i, \theta_0))' \to K,$$

$$n^{-1/2}\sum_{i=1}^{n}\nabla f(x_i, \theta_0)e_i \to_D N(0, K),$$

$$n^{-1}\sum_{i=1}^{n} r_i(\theta)e_i \to 0 \quad \text{a.s.},$$

then for any $\varepsilon > 0$ there is a constant C_ε such that the following holds:

$$0 \geq \hat{z}_n' K \hat{z}_n - 2\|z_n\|C_\varepsilon + o\left(n^{-1}\|\hat{z}_n\|^2\right),$$

with a probability larger than $1 - \varepsilon$, when n is large enough. By the positive definiteness of K, one can find a constant M_ε such that $\|\hat{z}_n\| \leq M_\varepsilon$ with a probability larger than $1 - \varepsilon$. This means that $\hat{z}_n$ is bounded in probability. □

4. Convergence of $\hat{z}_n$. In this section, we give the main result of this paper; that is, we show $\hat{z}_n$ converges in distribution to the optimal solution of problem (10). In the proof of this result, the weak convergence theory of probability measures will be used.

In Theorem 1 it is shown that for any fixed z the random variables $F_n(e, z)$ converge in distribution to $G(\xi, z)$. When z is varying over some connected set D, $\{F_n(e, z),\ z \in D\}$ and $\{G(\xi, z),\ z \in D\}$ can be viewed as stochastic processes. It is easy to see that all the sample functions of these stochastic processes are continuous functions on D. Denote by $C(D)$ the space of all continuous functions on D. We introduce a topology on $C(D)$ defined by the supremum norm. Then we can generate a Borel field $\mathscr{B}(D)$ on $C(D)$. Later on we will study convergence of these stochastic processes under this kind of topology. The following theorem shows the convergence in distribution of the sequence of these stochastic processes.

THEOREM 4. *Let D be the ball in R^m with center $z = 0$ and radius $d > 0$. Suppose the assumptions in Theorem 1 hold true. Then the stochastic processes $\{F_n(e, z), z \in D\}$ converge in distribution to $\{G(\xi, z), z \in D\}$.*

PROOF. According to the theory of stochastic processes with multidimensional parameter, $\{\zeta_n(t), t \in T\}$ converges in distribution to $\{\zeta(t), t \in T\}$ if and only if the following two conditions are satisfied [see Prakasa Rao (1975)]:

(a) any finite-dimensional distributions of $\{\zeta_n(t), t \in T\}$ converge weakly to the corresponding finite-dimensional distributions of $\{\zeta(t), t \in T\}$;

(b) for any $\varepsilon > 0$ it holds that

$$\limsup_{n\to\infty,\, h\to 0} P\left\{ \sup_{\|t_1 - t_2\| \le h} \|\zeta_n(t_1) - \zeta_n(t_2)\| > \varepsilon, t_1, t_2 \in T \right\} = 0.$$

Note that the conclusion of Theorem 1 is equivalent to weak convergence of the one-dimensional distribution of $\{F_n(e, z), z \in D\}$ to that of $\{G(\xi, z), z \in D\}$. To show the weak convergence of any finite-dimensional distributions, by the Cramér–Wold theorem it suffices to show the following: for any $c_1, \dots, c_r \in R$ and any values $z_1, \dots, z_r \in D$, we have

$$\sum_{j=1}^{r} c_j F_n(e, z_j) \to_D \sum_{j=1}^{r} c_j G(\xi, z_j).$$

It is equivalent to show [with $f_i(z) = -f(x_i, \theta_0) + f(x_i, \theta_0 + n^{-1/2} z)$]

$$\sum_{i=1}^{n} \sum_{j=1}^{r} \left[c_j f_i^2(z_j) - 2c_j f_i(z_j) e_i \right] \to_D \sum_{j=1}^{r} \left[c_j z_j' K z_j - 2c_j z_j' \xi \right].$$

In fact, this convergence result can be proved in the same way as Theorem 1. We will not repeat the procedure here. Thus condition (a) is satisfied for $\{F_n(e, z), z \in D\}$ and $\{G(\xi, z), z \in D\}$.

Next verify condition (b). For any z_1, z_2 in D, we have

$$|F_n(e, z_1) - F_n(e, z_2)| \le 2\left| \sum_{i=1}^{n} (f_i(z_1) - f_t(z_2)) e_i \right| + \left| \sum_{i=1}^{n} \left(\hat{f}_i^2(z_1) - \hat{f}_i^2(z_2) \right) \right|.$$

It is easy to see that

$$\sum_{i=1}^{n} (f_i(z_1) - f_i(z_2)) e_i \to_{\mathscr{D}} (z_1 - z_2)' \xi,$$

uniformly in $z_1, z_2 \in D$ because D is a compact ball. On the other hand, when n is large enough and $\|z_1 - z_2\|$ is small enough, for any given $\varepsilon > 0$, we have

$$\left| \sum_{i=1}^{n} \left(f_i^2(z_1) - f_i^2(z_2) \right) \right| \le |z_1' K z_1 - z_2' K z_2| + \tfrac{1}{4}\varepsilon < \tfrac{1}{2}\varepsilon.$$

Hence

$$\limsup_{n\to\infty,\, h\to 0} P\left\{ \sup_{\|z_1-z_2\|\le h} |F_n(e, z_1,) - F_n(e, z_2)| > \varepsilon,\, z_1, z_2 \in D \right\}$$

$$\le \limsup_{h\to 0} P\left\{ \sup_{\|z_1-z_2\|\le h} |(z_1 - z_2)'\xi| \ge \tfrac{1}{2}\varepsilon,\, z_1, z_2 \in D \right\} = 0.$$

The last equality holds because ξ has the $N(0, K)$ distribution. Then both conditions (a) and (b) are satisfied. Therefore the assertion of this theorem follows. □

Consider the following restricted optimization problems:

$$\begin{array}{ll} \text{Min} & F_n(e, z) \\ \text{s.t.} & z \in S_n,\, z \in B(M), \end{array} \tag{11}$$

$$\begin{array}{ll} \text{Min} & G(\xi, z) \\ \text{s.t.} & z \in S \cap B(M), \end{array} \tag{12}$$

where $B(M)$ is the ball $\{z\colon \|z\| < M\}$ and M is a large number. Denote the sets of optimal solutions of problems (11) and (12) by $A_n(M)$ and $A(M)$, respectively. Let $\hat{z}_n(M)$ and $\hat{z}(M)$ be their measurable selections. Then we have the following result.

THEOREM 5. *Suppose*:

(i) *the assumptions made in Theorems* 1 *and* 2 *hold true*;
(ii) *for each value of* ξ, $A(M)$ *is a singleton.*

Then any measurable selection $\hat{z}_n(M)$ *of* $A_n(M)$ *converges in distribution to* $\hat{z}(M)$.

PROOF. Observe that the sample functions of the stochastic processes $\{F_n(e, z),\ z \in B(M)\}$ and $\{G(\xi, z),\ z \in B(M)\}$ are continuous functions on $B(M)$. Let $C(M)$ be the space of all continuous functions over $B(M)$ and let $\mathscr{B}(M)$ be the Borel field on $C(M)$ as defined before. Then the stochastic processes $\{G(\xi, z),\ z \in B(M)\}$ and $\{F_n(e, z),\ z \in B(M)\}$ induce a family of probability measures $\{u, u_n,\ n = 1, 2, \dots\}$ on the measurable space $(C(M), \mathscr{B}(M))$. By Theorem 4, $\{F_n(e, z),\ z \in B(M)\}$ converges in distribution to $\{G(\xi, z),\ z \in B(M)\}$. This implies that $\{u_n\}$ converges weakly to u, written as $u_n \Rightarrow u$.

Define a collection of operators $H(\cdot)$ and $H_n(\cdot)$ on $(C(M), \mathscr{B}(M))$ as follows: for a function, $f(z) \in C(M)$, the image of f under $H(\cdot)$ is the optimal solution of the optimization problem

$$\begin{array}{ll} \text{Min} & f(z) \\ \text{s.t.} & z \in S \cap B(M) \end{array} \tag{13}$$

and the image of f under $H_n(\cdot)$ is the optimal solution of

$$\begin{array}{lll} (14) \qquad & \text{Min} & f(z) \\ & \text{s.t.} & z \in S_n \cap B(M), \end{array}$$

where S and S_n are the sets defined in Section 2. Without loss of generality we may assume that

$$(15) \qquad \begin{aligned} H(G(\xi, z), z \in B(M)) &= \hat{z}(M), \\ H_n(F_n(e, z), z \in B(M)) &= \hat{z}_n(M). \end{aligned}$$

We are going to show that

$$(16) \qquad \lim_{n \to \infty} H_n(f_n) = H(f)$$

for any f, f_n in $C(M)$ with $f_n \to f$ and f is such that problem (13) has a unique optimal solution. Note that the convergence of f_n to f in the space $(C(M), \mathscr{B}(M))$ means that

$$\max_{z \in B(M)} |f_n(z) - f(z)| \to 0$$

and this implies

$$(17) \qquad \lim_{n \to \infty} f_n(z_n) = f(z)$$

for any $z_n \to z$.

To show (16), first we show the following: if $\bar{z}_n$, $n = 1, 2, \dots,$ are optimal solutions of problem (14) and $\bar{z}$ is an accumulation point of $\{\bar{z}_n\}$, then $\bar{z}$ must be an optimal solution of problem (13). Suppose it is not true. Then there is a point z_0 in $S \cap B(M)$ such that $f(z_0) < f(\bar{z})$. Since f is continuous, we can find a neighborhood V of z_0 such that for all z in V it holds that $f(z) < f(\bar{z})$. Thus without loss of generality, we may assume that z_0 is an interior point of $B(M)$. On the other hand, by Theorem 2, there is a sequence z_n such $z_n \in S_n$ and $z_n \to z_0$. As z_0 is assumed to be an interior point of $B(M)$, so $z_n \in B(M)$ when n is large enough. Observing (17), we obtain

$$f(z_0) = \lim f_n(z_n) \geq \lim f_n(\bar{z}_n) = f(\bar{z}).$$

This contradicts the working assumption $f(z_0) < f(z)$. Hence $\bar{z}$ must be an optimal solution of problem (13).

Since $B(M)$ is compact and S, S_n are closed, the sequence $\bar{z}_n$ must have accumulation points. Moreover, the only possible accumulation point is $H(f)$, by the assumption on f. Therefore we get (16).

Combining (15), (16), $u_n \Rightarrow u$ and assumption (ii) of this theorem, we get by an extension of the continuous mapping theorem [cf. Theorem 5.5, Billingsley (1968)] that

$$\hat{z}_n(M) \to_{\mathscr{D}} \hat{z}(M).$$

This is the desired result. □

Now we come to the main result of this paper.

THEOREM 6. *Suppose the assumptions in Theorem* 5 *hold for any large* M. *Then the optimal solution*, $\hat{z}_n$ *of problem* (4) *converges in distribution to the optimal solution of problem* (10).

PROOF. As ξ has the $N(0, K)$ distribution, for any $\varepsilon > 0$ one can find C_ε such that $P(\|\xi\| \geq C_\varepsilon) \leq \varepsilon$. Observe that the optimal solution $\hat{z}$ of problem (10) satisfies

$$0 \geq \hat{z}'K\hat{z} - 2\hat{z}'\xi, \tag{18}$$

because $z = 0$ is a feasible solution of (10) and with $z = 0$ the value of the objective function is 0. From (18) we can see, since K is positive definite, when $\|\xi\| \leq C_\varepsilon$, there must be a constant M_ε such that $\|\hat{z}\| \leq M_\varepsilon$. Without loss of generality, we assume that this M_ε is the same as that M_ε in Theorem 3 (otherwise we may choose the larger one as the common M_ε).

Let $\hat{z}_n(M_\varepsilon)$ be the optimal solution of the following problem:

$$\begin{array}{ll} \text{Min} & z'Kz - 2z'\xi \\ \text{s.t.} & z \in S_n, \|z\| \leq M_\varepsilon, \end{array} \tag{19}$$

and $\hat{z}(M_\varepsilon)$ be the optimal solution of

$$\begin{array}{ll} \text{Min} & G(\xi, z) \\ \text{s.t.} & z \in S, \|z\| \leq M_\varepsilon. \end{array} \tag{20}$$

By Theorem 5 we have

$$\hat{z}_n(M_\varepsilon) \to_D \hat{z}(M_\varepsilon) \tag{21}$$

for any $\varepsilon > 0$ and corresponding C_ε, M_ε.

Note that when $\|\hat{z}\| \leq M_\varepsilon$, we have $\hat{z} = \hat{z}(M_\varepsilon)$. Thus

$$P(\hat{z} \neq \hat{z}(M_\varepsilon)) < \varepsilon.$$

Similarly (since $\hat{z}_n$ is bounded in probability),

$$P(\hat{z}_n \neq \hat{z}(M_\varepsilon)) < \varepsilon.$$

Therefore, for any $\varepsilon > 0$ and any open set G in R^m, we have

$$\begin{aligned} \liminf P(\hat{z}_n \in G) &> \liminf P(\hat{z}_n(M_\varepsilon) \in G) - \varepsilon \\ &\geq P(\hat{z}(M_\varepsilon) \in G) - \varepsilon \geq P(\hat{z} \in G) - 2\varepsilon, \end{aligned}$$

where the second inequality holds because of (21). Then, by the arbitrariness of ε, we obtain

$$\liminf P(\hat{z}_n \in G) > P(\hat{z} \in G).$$

This is equivalent to $\hat{z}_n \to_D \hat{z}$ and the proof is complete. □

Acknowledgment. The author wishes to thank an anonymous referee for his helpful comments and suggestions for improving the early version of this paper.

REFERENCES

BAZARAA, M. and SHETTY, C. (1979). *Nonlinear Programming*. Wiley, New York.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

DUPACOVA, J. and WETS, R. (1988). Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* **16** 1517–1549.

JENNRICH, R. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

LIEW, C. K. (1976). Inequality constrained least squares estimations. *J. Amer. Statist. Assoc.* **71** 746–751.

NAGARAJ, N. and FULLER, W. (1991). Estimation of the parameters of linear time series models subject to nonlinear regressions. *Ann. Statist.* **19** 1143–1154.

PRAKASA RAO, B. L. S. (1975). Tightness of probability measures generated by stochastic processes on metric spaces. *Bull. Inst. Math. Acad. Sinica* **3** 353–367.

PRAKASA RAO, B. L. S. (1987). *Asymptotic Theory of Statistical Inference*. Wiley, New York.

DEPARTMENT OF MATHEMATICS
NANJING UNIVERSITY
NANJING
CHINA