The Annals of Statistics
1989, Vol. 17, No. 4, 1472–1478

EDGEWORTH EXPANSIONS FOR BOOTSTRAPPING REGRESSION MODELS

By William Navidi

University of Southern California

The asymptotic performance of the bootstrap in linear regression models is studied. Edgeworth expansions show that asymptotically, the bootstrap is always at least as good as, and in some cases better than, the classical normal approximation. The performances of both the bootstrap and the normal approximation depend on the rate of increase in the elements of the design matrix.

1. Introduction. The use of the bootstrap to estimate the sampling distributions of parameter estimates in linear regression was first proposed by Efron (1979), and further developed by Freedman (1981). The process involves approximating the distribution of unobserved errors with the empirical distribution of the centered residuals.

We use Edgeworth expansions to examine the accuracy of bootstrap estimates of the distributions of linear combinations of regression parameter estimates. It turns out that asymptotically, the bootstrap is better than the normal approximation when the elements of the design matrix increase without bound. In other situations, the bootstrap is always at least as good as the normal approximation.

In Section 2 we describe the linear regression model in detail and make a one-term Edgeworth expansion for linear combinations of the coefficient estimates. In Section 3 we describe the bootstrap procedure, and derive an Edgeworth expansion for the bootstrap distribution. The accuracy of the bootstrap is assessed by comparing its Edgeworth expansion with that of the true distribution. Proofs of the theorems are given in Section 4.

2. The model. The model studied is $Y = X\beta + \varepsilon$, where X is an $n \times p$ matrix, β is a $p \times 1$ vector of unknown parameters and ε is a $n \times 1$ random vector whose components are independent and identically distributed with unknown distribution F. The number p of parameters is fixed. The distribution function F is nonlattice with mean 0, second moment σ^2 and a finite eighth moment. The matrix X is not random. Assume that the regression model is embedded in an infinite sequence of such models such that the number n of observations tends to ∞. We suppress the index for that sequence.

Consider the distribution of a linear combination of $\hat{\beta}$, i.e., a random variable of the form $\sqrt{n}\, c^{\mathrm{T}}(\hat{\beta} - \beta)$, where c is a $p \times 1$ vector and $\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y$ is the least-squares estimator of β. Without essential loss of generality, assume that the problem is scaled so that $nc^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c = 1$, and that the component of c

Received December 1987; revised November 1988.

AMS 1980 *subject classifications*. Primary 62E20; secondary 62J05.

Key words and phrases. Bootstrap, regression, asymptotic theory, Edgeworth expansion, Monte Carlo, empirical distribution, central limit theorem, resampling.

largest in absolute value is bounded away from both 0 and ∞. Let the coordinates of $\sqrt{n}\, c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ be $a_1, a_2, \ldots, a_n$. Now $\sqrt{n}\, c^{\mathrm{T}}(\hat{\beta} - \beta) = \sum_{i=1}^{n} a_i \varepsilon_i$. For $k = 1, 2, \ldots$, let $s_k = \sum_{j=1}^{n} a_i^k$. Then $s_2 = \sum_{j=1}^{n} a_j^2 = 1$ by the normalization.

The vector $\hat{\mathbf{y}}$ of fitted values is given by $\hat{\mathbf{y}} = HY$ with $H = X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$. The diagonal elements h_{ii} satisfy $0 \le h_{ii} \le 1$, and $\sum_{i=1}^{n} h_{ii} = \mathrm{tr}(H) = p$. Let $h_{\max} = \max_{1 \le i \le n} h_{ii}$. We assume

Assumption 1. $h_{\max} \to 0$.

This condition is necessary and sufficient for asymptotic normality of all linear combinations of the form $\sqrt{n}\, c^{\mathrm{T}}(\hat{\beta} - \beta)$ [see Huber (1973)].

We make an Edgeworth expansion of the distribution of $\sqrt{n}\, c^{\mathrm{T}}(\hat{\beta} - \beta)$. Recall that $\sqrt{n}\, c^{\mathrm{T}}(\hat{\beta} - \beta) = \sum_{i=1}^{n} a_i \varepsilon_i$. Define

$$m_n = \frac{1}{\max_{1 \le i \le n} a_i^2}.$$

The error rate in the expansion will turn out to depend on m_n.

By Huber (1981), Proposition 2.2, page 159,

$$\frac{1}{m_n} \le h_{\max}. \tag{2.1}$$

It follows from Assumption 1 that $m_n \to \infty$.

It is clear that

$$s_k = O\left(m_n^{-(k-2)/2}\right) \quad \text{and} \quad s_k = O\left(s_{k-1}/\sqrt{m_n}\right). \tag{2.2}$$

It follows from the scaling that the component of $\sqrt{n}\, c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}$ with largest absolute value is $O(1/\sqrt{n})$. Therefore the column vector $\mathbf{a}$ is a bounded linear combination of the columns of X, multiplied by $1/\sqrt{n}$. So, for example, in the sample mean case, $m_n = n$ and $s_k = n^{-(k-2)/2}$ for all k. If the elements of X are uniformly bounded, then $m_n = O(n)$ and $s_k = O(n^{-(k-2)/2})$ for all k. If the elements of X increase without bound, then the orders of m_n and s_k may depend on c, but they will generally be larger than in the bounded case.

To obtain an Edgeworth expansion for the distribution of $\sum_{i=1}^{n} a_i \varepsilon_i$, it is necessary to ensure that the behavior of the sum $\sum_{i=1}^{n} a_i \varepsilon_i$ will not be unduly influenced by its largest terms. To do this, we require that a sufficiently large number of the a_i be close in absolute value to the largest a_i. Specifically, for $n \ge 1$ and $\zeta > 0$, let n_ζ denote the number of $i \le n$ with $|a_i|\sqrt{m_n} > \zeta$. We assume

Assumption 2. $n_\zeta/(\log m_n) \to \infty$ for some $\zeta > 0$.

For example, if m_n is equal to a power of n, then $\log m_n = O(\log n)$. Then for some ζ, the number of a_i greater than $\zeta/\sqrt{m_n}$ in absolute value must grow faster than $\log n$.

THEOREM 2.1. *Suppose Assumptions* 1 *and* 2 *hold, and that* F *is a nonlattice distribution with finite third moment* μ_3. *Let* $\Phi(x)$ *and* $n(x)$ *denote the standard normal distribution and density functions, respectively. Let* F_n *denote the distribution function of* $\sum_{i=1}^n a_i\varepsilon_i$. *Let*

$$G(x) = \Phi(x/\sigma) + n(x/\sigma)\left[(s_3\mu_3)/(6\sigma^3)\right](1 - x^2/\sigma^2).$$

Then

$$\sup_x |F_n(x) - G(x)| = o\left(1/\sqrt{m_n}\right).$$

The theorem shows that the speed of convergence to normality of F_n depends on $\sqrt{m_n}$ and s_3. If the elements of X are uniformly bounded, with both μ_3 and $s_3 \neq 0$, then the speed of convergence is $O(1/\sqrt{n})$, which is the speed of convergence of the sample mean.

The value of the convergence rate $1/\sqrt{m_n}$ depends on the vector c. By (2.1), the largest possible value of $1/\sqrt{m_n}$ is $\sqrt{h_{\max}}$. The following proposition states that this rate is attained for some c.

PROPOSITION 2.1. *There exists a vector* c *such that* $1/m_n = h_{\max}$.

3. The bootstrap. Let F_n be as in Theorem 2.1, and let $\hat{\sigma}^2 = (n-p)^{-1}\sum_{i=1}^n e_i^2$ be the usual estimate of the error variance. Let $\mathbf{d}$ be the vector of residuals: $\mathbf{d} = Y - X\hat{\beta} = \varepsilon - X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\varepsilon$. Let $\mathbf{e} = (\mathbf{d} - \bar{\mathbf{d}})$, the vector of residuals centered to have mean 0. Define $\hat{F}$ to be the empirical distribution of the centered residuals, the distribution assigning mass $1/n$ to each of the points $e_1, e_2, \dots, e_n$.

Now assume we have on hand an observation $\mathbf{y}$, and that the estimate $\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\mathbf{y}$ of β and the residuals $e_1, \dots, e_n$ have been computed. Consider the model $Y^* = X\hat{\beta} + \varepsilon^*$, where $\varepsilon_1^*, \dots, \varepsilon_n^*$ are independent and identically distributed with distributed $\hat{F}$. Let $\hat{\beta}^*$ be the least-squares estimate of $\hat{\beta}$ in this model; that is, $\hat{\beta}^* = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Y^*$. The bootstrap approximation of the distribution F_n is the conditional distribution of $\sqrt{n}\,c^{\mathrm{T}}(\hat{\beta}^* - \hat{\beta})$ given $\mathbf{y}$. Denote this distribution by $\hat{F}_n$. Our purpose is to investigate the difference $|\hat{F}_n(x) - F_n(x)|$.

Denote the variance of $\hat{F}$ by $\hat{\sigma}^2$, and the kth moment by $\hat{\mu}_k$ for $k \geq 3$. The dependence of these moments on n is suppressed in the notation. We make an Edgeworth expansion of the distribution $\hat{F}_n(x)$. The following proposition ensures that the first four moments of $\hat{F}$ are well behaved.

PROPOSITION 3.1. *Suppose* F *has* $2m$ *moments. Then* $\hat{\mu}_k - \mu_k = O_p(1/\sqrt{n})$ *for* $k = 1, \dots, m$.

THEOREM 3.1. *Suppose Assumptions* 1 *and* 2 *hold, and that* F *is a nonlattice distribution with finite eighth moment. Let* $\Phi(x)$ *and* $n(x)$ *denote the standard normal distribution and density, respectively. Let* $\hat{F}_n$ *denote the distri-*

bution function of $\sum_{i=1}^{n} a_i \varepsilon_i^*$. *Let*

$$G(x) = \Phi(x/\hat{\sigma}) + n(x/\hat{\sigma})\big[(s_3\hat{\mu}_3)/(6\hat{\sigma}^3)\big](1 - x^2/\hat{\sigma}^2). \tag{3.1}$$

Then

$$\sup_x |\hat{F}_n(x) - G(x)| = o\big(1/\sqrt{m_n}\big). \tag{3.2}$$

We now examine the size of the difference $|F_n(x) - \hat{F}_n(x)|$. It follows from Theorems 2.1 and 3.1, and (2.2) that

$$\sup_x |F_n(x) - \hat{F}_n(x)| = O_p(1/\sqrt{n}) + o\big(1/\sqrt{m_n}\big), \tag{3.3}$$

while

$$\sup_x |F_n(x) - \Phi(x/\hat{\sigma})| = O_p(1/\sqrt{n}) + O(s_3) + o\big(1/\sqrt{m_n}\big). \tag{3.4}$$

Comparing (3.3) with (3.4) shows that the bootstrap is always asymptotically at least as accurate as the normal approximation, and better in situations where $m_n/n \to 0$ and $s_3 = O(1/\sqrt{m_n})$. In these situations the bootstrap error is $o(1/\sqrt{m_n})$ while the normal approximation error is $O(1/\sqrt{m_n})$.

A similar analysis of the standardized case shows that

$$\sup_x |F_n(\sigma x) - \hat{F}_n(\sigma x)| = o\big(1/\sqrt{m_n}\big), \tag{3.5}$$

while

$$\sup_x |F_n(\sigma x) - \Phi(x)| = O(s_3) + o\big(1/\sqrt{m_n}\big). \tag{3.6}$$

Comparing (3.5) with (3.6) shows again that the bootstrap is asymptotically at least as accurate as the normal approximation, and better whenever $s_3 = O(1/\sqrt{m_n})$.

Example 3.1. Suppose the elements of X are uniformly bounded. Then the coefficients a_i in $\sum_{i=1}^{n} a_i \varepsilon_i = \sqrt{n}\, c^{\mathrm{T}}(\hat{\beta} - \beta)$ are uniformly $O(1/\sqrt{n})$. It follows that s_3 is $O(1/\sqrt{n})$ and m_n is $O(n)$. The asymptotics of this situation are very much like those for the sample mean. The bootstrap and normal approximation each have an error $O(1/\sqrt{n})$ in the nonstandardized case, while in the standardized case, the bootstrap error is $o(1/\sqrt{n})$, and that of the normal approximation is $O(1/\sqrt{n})$.

Example 3.2. Let X be an $n \times p$ matrix whose first $n - \sqrt{n}$ rows are uniformly bounded, and whose last $\sqrt{n}$ rows each contain elements of order $n^{1/4}$. For most choices of contrast vector c, the first $n - \sqrt{n}$ of the a_i will be $O(1\sqrt{n})$, and the last $\sqrt{n}$ will be $O(n^{-1/4})$. Therefore $1/\sqrt{m_n}$ and s_3 are both $O(n^{-1/4})$. The bootstrap outperforms the normal approximation in both the standardized and nonstandardized cases; its error being $o(n^{-1/4})$ compared to $O(n^{-1/4})$ for the normal approximation.

EXAMPLE 3.3. This example shows that the order of s_3 can be less than its maximum possible value $1/\sqrt{m_n}$. Let X be a matrix whose first $n - n^{1/4}$ rows are uniformly bounded, and whose last $n^{1/4}$ rows each contain an element of order $n^{1/4}$. For most contrast vectors c, the first $n - n^{1/4}$ of the a_i will be $O(1/\sqrt{n})$, and the last $n^{1/4}$ will be $O(n^{-1/4})$. Thus the order of $1/\sqrt{m_n}$ will be $O(n^{-1/4})$, and the order of s_3 will be $O(1/\sqrt{n})$. The bootstrap and normal approximation are equally accurate in this situation, with errors $o(n^{-1/4})$ in both the standardized and nonstandardized cases.

EXAMPLE 3.4. Two independent samples of equal size are drawn from two distributions differing only in location. Consider estimating the difference in population means with the difference in sample means. This corresponds to the regression model in which the design matrix is a single column half of whose entries are 1 and half -1. It follows that $1/\sqrt{m_n} = 1/\sqrt{n}$, but $s_3 = 0$. Therefore the first Edgeworth correction term is 0 as well. This example illustrates the fact that symmetry in the design causes the same increase in efficiency as does symmetry in the underlying distribution.

4. Proofs.

PROOF OF THEOREM 2.1. The Fourier transform of $G'(x)$ is $\gamma(t) = e^{-\sigma^2 t^2/2}[1 + [(s_3\mu_3)/6](it)^3]$. Let $\phi(t)$ be the characteristic function of F. Let $\varepsilon > 0$. Berry's smoothing lemma [see Feller (1971), page 538] allows us to bound the difference $\sup_x |F_n(x) - G(x)|$ by an integral involving the Fourier transforms of F_n and G. Specifically, let A be a constant large enough so that $\varepsilon A \geq 24 \sup_x |G'(x)|$. Then

$$\sup_x |F_n(x) - G(x)| \leq \int_{-A\sqrt{m_n}}^{A\sqrt{m_n}} \frac{1}{|t|}\left|\prod_{j=1}^{n} \phi(a_j t) - \gamma(t)\right| dt + \frac{\varepsilon}{\sqrt{m_n}}.$$

It must be shown that the value of this integral is $o(1/\sqrt{m_n})$. To do this, we break the region of integration into two parts. First, for any $\delta > 0$, the contribution of the interval $\delta\sqrt{m_n} \leq |t| \leq A\sqrt{m_n}$ to the integral is less than

$$\int_{\delta\sqrt{m_n} \leq |t| \leq A\sqrt{m_n}} \frac{1}{|t|} \prod_{j=1}^{n} |\phi(a_j t)|\, dt + \int_{\delta\sqrt{m_n} \leq |t| \leq A\sqrt{m_n}} \frac{1}{|t|} |\gamma(t)|\, dt.$$

Clearly the right-hand term approaches 0 faster than any power of $1/\sqrt{m_n}$. Since F is nonlattice, $|\phi(t)|$ is bounded away from 1 on compact sets not containing the origin. Let ζ be as in Assumption 2 and let $v < 1$ be such that $|\phi(t)| < v$ for $\zeta\delta \leq |t| \leq A$. For large n, the left-hand integral is less than $v^{n_\zeta} \log(A\sqrt{m_n})$, which, by Assumption 2, approaches 0 faster than any power of $1/\sqrt{m_n}$.

Let $\psi(t) = \log \phi(t) + \sigma^2 t^2/2$. The moment conditions on F ensure that the third derivative of ψ is continuous at 0. Choose $\delta > 0$ so that, for $|t| < \delta$:

(i) $|\psi(t)| < \sigma^2 t^2/4$.
(ii) $|\psi'''(t) - \psi'''(0)| < \varepsilon$.
(iii) $|(s_3\mu_3/6)(it)^3| < \sigma^2 t^2/4$.

We are now ready to bound the integral over the region $|t| < \delta\sqrt{m_n}$. This integral is equal to

$$\int_{|t|<\delta\sqrt{m_n}} \frac{1}{|t|} e^{-\sigma^2 t^2/2} \left| \exp\left[\sum_{j=1}^{n} \psi(a_j t)\right] - 1 - \frac{s_3\mu_3}{6}(it)^3 \right| dt.$$

The inequality $|e^{\alpha} - 1 - \beta| \leq e^{\Lambda}(|\alpha - \beta| + \beta^2/2)$, where $\Lambda = \max(|\alpha|, |\beta|)$, is valid for all complex α and β. Using this inequality, the integrand above is seen to be less than or equal to

$$\frac{1}{|t|} e^{-\sigma^2 t^2/2} e^{\Lambda} \left(\left| \sum_{j=1}^{n} \psi(a_j t) - \frac{s_3\mu_3}{6}(it)^3 \right| - \frac{(s_3\mu_3)^2}{72}(it)^6 \right),$$

where

$$\Lambda = \max\left(\left| \sum_{j=1}^{n} \psi(a_j t) \right|, \left| \frac{s_3\mu_3}{6}(it)^3 \right| \right) \leq \tfrac{1}{4}\sigma^2 t^2.$$

Now $\psi'(0) = \psi''(0) = 0$ and $\psi'''(0) = i^3\mu_3$. It follows that for $|t| \leq \delta\sqrt{m_n}$, $|\sum_{j=1}^{n} \psi(a_j t) - (s_3\mu_3/6)(it)^3| < \varepsilon t^3 |s_3|$. Thus the integrand is less than

$$\frac{1}{|t|} e^{-\sigma^2 t^2/4} \left[\varepsilon t^3 s_3 + \frac{(s_3\mu_3)^2}{72}(it)^6 \right].$$

Since ε is arbitrary, and $s_3 \leq O(1/\sqrt{m_n})$, the integral $o(1/\sqrt{m_n})$. □

Proof of Proposition 2.1. Assume, without essential loss of generality, that $X^{\mathrm{T}}X = nI$, where I is the $p \times p$ identity matrix. Then $H = (1/n)XX^{\mathrm{T}}$. Let X_L be the row of X with the largest Euclidean norm. Then $h_{\max} = h_{LL} = (1/n)X_L X_L^{\mathrm{T}} = (1/n)|X_L|^2$. Let $c = X_L^{\mathrm{T}}/|X_L|$. Then $\mathbf{a} = \sqrt{n}X(X^{\mathrm{T}}X)^{-1}c = (1/\sqrt{n})Xc = (1/\sqrt{n})(XX_L^{\mathrm{T}})/|X_L|$, so $a_L = (1/\sqrt{n})(X_L X_L^{\mathrm{T}})/|X_L| = (1/\sqrt{n})|X_L|$. Thus $a_L^2 = (1/n)|X_L|^2 = h_{\max}$. □

Proof of Theorem 3.1. Let $\hat{\phi}(t)$ be the characteristic function of $\hat{F}$. The proof of Theorem 2.1 will go through, replacing ϕ with $\hat{\phi}$, ψ with $\hat{\psi}$, σ with $\hat{\sigma}$ and μ_3 with $\hat{\mu}_3$, if we show that δ can be chosen independently of n and that, for any constant A, the product $\prod_{j=1}^{n} |\hat{\phi}(a_j t)|$ approaches 0 faster than any power of $1/m_n$ for t satisfying $\delta \leq |t| \leq A$. By Freedman (1981), Lemma 2, $\hat{F} \Rightarrow F$ weakly, so $\hat{\phi}(t) \to \phi(t)$ uniformly on compact sets. Therefore $\prod_{j=1}^{n} |\hat{\phi}(a_j t)| \to 0$ faster than any power of $1/m_n$.

We now show that given $\varepsilon > 0$ there exists $\delta > 0$ such that conditions (i), (ii), and (iii) in the proof of Theorem 2.1 are satisfied for all n. This is clearly true for conditions (i) and (iii) by a.s. convergence of the sample moments. It is true for condition (ii) since the a.s. convergence of $\hat{\mu}_4$ shows that for small $|t|$,

$$|\hat{\phi}'''(t) - \hat{\phi}'''(0)| = \left|\int x^3(e^{itx} - 1)\hat{F}(dx)\right| \le |t|\left|\int x^4\hat{F}(dx)\right| < \varepsilon. \qquad \square$$

PROOF OF PROPOSITION 3.1. The difference $\hat{\mu}_k - \mu_k$ can be written as

$$\frac{1}{n}\sum_{j=1}^{n}\left(e_j^k - \mu_k\right) = \frac{1}{n}\sum_{j=1}^{n}\left(\varepsilon_j^k - \mu_k\right) + \frac{1}{n}\sum_{j=1}^{n}\left(e_j^k - \varepsilon_j^k\right).$$

It suffices to show that $(1/n)\Sigma_{j=1}^n(e_j^k - \varepsilon_j^k)$ is $O_p(1/\sqrt{n})$. Let $\mathbf{d} = \boldsymbol{\varepsilon} - \mathbf{e}$. Then

$$\frac{1}{n}\sum_{j=1}^{n}\left(e_j^k - \varepsilon_j^k\right) = \frac{1}{n}\sum_{j=1}^{n}\left[\left(\varepsilon_j - d_j\right)^k - \varepsilon_j^k\right] = \sum_{r=1}^{k}\left[\frac{c_r}{n}\sum_{j=1}^{n} d_j^r\varepsilon_j^{k-r}\right],$$

where the c_r are constants depending on r and k, but not on n.

For $r = 1, 2, \dots, k$,

$$\frac{1}{n}\sum_{j=1}^{n} d_j^r\varepsilon_j^{k-r} = \sum_{j=1}^{n}\frac{d_j^r}{n^{1/4}}\frac{\varepsilon_j^{k-r}}{n^{3/4}} \le \sum_{j=1}^{n}\frac{d_j^{2r}}{\sqrt{n}} + \sum_{j=1}^{n}\frac{\varepsilon_j^{2k-2r}}{n^{3/2}}.$$

By the strong law of large numbers, $\Sigma_{j=1}^n(\varepsilon_j^{2k-2r}/n^{3/2}) = O(1/\sqrt{n})$. It is clear that $E(\Sigma_{j=1}^n d_j^2) = p\sigma^2$. Since $\Sigma_{j=1}^n d_j^{2r} \le (\Sigma_{j=1}^n d_j^2)^r$, $\Sigma_{j=1}^n(d_j^{2r}/\sqrt{n}) = O_p(1/\sqrt{n})$. Therefore $(1/n)\Sigma_{j=1}^n d_j^r\varepsilon_j^{k-r} = O_p(1/\sqrt{n})$. This completes the proof. □

Acknowledgments. I would like to express my gratitude to David Freedman, for his guidance in bringing this paper to its present form. I would also like to thank Peter Bickel and Thomas Rothenberg for many helpful discussions.

REFERENCES

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

FREEDMAN, D. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

HUBER, P. (1973). Robust regression. *Ann. Statist.* **1** 799–821.

HUBER, P. (1981). *Robust Statistics*. Wiley, New York.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1113