MULTIVARIATE REGRESSION WITH ONE
STOCHASTIC PREDICTOR VARIABLE

BY

STANLEY L. SCLOVE

TECHNICAL REPORT NO. 124
DECEMBER 30, 1966

THIS RESEARCH WAS SPONSORED BY THE ARMY RESEARCH OFFICE,
OFFICE OF NAVAL RESEARCH, AND AIR FORCE OFFICE OF
SCIENTIFIC RESEARCH BY CONTRACT NO.
Nonr-225(52) (NR 342-022)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
MULTIVARIATE REGRESSION WITH ONE
STOCHASTIC PREDICTOR VARIABLE

by

Stanley L. Sclove

TECHNICAL REPORT NO. 124

December 30, 1966

PREPARED UNDER CONTRACT Nonr-225(52)
(NR-342-022) FOR
OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Multivariate Regression With One
Stochastic Predictor Variable*

by

Stanley L. Sclove

1. Introduction and summary.

In [2], we considered the following problem: given a random sample
of \( N \) observations from the joint normal distribution of \( q \) predic-
cands and \( p \) predictors and an additional independent observation on the
predictors, to predict the corresponding value of the predictands, when
the loss function is the conditional mean square of the distance between
the predicted and the actual values in the metric of the residual covari-
ance matrix, given the sample of \( N \) observations. Relations between
the prediction problem and the problem of estimating the regression
coefficients were discussed. The present paper treats the problem of
estimating the regression coefficients in the special case of one pre-
dictor variable (\( p=1 \)). In Section 2 the problem is stated formally;
in Section 3 we obtain a class of estimators, each of which is better
than the maximum likelihood estimator (MLE) for \( q \geq 3 \); and in Section
4 we prove that the MLE is admissible for estimating the regression
coefficients when \( q = 2 \) and the means of the variables are known.

* This work was supported in part under Contract Nonr-4259(08) (NR-042-034)
at Columbia University and Contract Nonr-225(52) (NR-342-022) at Stanford
University.
2. The problem.

We consider a sample of \( N \) independent observations \( \begin{pmatrix} Y_k \\ X_k \end{pmatrix}, k = 1, \ldots, N, \)
where \( Y_k \) is a \( q \)-vector, \( X_k \) is a scalar, and \( \begin{pmatrix} Y_k \\ X_k \end{pmatrix} \) is normally distributed with mean

\[
(2.1) \quad \mu = \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}
\]

and covariance matrix

\[
(2.2) \quad \Sigma^* = \begin{pmatrix} \Sigma_Y & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_X \end{pmatrix}.
\]

The regression function is

\[
\rho(x_k) = E[Y_k \mid X_k = x_k] = \alpha + \beta' X_k,
\]
where \( \beta' \) is a column vector of \( q \) components (This notation is used to provide compatibility with [2].) given by

\[
(2.3) \quad \beta = \frac{\Sigma_{XY}}{\Sigma_X},
\]

and the \( q \)-vector

\[
(2.4) \quad \alpha = \mu_Y - \beta' \mu_X.
\]

The \( q \times q \) conditional covariance matrix (residual covariance matrix) of \( Y_k \) given \( X_k \) is

\[
(2.5) \quad \Sigma = \Sigma_Y - \frac{\Sigma_{YX} \Sigma_{XY}'}{\Sigma_X}.
\]

(We assume \( \Sigma \) is nonsingular; this holds if \( \Sigma^* \) is nonsingular.)

Thus the conditional distribution of \( Y_k \), given \( X_k = x_k \), is

\[
\mathcal{N}(\alpha + \beta' x_k', \Sigma).
\]
We consider the problem of estimating $\beta$ when the loss function is

$$(2.6) \quad \text{tr} \Sigma_X(\hat{\beta}-\beta)\Sigma^{-1}(\hat{\beta}-\beta)'$$

and the problem of estimating $(\alpha, \beta)$ when the loss function is

$$(2.7) \quad [(\hat{\alpha}-\alpha)+(\hat{\beta}-\beta)\mu_\beta] \Sigma^{-1}[(\hat{\alpha}-\alpha)+(\hat{\beta}-\beta)\mu_\beta] + \text{tr} \Sigma_X(\hat{\beta}-\beta)\Sigma^{-1}(\hat{\beta}-\beta)' ,$$

which was shown to be a natural loss function in Section 2 of [2].

Define the statistics

$$(2.8) \quad \overline{Y} = \sum_{k=1}^{N} Y_k/N$$

and

$$(2.9) \quad \overline{X} = \sum_{k=1}^{N} X_k/N ,$$

$$(2.10) \quad T = \sum_{k=1}^{N} (Y_k-\overline{Y})(Y_k-\overline{Y})' ,$$

$$(2.11) \quad U = \sum_{k=1}^{N} (X_k-\overline{X})(X_k-\overline{X})' ,$$

and

$$(2.12) \quad V = \sum_{k=1}^{N} (X_k-\overline{X})^2 .$$

The MLE's are: for $\beta$,

$$(2.13) \quad b = U/V ;$$

for $\alpha$,

$$(2.14) \quad a = \overline{Y}-b'\overline{X} ;$$
for $\Sigma^*$,

$$\Sigma^* = \frac{1}{N} \begin{pmatrix} T & U' \\ U & V \end{pmatrix} .$$

and for $\Sigma, S/N$, where

$$S = T - U'U/V = T - b'b/V .$$

If $\mu_Y$ and $\mu_X$ are known, then $T, U$ and $V$ are redefined by replacing $\bar{Y}$ by $\mu_Y$ and $\bar{X}$ by $\mu_X$.

3. Estimators better than the MLE.

Theorem 1. For estimating $\beta$ when the loss function is (2.6) and $q \geq 3$, each of the estimators

$$\tilde{b}' = [1-a(n)/bS^{-1}b']b'$$

dominates the MLE, $b$, where $a(n)$ is any constant between 0 and $2(q-2)/(n-q+3)$, $n$ being the number of degrees of freedom associated with $S$.

Remark. $n = N-1$ if $\mu$ is known and $N-2$ otherwise.

Proof. Consider first the problem of estimating $\theta$ when the data are the $q$-vector $Z$, which is $\mathcal{L}(\theta, \Sigma)$, and $S$, which is distributed according to the Wishart distribution $\mathcal{W}(\Sigma, n)$ and independent of $Z$, and the loss function is $(\hat{\theta} - \theta')S^{-1}(\hat{\theta} - \theta)$. James and Stein [1] showed that the estimators

$$\varphi(Z, S) = [1-a(n)/Z'S^{-1}Z]Z ,$$

where $a(n)$ is any constant between 0 and $2(q-2)/(n-q+3)$, dominate $Z$. That is, they proved that, for all $(\theta, \Sigma)$,
(3.3) \[ E_{\theta, \Sigma} [\varphi(Z, S) - \theta]' \Sigma^{-1} [\varphi(Z, S) - \theta] \leq E_{\theta, \Sigma} (Z - \theta)' \Sigma^{-1} (Z - \theta). \]

Note that, given \( V, b' \) is \( \mathcal{W}(b', \Sigma/V) \) and \( S = T - bVb' \) is \( \mathcal{W}(\Sigma, n) \) and independent of \( b' \). Thus, under the change of variables

\[ \theta = \sqrt{b'}, \; Z = \sqrt{b'}, \]

(3.3) becomes

\[ E_{\Sigma^*}[\varphi(\sqrt{b'}, S)|\theta = \sqrt{b'}, S = \sqrt{b'}]|\Sigma] \leq E_{\Sigma^*}[\varphi(\sqrt{b'}, S)|\theta = \sqrt{b'}, S = \sqrt{b'}]|\Sigma], \]

for all \( \beta, \Sigma^* \). Multiplying both sides by \( \Sigma^{-1} \) and taking the expectation with respect to \( V \) gives

\[ E_{\Sigma^*}[\Sigma^{-1}(\sqrt{b'}, S)|\theta = \sqrt{b'}, S = \sqrt{b'}]|\Sigma] \leq E_{\Sigma^*}[\Sigma^{-1}(\sqrt{b'}, S)|\theta = \sqrt{b'}, S = \sqrt{b'}]|\Sigma], \]

for all \( \Sigma^* \). But \( \sqrt{b'} \varphi(\sqrt{b'}, S) \) is just \( \tilde{b}' \), and the theorem is proved.

The risk of \( \varphi \) is shown in [1] to be

\[ (3.4) \quad q - 2[a(n)]n(q-2)\tau + [a(n)]^2 n(n+2)\tau, \]

where \( \tau = E[(q-2+2k)^{-1}] \), \( k \) being a Poisson r.v. with parameter \( \theta' \Sigma^{-1} \theta/2 \).

Since \( \theta' \Sigma^{-1} \theta/2 \geq \Omega \theta' \Sigma^{-1} \theta/2 \), denote (3.4) by \( f(a, V) \). The function \( f(a, V) \) is a quadratic in \( a(n) \) and is minimized for all parameter values by setting \( a = a_o \), where \( a_o = (q-2)/(n-q+3) \). Thus \( f(a_o, V) \leq f(a, V) \), for all \( V \). From the proof of Theorem 1 we see that the risk of \( \tilde{b}' \)
is $E[(\Sigma_X/V)f(a,V)]$; consequently the choice $a = a_0$ minimizes the risk of $\tilde{b}$ for all parameter values.

Next we obtain upper and lower bounds for the risk of $\tilde{b}$. Using the fact that a quadratic $ax^2 + bx + c$ with $a > 0$ has a minimum of $c-b^2/4a$, we obtain $f(a_0,V) = q-[n(q-2)^2/(n+2)]\tau$. Thus the risk of $\tilde{b}$ when $a = a_0$ is

$$R(\tilde{b}; \Sigma^*) = E_{\Sigma^*}[\Sigma_X f(a_0,V)/V] = E_{\Sigma^*}[\Sigma_X (q-[n(q-2)^2/(n+2)]\tau)/V]$$

$$= q\Sigma_X E_{\Sigma^*}(V^{-1})\Sigma_X [n(q-2)^2/(n+2)]E_{\Sigma^*}[\tau/V].$$

Let $m$ be the number of degrees of freedom of $V$, which is $\Sigma_X$ times a chi-square r.v. ($m = N$ if $\mu$ is known and $N-1$ otherwise.) According to [3], we have $E(V^{-1}) = \Sigma_X/(m-2)$; hence

$$R(\tilde{b}; \Sigma^*) = \frac{q}{(m-2)\Sigma_X [n(q-2)^2/(n+2)]E_{\Sigma^*}[\tau/V]}.$$

The minimum risk is attained when $\beta = 0$ and is

$$q/(m-2)\Sigma_X [n(q-2)^2/(n+2)]E_{\Sigma^*}[1/(q-2)\Sigma^*] = [q-n(q-2)/(n+2)]/(m-2).$$

An upper bound for the risk is obtained as follows. Since $(q-2+2K)^{-1}$ is a convex function of $K$, we have, applying Jensen's inequality

$$\tau = E[(q-2+2K)^{-1}] \geq [(q-2+2E(K))^{-1} = [(q-2) + 2E[\beta^{-1}\beta']]^{-1}.$$ 

Thus

$$f(a_0,V) \leq q-[n(q-2)^2/(n+2)][(q-2)+2E[\beta^{-1}\beta']]^{-1} = f_1(V),$$

say, and this gives
\[ R(\hat{b}; \Sigma^*) \leq E_{\Sigma^*}[\Sigma^* f_1(V)/V] \]
\[ = q/(m-2) - \Sigma^*[n(q-2)^2/(n+2)]E_{\Sigma^*}[V(q-2)^2 + V^2 \beta \Sigma^{-1} \beta']^{-1} . \]

Since \((V(q-2)^2 + V^2 \beta \Sigma^{-1} \beta')^{-1}\) is a convex function of \(V\), we have, again applying Jensen's inequality,

\[(3.6) \quad R(\tilde{b}; \Sigma^*) \leq q/(m-2) - \Sigma^*[n(q-2)^2/(n+2)][(q-2)E(V) + \beta \Sigma^{-1} \beta' E(V^2)]^{-1} \]
\[ = q/(m-2) - [n(q-2)^2/(n+2)][(q-2)m + \Sigma^* \beta \Sigma^{-1} \beta' m(m+2)]^{-1} , \]

where we have used the fact that \(E(V^2) = \Sigma^2 X m(m+2)\). Note that the right hand side of (3.6) increases to \(q/(m-2)\) as \(\Sigma^* \beta \Sigma^{-1} \beta' \to \infty\). The quantity \(q/(m-2)\) is the risk of the MLE, for

\[ R(b; \Sigma^*) = E_{\Sigma^*}(b-\beta)\Sigma^{-1}(b-\beta)' \]
\[ = E_{\Sigma^*}((\Sigma^*/V)E_{\Sigma^*}[V(b-\beta)\Sigma^{-1}(b-\beta)' | V]) \]
\[ = E_{\Sigma^*}((\Sigma^*/V)E_{\theta, \Sigma}[(Z-\theta)\Sigma^{-1}(Z-\theta)] \]
\[ = q\Sigma^* E(V^{-1}) \]
\[ = q/(m-2) . \]

To obtain an estimator that is better than the MLE for estimating \((\alpha, \beta')\) we apply the theorem of Section 3 of [2]. The estimator \(\tilde{b}'\) is invariant under the transformations \((Y', X_k) \to (Y'K', MX_k)\), \(K\) non-singular, \(M \neq 0\), which operate transitively on the spaces of \(\Sigma^* X\) and \(\Sigma\). This proves
Theorem 2. The estimator \((\bar{Y} - \tilde{b}'\bar{X}, \tilde{b})\) dominates the MLE, \((a, b)\), for estimating \((a, \beta)\) when the loss function is (2.7).

One may obtain bounds for the risk of this estimator by applying formula (3.14) of [2] and expressions (3.5) and (3.6).

4. A case in which the MLE is admissible.

In [3] it was shown that for the univariate regression problem with the means of the predictands and predictors known, the MLE for \(\beta\) is admissible when \(p = 1\) or \(2\), and inadmissible when \(p \geq 3\). In a similar way we obtain an admissibility result for the present problem when \(q = 2\).

Theorem 3. When \(q = 2\), \(N \geq 5\), and the mean \((E(Y_k), E(X_k)) = (\mu_Y, \mu_X)\) is known, the MLE, \(b\), is admissible for estimating \(\beta\) with the loss function (2.6).

The proof is accomplished by applying a general result due to Stein (see lemma 2 of [3]), which we restate here:

**Lemma.** Let \(\mathcal{B}\) be the \(\sigma\)-algebra of all Borel subsets of the two-dimensional real coordinate space \(\mathbb{R}^2\) and \(\mathcal{G}\) a \(\sigma\)-algebra of subsets of a set \(\mathcal{Y}\). Let \(\mu\) be Lebesgue measure on \(\mathcal{B}\) and \(\nu\) a probability measure on \(\mathcal{G}\). Let \(f\) be a non-negative-valued \(\mathcal{B}\mathcal{G}\) measurable function on \(\mathbb{R}^2 \times \mathcal{Y}\) such that

\[
(4.1) \quad \int f(x,y)dx = 1, \text{ for all } y,
\]

\[
(4.2) \quad \int xf(x,y)dx = 0, \quad i = 1,2, \text{ for all } y,
\]
and

\[(4.3) \quad \int d\nu(y) \left[ \int (x'x)^{1+\epsilon} f(x,y) dx \right]^2 < \infty, \]

for some \( \epsilon > 0 \), where \( dx = d\mu(x) \). Then, if we observe \((X,Y)\) distributed so that, for some unknown \( \xi \), \((X-\xi,Y)\) has probability density \( f(x,y) \) with respect to \( \mu \nu \), the estimator \( \hat{\xi} \) is an admissible estimator of \( \xi \) with loss function

\[(4.4) \quad (\hat{\xi} - \xi)' A (\hat{\xi} - \xi), \]

for any constant positive semidefinite matrix \( A \).

**Remark.** \((4.2)\) states

\[(4.5) \quad \mathbb{E}[X-\xi | Y] = 0 \]

and \((4.3)\) states

\[(4.6) \quad \mathbb{E}E^2[((X-\xi)'(X-\xi))^{1+\epsilon} | Y] < \infty. \]

**Proof of Theorem 3.** Since \( \mu_Y \) and \( \mu_X \) are known, the sufficient statistic is the triplet \((b',V,S)\). To apply the lemma, take \( X = b' \), \( \xi = \beta' \), and \( Y = (V,S) \); take \( A = \Sigma_X^{-1} \Sigma^{-1} \) so that \((4.4)\) becomes \((2.6)\).

Given \( V, b' \) is \( \mathcal{N}(\beta', V^{-1} \Sigma) \); the statistic \( V \) is \( \Sigma_X^2 \); \( S \) is \( \mathcal{W}(\Sigma, N-1) \) and independent of \( b' \) and \( V \). The density \( f(x,y) \) is the conditional density of \( X | \xi = b', \beta' \) given \((V,S) = y \). Since \( \Sigma \) is independent of \( b' \) and \( V \), this is the conditional density of \( b' - \beta' \) given \( V \), which is \( \mathcal{N}(0, V^{-1} \Sigma) \). First suppose that \( \Sigma \) and \( \Sigma_X \) are known. Then \( f(x,y) \) and \( V \) are specified, the conditions \((4.1)\) and \((4.2)\) are met.
It remains to show that condition (4.6) is satisfied. Define \( U^* = U - V \beta \). Then we have \((X - \xi)' = b - \beta = V^{-1}U - \beta = V^{-1}U^*\), and

\[(X - \xi)'(X - \xi) = U^*U^*/V^2.\]

The finiteness of the left-hand side of (4.6) does not depend upon the values of \( \Sigma \) and \( \Sigma_X \); we assume \( \Sigma = I \) and \( \Sigma_X = 1 \). Then the conditional distribution of \( U^* \) given \( V \) is \( \mathcal{N}(0, VI) \) and that of \( U^*/V^{1/2} \) is \( \mathcal{N}(0, I) \), so that the conditional distribution of \( W = U^*U^*/V \) is \( \chi_2^2 \). Since this conditional distribution does not depend upon \( V \), it is also the unconditional distribution. The statistic \( U^*U^*/V^2 = \bar{w}/V \). We have, for any fixed \( \epsilon > 0 \),

\[
E^2[(\bar{w}/V)^{1+\epsilon}/V] = E^2(\bar{w}^{1+\epsilon}/V)/V^{2+2\epsilon} \\
= E^2(\bar{w}^{1+\epsilon})/V^{2+2\epsilon} \\
= E^2[(\chi_2^2)^{1+\epsilon}]/V^{2+2\epsilon} \\
\leq MV^{-2-2\epsilon}
\]

where \( 0 < E^2[(\chi_2^2)^{1+\epsilon}] < M \). Thus

\[
E(E^2[(U^*U^*/V^2)^{1+\epsilon}/V]) \leq ME(V^{-2-2\epsilon}) \\
= ME[(\chi_2^2)^{-2-2\epsilon}] \\
= M \int_0^\infty x^{N/2-1-2-2\epsilon} e^{-x/2} \, dx,
\]

which is finite for \( N/2-1-2-2\epsilon > -1 \), that is, for \( N \geq 5 \), if \( \epsilon < \frac{1}{4} \).

This proves that the MLE is admissible if \( \Sigma_X \) and \( \Sigma \) are known; consequently, it is also admissible when they are unknown. (An estimator admissible over each set of a partition of the parameter space is admissible.)
Strictly speaking, we have demonstrated only the admissibility of the MLE in the class of estimators which depend on the sufficient statistic, for we have taken \((b', (\nu, S))\) as the basic observation. However, by the complete class theorem of Section 4 of [2], these estimators constitute a complete class, so the MLE is admissible in the class of all estimators.

REFERENCES


1. ORIGINATING ACTIVITY (Corporate author)  
Department of Statistics  
Stanford University  
Stanford, California

2a. REPORT SECURITY CLASSIFICATION  
Unclassified

2b. GROUP  
Unclassified

3. REPORT TITLE  
Multivariate Regression With One Stochastic Predictor Variable

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)  
Technical Report

5. AUTHOR(S) (Last name, first name, initial)  
SCLOVE, Stanley, L.

6. REPORT DATE  
December 30, 1966

7a. TOTAL NO. OF PAGES  
11

7b. NO. OF REPS  
5

8a. CONTRACT OR GRANT NO.  
Monr 225(52)

8b. PROJECT NO.  
NR 342-022

d.

9a. ORIGINATOR’S REPORT NUMBER(S)  
Technical Report No. 124

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

d.

10. AVAILABILITY/LIMITATION NOTICES  
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY  
Logistics and Mathematical Sciences Branch  
Office of Naval Research  
Washington, D.C. 20360

13. ABSTRACT  
Estimators are obtained which dominate the maximum likelihood estimator for the parameters in the regression of at least three dependent variables on one stochastic independent variable, these variables being jointly normally distributed. The loss function corresponds to that for the following problem: given a random sample of N observations from the joint normal distribution of all the variables and an additional independent observation on the independent variable, to predict the corresponding value of the dependent variables, when the loss function is the conditional mean square of the distance between the predicted and the actual values in the metric of the residual covariance matrix, given the sample of N observations. It is proved that the maximum likelihood estimator is admissible when there are only two dependent variables and the means of the dependent variables and independent variable are known.
Multivariate regression
Maximum likelihood estimator
Admissibility