Nonparametric estimation of multiple regression parameters

by

Ronald C. Pruitt
University of Minnesota
Technical Report #508

Nonparametric estimation of multiple regression parameters

by

Ronald C. Pruitt
University of Minnesota
Technical Report #508

Keywords and phrases: U-statistics, concordance, algorithm, simulation.

Abstract

Nonparametric estimates for the parameters in a multiple linear regression are introduced and their large sample behavior studied. The estimators are generalizations of the median of the pairwise slopes estimate for a linear regression. With a two predictor model the estimates are asymptotically normal and algorithms are presented to compute the estimates, but extension to more than two predictors presents new problems which are indicated.

1. Introduction. The estimates we consider in this paper are generalizations to a multiple linear regression model of nonparametric estimates of the slope in a linear regression given by Theil (1950) and further studied by Sen (1968). The simple regression estimate is described as follows. Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ denote independent observations from $Y_i = \beta X_i + V_i$ where the V_i are i.i.d. but otherwise completely arbitrary. The estimate, $\hat{\beta}_n$, is then given as the median of the slopes between pairs of observations,

$$\hat{\beta}_n = \underset{1\le i<j\le n}{\text{med}} \left[\frac{Y_i - Y_j}{X_i - X_j}\right] . \qquad (1.1)$$

This estimate is more robust than the least squares estimate to outliers and heavy tails in the error distribution. For example, this estimate is consistent if the errors have a Cauchy distribution. The estimate, $\hat{\beta}_n$, is known to be asymptotically normal under mild conditions on the distribution of V_1 (Sen 1968), a fact which is proved using U-statistic theory. The estimate may be equivalently defined as a zero crossing of the U-statistic

$$U_n(b) = \binom{n}{2}^{-1} \sum_{1\le i<j\le n}\sum \text{sgn}(X_i-X_j)\ \text{sgn}[(Y_i-bX_i)-(Y_j-bX_j)], \quad (1.2)$$

where a zero crossing is a point z such that $U_n(z-\delta)\ U_n(z+\delta)\le 0$ for all $\delta>0$ (see Sen (1968)). We note that (1.1) is a measure of association between the independent variable in the regression and an estimated residual, and that the independent variable is independent of the true residual. The least squares estimate (LSE) can be viewed in the same spirit as (1.1) and (1.2): the LSE is a weighted average of the pairwise slopes and the zero of the numerator of the sample correlation coefficient between the independent variable and an estimated residual. Viewed in both of these ways, $\hat{\beta}_n$ is seen to be a more robust estimate.

Here we generalize $\hat{\beta}_n$ to a model with more than one predictor. The definition is similar to that in (1.2) and is given in Section 2 where the estimator is shown to be asymptotically normal under mild conditions for a two predictor model. The proof of this result draws on some geometrical ideas and U-statistic theory. We present two algorithms for obtaining a solution which are useful under different types of conditions and present a simple numerical example of the ideas. Finally we unsuccessfully try to extend the results to models with more than two predictors. Showing that the estimator exists and is asymptotically normal if it is consistent can be handled analogously to the two predictor case, but we have been unable to show that the estimator is consistent. We can show consistency based on an auxiliary lemma which is true in the examples we have looked at but which we have been unable to prove.

2. Asymptotics for two predictors

In this paper a simple linear regression model with multiple predictors is assumed. A two predictor model is considered first and then extensions to more predictors are indicated. The model considered is

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + V_i \qquad (2.1)$$

where the V_i are i.i.d. from an arbitrary distribution G, and the (X_{1i},X_{2i}) pairs are i.i.d. from a distribution F. The triple $Z_i = (Y_i,X_{1i},X_{2i})$ is observed for $i = 1,2,\ldots,n$. Interest is centered on elucidating the relationship between the dependent and independent variables and therefore estimating β_1 and β_2.

As before we are interested in an estimator which provides protection against heavy tailed errors and outliers. It is not immediately apparent how to generalize the one predictor estimate which was obtained as the median of the pairwise slopes. If, however, we consider the estimator as a zero of measure of association between the independent variable and the estimated residual given in (1.2), a natural generalization is apparent. In the multiple prediction problem each of the independent variables is independent of the true residual and the same scheme can be used, now with one measure of association for each independent variable. To define the estimator $(\hat{\beta}_{1n}, \hat{\beta}_{2n})$, following (1.2), first define an estimated residual by

$$V_i(b) = Y_i - b_1X_{1i} - b_2X_{2i} = V_i + (\beta_1-b_1)X_{1i} + (\beta_2-b_2)X_{2i}. \quad (2.2)$$

Then define two U-statistics

$$\begin{aligned} U_{1n}(b) &= \binom{n}{2}^{-1}\sum_{i<j}\sum \operatorname{sgn}(X_{1i}-X_{1j})\operatorname{sgn}(V_i(b)-V_j(b)) \\ &= \binom{n}{2}^{-1}\sum_{i<j}\sum h(1,b_1,b_2,Z_i,Z_j), \end{aligned}$$

and $U_{2n}(b) = \binom{n}{2}^{-1}\sum_{i<j}\sum \mathrm{sgn}(X_{2i}-X_{2j})\mathrm{sgn}(V_i(b)-V_j(b))$ (2.3)

$$= \binom{n}{2}^{-1}\sum_{i<j}\sum h(2,b_1,b_2,Z_i,Z_j)$$

which are Kendall's tau statistics between each of the independent variables and the estimated residual. From (2.2) and (2.3), $U_{1n}(b_1,b_2)$ is decreasing in b_1 for fixed b_2, and $U_{2n}(b_1,b_2)$ is decreasing in b_2 for fixed b_1. This does not even guarantee the existence of a solution to this system of equations. In fact a solution may not exist as a zero of both of these functions, but a solution exists in the following weaker sense. Define a zero crossing of (2.3) as any point (b_1,b_2) such that for any $\delta > 0$,

$$U_{1n}(b_1-\delta,b_2)\cdot U_{1n}(b_1+\delta,b_2) \le 0, \qquad (2.4a)$$

$$\text{and } U_{2n}(b_1,b_2-\delta)\cdot U_{2n}(b_1,b_2+\delta) \le 0, \qquad (2.4b)$$

We will show there does exist a zero crossing of (2.3), say $(\hat\beta_{1n},\hat\beta_{2n})$ and determine its asymptotic behavior, making extensive use of U-statistics.

In the following theorem we assume that X_{11} and X_{21} have absolutely continuous distributions to simplify the exposition. The elimination of this assumption is covered in Section 4. For the asymptotic distribution of the estimate we having the following theorem.

Theorem 2.1. Assume the model (2.1) holds, and that V_1 has a bounded continuous density g(v) with respect to Lebesgue measure. Assume X_{11} and X_{21} have absolutely continuous distributions, $E[|X_{11}-X_{12}|] < \infty$, and $E[|X_{21}-X_{22}|] <$

∞. Also assume $\gamma = P(X_{11}-X_{12})(X_{21}-X_{22}) > 0]$ is not equal to zero or one. Then there exists a zero crossing of (2.3), and with $(\hat{\beta}_{1n},\hat{\beta}_{2n})$ any such zero crossing,

$$n^{1/2} (\hat{\beta}_n-\beta) \xrightarrow{d} N(0,\Sigma), \qquad (2.5)$$

where Σ is given by (3.12). □

Before we proceed with the proof, we make the following preliminary comments. With least squares estimates, multicollinearity problems arise when the predictor variables are highly correlated and the solution becomes non-unique when the absolute value of the correlation between variables is one. Similar problems arise with this estimator although the association between variables is now measured by probability of concordance between sample pairs rather than by correlation. We will show that the estimation gets worse as the probability of concordant pairs reaches an extreme, and Theorem 2.1 holds only if this probability is bounded away from zero and one. If the concordance probability is zero or one, we will show that the set of all possible solutions is unbounded, and $\hat{\beta}_n$ need not converge to β, that is, γ being unequal to zero or one is necessary for the convergence in (2.5) to hold.

<u>3. Proof of Theorem 2.1.</u> The proof is broken into several steps. First a solution is shown to exist. Second we show that $n^{1/2-\epsilon}(\hat{\beta}_{sn}-\beta_s) \xrightarrow{P} 0$ for any $\epsilon > 0$, $s = 1,2$ and any solution. Third a U-statistic decomposition around β is

used to fine the asymptotic distribution of $(\hat{\beta}_{1n}, \hat{\beta}_{2n})$.

First we show that a solutiuon exists. The strategy is to collapse the (b_1, b_2) plane into one dimension by looking at specific points where $U_{1n}(b) = 0$, and then to show that U_{2n} inherits its monotonicity properties on the reduced one-dimensional space and crosses zero.

Let for each fixed b_2

$$\hat{b}_1(b_2) = \underset{i<j}{\text{med}} \left[\frac{Y_i - Y_j - b_2(X_{2i} - X_{2j})}{X_{1i} - X_{1j}}\right] . \qquad (3.1)$$

We assert that the graph of $(\hat{b}_1(b_2), b_2)$ is a connected sequence of line segments in the (b_1, b_2)-plane, and (2.4a) is satisfied for every point on the graph. For the latter claim see the argument in Sen (1968) showing that the medians of the pairwise slopes and zero crossings of (2.4a) are equivalent for fixed b_2. For the first claim, consider the lines $V_i(b) - V_j(b) = 0$. Let $H_k = \{i_s = (b_{1s}, b_{2s}): s = 1, \ldots, I_k\}$ be the set of all intersections of these $\binom{k}{2}$ lines in the plane for $1 \le i < j \le k$. $(\#H_n)$ is clearly finite. Choose $i_1 = (b_{11}, b_{21}) \in H_n$ and note that $P[b_{21} = b_{2s}] = 0$ for any $s = 2, \ldots, I_n$ since V_1 has an absolutely continuous distribution. To see this let $L_k = \{b: V_{k+1}(b) - V_j(b) = 0, j = 1, \ldots, k\} \cap \{b: V_i(b) - V_j(b) = 0,\ 1 \le i < j \le k\}$. Say $L_k = \{j_s = (b^*_{1s}, b^*_{2s}): s = 1, \ldots, J_k\}$. It then suffices to show $P[A_k] = P[b^*_{2s} = b_{2t}$: for some $1 \le s \le J_k$, $1 \le t \le I_k] = 0$. Now

$$P[A_k] = \sum_{s=1}^{J_k} \sum_{t=1}^{I_k} P[b^*_{2s}=b_{2t}]$$

$$= \sum_{s=1}^{J_k} \sum_{t=1}^{I_k} E\{ P[b^*_{2s}=b_{2t}) | Z_1,\ldots,Z_k,X_{1k+1},X_{2k+1}]\}$$

$$= \sum_{s=1}^{J_k} \sum_{t=1}^{I_k} E\{ P[V_{k+1} = v(s,t) | Z_1,\ldots,Z_k,X_{1k+1},X_{2k+1}]\}$$

$$= 0,$$

where $v(s,t)$ is a real number which depends on s and t. For example if $j^*_s = (V_{k+1}(b)-V_a(b)=0) \cap (V_b(b)-V_c(b)=0)$ and $i_t = (V_d(b)-V_e(b)=0) \cap (V_f(b)-V_g(b)=0)$ then

$$v(s,t) = Y_a-\beta_1X_{1k+1}-\beta_2X_{2k+1}+(X_{1b}-X_{1c})^{-1}[X_{1k+1}-X_{1a})(Y_b-Y_c)+AB/C],$$

where $A = [(X_{1f}-X_{1g})(Y_d-Y_e) - (X_{1d}-X_{1e})(Y_f-Y_g)]$;
$B = [X_{1b}-X_{1c})(X_{2k+1}-X_{2a}) - (X_{1k+1}-X_{1a})(X_{2b}-X_{2c})]$;
and $C = [(X_{1f}-X_{1g})(X_{2d}-X_{2e}) - (X_{1d}-X_{1e})(X_{2f}-X_{2g})]$. Now, suppressing the argument n, consider $b_{21},\ldots,b_{2I}$, and let $b_{2(1)},\ldots,b_{2(I)}$ be their order statistics. For $b_{2(s)} < b_2 < b_{2(s+1)}$, the ordering of $[(Y_i-Y_j - b_2(X_{2i}-X_{2j})](X_{1i}-X_{1j})^{-1}$ stays constant, and the median occurs at the same (i,j) pair or as the average between the same two (i,j) pairs. Hence $\{(\hat{b}_1(b_2),b_2): b_{2(s)} < b_2 < b_{2(s+1)}\}$ is a line segment. The fact that only one intersection occurs at $b_2 = b_{2(s)}$ implies that these line segments are connected. Hence $(\hat{b}_1(b_2),b_2)$ is a connected sequence of line segments of which every point satisfies (2.4a).

Now consider the values of U_{2n} in this one dimensional space. Considered as a function of two variables we know U_{2n} is decreasing in b_2 for fixed b_1: we wish to show this remains true for b_2 in the restricted space; that is, U_{2n} is decreasing in b_2 along the line segments $(\hat{b}_1(b_2),b_2)$. U_{2n} is clearly constant along the line segments and only changes value at the vertices. Consider a vertex $i_j = (b_{1j},b_{2j})$ and suppose $b_{2j} = b_{2(s)}$. Choose b_2^1 and b_2^u so that $b_{2(s-1)} < b_2^1 < b_{2(s)} < b_2^u < b_{2(s+1)}$. Let $b_1^1 = \hat{b}_1(b_2^1)$ and consider three cases. If $\mathrm{sgn}(b_1^1-b_{1j}) = \mathrm{sgn}(b_1^u-b_{1j})$ then clearly $U_{2n}(b^u) \le U_{2n}(b^1)$ since U_{2n} is decreasing in b_2 for fixed b_1. For the other two cases let $V_{1n}(b) = (U_{1n}(b) + U_{2n}(b))/2$ and $V_{2n}(b) = (U_{1n}(b) - U_{2n}(b))/2$. Note that

$$V_{1n}(b) = \binom{n}{2}^{-1} \Sigma_A\Sigma\ \mathrm{sgn}(X_{1i}-X_{1j})\mathrm{sgn}[V_i(b)-V_j(b)], \qquad (3.2)$$

where $A = \{(i,j): (X_{1j}-X_{1j})(X_{2i}-X_{2j}) > 0\}$. V_{1n} is decreasing in b_1 and b_2, while V_{2n} is decreasing in b_1 and increasing in b_2. If $b_1^1 < b_{1j}$ and $b_1^u > b_{1j}$, then $V_{1n}(b^u) = U_{1n}(b^1) = 0$ by the absolute continuity of V_1. Finally if $b_1^1 > b_{1j}$ and $b_1^u < b_{1j}$ the same argument can be used with V_{2n} in place of V_{1n}. We have shown that U_{2n} is decreasing along $(\hat{b}_1(b_2),b_2)$. It is easily checked that for small b_2, $U_2(\hat{b}_1(b_2),b_2) \ge 0$, and for large b_2, $U_{2n}(\hat{b}_1(b_2),b_2) \le 0$. Just note that along $(\hat{b}_1(b_2),b_2)$, $U_{1n} = V_{1n} + V_{2n} = 0$ and as b_2 decreases V_{2n} decreases. In fact, as $b_2 \to -\infty$, eventually $\binom{n}{2}V_{2n} = -\min[n_1, \binom{n}{2}-n_1]$ where $n_1 = \#A$, since the magnitude of V_{2n} is restricted by $V_{2n} = -V_{1n}$. When $\binom{n}{2}V_{2n} > -\min[n_1, \binom{n}{2}-n_1]$, V_{2n} can decrease and V_{1n} can increase, but when $\binom{n}{2}V_{2n} = -\min[n_1, \binom{n}{2}-n_1]$, one of V_{1n} and V_2 has reached its extreme value. Now, since $U_{2n} = V_{1n} - V_{2n}$, there exists b_2 with $\binom{n}{2}U_{2n}(\hat{b}_1(b_2), b_2) = 2\min[n_1,\binom{n}{2}-n_1] \ge$

0. Similarly there exists b_2 with $\binom{n}{2}U_{2n}(\hat{b}_1(b_2),b_2) = -2\min[n_1, \binom{n}{2}-n_1] \le 0$. Hence there must be a zero crossing of U_{2n} along $(\hat{b}_1(b_2),b_2)$ and such a zero crossing must satisfy (2.4b). If U_{2n} is zero on a line segment clearly (2.4b) is satisfied; and if U_{2n} crosses zero at a vertex i_j then $U_{2n}(b_{1j},b_{2j}-\delta) \ge U_{2n}(b^l) \ge 0 \ge U_{2n}(b^u) \ge U_{2n}(b_{1j},b_{2j}+\delta)$ for any $\delta > 0$. Therefore a solution exists.

Tacitly we have assumed that there are both concordant and discordant pairs in the sample. If this is not the case, then consider the case with all concordant pairs. Here $U_{1n} = U_{2n}$ so we need only consider the zero crossings of U_{1n}. This solution region is infinite in extent. This is not surprising since our method is rank based and the two samples are indistinguishable by ranks alone. In order to get an estimate that converges to the true value, all sample pairs cannot be concordant for large samples. Hence in Theorem 2.1 the condition that $\gamma \ne 0$ or 1 is required. Finally note that we have found a unique solution by this method. Of course there may be other solutions which are zero crossing of U_{1n} and U_{2n} if the number of sample pairs is even, since then the sample median in (3.1) need not be uniquely defined.

Second we show that $n^{1/2-\epsilon}(\hat{\beta}_{sn}-\beta_s) \xrightarrow{P} 0$ for s=1,2, any $\epsilon > 0$, and any solution. It suffices to show $P(\hat{\beta}_{1n} < \beta_1 - tn^{-1/2+\epsilon}) \to 0$ for any fixed $t > 0$ by symmetry. Fix $\epsilon > 0$. Recall

$$V_{1n}(b) = \binom{n}{2}^{-1}\Sigma_A\Sigma \, \mathrm{sgn}(X_{1i}-X_{1j})\mathrm{sgn}[Y_i-Y_j-b_1(X_{1i}-X_{1j})-b_2(X_{2i}-X_{2j})]$$

is decreasing in b_1 and decreasing in b_2. V_{2n} is decreasing in b_1 and increasing in b_2. Hence letting $v = tn^{-1/2+\epsilon}$

$$
\begin{aligned}
P[\hat{\beta}_{1n}<\beta_1-v] &\le \sup_{b_2} P[V_{1n}(\beta_1-v,b_2)\le 0,\ V_{2n}(\beta_1-v,b_2)\le 0] \\
&\le \max\{\sup_{b_2\ge\beta_2} P[V_{2n}(\beta_1-v,b_2)\le 0],\ \sup_{b_2\le\beta_2} P[V_{1n}(\beta_1-v,b_2)\le 0\} \\
&= \max\{P[V_{1n}(\beta_1-v,\beta_2)\le 0],\ P[V_{2n}(\beta_1-v,\beta_2)\le 0]\}. \quad (3.3)
\end{aligned}
$$

The normalized $V_{in}(\beta_1-tn^{-1/2+\epsilon},\beta_2)$ are diagonal elements of triangular arrays of U-statistics which we will prove to be asymptotically normal after we develop some notation. Recall

$$U_{in}(b) = \binom{n}{2}^{-1}\sum_{j<k} h(i,b_1,b_2,Z_j,Z_k),$$

and define

$$h(3,b_1,b_2,Z_j,Z_k) = [h(1,b_1,b_2,Z_j,Z_k) + h(2,b_1,b_2,Z_j,Z_k)]/2,$$

$$\text{and } h(4,b_1,b_2,Z_{jj},Z_k) = [h(1,b_1,b_2,Z_j,Z_k) - h(2,b_1,b_2,Z_j,Z_k)]/2,$$

so that

$$V_{in}(b) = \binom{n}{2}^{-1}\sum_{j<k}\sum h(i+2,b_1,b_2,Z_j,Z_k).$$

Let, for i=1,2,3,4,

$$h_1(i,b_1,b_2,Z_j) = E[h(i,b_1,b_2,Z_j,Z_k)|Z_j],$$

$$\text{and } \sigma_i^2(b_1,b_2) = 4\{\mathrm{Var}(h_1(i,b_1,b_2,Z_j))\}.$$

Finally, let

$$\mu_i(s_1,s_2) = E[h(i,\beta_1+s_1,\beta_2+s_2,Z_j,Z_k)], \text{ for } i = 1,2,$$
$$\nu_1(s) = (\mu_1(s,0) + \mu_2(s,0))/2,$$
$$\text{and } \nu_2(s) = (\mu_1(s,0) - \mu_2(s,0)/2. \qquad (3.4)$$

Then $\mu_i(s)$ and $\sigma_i^2(\beta_1+s_1,\beta_2+s_2)$ are the asymptotic mean and variance of $U_{in}(\beta_1+s_1,\beta_2+s_2)$ for fixed s, and $\nu_i(s)$ and $\sigma_{i+2}^2(\beta_1+s,\beta_2)$ are the asymptotic mean and variance of $V_{in}(\beta_1+s,\beta_2)$ for fixed s from U-statistic theory (Puri and Sen (1971), Theorem 3.2.1). Finally let $T_{in}(s) = n^{1/2}[V_{in}(\beta_1-s,\beta_2) - \nu_i(-s)]$. Then we assert

<u>Lemma 3.1</u> $T_{in}(n^{-\gamma}t) \xrightarrow{d} N(0,\sigma_{i+2}{}^2(\beta_1,\beta_2))$ for any $\gamma > 0$. □

<u>Proof</u>: Fix γ and t and consider $T_{in}(m^{-\gamma}t)$ as a triangular array of random variables of which we wish to show the diagonal elements converge. From standard theory we know $T_{in}(m^{-\gamma}t) \xrightarrow{d} T_i(m^{-\gamma}t) \sim N(0, \sigma^2{}_{i+2}(\beta_1-m^{-\gamma}t,\beta_2)$ as $n\to\infty$ and $T_i(m^{-\gamma}t) \xrightarrow{d} T \sim N(0, \sigma^2{}_{i+2}(\beta_1,\beta_2))$ as $m \to \infty$. To show the diagonal elements coverge we need only check the condition in Theorem 4.2 of Billingsley (1968),

$$\lim_{m\to\infty} \limsup_{n\to\infty} P[|T_{in}(m^{-\gamma}t)-T_{in}(n^{-\gamma}t)|>\epsilon] = 0 \text{ for any } \epsilon > 0.$$

Now

$$P[|T_{in}(m^{-\gamma}t)-T_{in}(n^{-\gamma}t)|>\epsilon] \le P[|S_{in}(m^{-\gamma}t)| > \epsilon/2]+P[|S_{in}(n^{-\gamma}t)|>\epsilon/2]$$
$$\le 4\epsilon^{-2} \quad E[S_{in}^{2}(m^{-\gamma}t)+S_{in}^{2}(n^{-\gamma}t)] \quad (3.5)$$

where $S_{in}(z) = T_{in}(z)-T_{in}(0)$. Using Hoeffding's (1948) projection method we can decompose the U-statistic $S_{in}(z)$ as

$$S_{in}(m^{-\gamma}t) = n^{1/2}\{2n^{-1}\sum_{j=1}^{n} k_{i1}(m^{-\gamma}t,Z_j) + r_{in}(m^{-\gamma}t)\}$$

where $k_{i1}(s,Z_j) = E[k_i(s,Z_j,Z_1)]$ and $k_i(s,Z,Z_1)$ $= h(i+2,\beta_1-s,\beta_2,Z_j,Z_1) - h(i+2,\beta_1,\beta_2,Z_j,Z_1) - \nu_i(-s)$. Then by the Cauchy-Schwarz inequality

$$\begin{aligned} E[S_{in}^{2}(m^{-\gamma}t)] &\le 2n\{E[\{2n^{-1}\sum_{j=1}^{n} k_{i1}(m^{-\gamma}t,Z_j)\}^2]+E[r_{in}^2(m^{-\gamma}t)]\} \\ &\le E[k_{i1}^2(m^{-\gamma}t,Z_i)] + 2nE[r_{in}^2(m^{-\gamma}t)] \qquad (3.6) \\ &= C(m,n). \end{aligned}$$

Note $\lim_{m\to\infty} \lim_{n\to\infty} C(m,n) = 0$ and $\lim_{n\to\infty} C(n,n) = 0$ since $\lim_{s\to 0} k_i(s,z) = 0$ for all z by the Dominated Convergence Theorem and $E[r_{in}^{2}(s)] \le \Gamma_\epsilon n^{-2+\epsilon}$ for the remainder of a U-statistic with any bounded kernel (see for example Denker and Keller (1983), Proposition 2). Combining (3.5) and (3.6) completes the proof of Lemma 3.1. □

Taking $\gamma = 1/2-\epsilon$ in Lemma 3.1 and returning to (3.3),

$$P[V_{1n}(\beta_1 - tn^{-1/2+\epsilon}, \beta_2) \le 0] \to \lim_{n\to\infty} \Phi\,(-n^{1/2}\sigma_3^{-1}(\beta_1,\beta_2)\nu_1(-tn^{-1/2+\epsilon})).$$

The asymptotic variance, $\sigma_3^2(\beta_1,\beta_2)$, is bounded away from zero and ∞ since $|h_1|$ is not constant and is bounded. After noting $\mu_i(0,0) = 0$ by symmetry, and recalling (2.3),

$$\mu_i(t) = E[\mathrm{sgn}(X_{i1}-X_{i2})\{\mathrm{sgn}(V_1-V_2-t_1(X_{11}-X_{12})-t_2(X_{21}-X_{22}))-\mathrm{sgn}(V_1-V_2)\}]$$

$$= -2E[\mathrm{sgn}(X_{i1}-X_{i2})\{G(V_2+t_1(X_{11}-X_{12})+t_2(X_{21}-X_{22}))-G(V_2)\}] \qquad (3.7)$$

so that, from (3.4), $\nu_1'(s) \le 0$ and $|\nu_1'(s)| \le 2E[|X_{11}-X_{12}|]\cdot\sup_v\{g(v)\}$.

This gives $\lim_{n\to\infty} -n^{1/2}\sigma_3^{-1}(\beta_1,\beta_2)\nu_1(-tn^{-1/2+\epsilon}) = -\infty$, and $\lim_{n\to\infty} P[V_{1n}(\beta_1 - tn^{-1/2+\epsilon}, \beta_2) \le 0] = 0$. Similarly for V_{2n}. Hence $n^{1/2-\epsilon}(\hat\beta_{in}-\beta_i) \xrightarrow{P} 0$.

Third we find the asymptotic distribution of the parameter vector estimate. Make the following decomposition,

$$\begin{bmatrix} U_{1n}(\hat\beta_n) \\ U_{2n}(\hat\beta_n) \end{bmatrix} = \begin{bmatrix} U_{1n}(\beta) \\ U_{2n}(\beta) \end{bmatrix} + \begin{bmatrix} \mu_1(\hat\beta_n-\beta) \\ \mu_2(\hat\beta_n-\beta) \end{bmatrix} + \begin{bmatrix} R_{1n}(\hat\beta_n) \\ R_{2n}(\hat\beta_n) \end{bmatrix}, \qquad (3.8)$$

where μ_i (s) is given by (3.4). The strategy is as follows. First we show that $R_{1n}(\hat\beta_{1n},\hat\beta_{2n}) = o_p(n^{-1/2})$ while $U_{1n}(\beta_1,\beta_2)$ is $O_p(n^{-1/2})$ but not $o_p(n^{-1/2})$. The asymptotic mean term can be written as $A_n(\hat\beta_n-\beta)$ using the Fundamental

Theorem of Calculus. Then if $A_n \to A$ smoothly and A is invertible we get

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, A^{-1}BA^{t}), \qquad (3.9)$$

where B is the asymptotic covariance matrix of $[U_{1n}(\beta), U_{2n}(\beta)]^t$ and $A^{-t} = (A^{-1})^t$.

To handle the remainder term we use the following lemma.

Lemma 3.2 Let $R_{in}(b_1, b_2)$ be given by (3.8). Then for $a > 1/3$

$$\sup_{\substack{0 \le |t_j| \le n^{-a} \\ j,i = 1,2}} |R_{in}(\beta_1 + t_1, \beta_2 + t_2)| = o_p(n^{-1/2}). \qquad \square$$

The proof of this lemma is given in Section 6. The proof is a modification of Lemma 5 of Bhattacharya, Chernoff, and Yang (1983) to the case of more than one dimension.

With a one predictor model, we can take advantage of the monotonicity of $U_n(b)$ to prove the asymptotic normality of $\hat{\beta}_n$ directly by noting that the events $\{\hat{\beta}_n < \beta + t\}$ and $\{U_n(\beta + t) < 0\}$ are identical. This allows us to deal only with triangular arrays of U-statistics evaluated at non-random arguments. Manipulation of these variables is fairly routine. For examples of the details, see Sen (1968). In the multiple prediction problem no such simple proof is possible for these estimators. Reduction of the multivariate asymptotic distribution to univariate distributions using the Cramer-Wold device does not appear to be possible. There is not a convenient definition of $c_1\hat{\beta}_{1n} + c_2\hat{\beta}_{2n}$ in terms of a zero crossing of a U-statistic. Also, direct examination of $c_1U_{1n} + c_2U_{2n}$, which might be useful since U_{1n} and U_{2n} are zero if and only if all linear combinations of them are zero, does not appear fruitful since this U-statistic does not have any nice monotonicity properties. In the present problem we need to use the decomposition given in (3.8). Now the remainder term is a U-statistic evaluated at a random point and so must be bounded uniformly over the possible region of solution. This is more difficult, but the relatively unsharp bound given by Lemma 3.2 is sufficient for our purpose since we have shown $|\hat{\beta}_{1n}-\beta_1| = o_p(n^{-1/2+z})$ for any $z > 0$. Hence $R_{1n}(\hat{\beta}_{1n},\hat{\beta}_{2n}) = o_p(n^{-1/2})$, since $P(|\hat{\beta}_{1n} - \beta_1| < n^{-1/2+z}) \to 1$ as $n \to \infty$ for any $z > 0$.

Next examine the asymptotic mean term. Let

$$d_{ij}(s) = \frac{\partial}{\partial t_j} \mu_i(t) \Big|_{t=s}.$$

Now write

$$\begin{pmatrix} \mu_1(\hat{\beta}_n-\beta) \\ \mu_2(\hat{\beta}_n-\beta) \end{pmatrix} = A_n(\hat{\beta}_n-\beta),$$

where $a_{ijn} = d_{ij}(s_{ijn})$ and $s_{ijn} \to (0,0)$ as $n \to \infty$. Let $A = (d_{ij}(0))$. If A is invertible and the d_{ij} are continuous, then A_n is invertible for large enough n. Recalling $\mu_i(t)$ from (3.7), it follows that,

$$d_{ij}(s) = -2E[(X_{j1}-X_{j2})\mathrm{sgn}(X_{i1}-X_{i2})g(V_2+s_1(X_{11}-X_{12})+s_2(X_{21}-X_{22}))].$$

This is continuous since g is continuous and

$$a_{ij} = d_{ij}(0) = -2E[X_{j1}-X_{j2})\mathrm{sgn}(X_{i1}-X_{i2})] \int g^2(v)dv. \quad (3.10)$$

This gives $a_{12}a_{21} < a_{11}\, a_{22}$ and hence A invertible, since $P[(X_{11}-X_{12})(X_{21}-X_{22}) > 0]$ is not equal to zero or one. Returning to (3.8) we have for any solution $(\hat{\beta}_{1n},\hat{\beta}_{2n})$

$$\begin{pmatrix} U_{1n}(\hat{\beta}_n) \\ U_{2n}(\hat{\beta}_n) \end{pmatrix} = \begin{pmatrix} U_{1n}(\beta) \\ U_{2n}(\beta) \end{pmatrix} + A_n(\hat{\beta}_n-\beta) + o_p \begin{pmatrix} U_{1n}(\beta) \\ U_{2n}(\beta) \end{pmatrix},$$

and

$$n^{1/2}(\hat{\beta}_n-\beta) = -A_n^{-1} n^{1/2} \begin{pmatrix} U_{1n}(\beta) \\ U_{2n}(\beta) \end{pmatrix} \begin{pmatrix} 1+o_p(1) & 0 \\ 0 & 1+o_p(1) \end{pmatrix},$$

which gives (3.9) after it is shown that the U-statistic vector is asymptotically normal. This asymptotic normality is obtained from standard U-statistic results (Puri and Sen (1971), Corollary 3.2.3.2) which shows, referring to (2.3) and (3.3),

$$n^{1/2} \begin{pmatrix} U_{1n}(\beta) \\ U_{2n}(\beta) \end{pmatrix} \to N(0,B),$$

where $B = (b_{ij})$ and $b_{ij} = 4\,\mathrm{Cov}[h_1(i,\beta_1,\beta_2,Z_1),h_1(j,\beta_1,\beta_2,Z_1)]$. It is easily shown that

$$B = \frac{4}{9} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \qquad (3.11)$$

where $\rho = 3\{4E[P(X_{11} < X_{12})P(X_{21} < X_{22})] - 1\}$. To determine ρ, note from (2.3) that

$$\begin{aligned} h_1(j,\beta_1,\beta_2,Z_1) &= E[\mathrm{sgn}(X_{j1}-X_{j2})\mathrm{sgn}(V_1-V_2) \mid (X_{11},X_{21},V_1)] \\ &= \{2P(X_{j2} < X_{j1}) - 1\}\{2P(V_2 < V_1) - 1\}. \end{aligned}$$

Now use the fact that (X_1,X_2) is independent of V and that $P(V_2 < V_1)$ and $P(X_{j2} < X_{j1})$ are uniform random variables.

This finishes the proof of Theorem 2.1 with

$$\Sigma = A^{-1}BA^{-t}, \qquad (3.12)$$

where $A = (a_{ij})$ is given in (3.10), and B is given in (3.11). □

In the proof of Theorem 2.1 $E[|X_{i1}-X_{i2}|] < \infty$ was used only in showing the asymptotic normality of the estimates. If this assumption is violated, the same techniques show $n^{1/2}(\hat{\beta}_{in}-\beta_i) \xrightarrow{P} 0$, if $E[|X_{j1}-X_{j2}|] = \infty$ for j=1,2. The same techniques for handling heavy tails as used by Pruitt (1987) in proving asymptotic convergence rates will work in these problems; we have not pursued the details.

4. Comments, algorithms, and an example. In Theorem 2.1, the assumption that X_{11} and X_{21} have absolutely continuous distributions is not necessary and was made to simplify the exposition. A more general form is Theorem 2.1a.

Theorem 2.1a. Assume the model (2.1) holds, and that V_1 has a bounded continuous density g(v) with respect to Lebesgue measure. Assume $0 < E[|X_{11}-X_{12}|] < \infty$, $0 < E[|X_{21}-X_{22}|] < \infty$, $\gamma_1 = P[(X_{11}-X_{12})(X_{21}-X_{22}) > 0] > 0$, and $\gamma_2 = P[(X_{11}-X_{12})(X_{21}-X_{22}) < 0] > 0$. Then there exists a zero crossing of (2.3), and with $\hat{\beta}_n$ being any such zero crossing, $n^{1/2}(\hat{\beta}_n-\beta) \xrightarrow{d} N(0,\Sigma)$. □

The only change in the proof is that now some of the lines $V_i(b)- V_j(b)=0$ may be parallel to the axes and the nice form given by (3.2) does not exist. $V_i(b)$ is still a U-statistic but the kernel is more complicated. The fact that some of the $V_i(b)-V_j(b)=0$ lines are parallel to the axes has no material effect on the proof.

To construct a confidence region for β we may use (2.5) if we have a large sample. For small samples an exact region is not possible unlike the one predictor case. For one predictor one could use the monotonicity of $U_n(b)$ and the tabled small sample distribution of $U_n(\beta)$ to construct a confidence interval (see Sen (1968)). With two predictors the distribution of $(U_{1n}(\beta),U_{2n}(\beta))^t$ [or $(V_{1n}(\beta),V_{2n}(\beta))^t$] depends on the unknown value ρ (see (3.11)), and these small sample distributions have not been tabled. A conservative confidence region may be obtained from the one predictor methods and Bonferroni's inequality. For example, if $R_n = \{b: |\binom{n}{2}U_{1n}(b)| < u(\epsilon)$, $|\binom{n}{2}U_{1n}(b) < u(\epsilon)\}$, then $P(R_n) \geq 1 - \epsilon$ where $u(\epsilon)$ is determined from the small sample Kendall's tau distribution by $P[|\binom{n}{2}U_{1n}(\beta)| < u(\epsilon) = 1-\epsilon/2$. Bonferrroni's inequality is sharper if U_{1n} and U_{2n} are nearly independent. V_{1n} and V_{2n} are more nearly independent but a small sample confidence region from them is not possible since the distribution of each depends on the unknown probability of concordance and the small sample distribution is not tabled. A sample confidence region is provided in the example.

We have a computer program to find a zero crossing of U_{1n} and U_{2n}. The entire solution region may be explored by examining the (U_{1n}, U_{2n}) surface near this solution point if desired. To simplify the language in this paragraph, when we say U_{in} (b) $\triangleq 0$, we shall mean that b is a zero crossing of U_{in}. Solution method A(alternating) is that of iteratively solving $U_{jn}(b) \triangleq 0$ for b_j, with j alternating between 1 and 2. We have had no convergence problems with this algorithm but it may be very slow especially if X_{11} and X_{21} are strongly associated. If they are strongly associated, the solution regions of U_{1n} (b) $\triangleq 0$ and U_{2n} (b) $\triangleq 0$ are separated by a small angle and convergence can be slow. Numerical results on this are given later. Algorithm S(splitting) is to first find $b_2^{(1)}$ and $b_2^{(2)}$ so that $U_{1n}(b^{(1)})\triangleq 0$, $U_{2n}(b^{(1)})>0$ and $U_{1n}(b^{(2)})\triangleq 0$, $U_{2n}(b^{(2)})<0$. We may then take $b_2^{(3)} =$

$[b_2^{(1)}+b_2^{(2)}]/2$ (or with $c_j = |U_{2n}(b^{(j)})|[|U_{2n}(b^{(1)})|+|U_{2n}(b^{(2)})|]^{-1}$, $b_2^{(3)} = c_2 b_2^{(1)}+c_1 b_2^{(2)}$) and find $b_1^{(3)}$ so that $U_{1n}(b^{(3)}) \triangleq 0$. If $U_{2n}(b^{(3)}) > 0$ discard $b^{(1)}$, otherwise discard $b^{(2)}$, and repeat this splitting process until either $U_{2n}(b^{(k)}) = 0$ or $\|b^{(k-1)}-b^{(k)}\|^2$ is small. Convergence is guaranteed since U_{2n} is monotone on the set $U_{1n}(b) \triangleq 0$. This algorithm may also be implemented using V_{1n} and V_{2n} in place of U_{1n} and U_{2n}. The advantage is that time may be saved especially if the number of concordant pairs is near $\binom{n}{2}$ or zero. Then solving one of $V_{jn}(b) \triangleq 0$ will be simplified. This simplification occurs only if a list is kept of concordant pairs so that the V_{jn} do not need to be summed over all pairs. The difference due to the often greater orthogonality of V_{1n} and V_{2n} compared to U_{1n} and U_{2n} does not seem to have much effect. In Table 4.1, we have tabulated the average number of U-statistics which must be evaluated using the alternating and splitting algorithms. The estimator β_n^* is the least squares estimate. The parameters are as follows:

$$X_j = (X_{1j},X_{2j})^t \sim N(0,\Sigma_i) \text{ with } \Sigma_1 = \begin{pmatrix}5 & 1\\ 1 & 5\end{pmatrix} \text{ and } \Sigma_2 = \begin{pmatrix}5 & 4\\ 4 & 5\end{pmatrix} ;$$

$\beta_1 = 1$, $\beta_2 = 2$; $V_j \sim V(0,1)$; and $n = 20$ or 70.

TABLE 4.1

Simulation for estimation of parameters from $Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + V_i$.

Σ	n	average number of U-statistics algorithm A	S	estimated variances $\hat{\beta}_n$	β^*_n	efficiency	fixed design theory
Σ_1	20	61	49	0.0151	0.0117	0.77	0.0104
Σ_1	70	57	72	0.00270	0.002338	0.88	0.00298
Σ_2	20	96	56	0.0535	0.0317	0.59	0.0278
Σ_2	70	100	76	0.00919	0.00646	0.70	0.00794

The simulations were done with 1000 trials for each combination of n and Σ. The estimated variances are the sample variances for the 1000 trials averaged over both parameters, and the estimated efficiences are just the ratio of these values. The values in the fixed design theory column are a diagonal element (both are equal) of $n^{-1}\Sigma^{-1}$. The rate of convergence of the alternating algorithm seems to depend largely on the covariance matrix of the observations and little on sample size while the reverse is true for the splitting algorithm. The alternating algorithm can probably be used unless the corvariance matrix shows the independent variables to be highly correlated. It appears the efficiency increases with sample size, and it does so more slowly will ill-conditioned matrices.

Example

Suppose n = 5 and the following triples are observed:

i	X_{1i}	X_{2i}	Y_i
1	-1.41	0.59	-1.93
2	-2.29	-0.72	-4.18
3	-0.53	-1.59	-4.21
4	0.02	0.52	0.26
5	0.73	0.75	2.97

Figure 4.1 about here

The resulting 10 $V_i(b)-V_j(b)=0$ lines are drawn in Figure 4.1 and in finer detail in Figure 4.2. The coefficients of these lines are given below:

line	i	j	equation
1	1	2	$-0.88b_1 + -1.31b_2 = -2.25$
2	1	3	$0.88b_1 + -2.18b_2 = -2.28$
3	1	4	$1.43b_1 + -0.07b_2 = 2.19$
4	1	5	$2.14b_1 + 0.16b_2 = 4.90$
5	2	3	$1.76b_1 + -0.87b_2 = -0.03$
6	2	4	$2.31b_1 + 1.24b_2 = 4.44$
7	2	5	$3.02b_1 + 1.47b_2 = 7.15$
8	3	4	$0.55b_1 + 2.11b_2 = 4.47$
9	3	5	$1.26b_1 + 2.24b_2 = 7.18$
10	4	5	$0.71b_1 + 0.23b_2 = 2.71$

The point estimate $(\hat{\beta}_{15}, \hat{\beta}_{25})$ occurs at the intersection of lines 2,3, and 8 is (1.61,1.70). From Kendall (1970), $P[|10U_{15}(\beta_1,\beta_2)| \le 0.516$ and $P[|10U_{15}(\beta_1,\beta_2)| \le 4] = 0.766$.

Figure 4.2 about here

Hence we can construct (for example) a 3.2% confidence region as $\{(b_1,b_2): |10U_{j5}(b_1,b_2)| \le 2, j=1,2\}$ or a 53.2% confidence region as $\{(b_1,b_2): |10U_{j5}(b_1,b_2)| \le 4, j=1,2\}$. In Figure 4.2 these confidence regions are indicated. The confidence regions suffer from conflicting aims. If $\eta = P[|\binom{n}{2}U_{1n}(\beta_1,\beta_2)| \le k]$ is small, Bonferroni's inequality gives an extremely crude bound and is useless if $\eta < 0.5$. If η is large, k is large and for small samples the confidence region is infinite in extent. However for samples as small as 10 this method can be used if X_{11} and X_{21} are nearly independent. For example if the number of concordant pairs is 20(out of 45), then $R = \{(b_1,b_2): |45U_{j10}(b_1,b_2) \le 17, j=1,2\}$ is not infinite in extent and $P[R\} \ge 0.784$. The region is not infinite in extent, since $\binom{n}{2}[|U_{1n}|+|U_{2n}|] = \binom{n}{2}2\max[|V_{1n}|, |V_{2n}|]$, and outside a finite region this is always at least $2\min(n_1,\binom{n}{2}-n_1)$, where n_1 is the number of concordant pairs, since one of $|V_{1n}|$ and $|V_{2n}|$ must be at an extreme. Hence $\max|U_{1n}|, |U_{2n}|] \ge \min(n_1,\binom{n}{2}-n_1) = 20$. Explicit representation of the region does become difficult as n increases since the number of lines increases as n^2.

5. Multiple regression. The model we consider is

$$Y_i = \sum_{j=1}^{r} \beta_j X_{ji} + V_i, \tag{5.1}$$

where the V_i are i.i.d. random variables from an arbitrary distribution G, and the r-tuples $(X_{1i},X_{2i},\ldots,X_{ri})$ are i.i.d. random vectors. The (r+1)-tuple Z_i

$= (Y_i, X_{1i}, \ldots, X_{ri})$ is observed for $i=1,2,\ldots,n$. Following (2.2) and (2.3) analogously define estimated residuals $V_i(b)$ and r U-statistics $U_{sn}(b)$ by

$$V_i(b) = Y_i - \sum_{j=1}^{r} b_j X_{ji}, \qquad (5.2)$$

and

$$U_{sn}(b) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \sum \operatorname{sgn}(X_{si} - X_{sj}) \operatorname{sgn}(V_i(b) - V_j(b)), \qquad (5.3)$$

where $b = (b_1, b_2, \ldots, b_r)$.

Conjecture 5.1. Assume the model (5.1) holds, and that V_1 has a bounded continuous density $g(v)$ with respect to Lebesgue measure. Assume the X_{j1} have absolutely continuous distribution, $E[X_{j1} - X_{j2}|] < \infty$ for $j=1,\ldots,r$, and that the r x r matrix A given by $a_{ij} = E[(X_{j1} - X_{j2})\operatorname{sgn}(X_{i1} - X_{i2})]$ is invertible. Then there exists a zero crossing of (5.3), and with $\hat{\beta}_n = (\hat{\beta}_{1n}, \ldots, \hat{\beta}_{rn})$ being any such zero crossing,

$$n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \Sigma), \qquad (5.4)$$

where Σ is given by (5.7). □

We have been unable to prove this conjecture. We indicate how to prove this conjecture making an extra assumption on the distribution of

$(X_{11},\ldots,X_{r1})^t$ which does not seem unreasonable, and using a lemma which we believe to be true.

The extra assumption is that each of the 2^r orthants generated by independent differences of the r-tuple $(X_{11},\ldots,X_{r1})^t$ contain positive probability. We require that

$$P[\mathrm{sgn}(X_{i1}-X_{i2}) = e_i : i=1,\ldots,r] > 0, \ \forall\, e = (e_1,\ldots,e_r), \qquad (5.5)$$

where e_i can be either -1 or 1. From symmetry there are only 2^{r-1} quantities involved since $P[\mathrm{sgn}(X_{i1}-X_{i2})=e_i : i=1,\ldots,r] = P[\mathrm{sgn}(X_{i1}-X_{i2})= -e_i : i=1,\ldots,r]$. This says that the conditional distribution of the r^{th} sign given the other r-1 is not degenerate. In two dimensions this requirement is implied by the matrix A of (3.10) being invertible. In higher dimensions the corresponding A matrix being invertible does not imply (5.5).

The lemma involves the analogues of V_{1n} and V_{2n} considered in (3.2). Define 2^{r-1} U-statistics by

$$V_{kn}(b) = \binom{n}{2}^{-1} \Sigma_{C_k}\Sigma\ \mathrm{sgn}(X_{1i}-X_{1j})\mathrm{sgn}(V_i(b)-V_j(b)), \qquad (5.6)$$

where $C_k = \{(i,j): \mathrm{sgn}(X_{si}-X_{sij}) = e_{ks}$ or $\mathrm{sgn}\,(X_{si} - X_{sj}) = -e_{ks}$ for $s=1,\ldots,r\}$ and e_k ranges over all 2^{r-1} possibilities where $e_{k1} = 1$. Each of the V_{kn} are monotone in all r arguments and decreasing in b_1.

<u>Conjecture 5.2.</u> If the r U-statistics in (5.3) have a zero crossing at b, then the 2^{r-1} U-statistics given in (5.6) also have a zero crossing at b. □

This conjecture was simple to prove with $r = 2$. We have been unable to prove it with higher values for r. The examples we have looked at with $r = 3$ follow this conjecture. Assuming this allows us to prove Conjecture 5.1.

<u>Theorem 5.1.</u> Assume the conditions of Conjecture 5.1 hold, along with condition (5.5) and Conjecture 5.2. Then there exists a zero crossing of (5.3), and with $\hat{\beta}_n$ any such zero crossing, $n^{1/2}(\hat{\beta}_n - \beta) \to N(0,\Sigma)$ where Σ is given by (5.7). □

<u>Proof:</u> The proof is essentially the same as that of Theorem 2.1. First we show a solution exists. The idea is the same as the existence proof in Theorem 2.1. We show how to collapse the $(b_1,\ldots b_r)$-space into an r-1 dimensional region $(\hat{b}_1(b_2,\ldots,b_r),b_2,\ldots,b_r)$ on which each of the r-1 U-statistics U_{jn} inherits its monotonicity properties (i.e. U_{jn} is decreasing in b_j for all of the other arguments fixed). We then show that there exists points (b_{j1},b_{j2}) on $(\hat{b}_1,b_2,\ldots,b_r)$ for each U_{jn} with $U_{jn}(b_{j1}) \geq 0$, $U_{jn}(b_{j2}) \leq 0$. Repetition of this argument r-1 more times shows the existence of a solution.

The argument is as before. Now $V_i(b)-V_j(b)=0$ are hyperplanes of dimension r-1 and $(\hat{b}_1,b_2,\ldots,b_r)$ is a connected sequence of hyperplane segments. We show U_{2n} is decreasing in b_2 for fixed $b_3,\ldots,b_r$ in the r-1 dimensional space. For fixed $b_3,\ldots,b_r$, $(\hat{b}_1,b_2,\ldots,b_r)$ is a sequence of connected line segments in the (b_1,b_2) plane and the same argument as in the case $r = 2$ (using $W_{1n} = U_{1n} + U_{2n}$ and $W_{2n} = U_{1n} - U_{2n}$ in place of V_{1n} and V_{2n}) shows that U_{2n} is decreasing along $(\hat{b}_1, b_2, \ldots, b_r)$ for fixed $(b_3, \ldots, b_r)$. The same argument as for $r = 2$ also shows the existence of points b_{21} and b_{22} on $(\hat{b}_1, b_2, \ldots, b_r)$ with $U_{2n}(b_{21}) \geq 0$ and $U_{2n}(b_{22}) \leq 0$. This completes the existence proof.

Second we show that $n^{1/2-\epsilon}(\hat{\beta}_{sb} - \beta_s) \xrightarrow{P} 0$ for $s=1,\ldots,r$, $\epsilon > 0$, and any solution. It suffices to show $P[\hat{\beta}_{1n} < \beta_1 - tn^{-1/2+\epsilon}] \to 0$ for any fixed $t > 0$ by symmetry. Fix $\epsilon > 0$. Letting $v = tn^{-1/2+\epsilon}$,

$$
\begin{aligned}
P(\hat{\beta}_{1n} < \beta_1 - v) &\leq \sup_{b_2 \ldots b_r} P[V_{in}(\beta_1 - v, b_2, \ldots, b_r) \leq 0,\ i=1,\ldots,2^{r-1}] \\
&\leq \max_i \{ P[V_{in}(\beta_1 - v, \beta_2, \ldots, \beta_r) \leq 0] \}
\end{aligned}
$$

since in each of the 2^{r-1} orthants for $(b_2, \ldots, b_r)$ created by the axes $(b_j = \beta_j)$ we can bound the first probability on the right hand side by the second. For example

$$
\begin{aligned}
&\sup_{\substack{b_j \geq \beta_j \\ 2 \leq j \leq r}} P[V_{in}(\beta_1 - v, b_2, \ldots, b_r) \leq 0,\ i=1,\ldots,2^{r-1}] \\
&\qquad \leq \sup_{\substack{b_j \geq \beta_j \\ 2 \leq j \leq r}} P[V_{fn}(\beta_1 - v, b_2, \ldots, b_r) \leq 0] \\
&\qquad = P[V_{fn}(\beta_1 - v, \beta_2, \ldots, \beta_2) \leq 0],
\end{aligned}
$$

where V_{fn} is that V_{in} which is increasing in $b_2, \ldots, b_r$. Note that (5.5) guarantees that each of the V_{in} has the number of non-zero summands going to infinity and the same argument as that given in Section 3 shows that $P(\hat{\beta}_{in} < \beta - v) \to 0$ and hence $n^{1/2-\epsilon}(\hat{\beta}_{sn} - \beta_s) \xrightarrow{P} 0$.

The asymptotic distribution is found as in Section 3 using Lemma 5.3.

Lemma 5.3. Let $R_{sn}(b) = U_{sn}(b) - U_{sn}(\beta) - \mu_s(b-\beta)$, where $U_{sn}(b)$ is given in (5.3) and $\mu_s(b-\beta) = E[U_{sn}(b)]$. Then for $a > r/(2r+2)$

$$\sup_{\substack{0 \le |t_j| \le n^{-a} \\ 1 \le j, s \le r}} |R_{sn}(\beta+t)| = o_p(n^{-1/2}). \quad \square$$

The proof is given in Section 6.

Thus the asymptotic distribution of $\hat{\beta}_n$ is given in (5.4) with

$$\Sigma = A^{-1}BA^{-t}, \qquad (5.7)$$

where A is given in the statement of Theorem 5.1, and $B = 4/9(\rho_{ij})$ where $\rho_{ij} = 3\{4E[P(X_{i1} < X_{i2})P(X_{j1} < X_{j2})] - 1\}$. $\square$

6. Proof of Lemma 5.3. Lemma 3.2 is a special case of Lemma 5.3 with r = 2. The lemma is motivated by Lemma 5 of Bhattacharya, Chernoff, and Yang (1983). The idea is to first write the remainder as a linear combination of errors with respect to the V_{jn} U-statistics which are monotone in each argument. The sup of V_{jn} over a region can be bounded above by breaking the region into smaller regions over which V_{jn} varies negligibly and looking at the maximum over the centers of these subregions. The approximations needed are quite crude.

Proof of Lemma 5.3: Let $T_{sn}(b) = V_{sn}(\beta) - \nu_s(b-\beta)$, for $V_{sn}(b)$ given by (5.6) and $\nu_s(b-\beta) = E[V_{sn}(b)]$. Then

$$\sup |R_{sn}(\beta+t)| = \sup \left|\sum_{i=1}^{2^{r-1}} k_i T_{in}(\beta+t)\right|$$

$$\leq \sup | T_{sn}(\beta+t)| \cdot \sum_{i=1}^{2^{r-1}} |k_i|,$$

and since $|k_i| \leq 1$ it suffices to show $\sup |T_{sn}(\beta+t)| = o_p(n^{-1/2})$ where all the sups above are over the same region as stated in Lemma 5.3. For specificity we examine T_{1n} where V_{1n} is decreasing in each of its r arguments. The other T's are similar.

Divide the region $B = \prod_{j=1}^{r} (\beta_j - n^{-a}, \beta_j + n^{-a})$ into $(2z)^r$ subregions $\prod_{j=1}^{r} (\beta_j + i_j n^{-a} z^{-1}, \beta_j + (i_j+1)n^{-a}z^{-1})$ where $-z \leq i_j < z$ for $j = 1,\ldots,r$. Denote these regions by B_m; the centers by b_m^c; the lower corners (at which each component takes its smallest value) by b_m^l; and the upper corners by b_m^u; where m varies from 1 to $(2z)^r$. Then for $b \in B_m$

$$|T_{1n}(b)| \leq |T_{1n}(b_m^c)| + |V_{1n}(b)-V_{1n}(b_m^c)| + |\nu_1(b-\beta)-\nu_1(b_m^c-\beta)|$$

$$\leq |T_{1n}(b_m^c)| + [V_{1n}(b_m^l)-V_{1n}(b_m^u)] + E[V_{1n}(b_m^l)-V_{1n}(b_m^u)].$$

Letting $W_{nm} = V_{1n}(b_m^l) - V_{1n}(b_m^u)$,

$$P[\sup_{b \in B_m} |T_{1n}(b)|>4d] \leq P[|T_{1n}(b_m^c)|>d] + P[W_{nm}>2d] + 1[E(W_{nm})>d],$$

and if $E[W_{nm}] \leq d$,

$$\leq d^{-2}\{Var(T_{1n}(b_m^c)) + Var(W_{nm})\},$$

and

$$P[\sup_{b\in B_m} |T_{1n}(b)|>4d] \leq 2^r z^r d^{-2}\{\sup_m Var(T_{1n}(b_m^c)) + \sup_m Var(W_{nm})\}.$$

Hence if we can show Var $T_{1n}(b_m^c) \leq Kn^{-1-a}$, $Var(W_{nm}) \leq Kn^{-1-a}$, and $E(W_{nm}) \leq Kn^{-a}z^{-1}$; then by taking $d = 2Kn^{-1/2-\epsilon}$ and $z = n^{1/2-a+\epsilon}$ for small ϵ we will have proved Lemma 5.3.

In fact we show that for $b_1 = (b_{11},\ldots,b_{r1})^t$ and $b_2 = (b_{12},\ldots,b_{r2})^t$

$$E|V_{1n}(b_1)-V_{1n}(b_2)\ | \leq K \sup_j\{|b_{j1}-b_{j2}|\},$$

$$\text{and } Var(V_{1n}(b_1)-V_{1n}(b_2)+ \leq Kn^{-1} \sup_j\{|b_{j1}-b_{j2}|\}. \qquad (6.1)$$

Recalling V_{1n} is decreasing in all its arguments and letting D_{ij} be the event that $sgn(X_{si}-X_{sj})$ is constant in s,

$$V_{1n}(b_1)-V_{1n}(b_2) = \binom{n}{2}^{-1} \sum_{1\leq i<j\leq n}\sum h(Z_i,Z_j),$$

where $h(Z_i,Z_j) = 1(D_{ij})sgn(X_{1i}-X_{1j})[sgn(V_i(b_1)-V_j(b_1))-sgn(V_i(b_2)-V_j(b_2))]$. Standard U-statistic theory gives $nVar[V_{1n}(b_1)-V_{1n}(b_2)] \rightarrow 4Var(h_1(Z_1))$ where $h_1(Z_1) = E[h(Z_1,Z_2)|Z_1]$. Note $Var(h_1(Z_1)) \leq Var(h(Z_1,Z_2))$, since

$\mathrm{Var}[h(Z_1,Z_2)]-\mathrm{Var}[h_1(Z_1)] = E\{\mathrm{Var}[h(Z_1,Z_2)|Z_1]\} \geq 0$, and $|h| \leq 2$ so to show (6.1) it suffices to show $E[|h(Z_1,Z_2)|] \leq K \sup_j\{|b_{j1} - b_{j2}|\}$. Let $t_j = b_j-\beta_j$ and $X_j = (X_{1j},\ldots,X_{rj})^t$. Then, integrating with respect to v_1, $E[|h(Z_1,Z_2)|]$ equals

$$= 2E\{1(D_{ij})\mathrm{sgn}(X_{11}-X_{12})[|G(V_2+t_1(X_1-X_2))-G(V_2+t_2(X_1-X_2))|]\}$$
$$\leq 2r \sup_v \{g(v)\} \sup_j\{E[|X_{j1}-X_{j2}|]\} \sup_j\{|b_{j1}-b_{j2}|\},$$

and Lemma 5.3 is proved. □

Acknowledgement. I would like to thank my advisor P.K. Bhattacharya for suggesting this topic and for useful discussions.

References

Bhattacharya, P.K., Chernoff, H., and Yang, S.S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* 11 505-514.

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

Denker, M. and Keller, G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. verw. Gebiete* 64 505-522.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Mat. Statist.* 19 293-325.

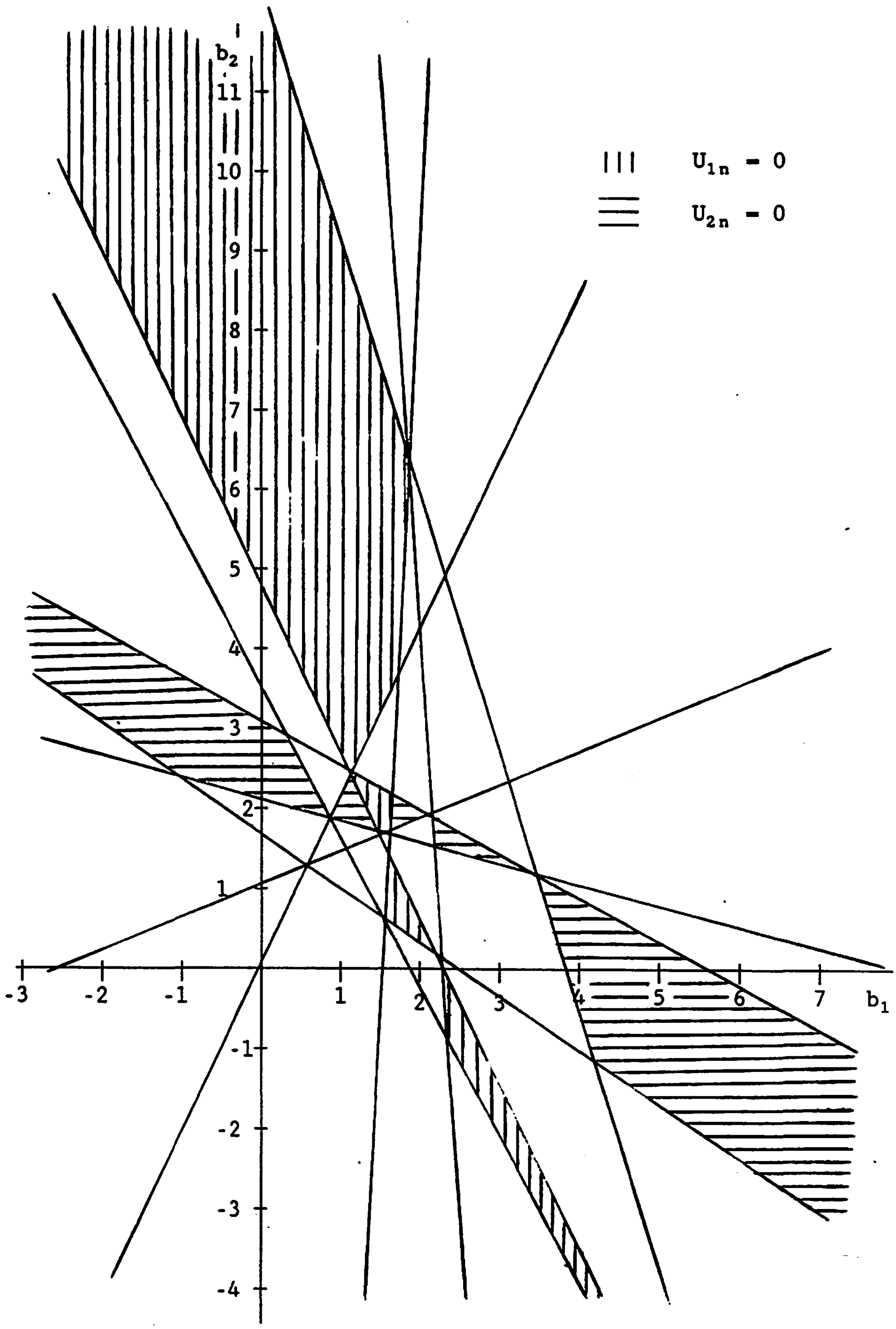
Kendall, M.G. (1970). *Rank Correlation Methods*. Griffin, London.

Pruitt. R.C. (1987). *Nonparametric Estimation of Autoregression and Multiple Regression Parameters*. Unpublished Ph.D. Thesis, University of California, Davis.

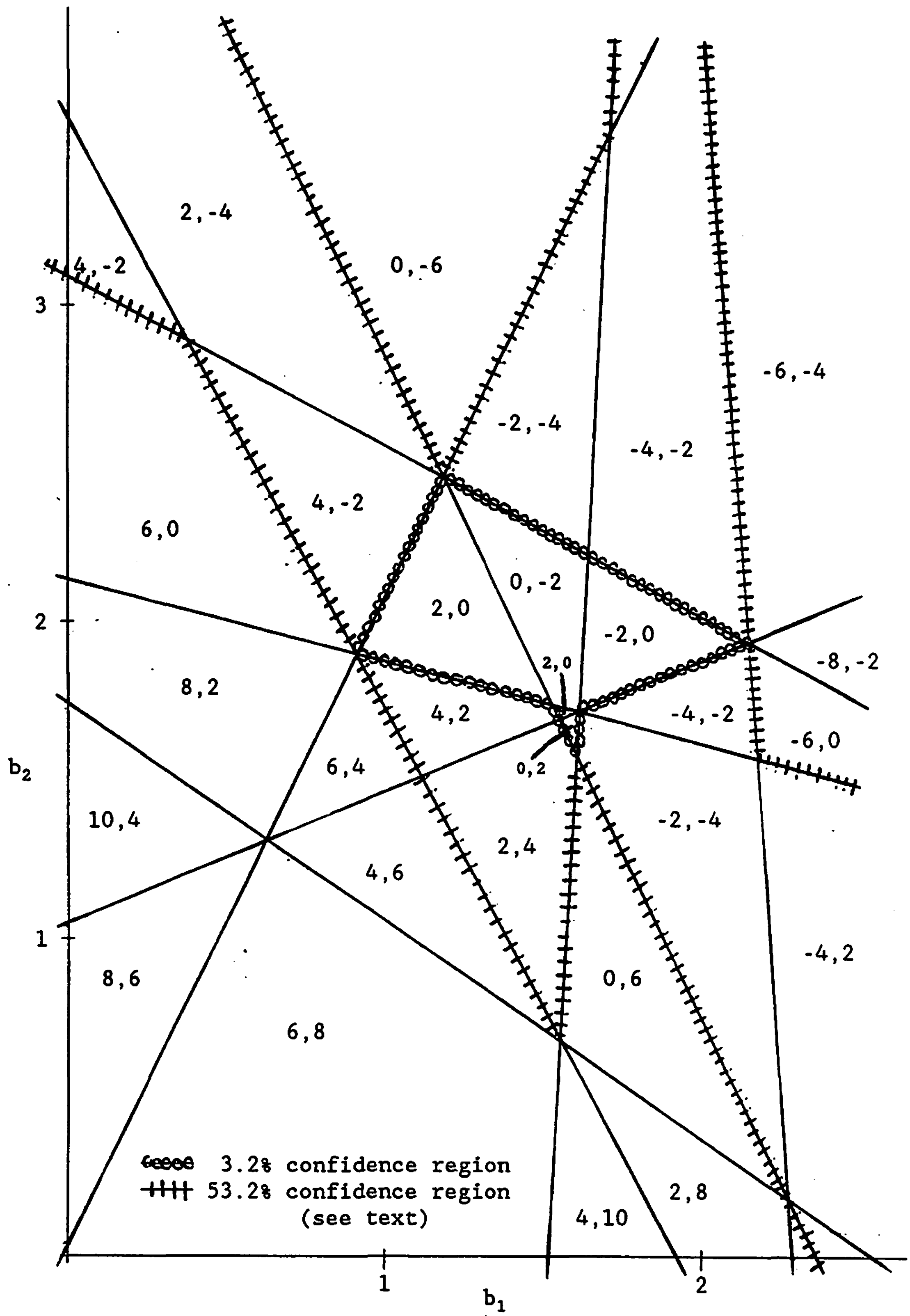
Puri, M.L. and Sen, P.K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

Sen, P.K. (1986). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* 63 1379-1389.

Theil, H. (1950). A rank invariant method of linear and polynomial regression analysis, I, II, III. *Nederl. Akad. Wetensch Proc.* 53 386-392, 521-525, 1397-1412.



<u>Figure 4.1.</u> The lines $V_i(b) - V_j(b) = 0$ for $1 \le i < j \le 5$.



<u>Figure 4.2.</u> Values of $10(U_{15}(\mathbf{b}), U_{25}(\mathbf{b}))$ for example.