The Annals of Statistics
1983, Vol. 11, No 2, 505–514

NONPARAMETRIC ESTIMATION OF THE SLOPE OF A TRUNCATED REGRESSION[1]

BY P. K. BHATTACHARYA, HERMAN CHERNOFF AND S. S. YANG

University of California, Davis; Massachusetts Institute of Technology; and Kansas State University

A nonparametric estimate β^* is presented for the slope of a regression line $Y = \beta_0 X + V$ subject to the truncation $Y \le y_0$. This model is relevant to a cosmological controversy which concerns Hubble's Law in Astronomy. The estimate β^* corresponds to the zero-crossing of a random function $S_n(\beta)$, which for each β is a Mann-Whitney type of statistic designed to measure heterogeneity among the calculated residuals $Y - \beta X$. The asymptotic distribution of β^* is derived making extensive use of U-statistics to show that $S_n(\beta_0)$ is asymptotically normal and then showing that $S_n(\beta)$ behaves like $S_n(\beta_0)$ plus a deterministic term which is locally linear. Results on asymptotic efficiency are compared with finite sample size results by simulation.

1. Introduction. We consider nonparametric estimates for the truncated regression model described as follows. Let $Y = \beta_0 X + V$ where V is independent of X and β_0 is positive. We observe (X, Y) only if $Y \le y_0$. On the basis of n independent observations (X_i, Y_i), $i = 1, \cdots, n$ it is desired to estimate β_0 and the distribution of V.

This problem is relevant to a current controversy in cosmology involving Hubble's Law and I. E. Segal's Chronometric Theory which predict different values for the slope β_0 of the straight line relating magnitude (negative log of luminosity) and log of velocity as measured by red shift, for distant celestial objects. It is generally agreed that the residual V, which represents *intrinsic luminosity*, is independent of red shift. However it is not agreed that the distribution of V, which is of interest in itself, is of any special form, nor even that it has finite second moments. Hence nonparametric methods are sought to estimate the slope β_0 and the distribution of V. The problem is complicated by the truncation due to the fact that objects of high magnitude are not visible, and hence unobserved.

We present a nonparametric Mann-Whitney type of estimate β^* of the slope β. This is a generalization, to the truncated case, of an estimate based on Kendall's tau, introduced by Theil (1950) and studied by Sen (1968). It may be described as follows. If the residuals $V_i(\beta) = Y_i - \beta X_i$ are calculated for β larger than the true value β_0, then the calculated residuals for small X_i will tend to be larger than for large X_i. The Theil procedure selects the estimate of β to balance the number of $V_i(\beta)$ greater than $V_j(\beta)$ for $X_i < X_j$ with the number of such $V_i(\beta)$ less than $V_j(\beta)$. In a modified version of this procedure considered by Adichie (1967) and Sievers (1978), the above comparisons between $V_i(\beta)$ and $V_j(\beta)$ for $X_i < X_j$ are weighted by $X_j - X_i$ (see also Lehmann, 1975, page 291). To adapt this modified version of the Theil procedure to the truncated case, observe that it is impossible for $V_j(\beta)$ to exceed $V_i(\beta)$ if $V_i(\beta) > y_0 - \beta X_j$, and so we call $V_i(\beta)$ and $V_j(\beta)$ *comparable* only if $V_i(\beta) \le y_0 - \beta X_j$. Our estimate β^* corresponds to the zero-crossing of $S_n(\beta)$ where $S_n(\beta)$ is the sum of the weights $\pm (X_j - X_i)$ applied to all comparable pairs, with the sign depending on whether $V_i(\beta) \le V_j(\beta)$ or not.

Received March 1981; revised January 1982.

[1] This paper is dedicated to Wassily Hoeffding who is responsible for the development of the theory of U-statistics which played a pivotal role in the derivations here and in the theory of Nonparametric Inference in general. It was prepared under Contract N00014-75-C-0555 (NR-042-331) for the Office of Naval Research.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62J05, 85A40.

Key words and phrases. Nonparametric, truncated regression, U-statistics, cosmology, Hubble's Law, chronometric theory, efficiency, simulation.

A nonparametric maximum likelihood estimate of the distribution of the residual V is also evaluated.

2. Asymptotic distribution of the estimate β^*. To state Theorem 1, which describes the asymptotic distribution of the estimate of β under regularity conditions, requires some notation. Our description of the sample observations and all our arguments are conditional on the X-values $x_1 < \cdots < x_m$ and their respective frequencies $n_1, \cdots, n_m$ among those observations (from a possibly larger set of data) which escaped truncation. Let (x_i, Y_{hi}), $h = 1, \cdots, n_i$ be the $n_i > 0$ observations corresponding to $X = x_i$ where $x_1 < x_2 < \cdots < x_m$ and $\sum n_i = n$. For $\beta > 0$, let

$$w(i, \beta) = y_0 - \beta x_i, \quad w_i = y_0 - \beta_0 x_i = w(i, \beta_0), \quad F_i = F(w_i),$$

$$V_h(i, \beta) = Y_{hi} - \beta x_i, \quad \text{and} \quad V_{hi} = Y_{hi} - \beta_0 x_i = V_h(i, \beta_0).$$

Then the *calculated residuals* $V_h(i, \beta)$ and $V_k(j, \beta)$ are *comparable* if $i < j$ and $V_h(i, \beta) \le w(j, \beta)$. If $V_h(i, \beta)$ and $V_k(j, \beta)$ are comparable, let

$$g_{hk}(i, j, \beta) = (x_j - x_i) \quad \text{for} \quad V_h(i, \beta) \le V_k(j, \beta)$$

and

$$g_{hk}(i, j, \beta) = -(x_j - x_i) \quad \text{for} \quad V_h(i, \beta) > V_k(j, \beta).$$

If $V_h(i, \beta)$ and $V_k(j, \beta)$ are not comparable, then $g_{hk}(i, j, \beta) = 0$. Let

$$S_n(\beta) = \textstyle\sum_{i<j} \sum_{h=1}^{n_i} \sum_{k=1}^{n_j} g_{hk}(i, j, \beta). \tag{2.1}$$

We estimate β in terms of the zero crossings of $S_n(\beta)$ which is a left-continuous function of β with ordinary (jump) discontinuities. We say that β is a *zero crossing* of S_n if the right and left-hand limits $S_n(\beta+)$ and $S_n(\beta-)$ do not both have the same sign, i.e., if $S_n(\beta+)S_n(\beta-) \le 0$. Generally, there may be several zero crossings of $S_n(\beta)$. Theorem 1 states that for β close to β_0 all zero crossings are very close to one another and are essentially equal. In fact it states that there exists a random variable β^* so that all zero crossings in some neighborhood of β_0 are within $o_p(n^{-1/2})$ of β^* and that $\sqrt{n}\,(\beta^* - \beta_0)$ has a limiting normal distribution.

Now let $H_n(x)$ be the sample cdf of X, f and F be the density and continuous cdf of $V = Y - \beta_0 X$, $w(x) = y_0 - \beta_0 x$, and

$$\sigma_n^2 = \frac{1}{3} \int \int \int (x - x')(x - x'') \frac{\tilde{F}^3}{FF'F''} \, dH_n(x)\, dH_n(x')\, dH_n(x'') \tag{2.2}$$

where F, F', F'', and $\tilde{F}$ are abbreviations for $F[w(x)]$, $F[w(x')]$, $F[w(x'')]$ and $\min(F, F', F'')$ respectively. Denote the mean and variance of H_n by $\mu(H_n)$ and $\sigma^2(H_n)$ and let $\gamma(H_n) = \int [\int |x - x'| \, dH_n(x')]^3 \, dH_n(x)$. Then $\gamma(H_n) \le \gamma^*(H_n) = \int \int |x - x'|^3 \, dH_n(x')\, dH_n(x)$. In view of the conditional nature of our arguments, as mentioned above, H_n is a fixed sequence.

Finally,

$$\alpha_n(\beta_0 + t) = \int_{x<x'} \frac{x' - x}{FF'} \left\{ 2 \int_{-\infty}^{w(x')} F(v - t(x' - x))\, dF(v) - F'F(w(x') - t(x' - x)) \right\} dH_n(x)\, dH_n(x') \tag{2.3}$$

and consequently

$$\alpha_n'(\beta_0 + t) = -\int_{x<x'} \frac{(x' - x)^2}{FF'} \left\{ 2 \int_{-\infty}^{w(x')} f(v - t(x' - x))\, dF(v) - F'f(w(x') - t(x' - x)) \right\} dH_n(x)\, dH_n(x'). \tag{2.4}$$

Compare $\alpha_n(\beta)$ with $\mu_n(\beta) = ES_n(\beta)$ given in Section 5 to see that $\alpha_n(\beta) = n^{-2}\mu_n(\beta)$. Note that σ_n and $\alpha_n(\beta)$ depend, implicitly, on β_0 through $w(x) = y_0 - \beta_0 x$. We may now state

THEOREM 1. *Under the regularity conditions* (i) *to* (v) *below, there is a random variable* β^* *such that*

$$\mathscr{L}[\sqrt{n}(\beta^* - \beta_0)/\tau_n] \to N(0, 1) \tag{2.5}$$

where

$$\tau_n = \sigma_n/\alpha_n'(\beta_0) \tag{2.6}$$

and each zero crossing of $S_n(\beta)$ *in* $(\beta_0 - t_0, \beta_0 + t_0)$ *is within* $o_p(n^{-1/2})$ *of* β^* *and there is at least one such zero crossing with probability approaching one. The conditions are* (i) $\gamma^*(H_n) = O(1)$, (ii) σ_n *is bounded away from* 0 *and* ∞, (iii) $\alpha_n'(\beta_0)$ *is bounded away from* 0, (iv) $\alpha_n(\beta_0 + t)$ *is bounded away from* 0 *for* $|t| \le t_0$ *and* t *bounded away from* 0, *and* (v) $f'(v)/F(y_0 - \beta_0 x)$ *and* $f(v)/F(y_0 - \beta_0 x)$ *are bounded on* $\{(x, v): v \le y_0 - \beta_0 x, 0 < H_n(x) < 1\}$.

Another estimate is determined by a variation of $S_n(\beta)$ where g_{hk} is $\pm$ 1 in place of $\pm (x_j - x_i)$. This variation, called the *unweighted* sum, leads to a similar theorem with β^* replaced by its analogue β^{**}. In that theorem, condition (i) is replaced by $\sigma^2(H_n) = O(1)$, the integral in σ_n^2 has the term $(x - x')(x - x'')$ replaced by its sign, the first factor $(x' - x)$ in $\alpha_n(\beta_0 + t)$ is removed, and $(x' - x)^2$ is replaced by $(x' - x)$ in $\alpha_n'(\beta_0 + t)$.

In the nontruncated case the asymptotic variances are relatively easy to calculate. One consequence, relevant to Table 3.1 of the efficiency computations in Section 3, where our examples were selected so that H_n is approximately uniform, is a result pointed out by Sen (1968). That is, the weighted and unweighted estimates have equal efficiency in the nontruncated case when H_n is uniform.

3. Information and efficiency. To compute the efficiency of the estimates of β, we require the Fisher information matrix for estimating $\theta = (\mu, \sigma, \beta)$ of the model $Y = \beta x + \mu + \sigma V$, $Y \le y_0$ where $\mathscr{L}(V) = F$ before truncation. The information matrix at $\theta = (0, 1, \beta_0)$ is

$$J(x) = \begin{bmatrix} I_0 - e_0^2 & I_1 - e_0 e_1 & x(I_0 - e_0^2) \\ I_1 - e_0 e_1 & I_2 - e_1^2 & x(I_1 - e_0 e_1) \\ x(I_0 - e_0^2) & x(I_1 - e_0 e_1) & x^2(I_0 - e_0^2) \end{bmatrix},$$

where

$$I_j = \int_{-\infty}^{w} v^j \left\{\frac{f'(v)}{f(v)}\right\}^2 \frac{f(v)}{F(w)}\, dv, \quad j = 0, 1, 2,$$

$w = y_0 - \beta_0 x$, $e_0 = f(w)/F(w)$, and $e_1 = we_0 - 1$, I_j and e_j depending on x through $w(x)$. If the average information based on all the observations, $\bar{J}_n = n^{-1} \sum J(x_i) = \int J(x)dH_n(x) \to J_0$, then $\mathscr{L}[\sqrt{n}(\hat{\beta}_n - \beta_0)] \to N(0, \sigma_0^2)$ where $\hat{\beta}_n$ is the maximum likelihood estimate of β, and σ_0^2 is the lower right diagonal element of J_0^{-1}.

Comparing σ_0^2 with τ_n^2 of Equation (2.6) one may compute the asymptotic efficiences of β^* and β^{**}. In Table 3.1 we tabulate these results for various values of σ and y_0 for the case where H_n is uniformly distributed at 10 equally spaced points between 0 and 1, and where the residuals V have the standard normal and Cauchy distributions.

When y_0 is close to 0.75 in the Cauchy case the efficiency is very low. This is due to the fact that $\alpha_n'(\beta_0)$ changes from negative to positive values. If we did not have truncation, $S_n(\beta)$ would be monotonic decreasing and this shift would not occur. As long as α_n' increases in magnitude, even though it has the "wrong" sign, the efficiency goes up, as for example when $y_0 = 0.5$. It should be remarked that under severe truncation, as in the cases

TABLE 3.1

Asymptotic Standard Deviation of $\hat{\beta}$ and Relative Efficiencies of β^* and β^{**} for the model $Y = \beta_0 x + \mu + \sigma V$, $Y \le y_0$, $(\mu, \beta_0) = (0, 1)$. e^* = efficiency of β^*; e^{**} = efficiency of β^{**}; $\hat{\beta}$ = m.l.e. of β_0, σ_0^2 = asymptotic variance of $\sqrt{n}\,(\hat{\beta} - \beta_0)$. All computations carried out on the assumption that the observed x values are uniformly distributed at $x = 0, 1/9, \cdots, 8/9, 1$.

σ	y_0	$\mathscr{L}(V)$ = Standard Normal			$\mathscr{L}(V)$ = Standard Cauchy		
		σ_0	e*	e**	σ_0	e*	e**
0.2	2.00	0.627	0.955	0.955	0.867	0.684	0.679
	1.25	0.668	0.932	0.946	0.869	0.679	0.673
	0.75	1.125	0.915	0.898	1.316	0.012	0.074
	0.50	1.841	0.877	0.835	1.682	0.086	0.018
0.5	2.00	1.607	0.941	0.947	2.136	0.698	0.695
	1.25	1.994	0.866	0.882	2.398	0.468	0.503
	0.75	2.817	0.847	0.839	3.601	0.006	0.004
	0.50	3.523	0.849	0.819	4.074	0.243	0.134
1.5	2.00	6.067	0.856	0.862	7.167	0.475	0.492
	1.25	6.994	0.816	0.818	9.385	0.151	0.185
	0.75	7.785	0.792	0.789	12.767	0.006	0.000
	0.50	8.234	0.782	0.776	14.574	0.146	0.097

where $y_0 = 0.5$ and σ_0 is small, there is a loss by truncation of vast amounts of (x, Y) points because of our choice of a uniform H_n for Table 3.1.

The asymptotic results of Table 3.1 were compared with finite sample size results by simulation. One hundred trials with sample sizes of $n = 100$ and $n = 30$ were observed for almost each set of parameters of Table 3.1. For a couple of cases $n = 900$ was taken. The astronomical catalogs have approximately 1200 galaxies but it is not clear which, if any, of them should be excluded. When asymptotic theory predicted standard deviations of β^* comparable to β_0, many of the trials gave poor results. These cases were usually accompanied by a diagnostic signal. That was that $S_n(\beta)$ was very ragged or had no zero crossings for any reasonable candidate for β. However, after censoring these bad trials, the remaining values of β^* fit the asymptotic distribution well.

Some of the results of this simulation appear in Table 3.2. There are partial lists of the average bias $\overline{\beta^* - \beta}$, and standard deviation s_{β^*} of the estimates to be compared with asymptotic theory. We list under "out" the number of trials in which zero crossings were difficult to obtain or β^* was an extreme outlier. In the presence of such outliers, $\overline{\beta^* - \beta}$ and s_{β^*} are not meaningful and are usually not presented.

4. Maximum likelihood estimate of distribution of residual. Let $v_1 > v_2 > \cdots > v_r$ be r distinct specified values at which it is desired to estimate F. For a given value of $X = x$, $w(x) = y_0 - \beta_0 x$ is the largest possible value of V. We assume that v_1 is less than $w_1 = \max_{1 \le i \le n} w(X_i)$. Clearly it is not possible to estimate $F(v_i)$ in this truncated problem but it should be possible to estimate $F(v_i)/F(w_1)$ for $i = 1, 2, \cdots, r$. In this section we shall describe the maximum-likelihood estimates of these ratios, assuming β_0 is known.

Merge the distinct values of $w(X_i)$ and the v_j and arrange them in decreasing order to obtain $w_1 > w_2 > \cdots > w_k$. We may regard each w_j as $y_0 - \beta_0 x_j$ for the value of x_j of X corresponding to which there exist $n_j \ge 0$ observations on Y and $V = Y - \beta_0 x_j$. Here $n_j = 0$ if w_j corresponds to one of the specified v values and there are no observations with $X = x_j = (y_0 - w_j)/\beta_0$. Let

$$\pi_i = F(w_{i+1})/F(w_i), \quad 1 \le i \le k - 1 \tag{4.1}$$

$$\theta_i = F(w_{i+1})/F(w_1) = \prod_{j=1}^{i} \pi_j, \quad 1 \le i \le k - 1, \ \theta_0 = 1, \tag{4.2}$$

and $N_i(v) = \#\{X = x_i, V \le v\}$, and $M_i(v) = \sum_{j=1}^{i} N_j(v) = \#\{X \le x_i, V \le v\}$ for $1 \le i \le k$, where $\#A$ represents the number of observations for which the event A is satisfied.

TABLE 3.2

Simulation of $n_e = 100$ trials of estimation of β^* for samples of size n for the model $Y = \beta x + \sigma V$, $Y \le y_0$, $\beta = 1$.

σ	y_0	n	$\mathscr{L}(V)$ = Normal				$\mathscr{L}(V)$ = Cauchy			
			$\overline{\beta^* - \beta}$	s_{β^*}	σ_{β^*}	out	$\overline{\beta^* - \beta}$	s_{β^*}	σ_{β^*}	out
0.2	2.00	100	0.001	0.063	0.064		−0.011	0.099	0.105	
		30	0.004	0.127	0.117		−0.044	0.234	0.191	
	1.25	900	—	—	0.023		−0.005	0.036	0.035	
		100	0.002	0.070	0.069		0.004	0.133	0.106	
		30	0.005	0.122	0.126		—	—	0.193	2
	0.75	100	0.001	0.120	0.118		—	—	1.202	6*
		30	0.026	0.263	0.215		—	—	2.195	8*
	0.50	100	0.038	0.225	0.197		—	—	0.573	11**
		30	0.116	1.331	0.359	1	—	—	1.046	15**
0.5	2.00	100	0.004	0.174	0.166		−0.024	0.281	0.256	
		30	−0.070	0.297	0.303		0.123	0.642	0.467	8
	1.25	900	—	—	0.071		−0.004	0.132	0.117	
		100	0.009	0.220	0.214		—	—	0.351	12
		30	0.020	0.465	0.391	2	—	—	0.640	20
	0.75	100	−0.011	0.282	0.306		—	—	4.592	15*
		30	0.320	1.752	0.559	2	—	—	8.383	11*
	0.50	100	0.114	0.430	0.382		—	—	0.827	17**
		30	0.338	1.448	0.698	1	—	—	1.510	15**
1.5	2.00	100	0.085	0.699	0.656	9	—	—	1.040	20
		30	0.082	0.973	1.198	20	—	—	1.899	29
	1.25	900	—	—	0.258	—	—	—	0.804	35
		100	−0.060	0.741	0.774	10	—	—	2.412	57
		30	0.475	1.620	1.413	21	—	—	4.404	—
	0.75	100	0.095	0.946	0.875	11	—	—	16.762	15*
		30	0.480	1.718	1.597	20	—	—	30.603	18*
	0.50	100	0.118	0.892	0.931	12	—	—	3.813	15**
		30	0.339	1.460	1.700	17	—	—	6.962	16**

* in out column represents $n_e = 25$ trials, ** in out column represents $n_e = 30$ trials, – means not computed. The x are distributed evenly among 0,1/9,2/9, $\cdots$, 1; σ_{β^*} is derived from asymptotic theory; $\overline{\beta^* - \beta}$ and s_{β^*} are based on samples of $n_e = 100$ trials. Out represents the number of outliers related to poor convergence. In most examples the presence of these outliers rendered $\overline{\beta^* - \beta}$ and s_{β^*} meaningless.

The likelihood function based on the $N_i(w_j)$ is

$$L = \prod_{i=1}^{k-1} \pi_i^{M_i(w_{i+1})}(1 - \pi_i)^{M_i(w_i) - M_i(w_{i+1})}.$$

Hence the maximum likelihood estimates are given by

$$\hat{\pi}_i = \frac{M_i(w_{i+1})}{M_i(w_i)}, \quad 1 \le i \le k - 1 \tag{4.3}$$

and

$$\hat{\theta}_i = \prod_{j=1}^{i} \frac{M_j(w_{j+1})}{M_j(w_j)}, \quad 1 \le i \le k - 1. \tag{4.4}$$

The information matrix $I(\pi)$ corresponding to the vector $\pi = (\pi_1, \cdots, \pi_{k-1})$ is the diagonal matrix whose ith diagonal element is

$$\begin{aligned} I_{ii}(\pi) &= E\left\{\frac{M_i(w_{i+1})}{\pi_i^2} + \frac{M_i(w_i) - M_i(w_{i+1})}{(1 - \pi_i)^2}\right\} \\ &= \frac{\theta_{i-1}}{\pi_i(1 - \pi_i)} \sum_{j=1}^{i} \frac{n_j}{\theta_{j-1}} = \frac{1}{\pi_i(1 - \pi_i)} \sum_{j=1}^{i} n_j \frac{F(w_i)}{F(w_j)}, \quad 1 \le i \le k - 1. \end{aligned}$$

The relationship between the θ_i and the π_i suggests that we introduce $\xi_i = \log \pi_i$ and

(4.5) $$\varphi_i = \log \theta_i = \sum_{j=1}^{i} \xi_j, \quad 0 \le i \le k-1.$$

The asymptotic variance of $\hat{\pi}_i$ is given by $I_{ii}^{-1}(\pi)$ and hence that of $\hat{\xi}_i = \log \hat{\pi}_i$ is simply $(1-\pi_i)/\pi_i a_i$, where

$$a_i = \sum_{j=1}^{i} n_j F(w_i)/F(w_j) = \sum_{j=1}^{i} n_j \theta_{i-1}/\theta_{j-1}.$$

Hence the asymptotic variance of φ_i is simply $\sum_{j=1}^{i}(1-\pi_i)/\pi_i a_i$.

5. Proof of Theorem 1, first part. The proof has two main parts. The first consists of a U-statistic type of decomposition to obtain

(5.1) $$S_n(\beta) = \mu_n(\beta) + \hat{S}_n(\beta) + R_n(\beta)$$

where $\mu_n(\beta) = ES_n(\beta)$ and $\hat{S}_n(\beta)$ may be expressed as a sum of independent terms with mean 0, i.e.,

(5.2) $$\hat{S}_n(\beta) = \sum_{i=1}^{m} \sum_{h=1}^{n_i} T_h(i, \beta).$$

We apply the Central Limit Theorem to $\hat{S}_n(\beta_0)$ and show that the remainder, $R_n(\beta_0)$, is relatively small, and hence $S_n(\beta_0)$ is approximately normally distributed with mean 0.

The second part consists of showing that for β close to β_0,

(5.3) $$D_n(\beta) = \{S_n(\beta) - \mu_n(\beta)\} - \{S_n(\beta_0) - \mu_n(\beta_0)\}$$

is negligible uniformly in β. Consequently $S_n(\beta)$ behaves like $S_n(\beta_0)$ plus a deterministic term which is locally linear. Thus the zero crossing of $S_n(\beta)$ can be approximated by β^*, a linear function of the approximately normal $S_n(\beta_0)$.

The first part requires some notation and lemmas. For $i < j$, let $\mu(i, j, \beta) = Eg_{hk}(i, j, \beta)$, $g^*_{hk}(i, j, \beta) = g_{hk}(i, j, \beta) - \mu(i, j, \beta)$,

(5.4) $$\underline{U}_h(i, j, \beta) = E\{g^*_{hk}(i, j, \beta) \mid V_{hi}\},$$

(5.5) $$\overline{U}_k(i, j, \beta) = E\{g^*_{hk}(i, j, \beta) \mid V_{kj}\},$$

and $R_{hk}(i, j, \beta) = g^*_{hk}(i, j, \beta) - \underline{U}_h(i, j, \beta) - \overline{U}_k(i, j, \beta)$. Then (5.1) and (5.2) hold with $\mu_n(\beta) = ES_n(\beta) = \sum_{i<j} n_i n_j \mu(i, j, \beta)$,

(5.6) $$T_h(i, \beta) = \sum_{j<i} n_j \overline{U}_h(j, i, \beta) + \sum_{j>i} n_j \underline{U}_h(i, j, \beta)$$

and $R_n(\beta) = \sum_{i<j} \sum_{h=1}^{n_i} \sum_{k=1}^{n_j} R_{hk}(i, j, \beta)$. Lemma 1 lists some of the U-statistic properties. We abbreviate by omitting some of the arguments in $g, \underline{U}, \overline{U}$, etc. The proof is moderately routine and is omitted.

LEMMA 1. $E\underline{U}_h = E\overline{U}_k = ER_{hk} = 0$,

$$\operatorname{Var} g_{hk} = \operatorname{Var} \underline{U}_h + \operatorname{Var} \overline{U}_k + \operatorname{Var} R_{hk}, \quad \operatorname{Var} S_n(\beta) = \operatorname{Var} \hat{S}_n(\beta) + \operatorname{Var} R_n(\beta),$$

$$\operatorname{Var} R_n(\beta) = \sum_{i<j} n_i n_j \operatorname{Var} R_{hk}(i, j, \beta), \quad \operatorname{Var} \hat{S}_n(\beta) = \sum_i n_i \operatorname{Var} T_h(i, \beta).$$

To apply the Central Limit Theorem, we require bounds on the moments of $\hat{S}_n(\beta)$ and $R_n(\beta)$. The Minkowski Inequality (Ash, 1972), states that $\{E|\sum Z_i|^r\}^{1/r} \le \sum[E|Z_i|^r]^{1/r}$ for $r \ge 1$, and implies

$$\operatorname{Var} T_h(i, \beta) \le 4[\sum_j n_j |x_j - x_i|]^2, \quad E|T_h(i, \beta)|^3 \le 8[\sum_j n_j |x_j - x_i|]^3.$$

Furthermore, using Lemma 1,

$$\operatorname{Var} R_n(\beta) \le 16 \sum_{i<j} n_i n_j (x_j - x_i)^2 = 16n \sum_{i=1}^{m} n_i (x_i - \bar{x})^2$$

where $\bar{x} = n^{-1} \sum n_i x_i$, and $n = \sum n_i$. Representing these sums in terms of integrals with respect to the empirical distribution function H_n of X, we have

LEMMA 2.

$$\operatorname{Var} \hat{S}_n(\beta) \le 4n^3 \int \left[\int |x' - x| \, dH_n(x') \right]^2 dH_n(x),$$

$$\operatorname{Var} \hat{S}_n(\beta) \le 4n^3 \iint (x' - x)^2 \, dH_n(x') \, dH_n(x) = 8n^3\sigma^2(H_n),$$

$$\textstyle\sum_{i=1}^{m} \sum_{h=1}^{n_i} E \,|\, T_h(i, \beta) \,|^3 \le 8n^4\gamma(H_n), \qquad \operatorname{Var} R_n(\beta) \le 16n^2\sigma^2(H_n).$$

We derive the asymptotic normality of $S_n(\beta_0)$ by showing that $\mu_n(\beta_0) = 0$, applying the Lyapounov conditions to $\hat{S}_n(\beta_0)$, and showing that $R_n(\beta_0)$ is negligible.

First, let $\beta = \beta_0 + t$. We have

$$\mu(i, j, \beta) = \frac{(x_j - x_i)}{F_i F_j} \left\{ 2 \int_{-\infty}^{w_j} F(v - t(x_j - x_i)) \, dF(v) - F(w_j - t(x_j - x_i)) F_j \right\}, \tag{5.7}$$

hence $\mu(i, j, \beta_0) = 0$, $\mu_n(\beta_0) = 0$ and $\mu_n(\beta) = n^2\alpha_n(\beta)$.

To study $\hat{S}_n(\beta_0)$ we obtain second and third moments of $T_h(i, \beta_0)$ and of $\underline{U}_h$ and $\overline{U}_k$. We have

$$\underline{U}_h(i, j, \beta_0) = \begin{cases} \dfrac{x_j - x_i}{F_j} \{F_j - 2F(V_{hi})\} & \text{if } V_{hi} \le w_j \\ 0 & \text{otherwise} \end{cases}$$

$$\overline{U}_k(i, j, \beta_0) = \frac{x_j - x_i}{F_i} \{2F(V_{kj}) - F_j\},$$

$$\operatorname{Var} \underline{U}_h(i, j, \beta_0) = (x_j - x_i)^2 F_j / 3F_i,$$

$$\operatorname{Var} \overline{U}_k(i, j, \beta_0) = (x_j - x_i)^2 F_j^2 / 3F_i^2.$$

If $i < j < \ell$, then

$$\operatorname{Cov}\{\underline{U}_h(i, j, \beta_0) \underline{U}_h(i, \ell, \beta_0)\} = (x_i - x_j)(x_i - x_\ell) F_\ell^2 / 3F_i F_j,$$

$$\operatorname{Cov}\{\overline{U}_k(i, \ell, \beta_0) \overline{U}_k(j, \ell, \beta_0)\} = (x_\ell - x_i)(x_\ell - x_j) F_\ell^2 / 3F_i F_j,$$

$$\operatorname{Cov}\{\overline{U}_r(i, j, \beta_0) \underline{U}_r(j, \ell, \beta_0)\} = (x_j - x_i)(x_j - x_\ell) F_\ell^2 / 3F_i F_j.$$

Combining these terms we may compute the variance of $T_h(i, \beta_0)$ and from that the variance of $\hat{S}_n(\beta_0)$.

$$\operatorname{Var} \hat{S}_n(\beta_0) \doteq n^3\sigma_n^2. \tag{5.8}$$

Lemma 2 bounds the sum of the third moments of $T_h(i, \beta_0)$. Hence we may apply the Lyapounov version of the Central Limit Theorem (Cramér, 1951) to obtain:

LEMMA 3. *If σ_n is bounded away from* 0, *and $\gamma(H_n) = O(1)$, then*

$$\mathscr{L}[\hat{S}_n(\beta_0)/n^{3/2}\sigma_n] \to N(0, 1).$$

From Lemma 2 it follows that for any specified β, $R_n(\beta) = O_p(n)$ as long as $\sigma^2(H_n) = O(1)$. Thus, we have:

LEMMA 4. *If $\sigma^2(H_n) = O(1)$ and σ_n is bounded away from* 0 *and* ∞, *and $\gamma(H_n) = O(1)$, then*

$$\mathscr{L}[S_n(\beta_0)/n^{3/2}\sigma_n] \to N(0, 1).$$

The same derivation could be used to determine the asymptotic normality of $S_n(\beta)$ for $\beta \neq \beta_0$ or even for a sequence of $\beta_n \to \beta_0$. Then the result could be applied to determine the Pitman efficiency of the test statistic $S_n(\beta_0)$ for testing $\beta = \beta_0$.

A variation of $S_n(\beta)$ is the *unweighted* statistic where the $(x_j - x_i)$ weights are omitted in the comparisons to be summed. The same derivations apply to give analogues of Lemmas 3 and 4 where σ_n is replaced by an integral which differs from (2.2) in two related respects. First, the quadratic terms $(x - x')(x - x'')$ is replaced by its sign. Second, the region of integration is restricted to $A = \{(x, x', x''): x \neq x' \text{ and } x \neq x''\}$. The set of points thus excluded may be important if H_n is such that some value of x has a substantial proportion of the observations (as is the case in our example of Section 3).

Lemma 4 may be used to construct an asymptotic confidence interval for β_0 using the set of β for which $|S_n(\beta)/n^{3/2}\sigma_n(\beta)| \leq K$. The construction of this interval requires knowledge of F which can be estimated. We have not seriously addressed the problem of the simultaneous estimation of F and β. Note, however, that our proposed point estimate of β does not involve knowledge of F.

6. Proof of Theorem 1, second part. The second part of the proof of Theorem 1 consists of showing that $D_n(\beta) = \{S_n(\beta) - \mu_n(\beta)\} - \{S_n(\beta_0) - \mu_n(\beta_0)\}$ is uniformly negligible for β close to β_0. If we can neglect $D_n(\beta)$, then

$$S_n(\beta) = \mu_n(\beta) - \mu_n(\beta_0) + S_n(\beta_0) + D_n(\beta) \approx (\beta - \beta_0)\mu_n'(\beta_0) + S_n(\beta_0)$$

will vanish near

$$\beta^* = \beta_0 - S_n(\beta_0)/\mu_n'(\beta_0) \tag{6.1}$$

if $S_n(\beta_0)/\mu_n'(\beta_0) = n^{-1/2}\{n^{-3/2}S_n(\beta_0)/\sigma_n\}\{\sigma_n/\alpha_n'(\beta_0)\}$ is small. Moreover, Lemma 4 implies that $\mathscr{L}[n^{1/2}(\beta^* - \beta_0)/\tau_n] \to N(0, 1)$ with $\tau_n = \sigma_n/\alpha_n'(\beta_0)$, if σ_n and $\alpha_n'(\beta_0)$ are well behaved.

What constitutes negligibility on the part of $D_n(\beta)$? It can be neglected if it is small compared to $S_n(\beta_0) = O_p(n^{3/2})$ or if it is small compared to $(\beta - \beta_0)\mu_n'(\beta_0) = O(n^2(\beta - \beta_0))$. But neither of these bounds applies for the entire interval $\beta_0 \pm t_0$. However, Lemma 5 states conditions under which $D_n(\beta) = o_p(n^{-3/2})$ uniformly for $|\beta - \beta_0| \leq n^{-a}$ and $D_n(\beta)/|\beta - \beta_0| = o_p(n^2)$ uniformly for $n^{-b} \leq |\beta - \beta_0| \leq t_0$ where $0 < a < b$. With this Lemma stated below, Theorem 1 can be established using a reasonably routine analysis where o_p and O_p are used with properties analogous to those of o and O (Chernoff, 1956; Pratt, 1959).

LEMMA 5. *If* $|S_n(\beta_1') - S_n(\beta_2')| \leq W_n(\beta_1, \beta_2)$ *for* $\beta_0 - t_0 \leq \beta_1 \leq \beta_1' \leq \beta_2' \leq \beta_2 \leq \beta_0 + t_0$ *where* $EW_n(\beta_1, \beta_2) \leq K_1|\beta_2 - \beta_1|n^2$, $\operatorname{Var} W_n(\beta_1, \beta_2) \leq K_2|\beta_2 - \beta_1|n^3$, *and* $\operatorname{Var} D_n(\beta) \leq K_2|\beta - \beta_0|n^3$ *for* $|\beta - \beta_0| \leq t_0$, *then for* $a > 1/4$ *and* $0 < b < 1/3$,

$$\sup_{0 \leq |t| \leq n^{-a}} |D_n(\beta_0 + t)| = o_p(n^{3/2}) \tag{6.2}$$

and

$$\sup_{n^{-b} \leq |t| \leq t_0} |t^{-1}D_n(\beta_0 + t)| = o_p(n^2). \tag{6.3}$$

PROOF. For $0 \leq t_1 \leq t_0$, divide the interval $(\beta_0 - t_1, \beta_0 + t_1)$ into $2r$ subintervals $(\beta_0 + it_1/r, \beta_0 + (i + 1)t_1/r) = (\beta_i, \beta_{i+1})$, $-r \leq i < r$. Then if $\beta_i \leq \beta \leq \beta_{i+1}$,

$$D_n(\beta) = D_n(\beta_i) + \{\mu_n(\beta_i) - \mu_n(\beta)\} + \{S_n(\beta) - S_n(\beta_i)\}$$

$$|D_n(\beta)| \leq |D_n(\beta_i)| + EW_n(\beta_i, \beta_{i+1}) + W_n(\beta_i, \beta_{i+1}).$$

Suppose $d \geq K_1n^2t_1/r$. Then $EW_n(\beta_i, \beta_{i+1}) \leq d$ and by the Chebychev Inequality

$$P[W_n(\beta_i, \beta_{i+1}) \geq 2d] \leq K_2n^3t_1/rd^2$$

and

$$P[|D_n(\beta_i)| \geq d] \leq K_2 n^3 t_1/d^2.$$

Hence

(6.4) $$P[\sup_{|t|\leq t_1} |D_n(\beta_0 + t)| \geq 4d] \leq 4rK_2 n^3 t_1/d^2.$$

In Equation (6.4) substitute $t_1 = n^{-a}$, $r \sim n^c$, and $d = n^{2-a-c}$ where $a > 0$, $c > 0$, $a + c > 1/2$, and $a + 3c < 1$. For $a > 1/4$, these inequalities may be achieved and (6.2) follows. Now let $t_1 = t_0$, $r \sim n^e$, $d = n^{2-e}$ where $0 < b < e < 1/3$. Then (6.4) implies (6.3). □

With Lemma 5, only two additional steps are required to establish Theorem 1. First, we must demonstrate the existence of W_n satisfying the conditions of Lemma 5. Second, the informal argument of the first two paragraphs of this section should be formalized. We shall omit this step which is long, but reasonably routine, and refer simply to the more complete discussion in Bhattacharya, Chernoff and Yang (1980). All that remains is the existence of W_n, which is covererd in:

LEMMA 6. *If $\gamma^*(H_n)$ and the densities of V_{hi} (i.e., f/F_i) are uniformly bounded, then the conditions of Lemma 5 are satisfied.*

PROOF. We recall that $S_n(\beta) = \sum g_{hk}(i, j, \beta)$. Then as β increases from β_1' to β_2', $g_{hk}(i, j, \beta)$ might stay the same or decrease if $V_h(i, \beta)$ remains comparable to $V_k(j, \beta)$. But if $V_h(i, \beta)$ fails to remain comparable to $V_k(j, \beta)$, $g_{hk}(i, j, \beta)$ may increase from $-(x_j - x_i)$ to 0. Hence $|g_{hk}(i, j, \beta_1') - g_{hk}(i, j, \beta_2')| \leq g_{hk}^1(i, j, \beta_1, \beta_2)$, where

$$g_{hk}^1(i, j, \beta_1, \beta_2) = 2(x_j - x_i)(A + B)$$

and A and B are one or zero depending on whether or not $-\beta_2(x_j - x_i) \leq V_{hi} - w_j - \beta_0(x_j - x_i) \leq -\beta_1(x_j - x_i)$ and on whether or not $-\beta_2(x_j - x_i) \leq V_{hi} - V_{kj} - \beta_0(x_j - x_i) \leq -\beta_1(x_j - x_i)$ respectively. Also $|S_n(\beta_1') - S_n(\beta_2')| \leq W_n(\beta_1, \beta_2)$ where

(6.5) $$W_n(\beta_1, \beta_2) = \textstyle\sum_{i<j} g_{hk}^1(i, j, \beta_1, \beta_2).$$

The g_{hk}^1 terms may be analyzed precisely as were the g_{hk} terms in Section 5 giving rise to g_{hk}^{1*}, $\underline{U}_h^1$, $\overline{U}_k^1$, R_{hk}^1, μ_n^1, $\hat{S}_n^1$, T_h^1 and R_n^1 terms to which Lemma 1 applies.

Since the densities of the V_{hi} are uniformly bounded, there is a constant K such that

$$Eg_{hk}^1(i, j, \beta_1, \beta_2) \leq 2(x_j - x_i)^2 K |\beta_2 - \beta_1|$$

$$\text{Var } g_{hk}^1(i, j, \beta_1, \beta_2) \leq 4(x_j - x_i)^3 K |\beta_2 - \beta_1|,$$

and since $\text{Var } g_{hk}^1 = \text{Var } \underline{U}_h^1 + \text{Var } \overline{U}_k^1 + \text{Var } R_{hk}^1$, the terms on the right are bounded by $\text{Var } g_{hk}^1$. Then, applying the Minkowski Inequality

$$\text{Var } T_h^1 \leq \{\textstyle\sum_j n_j \cdot 2 |x_j - x_{i'}|^{3/2}(K|\beta_2 - \beta_1|)^{1/2}\}^2.$$

Moreover,

$$\text{Var } R_n^1 \leq 4 \textstyle\sum_{i<j} n_i n_j (x_j - x_i)^3 K |\beta_2 - \beta_1|,$$

$$EW_n(\beta_1, \beta_2) \leq 2K|\beta_2 - \beta_1| \textstyle\sum_{i<j} (x_j - x_i)^2 n_i n_j = 2Kn^2 |\beta_2 - \beta_1| \sigma^2(H_n)$$

and

$$\text{Var } W_n(\beta_1, \beta_2) \leq 2K(2n^3 + n^2)|\beta_2 - \beta_1| \gamma^*(H_n).$$

Hence we have established the bounds on the mean and variance of W_n. Essentially the same derivation yields the bound on the variance of $D_n(\beta)$. Finally noting that $\sigma^2(H_n) \leq \{\gamma^*(H_n)/2\}^{2/3}$, we have completed our proof.

REFERENCES

ADICHIE, J. N. (1967). Estimates of regression parameters based on rank tests. *Ann. Math. Statist.* **38** 894–904.

ASH, R. B. (1972). *Real Analysis and Probability.* Academic, New York.

BHATTACHARYA, P. K., CHERNOFF, H. and YANG, S. S. (1980). Nonparametric estimation of the slope of a truncated regression. Tech. Report No. 18, Massachusetts Institute of Technology.

CHERNOFF, H. (1956). Large sample theory, parametric case. *Ann. Math. Statist.* **27** 1–22.

CRAMÉR, H. (1951). *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

LEHMANN, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks.* Holden-Day, San Francisco.

PRATT, J. W. (1959). On a general concept of "In Probability." *Ann. Math. Statist.* **30** 549–558.

SEN, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* **63** 1379–1389.

SIEVERS, G. L. (1978). Weighted rank statistics for simple linear regression. *J. Amer. Statist. Assoc.* **73** 628–631.

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

P. K. BHATTACHARYA
DIVISION OF STATISTICS
UNIVERSITY OF CALIFORNIA
DAVIS, CALIFORNIA 95616

HERMANN CHERNOFF
STATISTICS CENTER E40-107
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139

S. S. YANG
DEPARTMENT OF STATISTICS
KANSAS STATE UNIVERSITY
MANAHATTAN, KANSAS 66506