The Annals of Statistics
1996, Vol. 24, No. 4, 1586–1601

REGRESSION RANK SCORES ESTIMATION IN ANOCOVA

BY PRANAB KUMAR SEN

University of North Carolina, Chapel Hill

In semiparametric ANOCOVA (mixed-effects) models, the role of regression rank scores in robust estimation of fixed-effects parameters as well as covariate regression functionals is critically appraised, and the relevant asymptotic theory is presented.

1. Introduction. To motivate semiparametric analysis of covariance (ANOCOVA) models, consider first the conventional model where for the ith observation, Y_i, $\mathbf{Z}_t$ and $\mathbf{t}_i$ stand for the primary, (stochastic) concomitant and (nonstochastic) design variates, respectively, and, conditional on $\mathbf{Z}_i = \mathbf{z}_i$,

$$Y_i = \boldsymbol{\beta}'\mathbf{t}_i + \boldsymbol{\gamma}'\mathbf{z}_i + e_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the regression parameter vectors for the fixed and random effects components, and the e_i are independent and identically distributed random variables (i.i.d.r.v.'s) having a normal distribution with mean 0 and a finite (conditional) variance σ^2. The $\mathbf{t}_i$ are given p-vectors not all equal, $\mathbf{T} = (\mathbf{t}_1, \dots, \mathbf{t}_n)'$ and the $\mathbf{Z}_i$ are stochastic q-vectors, so that there are $p + q$ regression parameters and an additional scale parameter σ. The assumed joint normality of $(\mathbf{Z}_i, e_i)$ yields homoscedasticity, linearity of regression as well as normality of the conditional distribution in (1.1). Without this joint normality, a breakdown may occur in each of these three basic postulations. On the other hand, the design vectors may still pertain to a linear regression function. Thus, there is a need to examine thoroughly the robustness aspects of mixed-effects models with due emphasis on all these factors. Motivated by these considerations, we consider the following two semiparametric models.

MODEL 1. Assume that (1.1) holds and the distribution function (d.f.) $G(e|\mathbf{z})$ of e given $\mathbf{Z} = \mathbf{z}$ is independent of $\mathbf{z}$ and is continuous almost everywhere (a.e.). Thus, only the normality part of the basic assumption is dropped.

MODEL 2. The linearity of the regression of Y on $\mathbf{Z}$ is dropped, and further, the independence of $\mathbf{Z}$ and e is relaxed, so that the conditional d.f. $G(\cdot|\mathbf{z})$ may no longer be homoscedastic.

Received August 1993; revised August 1995.

AMS 1991 *subject classifications*. 62G05, 62J05.

Key words and phrases. Asymptotic efficiency and representation, covariate regression functional, heteroscedasticity, regression quantiles, R-estimators, semiparametric models, stochastic predictors, weak invariance principles.

The main motivation for the second model is that, without the joint normality of $(\mathbf{Z}, e)$, there may not be enough justification for a linear regression of Y on $\mathbf{Z}$ or even the homoscedasticity. As an illustrative example, consider the following: let the conditional density of e, given $\mathbf{Z} = \mathbf{z}$, be $(1 - \eta)\phi(x/\sigma_0) + (\eta/h(\mathbf{z}))\phi(x/(h(\mathbf{z})\sigma_0))$, where $\eta(> 0)$ is small and $h(\cdot)$ is a nonconstant, nonnegative function, and let the marginal density of $\mathbf{Z}$ be a similar contaminated normal law. For this mixture (outliers in covariates as well as primary variate) model, one can only appeal to Model 2. In general, for outliers with elliptically symmetrical d.f.'s, Model 2 works out well.

In line with the usual nonparametric setup [namely Puri and Sen (1985), Chapter 8], we may assume that the $\mathbf{Z}_i$ are i.i.d.r.v.'s having a q-variate d.f. $F(\mathbf{z})$ and let

$$G_i(y|\mathbf{z}) = G(y - \boldsymbol{\beta}'\mathbf{t}_i|\mathbf{z}), \qquad i = 1, \dots, n, \tag{1.2}$$

where $G(\cdot)$ is continuous and quite arbitrary (as in Model 2). Thus, (1.2) conforms to a parametric form with respect to the fixed-effects parameters but to a nonparametric one for the concomitant variates. In order to quantify further this model in terms of suitable regression functionals, we define $\theta(G(\cdot|\mathbf{z}))$, a *translation-invariant* functional of the conditional d.f. G, such that $\theta(G(\cdot - c|\mathbf{z})) = \theta(G(\cdot|\mathbf{z})) - c$ for every real c. Here $G(y - c|\mathbf{z})$ stands for the d.f. of the translated variable $Y_i - c$. This allows the percolation of the linear component $\boldsymbol{\beta}'\mathbf{t}_i$ in a natural way. As such, we consider the following quasi-parametric model:

$$\theta(G_i(\cdot|\mathbf{z})) = \theta(G(\cdot|\mathbf{z})) + \boldsymbol{\beta}'\mathbf{t}_i, \qquad i = 1, \dots, n. \tag{1.3}$$

Thus, with respect to ANOCOVA Model 2, one has a finite-dimensional parameter $\boldsymbol{\beta}$ for the design variates but a functional with respect to the concomitant variates. It is in this generality, we like to study a general class of robust estimators of the functionals in (1.3).

It may be quite natural for us to allow G to be a member of a broad class, so that *robustness* has to be interpreted in a *global* sense (rather than in a *local* sense where M-estimators are more appropriate). This leads us to the preference of the so-called R-estimators (and related rank statistics) which are *scale-equivariant* and possess other asymptotic optimality properties. Due to significant contributions of Gutenbrunner and Jurečková (1992) and Jurečková (1992), the *regression rank scores* (RRS) estimators have emerged as strong competitors of the classical R-estimators, and certain asymptotic equivalence results have also been obtained by Jurečková and Sen (1993). As such, we shall find it more convenient to deal with such RRS estimators. To motivate the setup, along with the preliminary notions of RRS in Section 2, Model 1 is briefly introduced in Section 3. The main results of this study are reported in Sections 4 and 5 where Model 2 is treated in its full generality. The concluding section deals with some general remarks to facilitate further the use of such RRS in ANOCOVA models.

2. Preliminary notions. For the usual *linear model*

$$\mathbf{Y} = (Y_1, \dots, Y_n)' = \mathbf{X}_n^0 \boldsymbol{\beta}^0 + \mathbf{e}, \qquad \mathbf{e} = (e_1, \dots, e_n)', \tag{2.1}$$

where $\mathbf{X}^0$ is an $n \times r$ matrix of known regression constants (with $r = p + q$) and $\boldsymbol{\beta}^0$ is an r-vector of unknown regression parameters. Koenker and Bassett (1978) formulated *regression quantiles* as follows. Let $\alpha \in (0, 1)$, the ith row of $\mathbf{X}^0$ be denoted by $\mathbf{x}_i^0$, $i = 1, \dots, n$, and let

$$\rho_\alpha(x) = |x|\{\alpha I(x > 0) + (1 - \alpha) I(x < 0)\}, \qquad x \in \mathbf{R}. \tag{2.2}$$

Then the α-regression quantile of $\boldsymbol{\beta}^0$ is expressed as

$$\hat{\boldsymbol{\beta}}_n^0(\alpha) = \arg\min\left\{ \sum_{i=1}^n \rho_\alpha\big(Y_i - \mathbf{x}_i^0 \mathbf{b}\big) \colon \mathbf{b} \in \mathbf{R}^r \right\}. \tag{2.3}$$

Gutenbrunner and Jurečková (1992) considered a set $\hat{a}_{ni}(\alpha)$, $i = 1, \dots, n$, of scores, termed the RRS, which are linked to the solution in (2.3) by

$$\begin{aligned} \mathbf{Y}' \hat{\mathbf{a}}_n(\alpha) = \max, \qquad & \mathbf{X}_n^{0\prime} \hat{\mathbf{a}}_n(\alpha) = (1 - \alpha) \mathbf{X}_n^{0\prime} \mathbf{1}_n, \\ \hat{\mathbf{a}}_n \in [0, 1]^n, \qquad & \alpha \in (0, 1). \end{aligned} \tag{2.4}$$

These RRS are *regression-invariant* in the sense that, for every $\mathbf{b} \in \mathbf{R}^r$, $\hat{\mathbf{a}}_n(\alpha, \mathbf{Y} + \mathbf{X}_n^0 \mathbf{b}) = \hat{\mathbf{a}}_n(\alpha, \mathbf{Y})$, and this plays a basic role in the developments on RRS in the recent past. Statistical inference based on RRS generally involves (linear) LRRS statistics. We define these by taking a nondecreasing and square-integrable *score generating function* ϕ: $(0, 1) \to \mathbf{R}$ and letting

$$\hat{b}_{ni} = -\int_0^1 \phi(\alpha)\, d\hat{a}_{ni}(\alpha), \qquad i = 1, \dots, n. \tag{2.5}$$

Then LRRS statistics are the usual linear rank statistics with the scores given by (2.5). We may set, without any loss of generality, $\bar{\phi} = \int_0^1 \phi(t)\, dt = 0$, so that $n^{-1} \sum_{i=1}^n \hat{b}_{ni} = \bar{\phi} = 0$.

Let us consider the RRS estimation of $\boldsymbol{\beta}$ in (1.1) treating $\boldsymbol{\gamma}$ as nuisance parameter. Jurečková (1992) incorporated the dispersion measure

$$D_n(\mathbf{u}) = \sum_{i=1}^n (Y_i - \mathbf{t}_i' \mathbf{u}) \big[\hat{b}_{ni}(\mathbf{Y} - \mathbf{T}\mathbf{u}) - \bar{\phi} \big], \tag{2.6}$$

and proposed the estimator

$$\tilde{\boldsymbol{\beta}}_n(\phi) = \arg\min\{ D_n(\mathbf{u}) \colon \mathbf{u} \in \mathbf{R}^p \}. \tag{2.7}$$

Such RRS estimators coincide with R-estimators when there is no nuisance parameter, and hence, for the estimation of the intercept θ, she proposed the use of signed rank statistics. The regularity conditions for such RRS estimators are mostly adopted from Jurečková (1992) and Jurečková and Sen (1996), and will not be restated for brevity of presentation; only modifications will be introduced in the respective contexts.

3. ANOCOVA Model 1. Recall that here e_i and $\mathbf{Z}_i$ are independent, so that the conditional d.f. $G(e|\mathbf{z})$ is independent of $\mathbf{z}$. This enables us to reproduce virtually the basic techniques in Gutenbrunner and Jurečková (1992) and Jurečková (1992) to construct RRS estimators of $\boldsymbol{\beta}$. Since the $\mathbf{Z}_i$ are stochastic, there is a need to modify their regularity assumptions to suit working with the conditional distributions, given the $\mathbf{Z}_i$. Generally, in the literature, it is assumed that the d.f. $F(\cdot)$ of the $\mathbf{Z}_i$ has a compact support, which, in turn, ensures that the $\mathbf{Z}_i$ all belong to a compact set, with probability 1; this leads to the fulfillment of all the needed regularity conditions. However, such a compact support is not necessary, and it suffices to assume that

$$E\|\mathbf{Z}\|^r \text{ exists} \quad \text{for some } r > r_0 > 4, \tag{3.1}$$

where r_0 may depend on other regularity conditions. Without any loss of generality, we may set $\sum_{i=1}^n \mathbf{t}_i = \mathbf{0}$ and set $\mathbf{Q}_{ntt} = n^{-1}\sum_{i=1}^n \mathbf{t}_i\mathbf{t}_i'$. Also, let $\mathbf{Q}_{nzz}$ and $\mathbf{Q}_{ntz}$ be the sample dispersion matrix of the $\mathbf{Z}_i$ and the covariance matrix of the $\mathbf{Z}_i$ and the $\mathbf{t}_i$. Then, under (3.1), it follows that $\mathbf{Q}_{ntz}$ converges a.s. to $\mathbf{0}$ and $\mathbf{Q}_{nzz}$ to $\boldsymbol{\Sigma}_z$ as $n \to \infty$, where $\boldsymbol{\Sigma}_z$ is assumed to be positive definite. As such, as $n \to \infty$, $Q_{ntz}\mathbf{Q}_{nzz}^{-1}\mathbf{Q}_{nzt} \to \mathbf{0}$ a.s. Therefore, whenever $\lim_{n\to\infty} \mathbf{Q}_{ntt} = \mathbf{Q}_{tt}$ exists, we have $\mathbf{Q}_{ntt:z} = \mathbf{Q}_{ntt} - \mathbf{Q}_{ntz}\mathbf{Q}_{nzz}^{-1}\mathbf{Q}_{nzt} \to \mathbf{Q}_{tt}$, a.s. as $n \to \infty$. Hence, we may virtually repeat the proof of Theorem 1 and Lemma 1 of Jurečková (1992) with adaptations from Ghosh and Sen (1971) to justify the a.s. convergence of the conditional setup, and conclude that, as $n \to \infty$,

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}}_n(\phi) - \boldsymbol{\beta}\right) = n^{-1/2}\kappa^{-1}\mathbf{Q}_{tt}^{-1}\sum_{i=1}^n \mathbf{t}_i\phi(G(e_i)) + o_p(1), \tag{3.2}$$

where $\kappa = \int_0^1 g(G^{-1}(\alpha))\,d\phi(\alpha)$ is assumed to be positive and finite. This first-order asymptotic representation yields that

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}}_n(\phi) - \boldsymbol{\beta}\right) \text{ is asymptotically } \mathcal{N}\left(\mathbf{0}, \kappa^{-2}\mathbf{Q}_{tt}^{-1}A_\phi^2\right), \tag{3.3}$$

where $A_\phi^2 = \|\phi\|^2 = \int_0^1 \phi^2(u)\,du - \bar{\phi}^2$. In view of the fact that the conditional density $g(e|\mathbf{z})$ does not depend on $\mathbf{z}$, we may note that (3.3) is in complete agreement with Theorem 1 of Jurečková (1992). Hence, for ANOCOVA Model 1, RRS estimators work out well, and they possess the same asymptotic properties as in the ANOVA model. This picture is different in Model 2, and we mainly focus on that study.

4. ANOCOVA Model 2. As has been explained earlier, here we have a partial linear model where the fixed-effects parameter $\boldsymbol{\beta}$ provides a finite-dimensional linear setup, while the possible nonlinearity of the regression of Y on $\mathbf{Z}$ and nonhomogeneity of the conditional d.f. $G(\cdot|\mathbf{z})$ call for more general functionals in (1.3) for which *local smoothness* conditions appear to be more appropriate than *global linearity* or other parametric forms. On the other hand, the covariates are i.i.d.r.v.'s, and hence the marginal d.f.'s of the Y_i possess adequate information to provide the customary $n^{-1/2}$ rate of convergence for the estimators of $\boldsymbol{\beta}$. The rate of convergence for the estimators of

$\boldsymbol{\theta}(G(\cdot|\mathbf{z}))$ is presumably slower even if $q = 1$. To handle this differential picture, we proceed as follows.

First, by an appeal to the conventional ANOVA model (covariates ignored and without the normality of the errors), Jurečková's (1992) methodology is incorporated in deriving some RRS estimators of $\boldsymbol{\beta}$. Second, these estimators are incorporated in formulating suitable residuals (eliminating the fixed effects) which are to be used in a nonparametric (smooth) estimation of $\theta(G(\cdot|\mathbf{z}))$. Since such residuals may no longer be independent or identically distributed, there are some complications arising in this functional estimation problem. This problem is handled through an alternative approach to estimating conditional functionals based on perturbed observations, and this study is relegated to Section 5. Third, we may note that the ANOVA model based estimators of $\boldsymbol{\beta}$ do not utilize any information contained in the covariates and hence may not be fully efficient, although they still have the $n^{-1/2}$ rate of convergence. In the concluding section, we therefore discuss some possibilities to jack up the asymptotic efficiency of such estimators. In the rest of this section, we formulate and justify these ANOVA model based estimators for ANOCOVA Model 2.

Let $H_i(y)$ be the marginal d.f. of Y_i, so that, by (1.2),

$$H_i(y) = \int \cdots \int G_i(y|\mathbf{z})\, dF(\mathbf{z}) = \int \cdots \int G(y - \boldsymbol{\beta}'\mathbf{t}_i|\mathbf{z})\, dF(\mathbf{z}), \quad i = 1, \dots, n,\ y \in \mathbf{R}. \tag{4.1}$$

Thus, $H_i(y + \boldsymbol{\beta}'\mathbf{t}_i) = \int \cdots \int B(y|\mathbf{z})\, dF(\mathbf{z}) = H(y)$, $y \in \mathbf{R}$, $i = 1, \dots, n$. Hence

$$H_i(y) = H(y - \boldsymbol{\beta}'\mathbf{t}_i), \qquad i = 1, \dots, n,\ y \in \mathbf{R}. \tag{4.2}$$

This enables us to write formally

$$Y_i = \boldsymbol{\beta}'\mathbf{t}_i + \eta_i, \qquad i = 1, \dots, n, \tag{4.3}$$

where the η_i are i.i.d.r.v.'s with the d.f. $H(\cdot)$ defined above. This is the conventional nonparametric ANOVA model where the covariates are ignored. Thus, the η_i may as well be regarded as consisting of two random components: $\theta(G(\cdot|\mathbf{Z}_i))$ and the (conditional) error e_i. For this reason, under (1.1), the η_i will have generally larger dispersion than the e_i, and hence estimators of $\boldsymbol{\beta}$ based on this ANOVA model may have generally larger dispersion than the ones based on (1.1). On the other hand, if (1.1) does not hold, $\theta(G(\cdot|\mathbf{Z}_i))$ becomes a functional (infinite dimensional), and its estimation entails a slower rate of convergence. In this section, we present the ANOVA model in (4.3) side by side with ANOCOVA Model 2 and study the relative efficiency picture.

As in (2.7) we consider an RRS estimator $\tilde{\boldsymbol{\beta}}_n(\phi)$ and proceed as in Section 3 with adaptations from Jurečková (1992). Then we have

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\phi) - \boldsymbol{\beta}\right) = n^{-1/2}\kappa_0^{-1}\mathbf{Q}_{tt}^{-1} \sum_{i=1}^{n} \mathbf{t}_i \phi(H(\eta_i)) + o_p(1), \tag{4.4}$$

where $\kappa_0 = \int_0^1 h(H^{-1}(\alpha))\,d\phi(\alpha)$ and $h(y) = H'(y)$ is the marginal density of η. As a result, parallel to (3.3), here we have

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n(\phi) - \boldsymbol{\beta}\right) \to_{\mathscr{D}} \mathscr{N}\left(\mathbf{0}, \kappa_0^{-2} A_\phi^2 \mathbf{Q}_{tt}^{-1}\right). \tag{4.5}$$

Comparing (3.3) and (4.5), we conclude that for the ANOCOVA model in (1.1), the *asymptotic relative efficiency* (ARE) of $\hat{\boldsymbol{\beta}}_n(\phi)$ with respect to $\tilde{\boldsymbol{\beta}}_n(\phi)$ is given by

$$e\left(\hat{\boldsymbol{\beta}}_\phi; \tilde{\boldsymbol{\beta}}_\phi\right) = \kappa_0^2/\kappa^2, \tag{4.6}$$

which depends on the score function ϕ and the densities $g(\cdot)$ and $h(\cdot)$, but not on the design matrix $\mathbf{Q}_{tt}$. Keeping (1.1) in mind, we may write $W_i = \boldsymbol{\gamma}'\mathbf{Z}_i$, center them and assume that they are i.i.d.r.v.'s. Then, under (1.1), we have $\eta_i = W_i + e_i$, $i = 1, \dots, n$, where, conditionally on $W_i = w$, e_i has the d.f. $G(e)$ with density $g(e)$, and we denote the marginal d.f. and density of W by $F^*(w)$ and $f^*(w)$, respectively. In order to examine the convolution H without imposing the normality on either G or F^*, we make the following specific assumption:

(4.7) Both F^* and G are unimodal and symmetric about 0.

It follows by some routine steps that $h(0) \le g(0)$ and

(4.8) $G \prec_r H$, i.e., G is r-ordered with respect to H,

where we refer to Doksum (1969) for such ordering of d.f.'s. Therefore,

(4.9) $H^{-1}(G(x)) - x$ is nondecreasing,

so that G is tail-ordered with respect to H. The last ordering, in turn, implies that $(d/dx)H^{-1}(G(x))$ is greater than or equal to 1 for every $x \in \mathbf{R}$, so that

$$h\left(H^{-1}(t)\right) \le g\left(G^{-1}(t)\right) \quad \text{for every } t \in (0,1). \tag{4.10}$$

This tail-ordering property of convolutions of symmetric unimodal distributions (not necessarily of the same functional form) provides us with the desired clue. Note that, by definition,

$$\kappa_0/\kappa = \left\{\int_0^1 h\left(H^{-1}(t)\right) d\phi(t)\right\} \Big/ \left\{\int_0^1 g\left(G^{-1}(t)\right) d\phi(t)\right\}. \tag{4.11}$$

Therefore, whenever the score function $\phi(t)$ is monotone and square integrable and the d.f.'s F^* and G both satisfy (4.7), we have, by (4.10) and (4.11),

(4.12) $\kappa_0^2/\kappa^2 \le 1$, with equality holding only when F^* is degenerate at 0.

Let us study further the ARE in (4.6) when the density g possesses a finite *Fisher information* $I(g) = \|\psi_g\|^2 = \int_0^1 \{-g'(G^{-1}(t))/g(G^{-1}(t))\}^2\,dt$, where $\psi_g(u) = -g'(G^{-1}(u))/g(G^{-1}(u))$, $u \in (0,1)$. Defining ψ_h in a similar manner, we have, by a theorem in Hájek and Šidák (1967), page 17,

$$I(h) = \|\psi_h\|^2 \le I(g) < \infty, \tag{4.13}$$

where equality holds only when F^* is degenerate at 0. Thus, $\kappa_0/\kappa = \{\langle \phi, \psi_h \rangle\}/\{\langle \phi, \psi_g \rangle\}$. As the score functions $\phi(\cdot)$, $\psi_h(\cdot)$ and $\psi_g(\cdot)$ all belong to the L_2-space, we may consider the projections $\psi_h(\cdot) = \psi_{hg}\psi_g(\cdot) + (\psi_h(\cdot) - \psi_{hg}\psi_g(\cdot))$, $\phi(\cdot) = \delta_g \psi_g(\cdot) + \delta_{hg}[\psi_h(\cdot) - \gamma_{hg}(\cdot)\psi_g(\cdot)] + \phi(\cdot) - \delta_g \psi_g(\cdot) - \delta_{hg}[\psi_h(\cdot) - \gamma_{hg}\psi_g(\cdot)]$, where the components are mutually orthogonal and

$$\gamma_{hg} = \langle \psi_h, \psi_g \rangle / \langle \psi_g, \psi_g \rangle, \qquad \delta_g = \langle \phi, \psi_g \rangle / \langle \psi_g, \psi_g \rangle, \tag{4.14}$$

$$\delta_{hg} = \langle \phi, \psi_h - \gamma_{hg}\psi_g \rangle / \langle \psi_h - \gamma_{hg}\psi_g, \psi_h - \gamma_{hg}\psi_g \rangle \quad (= 0 \text{ if } \psi_h = \gamma_{hg}\psi_g). \tag{4.15}$$

Note that $\phi(\cdot)$ is *partially concordant/discordant* to ψ_h, given ψ_g, if δ_{hg} is $>(<)0$. It is *projectable* on ψ_g if $\delta_{hg} = 0$. Further, note that $\gamma_{hg}^2 = \rho^2(\psi_g, \psi_h)\{I(h)/I(g)\}$, where $\rho^2(\psi_g, \psi_h) = \langle \psi_g, \psi_h \rangle^2 / \{\|\psi_g\| \, \|\psi_h\|\}^2$ is less than or equal to 1 with the equality sign holding only when $\delta_{hg} = 0$.

THEOREM 4.1. *For a monotone ϕ, whenever $I(g)$ is $< \infty$,*

$$\kappa_0^2/\kappa^2 \text{ is } >, =, < \gamma_{hg}^2, \quad \textit{according as } \delta_{hg} \textit{ is } >, =, < 0. \tag{4.16}$$

For brevity of presentation, the proof is omitted [see Sen (1993a)].

REMARK. In the normal theory ANOCOVA model, both g and f^* are normal densities, so that h is also normal with scale parameter $\sigma_h \geq \sigma_g$, the scale parameter of g. In this case $I(h) = \sigma_h^{-2}$, $I(g) = \sigma_g^{-2}$ and $\psi_h(\cdot) = (\sigma_g/\sigma_h)\psi_g(\cdot)$, so that $\rho(\psi_h, \psi_g) = 1$. Therefore, the ARE of the ANOVA estimators with respect to the ANOCOVA estimators is given by $(\sigma_g/\sigma_h)^2 (\leq 1)$. In general, $\rho(\psi_h, \psi_g) = 1$ when

$$h(H^{-1}(u)) = \{I(h)/I(g)\}^{1/2} g(G^{-1}(u)), \qquad u \in (0,1). \tag{4.17}$$

If g and h both have the same functional form (with possibly different scale parameters), then (4.17) holds. If g and f^* both have infinitely divisible characteristic functions for which the canonical representations differ only by scale factors, then (4.17) holds. But such a characterization does not include most of the common nonnormal densities where g and f^* may not have a common functional form. Thus, a possible departure from normality of either g or f^* may distort (4.17), and hence the ANOVA model based analysis may become comparatively inefficient relative to ANOCOVA Model 1.

Let us consider next the case where linearity or the homoscedasticity condition may not hold. As before, letting $\theta(G(\cdot|\mathbf{z})) = \theta(\mathbf{z})$, we may note that, by virtue of our centering, then $\theta(\mathbf{Z}_i) = W_i$ are i.i.d.r.v.'s having a d.f. (say) $F^*(w)$, where we assume that the W_i are centered, too. We denote the conditional d.f. of e, given $W = w$, by $G_w(e)$, $e \in \mathbf{R}$. The functional form of G_w may depend on w in a rather arbitrary manner. As a natural generalization of (4.7), we assume that the following holds:

(i) f^* is unimodal and symmetric about 0,
(ii) for each $w \in \mathbf{R}$, $g_w(\cdot)$ is unimodal and symmetric about 0, and
(iii) for each $y \in \mathbf{R}^+$, $G_w(y - w) + G_{-w}(y + w)$ is nonincreasing in $w \in \mathbf{R}^+$.

Let us elucidate condition (iii) a bit more. By reference to the usual heteroscedasticity of the η_i, conditioned on the W_i, we may set

$$g_w(e) = \sigma_w^{-1} g_0(e/\sigma_w), \qquad \sigma_w = \sigma_{-w}(>0), \qquad w, e \in \mathbf{R}. \tag{4.18}$$

Under homoscedasticity, $\sigma_w = \sigma_0$ for all $w \in \mathbf{R}$. On the other hand, under heteroscedasticity, we assume that $\sigma_w \geq \sigma_0$ for all w, so that the variability is a minimum at the center value of w. Suppose now that the σ_w satisfy the following *growth condition*:

$$w\left[(\partial/\partial w)\log \sigma_w\right] \leq 1 \quad \text{for every } w \geq 0. \tag{4.19}$$

Now (4.18) implies that

$$1 - (w - y)\sigma_w^{-1}(\partial/\partial w)\sigma_w \geq 1 + (w + y)\sigma_w^{-1}(\partial/\partial w)\sigma_w$$

for every $w, y \in \mathbf{R}^+$, while assumption (ii) implies that $g_0((y - w)/\sigma_w) \geq g_0((y + w)/\sigma_w)$ for every $w, y \in \mathbf{R}^+$. Therefore, we may conclude that (iii) holds. Note that these are sufficient but not necessary conditions for (iii) to hold. Nevertheless, they bring the relevance of heteroscedastic errors in ANOCOVA models, and (iii) is in a sense a representation for this. Then, parallel to (4.9), we conclude that, under the above conditions, G_0 is tail-ordered with respect to H, so that $h(H^{-1}(t)) \leq g_0(G_0^{-1}(t))$ for every $t \in (0, 1)$.

Note that this tail-ordering of G_0 and H can readily be incorporated in extending Theorem 4.1 to the general ANOCOVA Model 2, where we need to replace the homogeneous density $g(\cdot)$ by the dominating one $g_0(\cdot)$ (and ψ_g accordingly). In this setup, if we stick to the linearity of $\theta(\mathbf{z})$, but may not like to impose the homogeneity of the $g_w(\cdot)$, as has been done earlier, then this extension covers the RRS estimation of $\boldsymbol{\beta}$ for which such an extended Theorem 4.1 would be applicable. This covers the so-called *heteroscedastic linear models*. However, as has already been noticed earlier, in a general ANOCOVA model, the linearity of the regression on the covariates may not be universally true, and the very way the $\theta(\mathbf{z})$ have been introduced, we are faced with *nonparametric estimation* of such conditional functions. Since such estimators have typically slower rates of convergence, first, in Section 5, we treat $\boldsymbol{\beta}$ as a nuisance parameter and estimate $\theta(\mathbf{z})$, and then, in the concluding section, we present a method of improving the ANOVA model based estimators of $\boldsymbol{\beta}$ for a general ANOCOVA model.

5. Estimation of covariate functionals. Our primary goal is to provide nonparametric estimators of $\theta(\mathbf{z}) = \theta(G(\cdot|\mathbf{z}))$, treating $\boldsymbol{\beta}$ as a nuisance parameter. As in Section 3, denote the RRS estimator of $\boldsymbol{\beta}$ by $\tilde{\boldsymbol{\beta}}_n = \tilde{\boldsymbol{\beta}}_n(\phi)$, and consider the aligned observations:

$$\hat{Y}_{ni} = Y_i - \tilde{\boldsymbol{\beta}}_n' \mathbf{t}_i \quad \text{for } i = 1, \dots, n. \tag{5.1}$$

Recall that the residuals in (5.1) have perturbations of the order $n^{-1/2}$, even for the ANOCOVA model under consideration. Next, consider the set of stochastic $(q+1)$-vectors:

$$(5.2) \qquad \left(\hat{Y}_{ni}, \mathbf{Z}'_i\right), \qquad i = 1, \dots, n.$$

We intend to incorporate (5.2) in the estimation of $\theta(\mathbf{z})$. Among the possibilities of adopting a *kernel* and a *nearest-neighborhood* method of smoothing, we will pursue the latter one for its comparative simplicity and relatively wider scope of applicability for moderately large sample sizes. Both of these methods share the common asymptotic properties [namely Sen (1993b)] and hence, apparently, there is no need to present them side by side.

A basic consideration in this estimation problem is the identification of the stochastic nature of the concomitant variates $\mathbf{Z}_i$, and, in view of that, it is necessary to formulate the functional $\theta(\mathbf{z})$, allowing $\mathbf{z}$ to vary over an appropriate domain. Choosing a fixed set of points $\mathbf{z}_j$, $j = 1, \dots, k$, may not serve the estimation purpose well, and hence a stochastic process formulation appears to be far more rational. We therefore look into the problem of estimating $\theta(\mathbf{z})$ when $\mathbf{z}$ lies in a compact set $\mathscr{C}$, embedded in the essential support of the distribution of the concomitant vector $\mathbf{Z}$. Often, such a compactification can be achieved by means of suitable transformations on $\mathbf{z}$. However, in practice, practical considerations mostly guide us to the choice of such a compact $\mathscr{C}$, and, further, compactification may not be that crucial. Thus, we confine ourselves to the estimation of the *functional process*:

$$(5.3) \qquad \Theta(\mathscr{C}) = \left\{\theta\left(G(\cdot|\mathbf{z})\right) : \mathbf{z} \in \mathscr{C}\right\} \quad \text{for some compact } \mathscr{C}.$$

The nearest-neighborhood method can be readily adopted for the estimation of the functional in (5.3). The only point which merits a careful study in this context is the technical difficulty arising from the facts that the $\hat{Y}_{ni}$ in (5.1) are not independent or identically distributed, so that the usual asymptotic theory of *k-NN* methods developed earlier by Bhattacharya and Gangopadhyay (1990) and Gangopadhyay and Sen (1992, 1993) and others, may not be directly adoptable. In the current context certain uniform convergence results on perturbed empirical distributions provide the access to the desired asymptotics.

To motivate this approach, consider first the case of i.i.d.r.v.'s $(Y_i^0, \mathbf{Z}'_i)$, $i = 1, \dots, n$, where $Y_i^0 = Y_i - \boldsymbol{\beta}'\mathbf{t}_i$, $i = 1, \dots, n$, so that the Y_i^0 are not affected by the regressors $\mathbf{t}_i$ and are i.i.d. Consider a specific point $\mathbf{z}_0 \in \mathscr{C}$, and conceive of a suitable metric $\rho(\mathbf{z}, \mathbf{z}_0)$: $\mathbf{R}^q \times \mathbf{R}^q \to \mathbf{R}^+$. In particular, one may choose a quadratic norm for $\rho(\cdot)$. Consider then the nonnegative r.v.'s:

$$(5.4) \qquad D_i = \rho(\mathbf{Z}_i, \mathbf{z}_0), \qquad i = 1, \dots, n.$$

Note that the D_i are i.i.d.r.v.'s with a d.f. dependent on $\mathbf{z}_0$, $\rho(\cdot)$ and the d.f. on $\mathbf{Z}_i$. In a *k-NN* method, corresponding to the sequence $\{n\}$ of sample size, one may choose a nondecreasing sequence $\{k_n\}$, such that $n^{-1}k_n$ is nonincreasing in n. Typically, we set

$$k_n \sim a[n^{4/(q+4)}] \quad \text{for some } a \in \mathbf{R}^+. \tag{5.5}$$

Further, let $D_{n:1} \le \cdots \le D_{n:n}$ be the *order statistics* corresponding to the D_i defined by (5.4); by virtue of the assumed continuity of the d.f. of D, ties among the D_i can be neglected with probability 1. Then we define a *k-NN* (stochastic) neighborhood of $\mathbf{z}_0$ by

$$\operatorname{nhd}(\mathbf{z}_0; \rho, k_n) = \{\mathbf{z} \in \mathscr{C}: \rho(\mathbf{z}, \mathbf{z}_0) \le D_{n:k_n}\}. \tag{5.6}$$

Now, corresponding to the pivot $\mathbf{z}_0$, we define the *antiranks* $S_1, \dots, S_n$, by letting $D_{n:i} = D_{S_i}$, $i = 1, \dots, n$. Consider then the set of observations

$$\left(Y_{S_i}^0, \mathbf{Z}_S'\right), \qquad i = 1, \dots, k_n. \tag{5.7}$$

The *k-NN* empirical d.f. of the Y_i^0 with respect to the pivot $\mathbf{z}_0$ and metric $\rho(\cdot)$ is defined as

$$G_{n,k_n}(y) = k_n^{-1} \sum_{i=1}^{k_n} I\left(Y_{S_i}^0 \le y\right), \qquad y \in \mathbf{R}. \tag{5.8}$$

Let us denote the conditional d.f. on Y^0, given $\mathbf{Z} = \mathbf{z}$, by $G^0(y|\mathbf{z})$, $\mathbf{z} \in \mathscr{C}$ and $y \in \mathbf{R}$. Then we are interested in functionals $\theta(G^0(\cdot|\mathbf{z}_0))$ of the d.f. G^0 at various $\mathbf{z}_0 \in \mathscr{C}$. Typically, $\theta(\cdot)$ is a location parameter of the d.f. G^0, and hence, based on robustness considerations, we propose to use M-, L- and R-functionals. Gangopadhyay and Sen (1992, 1993) have considered some general asymptotics relating to two broad classes of such functionals, namely, (i) $\theta(\cdot)$ is a conditional quantile functional [namely $\theta(G^0(\cdot|\ \mathbf{z}_0)) = \inf\{y\colon G^0(y|\mathbf{z}_0) \ge \alpha\} = (G^0)^{-1}(\alpha)$, where $\alpha \in (0,1)$, and (ii) $\theta(\cdot)$ is a (typically, *Hadamard* or compact differentiable statistical functional. For the case of $q = 1$ (i.e., scalar Z_i), $\mathscr{C}$ is a compact interval on $\mathbf{R}$, and they have studied the asymptotics for stochastic processes relating to the estimator of $\Theta(\mathscr{C})$. Sen (1994) has a general treatment of this for $q \ge 1$. As such, we are naturally tempted to use a functional of the corresponding empirical d.f. in (5.8) as the desired estimator. We let

$$T_n(\mathbf{z}_0) = \theta\left(G_{n,k_n}^0\right), \qquad \mathbf{z}_0 \in \mathscr{C}. \tag{5.9}$$

In passing, we may remark that the empirical d.f. in (5.8) is specifically based on the pivot $\mathbf{z}_0$ in the sense that the antiranks S_i and hence the set of observations in (5.7) are all geared to a specific choice of $\mathbf{z}_0$. As $\mathbf{z}_0$ varies over $\mathscr{C}$, they also vary, resulting in different empirical d.f.'s. Based on this construction, we consider the stochastic process:

$$W_n(\mathbf{z}) = \sqrt{k_n}\{T_n(\mathbf{z}) - \theta(\mathbf{z})\}, \qquad \mathbf{z} \in \mathscr{C}. \tag{5.10}$$

The weak convergence of W_n to a Gaussian random function on $\mathscr{C}$ follows from general results of Gangopadhyay and Sen (1993) and Sen (1994). The interesting feature of (5.10) is that, if k_n satisfies (5.5), the asymptotic bias of

$W_n(\mathbf{z})$ may not be typically equal to 0, and hence the drift function of such a Gaussian function may not be null. On the other hand, if we choose $k_n = o(n^{4/(q+4)})$, then this asymptotic bias is 0, and hence the drift function is also null.

Let us now consider a compact $\mathscr{K} \in \mathbf{R}^p$, containing $\mathbf{0}$ as an inner point, and let

$$Y_{ni}^0(\mathbf{b}) = Y_i^0 - n^{-1/2}\mathbf{b}'\mathbf{t}_i, \qquad i = 1, \dots, n, \mathbf{b} \in \mathscr{K}. \tag{5.11}$$

We introduce the same notation as in (5.4) through (5.6), and, as in (5.7), we consider the set

$$\left(Y_{nS_i}^0(\mathbf{b}), \mathbf{Z}_i'\right), \qquad i = 1, \dots, k_n. \tag{5.12}$$

As such, as in (5.8), we arrive at the empirical d.f.'s

$$G_{n,k_n}(y;\mathbf{b}) = k_n^{-1} \sum_{i=1}^{k_n} I\left(Y_{nS_i}^0(\mathbf{b}) \le y\right), \qquad y \in \mathbf{R}, \mathbf{b} \in \mathscr{K}. \tag{5.13}$$

At this stage, we may assume that the d.f. of $\mathbf{Z}$, denoted by F, admits a continuously differentiable (up to the second order) density function $f(\mathbf{z})$, $\mathbf{z} \in \mathbf{R}^q$, and that $f(\mathbf{z})$ is positive for all $\mathbf{z} \in \mathscr{C}$. Note that, by definition, the density function for the D_i is given by

$$f_D(d;\mathbf{z}_0) = \int_{\{\mathbf{Z}:\ \rho(\mathbf{Z},\mathbf{Z}_0) = d\}} f(\mathbf{z})\, d\mathbf{z}, \qquad d \in \mathbf{R}^+, \tag{5.14}$$

so that the differentiability properties of the density $f(\mathbf{z})$ transmit onto $f_D(\cdot)$ as well. Also, note that the conditional d.f. and density of Y_i^0, given $D_i = d$ (at the pivot $\mathbf{z}_0$), are, respectively, given by

$$G_D^0(y|\mathbf{z}_0, d) = \left\{\int_{\{\mathbf{Z}:\ \rho(\mathbf{Z},\mathbf{Z}_0) = d\}} f(\mathbf{z}) G^0(y|\mathbf{z})\, d\mathbf{z}\right\} \Big/ f_D(d, \mathbf{z}_0), \tag{5.15}$$

$$g_d^0(y|\mathbf{z}_0, d) = \left\{\int_{\{\mathbf{Z}:\ \rho(\mathbf{Z},\mathbf{Z}_0) = d\}} f(\mathbf{z}) g^0(y|\mathbf{z})\, d\mathbf{z}\right\} \Big/ f_D(d, \mathbf{z}_0), \tag{5.16}$$

defined for $y \in \mathbf{R}$, $d \in \mathbf{R}^+$ and $\mathbf{z}_0 \in \mathscr{C}$. We assume that, for every $\mathbf{z}_0 \in \mathscr{C}$, $g_D^0(\cdot|d, \mathbf{z}_0)$ admits a continuously differentiable first derivative with respect to $d \in [0, \eta]$ for some $\eta > 0$. Also, in (5.13), to stress the dependence on the pivot $\mathbf{z}_0$, we write the empirical d.f. as $G_{n,k_n}(y;\mathbf{b},\mathbf{z}_0)$. Then we have the following result.

THEOREM 5.1. *Under the assumed regularity conditions on* $\mathbf{t}_k$, k_n *and* $f_D(\cdot)$, $g_D(\cdot)$,

$$\sup_{\mathbf{z}\in\mathscr{C}} \sup_{\mathbf{b}\in\mathscr{K}} \sup_{y\in\mathbf{R}} k_n^{1/2}\left|G_{n,k_n}(y;\mathbf{b},\mathbf{z}) - G_{n,k_n}(y;\mathbf{0},\mathbf{z})\right| \to_P 0 \quad \text{as } n \to \infty. \tag{5.17}$$

OUTLINE OF THE PROOF. A relatively stronger result for the marginal empirical d.f. is well known [namely Sen and Ghosh (1972)]. In that case, as

$k_n = n$, the factor $\mathbf{b}'\mathbf{t}_i$ in (5.11) may generally contribute a drift function in (5.17). However, in the current conditional case, $k_n = o(n)$ and the covariates $\mathbf{Z}_i$ are i.i.d.r.v.'s, so that their clusterings around a pivot does not stochastically affect the relative partitioning of the (centered) $\mathbf{t}_i$; thus, the drift function is null in the current case. Let us define the antiranks as in (5.7), and, as they are constructed by reference to the pivot $\mathbf{z}_0$, we write

$$\mathbf{S}_n(\mathbf{z}_0) = \{S_1(\mathbf{z}_0), \dots, S_{k_n}(\mathbf{z}_0)\} \quad \text{for every } \mathbf{z}_0 \in \mathscr{C}. \tag{5.18}$$

Then we may rewrite (5.13) as

$$G_{n,k_n}(y, \mathbf{b}, \mathbf{z}) = k_n^{-1} \sum_{i=l}^{n} I(i \in \mathbf{S}_n(\mathbf{z})) I(Y_{ni}^0(\mathbf{b}) \le y), \qquad y \in \mathbf{R}, \mathbf{b} \in \mathscr{K}, \mathbf{z} \in \mathscr{C}. \tag{5.19}$$

Let us also denote by

$$U_{ni}(y; \mathbf{b}) = G(y + n^{-1/2}\mathbf{b}'\mathbf{t}_i | \mathbf{Z}_i) - G(y | \mathbf{Z}_i), \qquad i = 1, \dots, n, \mathbf{b} \in \mathscr{K}, y \in \mathbf{R}. \tag{5.20}$$

Then, for every $\mathbf{b} \in \mathscr{K}$, $\mathbf{z} \in \mathscr{C}$ and $y \in \mathbf{R}$, we have

$$\begin{aligned} k_n^{1/2}&\big[G_{n,k_n}(y; \mathbf{b}, \mathbf{z}) - G_{n,k_n}(y; \mathbf{0}, \mathbf{z})\big] \\ &= k_n^{-1/2} \sum_{i=1}^{n} I(i \in \mathbf{S}_n(\mathbf{z})) U_{ni}(y, \mathbf{b}) \\ &\quad + k_n^{-1/2} \sum_{i=1}^{n} I(i \in \mathbf{S}_n(\mathbf{z})) \big[I(Y_{ni}^0(\mathbf{b}) \le y) \\ &\qquad\qquad - I(Y_i^0 \le y) - U_{ni}(y; \mathbf{b}) \big]. \end{aligned} \tag{5.21}$$

Since the cardinality of $\mathbf{S}_n(\mathbf{z})$ is equal to $k_n = o(n)$, the first term on the right-hand side of (5.21) is easily seen to be $O_p((n^{-1}k_n)^{1/2}) = o_p(1)$, while, conditional on the $\mathbf{Z}_i$, we may use the central limit theorem on the second term and show that it converges (in law) to a degenerate normal $(0, 0)$ variable. Therefore, the left-hand side of (5.21) converges in probability to 0. This shows that the finite-dimensional distributions of the $(p + q + 1)$-parameter stochastic process

$$W_n^0(y, \mathbf{b}, \mathbf{z}) = k_n^{1/2}\big[G_{n,k_n}(y, \mathbf{b}, \mathbf{z}) - G_{n,k_n}(y, \mathbf{0}, \mathbf{z})\big], \qquad y \in \mathbf{R}, \mathbf{b} \in \mathscr{K}, \mathbf{z} \in \mathscr{C}, \tag{5.22}$$

are all degenerate. Hence, to prove (5.17), it remains only to show that $W_n^0(\cdot)$ is *tight*. If we define by $\mathscr{T} = \mathbf{R} \times \mathscr{K} \times \mathscr{C}$ (a subset of $\mathbf{R}^{p+q+1}$), then $W_n^0(\cdot)$ belongs to the space $D[\mathscr{T}]$. Thus, if we define two *blocks*, say $\mathscr{B}_1$ and $\mathscr{B}_2$, belonging to $\mathscr{T}$, which are not overlapping, then proceeding as in Bickel and Wichura (1971) and using the decomposition in (5.21), it can be shown that the multiparameter Billingsley-type inequality holds for the increments $W_n^0(\mathscr{B}_1)$ and $W_n^0(\mathscr{B}_2)$ over these blocks. These, in turn, ensure the tightness.

Therefore, $W_n^0 \to_{\mathscr{D}}$ to a null functional on $D[\mathscr{T}]$, and this completes the proof of (5.17). $\square$

Consider now the functional formulation of (5.9), and, side by side, we write

$$T_n(\mathbf{z}_0;\mathbf{b}) = \theta\big(G_{n,k_n}(\cdot|\mathbf{b},\mathbf{z}_0)\big), \qquad \mathbf{b} \in \mathscr{K}, \mathbf{z}_0 \in \mathscr{C}. \tag{5.23}$$

Assume that the functional $\theta(H)$ is *Hadamard* (or compact) differentiable at $G(\cdot|\mathbf{Z}_0)$, uniformly in $\mathbf{z}_0 \in C$; its Hadamard derivative agrees with the *influence function* which we denote by $\theta_1(G(\cdot|\mathbf{z}_0), y)$ for $y \in \mathbf{R}$ and every $\mathbf{z}_0 \in \mathscr{C}$. Since (5.9) relates to a location functional, from robustness considerations, it may be quite appropriate to assume that this influence function is bounded and monotone in $y \in \mathbf{R}$. Theorem 5.1 may then be readily incorporated to show that, as $n \to \infty$,

$$\sup\big\{k_n^{1/2}|T_n(\mathbf{z};\mathbf{b}) - T_n(\mathbf{z},\mathbf{0})|: \mathbf{b} \in \mathscr{K}, \mathbf{z} \in \mathscr{C}\big\} \to_P 0. \tag{5.24}$$

Let us now define the residuals $\hat{Y}_{ni}$ as in (5.1) and note that $\hat{\mathbf{b}}_n = \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \in \mathscr{K}$, in probability, for a suitable compact $\mathscr{K}$. Therefore, if we let

$$\hat{T}_n(\mathbf{z}) = T_n\big(\mathbf{z},\hat{\mathbf{b}}_n\big), \qquad \mathbf{z} \in \mathscr{C}, \tag{5.25}$$

then by virtue of (5.24) and (5.25), we have, for $n \to \infty$,

$$\sup\big\{\sqrt{n}\,\big|\hat{T}_n(\mathbf{z}) - T_n(\mathbf{z},\mathbf{0})\big|: \mathbf{z} \in \mathscr{C}\big\} \to_P 0. \tag{5.26}$$

On the other hand, the weak invariance principle in (5.10) applies to $T_n(\mathbf{z},\mathbf{0})$, so that the same weak invariance principle holds for

$$\hat{W}_n(\mathbf{z}) = \sqrt{k_n}\,\big(\hat{T}_n(\mathbf{z}) - \theta(\mathbf{z})\big), \qquad \mathbf{z} \in \mathscr{C}. \tag{5.27}$$

The important feature of (5.27) is that the rate of convergence of the functional estimators does not depend on an initial estimator of $\boldsymbol{\beta}$, as long as the latter is $\sqrt{n}$-consistent, and the intricate structure of the ANOVA–ANOCOVA Model 2 and the i.i.d. nature of the covariates allow us to exploit this fully with the aid of the estimators of $\boldsymbol{\beta}$ considered in Section 3.

6. Improved estimation of $\boldsymbol{\beta}$ in ANOCOVA Model 2. Theorem 4.1 reveals that, in general, the ANOVA model based estimator of $\boldsymbol{\beta}$ is not fully efficient relative to the ANOCOVA Model 2 estimators (based on the common score function). This picture becomes even more complex in the case where the d.f. $G(\cdot|\mathbf{z})$ may violate the homogeneity (with respect to $\mathbf{z}$) and/or the linearity of the regression on the covariates. Faced with the basic fact that $\theta(\mathbf{z})$, $\mathbf{z} \in \mathscr{C}$, is an infinite-dimensional nuisance parameter, an optimal estimator of $\boldsymbol{\beta}$ in this general model is rather difficult to construct. On the other hand, since the ANOVA based estimator of $\boldsymbol{\beta}$ is $\sqrt{n}$-consistent and valid for ANOCOVA Model 2, too, it may be incorporated in some way or other to yield a better estimator of $\boldsymbol{\beta}$. We intend to pursue such an estimator here.

The basic idea is simple and is based on the *stratification* of the concomitant vectors into a number of relatively homogeneous subsets and then combining the ANOVA model based estimators from these subsets into a pooled one. The very basic reason that this simple idea of pooling works out well is that the covariates $\mathbf{Z}_i$ are themselves i.i.d.r.v.'s, independent of the design variates, so that their stratification does not affect the allocation of the design vectors $\mathbf{t}_i$ beyond the normal chance variation level, while the error distributions within the strata become relatively more homogeneous and thereby reduce the margin of errors for the estimates based on them. This stratification may even be made *data oriented*. For example, if the Z_i are real valued, one may use the *order statistics* to define a number of nonoverlapping strata (intervals). We consider here the more general case $\mathscr{C} \in \mathbf{R}^q$ and divide $\mathscr{C}$ into a number of nonoverlapping subsets, say $\mathscr{C}_j$, $j = 1, \dots, M$, where M is a fixed positive integer. While a few of these subsets may be partially unbounded (namely, the frontier ones when $\mathscr{C} = \mathbf{R}^q$), the rest of these are all compact and convex. Let there be n_j observations in the subset $\mathscr{C}_j$, $j = 1, \dots, M$. Generally, the n_j are stochastic, but under fairly general regularity conditions, we may assume that

$$n_j/n \to \rho_j : 0 < \rho_j < 1, \qquad 1 \le j \le M, \qquad \sum_{j=1}^{M} \rho_j = 1. \tag{6.1}$$

Treat these n_j as nonstochastic and, pertaining to the subset $\mathscr{C}_j$, frame an ANOVA model and use an RRS estimator as in (2.7); we term it $\hat{\boldsymbol{\beta}}_{n,j}(\phi)$ for $j = 1, \dots, M$. Subject to (6.1), the verification of the classical Anscombe (1952) *uniform continuity in probability* condition can be carried out as in Sen (1981), Chapter 10, so that, parallel to (4.5), we would have here

$$\sqrt{n_j}\left(\hat{\boldsymbol{\beta}}_{n,j}(\phi) - \boldsymbol{\beta}\right) \to_{\mathscr{D}} \mathscr{N}_p\left(\mathbf{0}, \kappa_{0j}^{-2} A_\phi^2 \mathbf{Q}_{tt}^{-1}\right), \qquad j = 1, \dots, M, \tag{6.2}$$

where $A_\phi^2, \mathbf{Q}_{tt}$ are defined as before,

$$\kappa_{0j} = \int_0^1 h_j\left(H_h^{-1}(u)\right) d\phi(u), \qquad j = 1, \dots, M, \tag{6.3}$$

and $H_j(\cdot)$ is the error d.f. pertaining to the set $\mathscr{C}_j$ for $j = 1, \dots, M$. Moreover, for each j, the Jurečková (1992) linearity of RRS processes can readily be incorporated to estimate the κ_{0j} consistently, and these estimators are denoted by $\hat{\kappa}_{0j}$, $j = 1, \dots, M$. We introduce a set of nonnegative weights:

$$w_{nj} = n_j \hat{\kappa}_{0j}^2 \Bigg/ \left(\sum_{r=1}^{M} n_r \hat{\kappa}_{0r}^2 \right), \qquad j = 1, \dots, M. \tag{6.4}$$

The pooled estimator of $\boldsymbol{\beta}$ is then proposed as

$$\hat{\boldsymbol{\beta}}_n^*(\phi) = \sum_{j=1}^{M} w_{nj} \hat{\boldsymbol{\beta}}_{n,j}(\phi). \tag{6.5}$$

Defining $\kappa^{*2} = \sum_{j=1}^{M} \rho_j \kappa_{0j}^2$, we have from (6.1), (6.2) and by some standard steps that, as $n \to \infty$,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^*(\phi) - \boldsymbol{\beta}\right) \to_{\mathscr{D}} \mathscr{N}\Big(\mathbf{0}, \left(A_\phi^2/\kappa^{*2}\right)\mathbf{Q}_{tt}^{-1}\Big). \tag{6.6}$$

Comparing (4.5) and (6.6), we obtain that

$$\mathrm{ARE}\big(\hat{\boldsymbol{\beta}}^*(\phi); \hat{\boldsymbol{\beta}}(\phi)\big) = \left(\kappa^{*2}/\kappa_0^2\right), \tag{6.7}$$

which is bounded from below by $(\sum_{j=1}^{M} \rho_j \kappa_{0j})^2/\kappa_0^2$. Therefore, whenever the densities $h_j(\cdot)$ satisfy the basic condition that $\sum_{j=1}^{M} \rho_j \kappa_{0j} \ge \kappa_0$, (6.7) is bounded from below by 1. As in Section 4, we may introduce the score functions $\psi_j(u)$, $y \in (0,1)$, corresponding to the densities h_j, and write $\kappa_{0j} = \langle \phi, \psi_j \rangle$, $j = 1, \dots, M$. Then, under the heteroscedastic and some other related models, it can be shown that $\langle \phi, \psi_j \rangle \ge \langle \phi, \psi_h \rangle$ for every $j = 1, \dots, M$, so that (6.7) remains bounded from below by 1. If M is chosen large and the individual ρ_j small, then the increase in the ARE in (6.7) is perceptible even more, but this may also invite a comparatively larger sample size to justify the adequacy of the asymptotics on which (6.7) is based. In any case, the proposed pooled estimator remains relevant for ANOCOVA Model 2 as well.

Acknowledgments. The author is grateful to the reviewers for their valuable comments on the manuscript and to Dr. Antonio Pedroso de Lima for his help in preparing the LaTeX version.

REFERENCES

ANSCOME, F. J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

BHATTACHARYA, P. K. and GANGOPADHYAY, A. K. (1990). Kernel and nearest neighbor estimation of a conditional quantile. *Ann. Statist.* **18** 1400–1415.

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

DOKSUM, K. (1969). Starshaped transformations and power of rank tests. *Ann. Math. Statist.* **40** 1167–1176.

GANGOPADHYAY, A. K. and SEN, P. K. (1992). Contiguity in nonparametric estimation of a conditional functional. In *Nonparametric Statistics and Related Topics* (A. K. M. E. Saleh, ed.) 141–162. North-Holland, Amsterdam.

GANGOPADHYAY, A. K. and SEN, P. K. (1993). Contiguity in Bahadur-type representations of a conditional quantile and application in conditional quantile processes. In *Statistics and Probability, a Raghu Raj Bahadur Festschrift* (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa Rao, eds.) 219–231. Wiley Eastern, New Delhi.

GHOSH, M. and SEN, P. K. (1971). On a class of rank order tests for regression with partially informed stochastic predictors. *Ann. Math. Statist.* **42** 650–661.

GUTENBRUNNER, C. and JUREČKOVÁ, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20** 305–330.

HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Tests*. Academia, Prague.

JUREČKOVÁ, J. (1992). Estimation in a linear model based on regression rank scores. *J. Nonparamet. Statist.* **1** 197–203.

JUREČKOVÁ, J. and SEN, P. K. (1993). Asymptotic equivalence of regression rank scores estimators and *R*-estimators in linear models. In *Statistics and Probability, a Raghu Raj Bahadur Festschrift* (J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa Rao, eds.) 279–291. Wiley Eastern, New Delhi.

JUREČKOVÁ, J. and SEN, P. K. (1996). *Robust Statistical Procedures*: *Asymptotics and Interrelations*. Wiley, New York.

KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

PURI, M. L. and SEN, P. K. (1985). *Nonparametric Methods in General Linear Models*. Wiley, New York.

SEN, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. Wiley, New York.

SEN, P. K. (1993a). Regression rank scores estimation in ANOCOVA. Technical Report 2117, Institute of Statistics, Univ. North Carolina.

SEN, P. K. (1993b). Perspectives in multivariate nonparametrics: conditional functionals and ANOCOVA models. *Sankhyā Ser*. *A* **55** 516–532.

SEN, P. K. (1994). Regression quantiles in nonparametric regression. *J*. *Nonparamet*. *Statist*. **3** 237–253.

SEN, P. K. and GHOSH, M. (1972). On strong convergence of regression rank statistics. *Sankhyā Ser*. *A*. **34** 335–348.

DEPARTMENTS OF BIOSTATISTICS AND STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-7400
E-MAIL: pksen@sphrax.sph.unc.edu