L_1-*Statistical Procedures and Related Topics*
IMS Lecture Notes – Monograph Series (1997) Volume 31

Regression rank statistics

C. Gutenbrunner

University of Marburg, Germany

Abstract: This article deals with a family of implicitly defined rank statistics, which are designed to make inference on general linear hypotheses in a large class of nonparametric extensions of the classical linear model. The new rank statistics are defined via the solutions of a continuous family of minimization problems. For simple designs, the procedure leads to the classical rank statistics.

Key words: Rank regression, logistic regression, robust estimation, transformation models, linear models.

AMS subject classification: Primary 62G05, 62G10, 62G20; secondary 62G30, 62J02.

1 Introduction

For a given known c.d.f. F_0 with continuous positive density f_0 and finite second moment, let us first consider the classical parametric linear model

$$M^{\mathrm{Par}}(F_0) : Y_i \;\sim\; F_0(t-\mu_i), \quad \mu_i = \boldsymbol{\beta}'\mathbf{x}_i \tag{1}$$

where Y_i, $1 \le i \le n$ are independent responses and the vectors $\mathbf{x}_i$ represent design conditions and covariables (we assume that the first component x_{i1} of the $\mathbf{x}_i$ is 1 corresponding to the intercept and denote with $\mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_n')'$ the design matrix). Usually, in such models one is interested in linear hypotheses of the form

$$H_0^{\mathrm{Par}}(F_0) : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}. \tag{2}$$

It is well known, that this model is not *invariant* w.r.t. nonlinear increasing transformations of the response, that is, if $m(t)$ is a nonlinear increasing function, then the transformed responses $m(Y_i)$ in general do not follow a linear model of the form described above (in the following, we use the terms *invariant* or *ordinal invariant* as short terms for *invariant against strictly increasing transformations*). On the other hand, in practice it often is obvious, that the observed responses do not come from a linear model (1), but there might be an unknown nonlinear monotone transformation, such that the transformed observations do.

The smallest invariant model containing $M^{\mathrm{Par}}(F_0)$ is the semiparametric transformation model

$$M^{\mathrm{SPar}}(F_0) : Y_i \sim F_0(a(t) - \mu_i), \quad \mu_i = \boldsymbol{\beta}' \mathbf{x}_i, \tag{3}$$

where the unknown increasing function $a(t)$ introduces a nonparametric component. $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ is still identifiable up to an intercept, i.e. a constant multiple of $\mathbf{1}_n$. The null hypothesis (2) in the context of model $M^{\mathrm{SPar}}(F_0)$ will be denoted with $H_0^{\mathrm{SPar}}(F_0)$.

Rank statistics for inference within this kind of model were studied for example in Pettitt (1982, 1983, 1987), Doksum (1987), Cuzick (1988), Tsukahara (1992), see also Bickel et al. (1993) and Chauduri et al. (1994). They are based on different approximations of Hoeffding's formula (Hoeffding, 1951) for the partial or marginal likelihood of the ranks,

$$n!\ \mathbb{P}_\beta(\mathbf{R}(\mathbf{Y}) = \mathbf{r}) = \mathbb{E}\Big(\prod_{i=1}^{n} \frac{f_0(F_0^{-1}(U_{n:r_i}) - \boldsymbol{\beta}' \mathbf{x}_i)}{f_0(F_0^{-1}(U_{n:r_i}))} \Big) \tag{4}$$

in which $U_{n:l}$ is the l-th order statistic in a sample of n i.i.d. variables $U_{n1}, \ldots, U_{nn}$ distributed uniformly on $(0,1)$, and $\mathbf{r} = (r_1, \ldots, r_n)'$.

Note, that in (4) only $\boldsymbol{\beta}$ is unknown, that is, reduction to the maximally invariant vector of ranks leads to a parametric likelihood, which however in general is numerically difficult. One prominent counter-example is the *proportional hazards model* (F_0 the extreme value c.d.f.), in which (4) becomes Cox's (1972) partial likelihood (for the case without censoring or ties)

$$n!\ \mathbb{P}_\beta(\mathbf{R}(\mathbf{Y}) = \mathbf{r}) = \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\beta}' \mathbf{x}_{(i)})}{\sum_{j=i}^{n} \exp(\boldsymbol{\beta}' \mathbf{x}_{(j)})},$$

where $\mathbf{x}_{(i)}$ corresponds to the i-th largest response.

Pettitt treated approximations to (4) for F_0 normal, whereas Doksum and Cuzick used different approximations for the general case, which however still are numerical quite difficult. For the $F_0 = \Phi$ (normal) case, which is one of the cases with an explicit expression for the expectation of the ranks, Pettitt also proposed an alternative to maximizing a likelihood, namely to minimize

$$\sum_i [R_i - \mathbb{E}_\beta(R_i)]^2, \quad \text{with} \quad \mathbb{E}_\beta(R_i) = 1 + \sum_{j \neq i} \Phi[(\mathbf{x}_i - \mathbf{x}_j)' \boldsymbol{\beta} / \sqrt{2}].$$

2 Statistical model

In the present paper, we shall generalize (3) to a nonparametric model $M^{\text{NPar}}(F_0)$, which may be regarded as the synthesis of (3) with a nonparametric model proposed by Akritas and Arnold (1994) in the case of $\mathbf{X}$ representing a factorial design. The extension compared to (4) consists in replacing inside F_0 the functions $\mu_i(t) = a(t) - \boldsymbol{\beta}'\mathbf{x}_i$ with the more general functions $\mu_i(t) = \mathbf{x}_i'\boldsymbol{\beta}(t)$ where $\boldsymbol{\beta}(t)$ is any smooth function of t, such that the $\mu_i(t)$ are strictly increasing:

$$M^{\text{NPar}}(F_0) : Y_i \;\sim\; F_0(\mathbf{x}_i'\boldsymbol{\beta}(t)). \tag{5}$$

The semiparametric transformation model (3) is contained as special case $\boldsymbol{\beta}(t) = a(t)\mathbf{e}_1 - \boldsymbol{\beta}$, the Akritas-Arnold (1994) model is obtained from (5) by choosing $F_0(t) = t$, the c.d.f. of the uniform distribution. However this latter choice does not satisfy our requirement that the density f_0 should be positive on the whole real line. This causes the compatibility conditions (8)-(11) to fail for the Akritas-Arnold model.

Introducing the *link function* $h = F_0^{-1}$, and denoting with F_i the c.d.f. of Y_i, $h \circ \mathbf{F} = (h \circ F_1, \ldots, h \circ F_n)'$, we may state (5) also in the form that $h \circ \mathbf{F}$ must be continuously differentiable and satisfy for all $t \in I\!R$ a linear condition

$$h \circ \mathbf{F}(t) \in \mathbf{L} = \mathbf{X}I\!R^p = \{\mathbf{X}\boldsymbol{\beta} \mid \boldsymbol{\beta} \in I\!R^p\}.$$

A natural extension of the linear hypothesis (2) is given by

$$H_0^{\text{NPar}}(F_0) : \mathbf{C}\boldsymbol{\beta}(t) = \mathbf{0} \quad \text{for all } t, \qquad \text{or equivalently} \tag{6}$$

$$H_0^{\text{NPar}}(F_0) : h \circ \mathbf{F}(t) \in \mathbf{L}_0 = \{\mathbf{X}\boldsymbol{\beta} \mid \boldsymbol{\beta} \in I\!R^p, \mathbf{C}\boldsymbol{\beta} = \mathbf{0}\} \text{ for all } t. \tag{7}$$

These definitions imply the *compatibility properties*

$$M^{\text{Par}}(F_0) \subset M^{\text{SPar}}(F_0) \subset M^{\text{NPar}}(F_0), \tag{8}$$

$$\begin{aligned} H_0^{\text{SPar}}(F_0) &= M^{\text{SPar}}(F_0) \cap H_0^{\text{NPar}}(F_0), & (9)\\ H_0^{\text{Par}}(F_0) &= M^{\text{Par}}(F_0) \cap H_0^{\text{SPar}}(F_0) & (10)\\ &= M^{\text{Par}}(F_0) \cap H_0^{\text{NPar}}(F_0) & (11) \end{aligned}$$

The models presented above should not be confused with the *semiparametric shift model* $M^{\text{SShft}} = \bigcup_{F_0 \in \mathcal{F}} M^{\text{Par}}(F_0)$, $\mathcal{F}$ a set of real c.d.f., i.e.

$$M^{\text{SShft}} : Y_i \;\sim\; F(t - \mu_i), \;\; F \in \mathcal{F} \text{ unknown}. \tag{12}$$

This model is not invariant in our sense (w.r.t. monotone transformations). It is closely related to a large class of rank tests, namely the procedures collected under the name "ranking after alignment" (overviews in Adichie (1984), Puri and Sen (1985), further references see there).

For example, in the k-sample problem, $M^{\mathrm{NPar}}(F_0)$ is larger then M^{SShft} (requiring only groupwise identical distribution but not the assumption that the between-group differences just shift the c.d.f.), whereas the hypothesis

$$H_0^{\mathrm{NPar}}(F_0) : \beta_2(t) = \ldots = \beta_p(t) = 0 \text{ for all } t$$

is identical with the classical hypothesis of no group differences.

Besides the property of being an invariant extension of the parametric linear model, $M^{\mathrm{SPar}}(F_0)$ and $M^{\mathrm{NPar}}(F_0)$ have another interesting feature: They are characterized by the property, that any discretization of the response Y leads to the class of *discrete ordinal regression models* commonly named "cumulative logit" models (but the link function need not be the logit link, it is just h.). $M^{\mathrm{SPar}}(F_0)$ corresponds to the model with the "parallel regression lines" assumption (which is tested e.g. by SAS procedure LOGISTIC). This model was considered e.g. by McCullagh (1980) and Anderson and Philips (1981). $M^{\mathrm{NPar}}(F_0)$ corresponds to the model without parallel regression lines assumption, which was considered e.g. by Williams and Grizzle (1972).

Example 1 *Let us consider the* 2×2 *factorial design with a continuous covariate* x. *The nonparametric transformation model (5) in this case is*

$$\mathbb{P}(Y_{ijk} \leq t) = F_0(\alpha(t) + \gamma_{1i}(t) + \gamma_{2j}(t) + \gamma_{3ij}(t) + \gamma_4(t)x_{ijk}),$$

α, $\gamma_{1i}, \gamma_{2j}, \gamma_{3ij}, \gamma_4$ *smooth functions of* t, *and*

$$H_0 : \gamma_{3ij} \equiv 0 \qquad \text{or} \qquad H_0 : \gamma_4 \equiv 0.$$

Within the semiparametric transformation model, only α *may depend on* t, *within the parametric submodel* $\alpha(t)$ *must be a linear function of* t. *The Akritas-Arnold-model corresponds to* $F_0(t) = t$ *on* $(0, 1)$.

3 Regression rank statistics

We define the *regression rank score* $\widehat{\boldsymbol{\beta}}(Y_i)$ of response Y_i by the nonlinear regression equation

$$\sum_{j=1}^{n} F_0(\mathbf{x}_j' \widehat{\boldsymbol{\beta}}(Y_i))\mathbf{x}_j = \sum_{j=1}^{n} c(Y_i - Y_j)\mathbf{x}_j, \tag{13}$$

where $c(t) = 0.5I[t = 0] + I[t > 0]$.

Proposition 1 *If f_0 is positive on $I\!R$ and has finite expectation, the solution of (13) exists and is unique.*

Proof: The solution characterizes the minimum of the strictly convex function of $\boldsymbol{\beta}$:

$$G_i(\boldsymbol{\beta}) = \sum_{j=1}^{n} \Big(\int_{h(0)}^{\mathbf{x}_j'\boldsymbol{\beta}} [F_0(t) - c(Y_i - Y_j)]dt \Big), \tag{14}$$

which is bounded from below because $\int |t|\, d\,F_0(t) < \infty$. $\sqcap$

The solution might be infinite if there exists $\boldsymbol{\mu} \in \mathbf{L}$ such that $\operatorname{sign}(\mu_j) = 2c(Y_i - Y_j) - 1$ for all j, that is, the $\mathbf{x}_j$ corresponding to $Y_j \leq Y_i$ are separated from those corresponding to $Y_j > Y_i$ by some hyperplane $\{\mathbf{x} \in I\!R^p \mid \mathbf{x}'\boldsymbol{\beta}^* = \mathbf{0}\}$. We may avoid infinite solutions by replacing F_0 with $[(n+1)F_0 - 0.5]/n$ in (13) and (14).

Only in the special case of h being the logit link, (13) defines the maximum likelihood estimator. In the one sample case (Y_i *i.i.d.*, $\mathbf{x}_i \equiv 1$), (13) reduces to

$$nF_0(\widehat{\boldsymbol{\beta}}(Y_i)) = \sum_{j=1}^{n} c(Y_i - Y_j) = R(Y_i), \qquad \text{hence}$$

$$\widehat{\boldsymbol{\beta}}(Y_i) = h(R(Y_i)/n), \quad R(Y_i) = \text{rank of } Y_i.$$

In the k-sample case, $\widehat{\boldsymbol{\beta}}(Y_i)$ is the vector of rank scores of Y_i within the k samples,

$$\widehat{\boldsymbol{\beta}}(Y_i) = \left(h\left[\frac{R_1(Y_i)}{n_1}\right], \ldots, h\left[\frac{R_k(Y_i)}{n_k}\right]\right)'. \tag{15}$$

In the general case, it might appear at the first glance that computation of all n regression rank scores could be a time consuming task. This is not true however, if they are computed sequentially in their natural order: If $Y_{n:1} < \ldots < Y_{n:n}$ denotes the ordered sample, one should compute the *ordered regression rank scores* $\widehat{\boldsymbol{\beta}}(Y_{n:i})$ sequentially: $\widehat{\boldsymbol{\beta}}(Y_{n:i})$ is computed from data $c(Y_{n:i} - Y_j)$, $\widehat{\boldsymbol{\beta}}(Y_{n:i+1})$ is computed from data $c(Y_{n:i+1} - Y_j)$, which differ from the first set only by adding 0.5 to two components. Consequently, only few iterations will be necessary to compute $\widehat{\boldsymbol{\beta}}(Y_{n:i+1})$ when using $\widehat{\boldsymbol{\beta}}(Y_{n:i})$ as starting value. Hence the actual effort for computation of all solutions is only the effort of computing one nonlinear regression plus $O(n)$. The regression rank statistics we define below, do not use the solutions at extreme order statistics. It is convenient to start the iteration process at the sample median and to proceed in both directions, until the solutions at the $\epsilon \times 100\%$- and the $(1-\epsilon) \times 100\%$- sample quantiles are obtained.

In general, it is useful for both motivation and investigation of (13) , to replace there Y_i respective $Y_{n:i}$ with a continuously varying parameter t:

$$\sum_{j=1}^{n} F_0(\mathbf{x}_j' \widehat{\boldsymbol{\beta}}(t))\mathbf{x}_j = \sum_{j=1}^{n} c(t - Y_j)\mathbf{x}_j. \tag{16}$$

The path $t \mapsto \widehat{\boldsymbol{\beta}}(t)$ jumps at the points $t_i = Y_{n:i} = \widehat{H}^{-1}(i/n)$, $u \mapsto \widehat{\boldsymbol{\beta}} \circ \widehat{H}^{-1}(u)$ jumps at $u_i = i/n$, where in the usual notation, $\widehat{H}$ is the empirical c.d.f. of the total (pooled) sample and $H = n^{-1} \sum_{i=1}^{n} F_i$. A key role will be played by the *regression rank score process* for the transformation model,

$$u \mapsto n^{1/2}(\widehat{\boldsymbol{\beta}} \circ \widehat{H}^{-1}(u) - \boldsymbol{\beta} \circ H^{-1}(u)). \tag{17}$$

Remark 1 *As it was mentioned already, (13) is the ML-equation only if h is the logistic link function. A worthwhile alternative to the present approach consists in replacing (13) by the ML-equation for any link function. This could improve efficiency, but at the cost of additional assumptions, like strong unimodality of f_0, to guarantee uniqueness of the solution. Note, that our assumptions, requiring only positivity, continuity and finite first moment for f_0, are very weak.*

Remark 2 *In Gutenbrunner and Jureckovà (1992), we defined regression rank scores for the semiparametric shift model M^{SShft} in a different way, due to the different nature of the statistical model and the associated invariance requirements (invariance w.r.t. p-dimensional affine transformations versus invariance w.r.t. componentwise monotone transformations). The common name is used because of the common purpose, namely to define rank statistics for the general linear model.*

We shall define now generalized rank statistics which are appropriate for testing linear hypotheses $H_0^{\mathrm{NPar}}(F_0)$ within model $M^{\mathrm{NPar}}(F_0)$. We call these statistics, which are linear combinations of ordered regression rank scores, shortly *regression rank statistics.*

More specifically, we focus here on weighted averages

$$\widehat{\mathbf{T}} = \sum_{i=1}^{n} w_{ni} \widehat{\boldsymbol{\beta}}(Y_{n:i}),$$

with weights satisfying $\sum_{i=1}^{n} w_{ni} = 1$, $w_{ni} \geq 0$, $w_{ni} = 0$ if not $\epsilon \leq i/n \leq 1-\epsilon$ for some $\epsilon > 0$.

We assume here, that the weights are generated by a score function J via

$$w_{ni} = J(\frac{i+1}{n}) - J(\frac{i}{n}),$$

and that J is the c.d.f. of a probability measure, which is concentrated on a compact subset of the open unit interval.

To stress the dependence of $\widehat{\mathbf{T}}$ on J and its implicit dependence on the link function h, we shall also write sometimes $\widehat{\mathbf{T}} = \widehat{\mathbf{T}}(h, J)$. Writing $\widehat{H}$ for the empirical c.d.f. of the total sample, $Y_{n:i} = \widehat{H}^{-1}(i/n)$ and $H = n^{-1}\sum_{i=1}^{n} F_i$, we have the representation

$$\widehat{\mathbf{T}}(h, J) = \sum_{i=1}^{n} w_{ni}\widehat{\boldsymbol{\beta}}(Y_{n:i}) = \int \widehat{\boldsymbol{\beta}} \circ \widehat{H}^{-1}(u)dJ(u) = \int \widehat{\boldsymbol{\beta}}(t)dJ \circ \widehat{H}(t) \quad (18)$$

corresponding to the functional

$$\mathbf{T}(h, J) = \int \boldsymbol{\beta} \circ H^{-1}(u)\, d\, J(u) = \int \boldsymbol{\beta}(t)\, d\, J \circ H(t). \quad (19)$$

Within $M^{\mathrm{SPar}}(F_0)$, (19) reduces to

$$\mathbf{T}(h, J) = \int a(t)\ d\ J \circ H(t)\mathbf{e}_1 - \boldsymbol{\beta}$$

From (19) it is clear, that the null hypothesis (6)/(7) implies $\mathbf{CT}(h, J) = \mathbf{0}$.

The two "score functions" h and J determine the rank statistic $\widehat{\mathbf{T}}(h, J)$ in an asymmetric way: While h is crucial for the statistical model and must be specified correctly in order to obtain consistent estimators and tests, J merely determines efficiency properties like the score function of ordinary linear rank statistics.

The basic idea, to replace one complicated estimating equation with a continuous family of simple equations and to take a weighted average of the family of solutions instead the one solution of the complicated equation, is not new: Koenker an Bassett (1978) extended sample quantiles to linear shift model M^{SShft} (12) using an analogous approach, Koenker and Portnoy (1987) and Gutenbrunner and Jurečková (1992) considered linear combinations of solutions.

Example 2 *Taking the design from Example 1, but without the continuous covariate, the regression ranks take the explicit form (15), since the model is saturated. The components of rank statistic of type $\widehat{\mathbf{T}}$ for testing H_0 : $\gamma_{3ij} \equiv 0$ hence can be expressed directly as*

$$\widehat{T}_{ij}(h, J) = \sum_{l=1}^{n} w_l h[R_{ij}(Y_{n:l})],$$

where $R_{ij}(Y_{n:l})$ is the rank of the l-th largest pooled observation w.r.t. the ij-th group.

For comparison with other rank statistics it is convenient to sum by parts and express the statistic as a function of the ranks $R_{ijk} = n\widehat{H}(Y_{ijk})$ within the combined sample. This leads to

$$\begin{aligned}\widehat{T}_{ij}(h,J) &= -\int (J\circ\widehat{H} - J\circ\widehat{F}_{ij})\, d\, h\circ\widehat{F}_{ij}\\ &= -\sum_{k=1}^{n_{ij}} v_{ijk}\left[J\Big(\frac{(n_{ij}/n)R_{ij:k}+0.5}{n_{ij}+1}\Big) - J\Big(\frac{k}{n_{ij}+1}\Big)\right],\end{aligned}$$

where $R_{ij:k}$ is the k-th largest rank in group ij and

$$v_{ijk} = h\Big(\frac{k+0.5}{n_{ij}+1}\Big) - h\Big(\frac{k-0.5}{n_{ij}+1}\Big),$$

using appropriate versions of empirical c.d.f. and ranks.

Example 3 ***(median scores):*** *The simplest score function in our context is the median score function $J(u) = I[u \geq 0.5]$. It is particularly interesting because our results in this case state that the proper generalization of the median test to tests for general linear hypotheses consists in a very pragmatic procedure: dichotomize the continuous response at the pooled median and compute a categorial regression (like logistic or probit regression). Our results in Section 4 on the asymptotics imply that not knowing the true median (e.g. the data dependent dichotomization point) introduces an additional random vector to the estimator of $\boldsymbol{\beta}(0.5)$, which is contained in the linear space generated by the null hypothesis and hence does not affect test statistics based on contrasts orthogonal to that space. In the semiparametric transformation model $M^{\mathrm{SPar}}(F_0)$ the additional random component asymptotically is proportional to the intercept, the slope components of $\widehat{\boldsymbol{\beta}}$ are not affected. In the nonparametric extension $M^{\mathrm{NPar}}(F_0)$, it is proportional to the derivative $\dot{\boldsymbol{\beta}}(t)$ of $t \mapsto \boldsymbol{\beta}(t)$ at $t = H^{-1}(0.5)$, as follows from (25)- (28).*

Example 4 ***(Steam data from Draper and Smith, 1981):*** *Doksum (1987) used these data to demonstrate his Monte Carlo approximation for the MPLE (approximately maximizing Hoeffding's partial likelihood (4)). The data are an example of a simple linear regression with a good fit of the parametric linear model, hence a case where the ordinary least squares estimator (LSE) is appropriate.*

The response Y is pounds of steam per month needed by a power plant, the regressor x average atmospheric temperature in degree Fahrenheit. In the usual notation

$$Y_i = \alpha + \tilde{\beta}x_i + \sigma\epsilon_i,$$

one has to be aware that our $\boldsymbol{\beta}' = (\beta_1, \beta_2)$ from the transformation model is $(\alpha/\sigma, \tilde{\beta}/\sigma)$, because $F_0((t - \alpha - x\tilde{\beta})/\sigma) = F_0(t/\sigma - \alpha/\sigma - x\tilde{\beta}/\sigma) = F_0(a(t) - \alpha/\sigma - x\tilde{\beta}/\sigma)$.

We compare examples of our regression rank statistics with the LSE and Doksum's likelihood sampling estimator. The second, third and fifth regression rank estimators correspond to the score function $J(u) = (-\frac{1}{2}) \vee 3(u - \frac{1}{2}) \wedge \frac{1}{2}$, defining a trimmed mean type statistic. There is a remarkable gap between the LSE and our estimators on one hand and Doksum's estimator on the other hand. Also replacing estimating equation (13) by the ML-equation for probit regression takes the estimator nearer to the LSE.

The estimates in descending order:

-.092 LSE (95%-confidence limits: $[-.136, -.048]$)

-.090 trimmed mean of regression rank scores, *probit link function*, using ML-equation for dichotomized data instead of (13) (trimming: 33% from both sides).

-.084 trimmed mean of regression rank scores, *logit link function* (trimming: 33% from both sides).

-.082 regression rank score median, *probit link function*, using (13).

-.081 trimmed mean of regression rank scores, *probit link function*, using (13) (trimming: 33% from both sides)

-.064 Doksum's likelihood sampling estimator

4 Asymptotic representations

In this section we show that the asymptotic representation of regression rank statistics $\widehat{\mathbf{T}}(h, J)$ (18) shares important properties of the corresponding representations for linear rank statistics as it was recently developed by Akritas and Arnold (1994), Akritas and Brunner (1996) and others.

According to the asymptotic behaviour of the design we assume

$$\|\mathbf{X}\|_\infty = o(n^{1/2}) \quad \text{as } n \to \infty, \tag{20}$$

$$\|(n^{-1}\mathbf{X}'\mathbf{X})^{-1}\|_2 = O(1) \quad \text{as } n \to \infty, \tag{21}$$

$$\sup_{n>0} n^{-1} \sum_{i=1}^{n} I[\|\mathbf{x}_i\| > K] \to 0 \quad \text{as } K \to \infty. \tag{22}$$

For $F_i(t) = F_0(\mathbf{x}_i'\boldsymbol{\beta}(t))$ we assume a continuous density $f_i(t)$. This assumption is equivalent to continuous differentiability of $\boldsymbol{\beta}(t)$ w.r.t. t. We denote the derivative w.r.t t as $\dot{\boldsymbol{\beta}}(t)$.

Much in the spirit of Pyke and Shorak (1968), we shall expand the regression rank process (17) into the difference of two empirical processes.

Writing $\mathbf{Z}(t)$ for the vector with components $c(t-Y_i)$, we may write (16) as

$$\mathbf{g}_t(\widehat{\boldsymbol{\beta}}(t)) = \mathbf{0}, \quad \text{where}$$
$$\mathbf{g}_t(\boldsymbol{\beta}) = \mathbf{X}'[F_0 \circ \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}(t)], \quad \boldsymbol{\beta} \in I\!R^p.$$

The derivative $\mathbf{D}_t = (\partial/\partial\boldsymbol{\beta})\mathbf{g}_t(\boldsymbol{\beta}(t))$ is

$$\mathbf{D}_t = \mathbf{X}' diag[f_0 \circ \mathbf{X}\boldsymbol{\beta}(t)]\mathbf{X}. \tag{23}$$

The following steps are standard calculations: Denoting with $o^*(\cdot)$ and $o_p^*(\cdot)$ convergence uniformly in t for t in compact sets,

$$\begin{aligned}\mathbf{0} &= \mathbf{g}_t \circ \widehat{\boldsymbol{\beta}}(t) = \mathbf{g}_t \circ \boldsymbol{\beta}(t) + \mathbf{D}_t[\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)] + o^*(\|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\|) \\ &= \mathbf{X}'(\mathbf{F}(t) - \mathbf{Z}(t)) + \mathbf{D}_t[\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)] + o^*(\|\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\|)\end{aligned}$$

hence

$$n^{1/2}(\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)) = \mathbf{W}_n(t) + o_p^*(1),$$

with the empirical process

$$\mathbf{W}_n(t) = n^{1/2}\mathbf{D}_t^{-1}\mathbf{X}'(\mathbf{Z}(t) - \mathbf{F}(t)). \tag{24}$$

We used here, that $f_0(\mathbf{x}_i'\boldsymbol{\beta}(t))$ is bounded away from zero for t in compact sets. The next routine step is

$$\begin{aligned}&\sqrt{n}(\widehat{\boldsymbol{\beta}} \circ \widehat{H}^{-1} - \boldsymbol{\beta} \circ H^{-1}) \\ &= \mathbf{W}_n \circ H^{-1} + \sqrt{n}(\boldsymbol{\beta} \circ \widehat{H}^{-1} - \boldsymbol{\beta} \circ H^{-1}) + o_p^*(1) \\ &= \mathbf{W}_n \circ H^{-1} + \sqrt{n}\dot{\boldsymbol{\beta}} \circ H^{-1}(\widehat{H}^{-1} - H^{-1}) + o_p^*(1) \\ &= \mathbf{W}_n \circ H^{-1} - \sqrt{n}\frac{\widehat{H} \circ H^{-1} - id}{\dot{H} \circ H^{-1}}\dot{\boldsymbol{\beta}} \circ H^{-1} + o_p^*(1) \\ &= \mathbf{W}_n \circ H^{-1} - \mathbf{V}_n \circ H^{-1} + o_p^*(1)\end{aligned} \tag{25}$$

with the second empirical process

$$\mathbf{V}_n(t) = \sqrt{n}\frac{\widehat{H}(t) - H(t)}{\dot{H}(t)}\dot{\boldsymbol{\beta}}(t). \tag{26}$$

In (25), the asymptotic replacement of $\sqrt{n}(\widehat{H}^{-1} - H^{-1})$ by $-\sqrt{n}\frac{(\widehat{H} \circ H^{-1} - id)}{\dot{H} \circ H^{-1}}$ follows the argument given e.g. in Serfling (1980), sect. 2.8.3., p. 112.

The vector $\dot{H}(t)^{-1}\dot{\boldsymbol{\beta}}(t)$ may also be written as $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{d}(t)$, $d_i(t) = (f_i(t)/\bar{f}.(t))h'(F_i(t))$. If $H_0^{\mathrm{NPar}}(F_0)$ is true, the derivative $\mathbf{C}\dot{\boldsymbol{\beta}}(t)$ of $\mathbf{0} \equiv \mathbf{C}\boldsymbol{\beta}(t)$ is zero, therefore $\mathbf{V}_n(t)$ has the important property

$$\mathbf{C}\mathbf{V}_n(t) \equiv \mathbf{0} \qquad \text{under } H_0^{\mathrm{NPar}}(F_0).$$

Hence, defining

$$\mathbf{M}_i(t) = \int (I[t \le s] - F_i(s))(\mathbf{D}_s^{-1}\mathbf{x}_i - \dot{H}(s)^{-1}\dot{\boldsymbol{\beta}}(s))\, d\, J \circ H(s) \quad \text{and}$$

$$\mathbf{M}_i^*(t) = \int (I[t \le s] - F_i(s))(\mathbf{D}_s^{-1}\mathbf{x}_i)\, d\, J \circ H(s),$$

we have sketched the proof of the asymptotic representation stated in the following

Theorem 1 *Within model $M^{\mathrm{NPar}}(F_0)$, if the density f_0 is continuous, positive on $I\!R$ and has finite expectation, the score function J is trimming and of bounded variation and conditions (20)-(22) are satisfied, then*

$$\widehat{\mathbf{T}}(h, J) = \mathbf{T}(h, J) + n^{-1}\sum_{i=1}^{n} \mathbf{M}_i(Y_i) + o_p(n^{-1/2}), \tag{27}$$

and, under $H_0^{\mathrm{NPar}}(F_0)$,

$$\mathbf{C}\widehat{\mathbf{T}}(h, J) = n^{-1}\sum_{i=1}^{n} \mathbf{C}\mathbf{M}_i^*(Y_i) + o_p(n^{-1/2}). \tag{28}$$

The regression rank score process (17) has the asymptotic representation

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} \circ \widehat{H}^{-1}(u) - \boldsymbol{\beta} \circ H^{-1}(u)) = \mathbf{W}_n \circ H^{-1}(u) - \mathbf{V}_n \circ H^{-1}(u) + o_p^*(1),$$

where $o_p^*(1)$ denotes approximation uniformly in u for u in compact subsets of (0,1) and the empirical processes $\mathbf{W}_n$ and $\mathbf{V}_n$ are given by (24) and (26). □

(27) and (28) imply multivariate asymptotic normality with an asymptotic covariance matrix that under $H_0^{\mathrm{NPar}}(F_0)$ has a simplified structure that follows from the covariance function $\mathbf{K}(s,t)$ of $\mathbf{W}_n$:

$$\mathbf{K}(s,t) = \mathbf{D}_s^{-1}\boldsymbol{\Delta}(s,t)\mathbf{D}_t^{-1}, \tag{29}$$

$$\boldsymbol{\Delta}(s,t) = \mathbf{X}' diag(F_0[\mathbf{x}_i'\boldsymbol{\beta}(s \wedge t)] - F_0[\mathbf{x}_i'\boldsymbol{\beta}(s)]F_0[\mathbf{x}_i'\boldsymbol{\beta}(t)])\mathbf{X}. \tag{30}$$

Corollary 1 *Under the assumptions of the theorem, $\widehat{\mathbf{T}}(h, J)$ asymptotically has a multivariate normal distribution. Under $H_0^{\mathrm{NPar}}(F_0)$, the covariance matrix of $\mathbf{C}\widehat{\mathbf{T}}(h, J)$ may be estimated consistently by*

$$\widehat{Cov}(\mathbf{C}\widehat{\mathbf{T}}(h, J)) = \mathbf{C}[\sum_{i,j} w_{ni} w_{nj} \widehat{\mathbf{K}}(Y_{n:i}, Y_{n:j})]\mathbf{C}',$$

where $\widehat{\mathbf{K}}(s,t)$ is obtained from (23), (29), (30), replacing $\boldsymbol{\beta}(t)$ with $\widehat{\boldsymbol{\beta}}(t)$.

5 Extension to nonlinear models and application to ROC analysis

In this section we briefly outline the extension of our method to nonlinear models and its application to the analysis of ROC (receiver operating characteristic) curves when covariables have to be taken into account.

Transformation models and ROC models are in close correspondence because of their invariance. For a given statistical model

$$Y_i \sim F_i,$$

the matrix $\rho(\mathbf{F}) = \{\rho_{ij}\}$ of ROC functions

$$\rho_{ij} = F_i \circ F_j^{-1}$$

extracts the ordinal invariant part of information contained in that model. As at the level of *statistics* the vector $\mathbf{R}(\mathbf{Y})$ of ranks is maximally invariant, at the level of *functionals*, the matrix of ROC functions has this property: Any ordinal invariant functional $T^*(\mathbf{F})$ may be written as a functional of $\rho(\mathbf{F})$.

Parametric ROC models correspond to semiparametric transformation models. For example, the parametric ROC model

$$\rho_{ij}(u) = F_0(F_0^{-1}(u) + \mu_j - \mu_i)$$

corresponds to $M^{\mathrm{SPar}}(F_0)$ (3).

On the other hand, starting instead of (3) with the heteroscedastic transformation model

$$Y_i \sim F_0\left(\frac{a(t) - \mu_i}{\sigma_i}\right), \quad \mu_i = \boldsymbol{\beta}'\mathbf{x}_i, \ \sigma_i = \boldsymbol{\gamma}'\mathbf{x}_i, \tag{31}$$

we arrive at the ROC model

$$\rho_{ij}(u) = F_0\left(\frac{\sigma_j}{\sigma_i}F_0^{-1}(u) + \frac{\mu_j - \mu_i}{\sigma_i}\right), \tag{32}$$

which e.g. was used as starting point in Hsieh (1996) (for F_0 the extreme value c.d.f.).

Tosteson and Begg (1988) proposed to analyze the discrete-response version of (31) respective (32) with the PLUM-program of McCullagh (they assumed F_0 to be the logistic c.d.f.). Considering the nonparametric version of (31) ($\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ depending on t) leads us to nonlinear nonparametric transformation model

$$\mathbb{P}(Y_i \leq t) = F_0 \circ g(\mathbf{x}_i, \boldsymbol{\vartheta}(t)), \qquad \boldsymbol{\vartheta} : \mathbb{R} \to \Theta \quad \text{unknown, smooth,}$$

where

$$g : \mathcal{X} \times \Theta \to I\!R$$

is a known, smooth second link function and Θ a Euclidean parameter space. In model (31) we would use $g(\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \boldsymbol{\beta}'\mathbf{x}_i/\boldsymbol{\gamma}'\mathbf{x}_i$. The corresponding family of estimating equations is

$$\sum_{j=1}^{n} F_0 \circ g(\mathbf{x}, \widehat{\boldsymbol{\vartheta}}(t))\mathbf{x}_j =^! \sum_{j=1}^{n} c(t - Y_j)\mathbf{x}_j,$$

which however in general is not the gradient condition of a convex function, hence some additional assumptions are necessary to guarantee consistent estimators.

In an analogous manner, we may define the nonlinear regression rank score process $u \mapsto n^{1/2}(\widehat{\boldsymbol{\vartheta}} \circ \widehat{H}^{-1}(u) - \boldsymbol{\vartheta} \circ H^{-1}(u))$, nonlinear regression rank statistics $\widehat{T}_x(h, J) = \int g(\mathbf{x}, \widehat{\boldsymbol{\vartheta}}(t))\, d\, J \circ \widehat{H}(t)$, corresponding to functionals $T_x(h, J) = \int g(\mathbf{x}, \boldsymbol{\vartheta}(t))\, d\, J \circ H(t)$, hypotheses $H_0 : g^*(\mathbf{x}, \boldsymbol{\vartheta}(t)) \equiv \mathbf{0}$ (g^* another known link function) and test statistics based on $\widehat{T}_x^*(h, J) = \int g^*(\mathbf{x}, \widehat{\boldsymbol{\vartheta}}(t))\, d\, J \circ \widehat{H}(t)$.

Acknowledgements

I would like to thank Roger Koenker for valuable hints to related literature.

References

[1] Adichie, J. N. (1984). Rank tests in linear models. In *Handbook of Statistics,* Vol.4, Eds. P.R. Krishnaiah and P. K. Sen, pp. 229-257.

[2] Akritas M. G. and Arnold S. F. (1994). Fully nonparametric hypotheses for factorial designs I: Multivariate repeated measurement designs. *J. Am. Statist. Assoc.* **89,** 336-343.

[3] Anderson, J. A. and Philips, P. R. (1981). Regression discrimination and measurement models for ordered categorial variables. *Appl. Statist.* **30**, 22–31.

[4] Bickel, P. J., Klaassen, C. A. J., Ritov Y. and Wellner J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* John Hopkins University Press.

[5] Brunner E. and Puri M. L. (1996). Nonparametric methods in design and analysis of experiments. In *Handbook of Statistics,* Vol.13, Eds. C.R. Rao and S. Ghosh.

[6] Chauduri P., Doksum K. and Samarov A. (1994). Regression estimators based on conditional quantiles. *J. Nonpar. Statist.* **3**, 317–321.

[7] Cox D. R. (1972). Regression models and life tables (with discussion). *J. R. Statist. Soc. B* **34**, 187–220.

[8] Cuzick J. (1988). Rank regression. *Ann. Statist.* **16**, 1369–1389.

[9] Doksum K. A. (1987). An extension of partial likelihood methods for proportional hazard models to general transformation models. *Ann. Statist.* **15**, 325–345.

[10] Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis.* New York: Wiley.

[11] Gutenbrunner, C. and Jurečková, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20**, 305-330.

[12] Hoeffding, W. (1951). "Optimum" nonparametric tests. *Second Berkeley symp. Math Statist. Probab.*, 83–92.

[13] Hsieh, F. (1996). A transformation model for two survival curves: an empirical process approach. *Biometrika* **83**, 519–528.

[14] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

[15] Koenker, R. and Portnoy, S. (1987). *L*-estimation for linear models. *J. Amer. Statist. Assoc.* **82**, 851–857.

[16] McCullagh P. (1980). Regression models for ordinal data. *J. R. Statist. Assoc.* B **42**, 109-127.

[17] Pettitt, A. N. (1982). Inference for the linear model using a likelihood based on ranks. *J. R. Statist. Assoc.* B **44**, 234–243.

[18] Pettitt, A. N. (1983). Approximate methods using ranks for regression with censored data. *Biometrika* **70**, 121–132.

[19] Pettitt, A. N. (1987). Estimates for a regression parameter using ranks. *J. R. Statist. Assoc.* B **49**, 58–67.

[20] Puri, M. L. and Sen, P. K. (1985). *Nonparametric Methods in General Linear Models.* New York: Wiley.

[21] Pyke, R. and Shorak, G. (1968). Weak convergence of a two-sample empirical process an a new approach to Chernoff-Savage theorems. *Ann. Statist.* **39**, 755-771.

[22] Serfling R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

[23] Tosteson A. N. A. and Begg C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making* **8,** 204-215.

[24] Tsukahara, H. (1992). A rank estimator in the two-sample transformation model with randomly censored data. *Ann. Inst. Statist. Math.* **44,** 313-333.

[25] Williams O. D. and Grizzle, J. E. (1972). Analysis of contingency tables having ordered categories. *J. Am. Statist. Assoc.* **67**, 55–63.