

Rank Regression with Estimated Scores

Joshua D. Naranjo* & Joseph W. McKean

Western Michigan University

Abstract

Rank-based estimates are asymptotically efficient when optimal scores are used. This paper describes a method for estimating the optimal score function based on residuals from an initial fit. The resulting adaptive estimate is shown to be asymptotically efficient.

Keywords: Rank regression; Score Estimation; Adaptive estimate; Asymptotic efficiency.

1 Introduction

Consider the linear model

$$y_i = \alpha^* + x_i' \beta^* + e_i, \quad i = 1, \dots, n, \quad (1)$$

where e_1, \dots, e_n are independent random variables with distribution function F and density f , x_i' is the i th row of a known $n \times p$ matrix of centered explanatory variables X , α^* is an intercept parameter, and β^* is a $p \times 1$ vector of slope parameters. Consider the estimate which minimizes

$$\sum_{i=1}^n a(R(y_i - x_i' \beta))(y_i - x_i' \beta), \quad (2)$$

*Corresponding author. Department of Mathematics and Statistics, Western Michigan University, Kalamazoo, MI 49008-5152, USA

where $R(y_j - x'_j\beta)$ is the rank of $y_j - x'_j\beta$ among $y_1 - x'_1\beta, \dots, y_n - x'_n\beta$, and $a(1) \leq \dots \leq a(n)$ is a nondecreasing set of scores. If the scores are chosen so that

$$a(j) = \varphi_F(j/(n+1)) = -\frac{f'(F^{-1}(j/(n+1)))}{f(F^{-1}(j/(n+1)))},$$

then the resulting estimate $\hat{\beta}_F$ is asymptotically efficient.

In practice, the data analyst only has approximate knowledge of φ_F . In this case, a preliminary analysis of residuals from an initial fit may provide information as to an appropriate choice of score function (see for example McKean and Sievers (1989), Kapenga and McKean (1989), and McKean, Vidmar and Sievers (1989)). Another approach would be to estimate the scores themselves. An asymptotically efficient estimate of location was proposed by van Eeden (1970), by estimating φ_F from a subset of the data. Dionne (1981) developed efficient estimates of linear model parameters, also by estimating the scores based on a small subset of the data. Beran (1974) introduced asymptotically efficient estimates in the one and two-sample situations, using the whole sample in estimating φ_F . This paper follows Beran's (1974) approach in score function estimation. An initial estimate $\hat{\beta}$ yields preliminary residuals $\hat{e}_1, \dots, \hat{e}_n$, which are used to construct estimates $\tilde{\varphi}(t)$ of $\varphi(t)$, which are then used to compute the adaptive rank estimate $\tilde{\beta}_R$.

2 Estimation of φ_F

Consider the Fourier expansion

$$\varphi_F(t) = \sum_{|k|=1}^{\infty} c_k e^{2\pi ikt}, \quad (3)$$

where

$$c_k = \int_0^1 \varphi_F(t) e^{-2\pi ikt} dt. \quad (4)$$

Express c_k as a more general functional

$$T(\phi) = \int_0^1 \varphi_F(t) \phi(t) dt = \int \left[\frac{d}{dx} \phi(F(x)) \right] dF(x), \quad (5)$$

where $\phi(t) = e^{-2\pi ikt}$ in the case of (4). Note that the second expression for $T(\phi)$ depends on f only through the cdf F . If we had a random sample Z_1, \dots, Z_n from the cdf F , then (5) suggests the estimator

$$T_n^Z(\phi) = \frac{1}{2n\theta_n} \sum_{i=1}^n [\phi(F_n(Z_i + \theta_n)) - \phi(F_n(Z_i - \theta_n))],$$

where $F_n(t) = (1/n) \sum_{i=1}^n \mathbf{I}(Z_i \leq t)$ and $\theta_n \rightarrow 0$ at an appropriate rate. Beran (1974) used $T_n^Z(\phi)$ for estimating scores in the two-sample problem.

In the absence of a random sample from F , we rely on the residuals from an initial fit, yielding the estimate

$$T_n(\phi) = \frac{1}{2n\theta_n} \sum_{i=1}^n [\phi(F_n^*(\hat{e}_i + \theta_n)) - \phi(F_n^*(\hat{e}_i - \theta_n))],$$

where F_n^* is the empirical cdf based on $\hat{e}_1, \dots, \hat{e}_n$. From (3) and (4), we can construct score estimates

$$\tilde{\varphi}_F(t) = \sum_{|k|=1}^{M_n} \tilde{c}_k e^{2\pi ikt}, \quad (6)$$

where $\tilde{c}_k = T_n(e^{-2\pi ik \cdot})$ and $M_n \rightarrow \infty$ at an appropriate rate.

Theorem 1 at the end of this section shows that the proposed scores are consistent. Without loss of generality, in the rest of the paper we will assume that $\beta^* = 0$ and $\alpha^* = 0$ in (1) so that y_1, \dots, y_n are i.i.d. F .

Let $\|\cdot\|$ denote the Euclidean norm.

Lemma 1 *Assume that X and F satisfy the following:*

$$(A1) \max_{1 \leq i \leq n} \|x_i\| / \sqrt{n} = o(1).$$

$$(A2) (1/n) \sum_{i=1}^n \|x_i\|^2 = O(1).$$

(A3) f is uniformly continuous and bounded with finite Fisher information, f' exists, and f'/f is monotone.

Then

$$\sup_{w, \|\sqrt{n}\hat{\beta}\| \leq B} |(1/\sqrt{n}) \sum_i \mathbf{I}(y_i - x_i' \hat{\beta} \leq w) - (1/\sqrt{n}) \sum_i \mathbf{I}(y_i \leq w) - \sqrt{n} \hat{\beta}' \bar{x} f(w)| = o_p(1).$$

The proof may be found in Koul (1992, Sec. 2.3). Now let $F_n^*(t) = (1/n) \sum_{i=1}^n \mathbf{I}(\hat{\epsilon}_i \leq t)$, where $\{\hat{\epsilon}_i = y_i - x_i' \hat{\beta}\}$ are residuals from an initial estimate $\hat{\beta}$. The next result says that F_n^* is consistent if $\hat{\beta}$ is tight.

Lemma 2 *Let (A1), (A2), (A3) hold. In addition assume*

$$(A4) \sqrt{n}\hat{\beta} = O_p(1).$$

Then

$$\sup_w |\sqrt{n}(F_n^*(w) - F(w))| = O_p(1).$$

Proof. Note that

$$\begin{aligned} & \sup_w |\sqrt{n}(F_n^*(w) - F(w))| \\ & \leq \sup_w |(1/\sqrt{n}) \sum \mathbf{I}(y_j - x_j' \hat{\beta} \leq w) - (1/\sqrt{n}) \sum \mathbf{I}(y_j \leq w) - \sqrt{n} \hat{\beta}' \bar{x} f(w)| \\ & \quad + \sup_w |\sqrt{n}[(1/n) \sum \mathbf{I}(y_j \leq w) - F(w)]| + \sup_w |\sqrt{n} \hat{\beta}' \bar{x} f(w)|. \end{aligned}$$

The first sup term on the right side is $o_p(1)$ by Lemma 1. The second term is $O_p(1)$ by standard asymptotic distribution results on the empirical cdf (see e.g. Serfling, 1980, Sec. 2.1.5). The last term is $O_p(1)$, by (A2)-(A3).

Lemma 3 *Assume that ϕ and F satisfy the following conditions:*

(A4) ϕ' , ϕ'' , and ϕ''' are bounded,

(A5) $\int |\phi''(F(w))| dF(w) < \infty$,

(A6) $\phi'(F)$, $(\phi(F))' F^{-1}$ are uniformly continuous with bounded first derivative.

Then, for any sequence of constants $M_n \rightarrow \infty$, $\theta_n \rightarrow 0$ such that

(A7) $M_n/(\theta_n\sqrt{n}) \rightarrow 0$, $M_n\theta_n^2 \rightarrow 0$,

it follows that

$$M_n(T_n(\phi) - T(\phi)) = o_p(1). \quad (7)$$

Proof. Write the left side of (7) as

$$M_n(T_n(\phi) - T(\phi)) = M_n(T_n(\phi) - T_{n1}) + M_n(T_{n1} - T(\phi)), \quad (8)$$

where

$$T_{n1} = \frac{1}{2n\theta_n} \sum_{j=1}^n [\phi(F(\hat{e}_j + \theta_n)) - \phi(F(\hat{e}_j - \theta_n))].$$

We will show that both terms on the right side of (8) are $o_p(1)$. Expanding $\phi(F_n^*)$ about $\phi(F)$ and using Lemma 2 and $M_n/(\theta_n\sqrt{n}) \rightarrow 0$, we have

$$\begin{aligned} M_n(T_n(\phi) - T_{n1}) &= (M_n/2\theta_n) \int [F_n^*(w + \theta_n) - F(w + \theta_n)] \phi'(F(w + \theta_n)) dF_n^*(w) \\ &\quad - (M_n/2\theta_n) \int [F_n^*(w - \theta_n) - F(w - \theta_n)] \phi'(F(w - \theta_n)) dF_n^*(w) \\ &\quad + (M_n/4\theta_n) \int [F_n^*(w + \theta_n) - F(w + \theta_n)]^2 \phi''(\xi_{1,n}(w)) dF_n^*(w) \\ &\quad - (M_n/4\theta_n) \int [F_n^*(w - \theta_n) - F(w - \theta_n)]^2 \phi''(\xi_{2,n}(w)) dF_n^*(w) \\ &= o_p(1), \end{aligned}$$

where $\xi_{1,n}(w)$ is between $F_n^*(w + \theta_n)$ and $F(w + \theta_n)$ and $\xi_{2,n}(w)$ is between $F_n^*(w - \theta_n)$ and $F(w - \theta_n)$. Now

$$\begin{aligned} M_n(T_{n1} - T(\phi)) &= (M_n/2\theta_n) \int [\phi(F(w + \theta_n)) - \phi(F(w - \theta_n))] dF_n^*(w) \\ &\quad - M_n \int (\phi(F))'(w) dF(w) \\ &= (M_n/2\theta_n) \int [(\phi(F))'(w)\theta_n - (\phi(F))'(w)(-\theta_n)] dF_n^*(w) \\ &\quad + (M_n/4\theta_n) \int [(\phi(F))''(w)(\theta_n^2) - (\phi(F))''(w)(-\theta_n)^2] dF_n^*(w) \\ &\quad + (M_n/12\theta_n) \int [(\phi(F))'''(\xi_{1,n}(w))(\theta_n^3) - (\phi(F))'''(\xi_{2,n}(w))(-\theta_n)^3] dF_n^*(w) \\ &\quad - M_n \int (\phi(F))'(w) dF(w) \end{aligned}$$

$$\begin{aligned}
&= M_n \left[\int (\phi(F))'(w) dF_n^*(w) - \int (\phi(F))'(w) dF(x) \right] + o_p(1) \\
&= M_n \left[\int (\phi(F))'(F^{-1} F F_n^{*-1}(t)) dt - \int (\phi(F))'(F^{-1}(t)) dt \right] + o_p(1).
\end{aligned}$$

Now $|F(F_n^{*-1}(t)) - t| = |F_n^*(F_n^{*-1}(t)) - t - F_n^*(F_n^{*-1}(t)) + F(F_n^{*-1}(t))| \leq 1/n + O_p(1/\sqrt{n}) = O_p(1/\sqrt{n})$. Expanding $(\phi(F))'(F^{-1} F F_n^{*-1}(t))$ about $(\phi(F))'(F^{-1}(t))$, we have $M_n(T_{n1} - T(\phi)) = o_p(1)$, which proves the lemma.

The following theorem states consistency of $\tilde{\varphi}_F$ defined in (6).

Theorem 1 *Under assumptions (A1)-(A7),*

$$\sup_{0 < t < 1} |\tilde{\varphi}_F(t) - \varphi_F(t)| \xrightarrow{p} 0.$$

Proof: Write

$$\begin{aligned}
\sup_{0 < t < 1} |\tilde{\varphi}_F(t) - \varphi_F(t)| &= \sup_{0 < t < 1} \left| \sum_{|k|=1}^{M_n} \tilde{c}_k e^{2\pi i k t} - \sum_{|k|=1}^{\infty} c_k e^{2\pi i k t} \right| \\
&\leq \sum_{|k|=1}^{M_n} |\tilde{c}_k - c_k| + \sum_{|k|=M_n+1}^{\infty} |c_k|.
\end{aligned}$$

From Lemma 3, $\tilde{c}_k - c_k = T_n(e^{2\pi i k \cdot}) - T(e^{2\pi i k \cdot}) = o_p(1/M_n)$ uniformly in k , since $\phi(s) = e^{2\pi i k s} = \cos(2\pi k s) + i \sin(2\pi k s)$ satisfies Assumptions (A4)-(A6) of Lemma 3 for all k . By absolute convergence of the Fourier series, we have $\sum_{|k|=M_n+1}^{\infty} |c_k| = o(1)$. The result follows.

Note that when F is continuous and $\phi(\cdot) = e^{-2\pi i k \cdot}$ then assumptions (A4)-(A6) are satisfied. Assumption (A7) is satisfied by, say, $M_n = \lceil n^{1/5} \rceil$ and $\theta_n = \theta/n^{1/5}$, for some $\theta > 0$.

3 Estimation

Let $T(\beta) = (1/\sqrt{n}) \sum_{i=1}^n x_i \varphi_F(R(y_i - x_i' \beta)/(n+1))$. Let $\tilde{T}(\beta)$ denote the same expression with φ_F replaced by $\tilde{\varphi}_F$. At the true value $\beta^* = 0$, we have the following normality result.

Theorem 2 Under assumptions (A1)-(A7),

$$\tilde{T}(0) \sim AN(0, \Sigma) ,$$

where $\Sigma = \lim_{n \rightarrow \infty} (1/n) \mathbf{X}' \mathbf{X}$.

Proof: We will show that $\tilde{T}(0) - T(0) \xrightarrow{p} 0$. The result will then follow from asymptotic normality of $T(0)$ (see Heiler and Willers, 1988). It is enough to show that $\tilde{T}(0) - T(0) \xrightarrow{p} 0$ elementwise, hence without loss of generality we may assume in this proof that T is scalar. Let R_j denote the rank of y_j among y_1, \dots, y_n .

$$\begin{aligned} \tilde{T}(0) - T(0) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n x_j \left[\tilde{\varphi} \left(\frac{R_j}{n+1} \right) - \varphi \left(\frac{R_j}{n+1} \right) \right] \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n x_j \left[\sum_{|k|=1}^{M_n} (\tilde{c}_k - c_k) \exp(2\pi i k \frac{R_j}{n+1}) - \sum_{|k|=M_n+1}^{\infty} c_k \exp(2\pi i k \frac{R_j}{n+1}) \right] \\ &= \sum_{|k|=1}^{M_n} (\tilde{c}_k - c_k) \frac{1}{\sqrt{n}} \sum_{j=1}^n x_j \exp(2\pi i k \frac{R_j}{n+1}) - \sum_{|k|=M_n+1}^{\infty} c_k \frac{1}{\sqrt{n}} \sum_{j=1}^n x_j \exp(2\pi i k \frac{R_j}{n+1}) \\ &= A_1 + A_2 , \end{aligned}$$

say. Since $\tilde{c}_k - c_k = o_p(1/M_n)$ by Lemma 3, it suffices to show that $\frac{1}{\sqrt{n}} \sum_{j=1}^n x_j \exp(2\pi i k \frac{R_j}{n+1}) = O_p(1)$ uniformly in k in order to prove that $A_1 \xrightarrow{p} 0$. By Chebyshev's inequality, it will suffice to show that

$$E \left(\frac{1}{\sqrt{n}} \sum_{j=1}^n x_j \exp(2\pi i k \frac{R_j}{n+1}) \right)^2 = O(1), \quad (9)$$

uniformly in k . Expanding the square, the left side of (9) may be written as

$$\begin{aligned} &E \left((1/n) \sum_{j=1}^n x_j^2 \exp(4\pi i k \frac{R_j}{n+1}) + (1/n) \sum_{j \neq l} \sum x_j x_l \exp(2\pi i k \frac{R_j + R_l}{n+1}) \right) \\ &= (1/n) \sum_{j=1}^n x_j^2 \left[E \exp(4\pi i k \frac{R_j}{n+1}) \right] + (1/n) \sum_{j \neq l} \sum x_j x_l \left[E \exp(2\pi i k \frac{R_j + R_l}{n+1}) \right] \end{aligned}$$

$$= \delta_{1,k} (1/n) \sum_{j=1}^n x_j^2 + \delta_{2,k} (1/n) \sum_{j \neq l} \sum x_j x_l$$

where $\delta_{1,k} = E(\exp(4\pi i k \frac{R_j}{n+1}))$ and $\delta_{2,k} = E(\exp(2\pi i k \frac{R_j+R_l}{n+1}))$ are bounded constants for all k . By Assumption (A2), $(1/n) \sum_{j=1}^n x_j^2 = O(1)$. It remains to show that $(1/n) \sum \sum_{j \neq l} x_j x_l = O(1)$. This follows from the identity $0 = (1/n)(\sum_{j=1}^n x_j)^2 = (1/n) \sum_{j=1}^n x_j^2 + (1/n) \sum \sum_{j \neq l} x_j x_l$ which implies that $|(1/n) \sum \sum_{j \neq l} x_j x_l| = |(1/n) \sum_{j=1}^n x_j^2| = O(1)$. Hence (9) is proved, which proves that $A_1 \rightarrow 0$. Now, to show $A_2 \xrightarrow{p} 0$, we will show that $E|A_2| \rightarrow 0$.

$$\begin{aligned} E|A_2| &\leq E \sum_{|k|=M_{n+1}}^{\infty} |c_k| \left| \frac{1}{\sqrt{n}} \sum_{j=1}^n x_j \exp(2\pi i k \frac{R_j}{n+1}) \right| \\ &\leq \sum_{|k|=M_{n+1}}^{\infty} |c_k| \left[E \left(\frac{1}{\sqrt{n}} \sum_{j=1}^n x_j \exp(2\pi i k \frac{R_j}{n+1}) \right)^2 \right]^{1/2} \\ &\rightarrow 0, \end{aligned}$$

by absolute convergence of the Fourier series and (9). This proves Theorem 2.

Theorem 3 *Under assumptions (A1)-(A7), we have for $\epsilon > 0$ and $B > 0$,*

$$\lim_{n \rightarrow \infty} P(\sup_{\|\Delta\| \leq B} \| \tilde{T}(\Delta) - \tilde{T}(0) + \tau^{-1}(n^{-1} X' X) \Delta \| \geq \epsilon) \rightarrow 0, \quad (10)$$

where $\tau^{-1} = \int_0^1 \varphi_F^2(u) du$.

Proof:

$$\begin{aligned} &P(\sup_{\|\Delta\| \leq B} \| \tilde{T}(\Delta) - \tilde{T}(0) + \tau^{-1}(n^{-1} X' X) \Delta \| \geq \epsilon) \\ &\leq P(\sup_{\|\Delta\| \leq B} \| T(\Delta) - T(0) + \tau^{-1}(n^{-1} X' X) \Delta \| \geq \epsilon/3) \\ &\quad + P(\sup_{\|\Delta\| \leq B} \| \tilde{T}(\Delta) - T(\Delta) \| \geq \epsilon/3) + P(\| \tilde{T}(0) - T(0) \| \geq \epsilon/3). \end{aligned}$$

The first term on the right side goes to 0 by the standard linearity result for the rank test statistic (Heiler and Willers, 1988). From the proof of Theorem 2, we have $P(\|\tilde{T}(0) - T(0)\| \geq \epsilon/3) \rightarrow 0$. Using a contiguity argument as in Heiler and Willers (1988), it can be shown that $P(\sup_{\|\Delta\| \leq B} \|\tilde{T}(\Delta) - T(\Delta)\| \geq \epsilon/3) \rightarrow 0$, which proves the theorem.

Finally, we define the adaptive estimate as

$$\tilde{\beta}_R = \hat{\beta} + \tau\sqrt{n}(X'X)^{-1}\tilde{T}(\hat{\beta}), \quad (11)$$

where $\hat{\beta}$ is the initial estimator. The following theorem shows that $\tilde{\beta}_R$ is asymptotically efficient.

Theorem 4 *Under assumptions (A1)-(A7),*

$$\sqrt{n}\tilde{\beta}_R \sim AN(0, \tau^2\Sigma^{-1}).$$

Proof: We have from (11) and (10) that

$$\begin{aligned} \sqrt{n}\tilde{\beta}_R &= \sqrt{n}\hat{\beta} + \tau n(X'X)^{-1} [\tilde{T}(0) - \tau^{-1}n^{-1/2}(X'X)\hat{\beta} + o_p(1)] \\ &= \tau(n^{-1}X'X)^{-1}\tilde{T}(0) + o_p(1). \end{aligned}$$

The result follows from Theorem 2.

4 Summary

We have proposed an asymptotically efficient adaptive rank estimator $\tilde{\beta}_R$ that estimates the optimal score function from the residuals of an initial estimate. The primary conditions on the error density f are uniform continuity and finite Fisher information, so $\tilde{\beta}_R$ is asymptotically efficient over a large class of distributions.

The construction of $\tilde{\beta}_R$ as a one-step estimator is necessary because the estimated scores $\tilde{\varphi}_F(t)$ are not monotone. The development of monotone

score estimates would allow a Jaeckel-type estimate that minimizes a rank dispersion function with estimated scores. This is a problem under current investigation.

REFERENCES

- Beran, R. (1974), Asymptotically Efficient Adaptive Rank Estimates in Location Models, *The Annals of Statistics*, **2**, No. 1, 63-74.
- Dionne, L. (1981), Efficient Nonparametric Estimators of Parameters in the General Linear Hypothesis, *The Annals of Statistics*, **9**, No. 2, 457-460.
- Heiler, S. and Willers, R. (1988), Asymptotic Normality of R-Estimates in the Linear Model, *Statistics*, **19**, 173-184.
- Kapenga, J. and McKean, J. W. (1989), Spline Estimation of the Optimal Score Function for Fitting Linear Models, in: *ASA Proceedings of the Stat. Comp. Section*, 227-232.
- Koul, H. L. (1992). *Weighted Empiricals and Linear Models* (Institute of Mathematical Statistics Lecture Notes-Monograph Series, Vol. 21).
- McKean, J. W. and Sievers, G. L. (1989), Rank Scores Suitable for Analyses of Linear Models Under Asymmetric Error Distributions, *Technometrics*, **31**, No. 2, 207-218.
- McKean, J. W., Vidmar, T. J. and Sievers, G. L. (1989), A Robust Two-Stage Multiple Comparison Procedure with Application to a Random Drug Screen, *Biometrics*, **45**, 1281-1297.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics* (John Wiley & Sons, New York).
- van Eeden, C. (1970), Efficiency-Robust Estimation of Location, *Annals of Mathematical Statistics*, **41**, 172-181.