EFFICIENT RANK REGRESSION WITH WAVELET ESTIMATED SCORES

E. Kwessi[1], A. Abebe[2] and G. De Souza[2]

[1]Department of Mathematics, Trinity University

One Trinity Place, San Antonio, TX78212,

ekwessi@trinity.edu

[2]Department of Mathematics and Statistics

Auburn University

221 Parker Hall, Auburn, AL36849

abebeas@aauburn.edu, desougs@auburn.edu

Key Words: Asymptotic efficiency; Score function, Besov spaces; Wavelets.

ABSTRACT

We provide an estimate of the score function for rank regression using compactly supported wavelets. This estimate is then used to find a rank-based asymptotically efficient estimator for the slope parameter of a linear model. We also provide a consistent estimator of the asymptotic variance of the rank estimator. For related mixed models, the asymptotic relative efficiency is also discussed

1 Introduction

Consider the linear model

$$y_i = \mu + \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \, , \qquad i = 1, \cdots , n \, , \tag{1.1}$$

[1]Corresponding author

where $\varepsilon_1, \cdots, \varepsilon_n$ are independent random variables with distribution function F and density f. Consider the estimator $\widehat{\boldsymbol{\beta}}_F$ of $\boldsymbol{\beta} \in \mathbb{R}^p$ that minimizes the dispersion function $\sum_{i=1}^n a(R(y_i - \mathbf{x}_i^T\boldsymbol{\beta}))(y_i - \mathbf{x}_i^T\boldsymbol{\beta})$, where $R(y_i - \mathbf{x}_i^T\boldsymbol{\beta})$ is the rank of $y_i - \mathbf{x}_i^T\boldsymbol{\beta}$ among $y_1 - \mathbf{x}_1^T\boldsymbol{\beta}, \cdots, y_n - \mathbf{x}_n^T\boldsymbol{\beta}$ and $a(1) \leq \cdots \leq a(n)$ are scores. The scores are usually chosen as $a(j) = h(j/(n+1))$, where $h : (0,1) \to \mathbb{R}^+$ is a nondecreasing score function. If the scores are chosen so that $a(j) = h_F(j/(n+1))$, where

$$h_F(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} , \tag{1.2}$$

then the estimator $\widehat{\boldsymbol{\beta}}_F$ of $\boldsymbol{\beta}$ is asymptotically efficient (Hettmansperger and McKean, 1998).

If the Fisher information $I(f) = \int_{-\infty}^{\infty} \left(f'(x)/f(x)\right)^2 f(x)dx$ is finite, then $h_F \in L^2(0,1)$. Therefore, under $I(f) < \infty$, there exist coefficients C_{jk}, such that

$$h_F(t) = \sum_{j=-\infty}^{\infty} \sum_{k \in \mathbb{Z}} C_{jk}\psi_{jk}(t) ,$$

where $\{\psi_{jk}\}_{j,k\in\mathbb{Z},}$ is an orthonormal system in $L^2(0,1)$ with

$$\psi_{jk}(t) = 2^{j/2}\psi(2^{j/2}t - k) \tag{1.3}$$

for some function ψ and

$$C_{jk} = \int_0^1 h_F(s)\psi_{jk}(s)ds . \tag{1.4}$$

An asymptotically efficient estimate of the coefficients C_{jk} will yield an asymptotically efficient estimate of h_F.

A common approach is to fix the score function apriori on the basis of robustness or simplicity considerations. However, for efficient results, a good approximation of h_F

based on an approximate knowledge of F from a sample is of some value. To that end, Van Eeden (1970) proposed an asymptotically efficient estimate of location parameters using an estimate of h_F based on a subset of the data. Dionne (1981) used a similar subset-based technique to develop estimators of linear model parameters. Beran (1974), for the location model, and Naranjo and McKean (1997), for the linear model, provided Fourier series estimators of h_F based on the whole sample. A different approach to aforementioned methods that uses density estimation and based on quantiles regression was proposed by Koenker and Basset (1978).

The estimator proposed in this paper also uses the whole sample to estimate h_F. Our approach differs from that of Naranjo and McKean (1997) in that:

1. we develop estimates of h_F that provide asymptotically efficient estimators $\widehat{\boldsymbol{\beta}}_F$ based on a large class of orthonormal basis in $L^2(0, 1)$,

2. we develop estimates based on second order approximations (Beran (1974) and Naranjo and McKean (1997) used first order approximations),

3. we eliminate restrictive assumptions on the data such as those in assumption $(A6)$ of Naranjo and McKean (1997) by using second order approximations, and

4. we provide a consistent, wavelet-based, estimator of the asymptotic variance of the estimator $\widehat{\boldsymbol{\beta}}_F$.

Zygmund (1945) pointed out that second differences of functions are much more useful than first differences in estimating smooth functions. This motivates our use of second order approximations. Also, the use of the second derivative gives us expressions of coefficients that are easier to manipulate than the ones in Naranjo and McKean (1997). This allows us to avoid making restrictive distributional assumptions such as assumption $(A6)$ of Naranjo and McKean (1997) that asserts the first derivative of $(\phi(F))'F^{-1}$ be bounded, where $\phi(t) = \exp(-2\pi ikt)$ and k is an integer. This excludes a wide range of distributions such as the normal and the logistic.

This paper is organized as follows. Section 2 contains preliminary results needed for the estimation of the score function. In Section 3, we provide our estimate of the score function which will be used in Section 4 to find an asymptotically efficient rank estimate of the slope parameter in the linear model.

2 Preliminaries

We will assume without loss of generality that $\mu = 0$, $\boldsymbol{\beta} = \mathbf{0}$, and $\mathbf{x}_i$ are centered to have mean $\mathbf{0}$ in (1.1). Moreover, we assume the following conditions:

(H_1) ψ has compact support in $(0,1)$ and is three times differentiable with bounded derivatives.

(H_2) $\frac{1}{\sqrt{n}} \max_{1\le i\le n} \|\mathbf{x}_i\| = o(1)$.

(H_3) $\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}_i\|^2 = O(1)$.

(H_4) f is absolutely continuous with $I(f) < \infty$ with f'/f monotone.

(H_5) There exists a sequence $\{\widehat{\boldsymbol{\beta}}_n\}$ in $\mathbb{R}^p$ such that $\sqrt{n}\widehat{\boldsymbol{\beta}}_n = O_p(1)$.

Remark 2.1. Assumption (H_1) implies that $\psi_{jk}^{(l)} = O(2^{2j})$ for all $k \in \mathbb{N}$ and $l = 0, 1, 2, 3$. Assumption (H_4) implies that $f \in L^2(\mathbb{R})$, uniformly continuous, and uniformly bounded. The quantity $\widehat{\boldsymbol{\beta}}_n$ in (H_5) will be used as the initial estimator of $\boldsymbol{\beta}$ in (1.1). There are a number of estimators that satisfy assumption (H_5) including the least squares estimator and the general rank estimator with a specified score function h (Jureckova, 1971; Jaeckel, 1972).

The following lemma contains an alternative representation of the orthonormal basis coefficients C_{jk} in the expansion of h_F.

Lemma 2.2. *Assume* (H_1) *and* (H_4). *Then*

$$\int_0^1 h_F(t)\psi(t)dt = -\int_{-\infty}^{\infty} \frac{d^2}{dz^2}\Big\{\psi\big[F(z)\big]\Big\}F(z)dz\ . \tag{2.1}$$

Proof. Eq. (5) of Naranjo and McKean (1997) gives

$$\int_0^1 h_F(t)\psi(t)dt = \int_{-\infty}^{\infty} \frac{d}{dz}\left\{\psi[F(z)]\right\}dF(z)\ . \tag{2.2}$$

Integrating by parts the right-hand side of equation (2.2), we have

$$\int_{-\infty}^{\infty} \frac{d}{dz}\left\{\psi[F(z)]\right\}dF(z) = \left[F(z)\frac{d}{dz}\left\{\psi\left[F(z)\right]\right\}\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{d^2}{dz^2}\left\{\psi[F(z)]\right\}F(z)dz\ .$$

We can write

$$\left[F(z)\frac{d}{dz}\left\{\psi\left[F(z)\right]\right\}\right]_{-\infty}^{\infty} = \lim_{z\to\infty}\left[F(z)f(z)\psi'[F(z)] - F(-z)f(-z)\psi'[F(-z)]\right]\ .$$

But $\lim_{z\to\infty} F(z) = 1$, $\lim_{z\to-\infty} F(z) = 0$, $\lim_{z\to\infty} \psi'(F(z)) = \psi'(1) = 0 = \lim_{z\to-\infty} \psi'(F(z)) = \psi'(0)$ since ψ is compactly supported. The Lemma follows. □

Remark 2.3. If we replace assumption (H_1) in Lemma 2.2 by the assumption that f is absolutely continuous and $\int_{-\infty}^{\infty} |f'(x)|dx < \infty$, then (2.1) still holds. In fact, these two conditions insure that both $\lim_{z\to-\infty} f(z)$ and $\lim_{z\to\infty} f(z)$ exist and are both equal to zero. They are restrictive though since they require a well behaved source for the sample data. In comparison, we have numerous functions ψ satisfying (H_1) such as the

Daubechies base functions, Symlets and coiflets.

In the remainder of the article, for $\boldsymbol{\alpha} \in \mathbb{R}^p$, we will let $F_n(\cdot;\boldsymbol{\alpha})$ represent the empirical distribution function of $y_1 - \mathbf{x}_1^T\boldsymbol{\alpha}, \ldots, y_n - \mathbf{x}_n^T\boldsymbol{\alpha}$; that is, $F_n(z;\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^n I(y_i - \mathbf{x}_i^T\boldsymbol{\alpha} \leq z)$. The following lemma is a combination of Lemma 1 and Lemma 2 of Naranjo and McKean (1997) and gives the asymptotic linearity of $F_n(w;\boldsymbol{\alpha}_n)$ for $\boldsymbol{\alpha}_n$ converging to $\mathbf{0}$ at a suitable rate. The proof follows from Section 2.3 of Koul (1992) and will not be given here for the sake of brevity.

Lemma 2.4. *Assume $(H_1) - (H_5)$. Then*

$$\sup_{z\in\mathbb{R}} \sqrt{n}|F_n(z;\widehat{\boldsymbol{\beta}}_n) - F(z)| = \mathrm{O}_p(1)\ .$$

We now discuss some basic results on wavelets. For more details, including the definition of Besov spaces $B_p^{sq}(\mathbb{R})$, the reader may refer to Härdle et al. (1998).

Definition 2.5. A function $\varphi \in L^2(\mathbb{R})$ such that $\{\varphi(x-k), k \in \mathbb{Z}\}$ is an othonormal system (ONS) is called a scaling function or "father wavelet".

Note that the C_{jk} in (1.4) and ψ in (1.3) can be written as $C_{jk} = 2^{j/2}\int \overline{\psi(2^jx-k)}f(x)dx$ and $\psi(x) = \sqrt{2}\sum_k \lambda_k\varphi(2x-k)$, respectively, for some coefficients λ_k, where $\overline{\psi(\cdot)}$ is the complex conjugate of $\psi(\cdot)$. Moreover, if the scaling function φ is compactly supported, then the number of coefficients λ_k is finite. The function ψ is referred to in the literature as the "mother wavelet".

Characterizations of ONS of $L^2(0,1)$ can be found in Meyer (1991), Cohen et al. (1993) and Andersson et al. (1994). If φ satisfies conditions given in Theorem 9.6 of Härdle et al. (1998) (for example certain Daubechies wavelets), then h_F belongs to the Besov Space $B_2^{sq}(\mathbb{R})$. Besov spaces can be characterized using wavelet coefficients; thus, they are the natural spaces for wavelet estimation of functions. Moreover, in some Besov spaces, wavelet coefficients decay faster than Fourier coefficients. For instance, it is shown in Zygmund (2002) that if a function belongs to the Zygmund space $B_{\infty}^{1\infty}(0,1)$, then its Fourier coefficients C_n are $O(n^{-1})$. It was proved in Meyer (1990) that the wavelet coefficients W_n of such a function are $O(2^{-3n/2})$.

3 Estimation of Score Function

Let $\{\theta_n\}_{n\in\mathbb{N}}$ and $\{M_n\}_{n\in\mathbb{N}}$ be sequences of real numbers such that $M_n = O(n^{\alpha}), 0 < \alpha < 1/4$ and $\frac{M_n}{\sqrt{n}\theta_n^2} \to 0, M_n\theta_n^2 \to 0$ as $n \to \infty$. This means that $\theta_n = O(n^{\lambda})$ where $\lambda = \alpha/2 - 1/4 - \gamma$ and $\alpha - 1/4 < \gamma < 0$. Given a scaling function φ with corresponding "mother wavelet" ψ and a $\sqrt{n}$-consistent estimator $\widehat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}$, equation (2.1) in Lemma 2.2 suggests an estimator

$$\widehat{C_{jk}^n} := \frac{1}{\theta_n^2}\int_{-\infty}^{\infty}\Big\{2\psi_{jk}\big[F_n(z;\widehat{\boldsymbol{\beta}}_n)\big] - \psi_{jk}\big[F_n(z+\theta_n;\widehat{\boldsymbol{\beta}}_n)\big] - \psi_{jk}\big[F_n(z-\theta_n;\widehat{\boldsymbol{\beta}}_n)\big]\Big\}F_n(z;\widehat{\boldsymbol{\beta}}_n)dz$$

of $C_{jk} = \int_0^1 h_F(s)\psi_{jk}(s)ds$. Note that $\widehat{C_{jk}^n}$ can be computed from the data as

$$\widehat{C_{jk}^n} = \frac{2}{n\theta_n}\sum_{i=1}^{n} i\cdot\Big\{2\phi\big[F_n(e_i;\hat{\boldsymbol{\beta}}_n)\big] - \phi\big[F_n(e_i+\theta_n;\hat{\boldsymbol{\beta}}_n)\big] - \phi\big[F_n(e_i-\theta_n;\hat{\boldsymbol{\beta}}_n)\big]\Big\}\,, \tag{3.1}$$

where $e_i = y_i - \mathbf{x}_i^T\widehat{\boldsymbol{\beta}}_n$. In fact, let $W(z) = \displaystyle\int_{\eta}^{z} F(s)ds$, where η a zero of W. Then

$$-\int_{-\infty}^{\infty}\frac{d^2}{dz^2}\Big\{\phi[F(z)]\Big\}F(z)dz = -\sum_{i=1}^{n}\int_{e_i-\theta_n}^{e_i+\theta_n}\frac{d^2}{dz^2}\Big\{\phi[F_n(z;\hat{\boldsymbol{\beta}}_n)]\Big\}dW(z)\,. \tag{3.2}$$

For θ_n small enough so that there is only one e_j between $e_i-\theta_n$ and $e_i+\theta_n$, namely e_i, an approximation of the opposite of the integral on the right-hand side of equation (3.2) is

$$\left\{\frac{\phi[F_n(e_i+\theta_n;\widehat{\boldsymbol{\beta}}_n)]-2\phi[F_n(e_i;\widehat{\boldsymbol{\beta}}_n)]+\phi[F_n(e_i-\theta_n;\widehat{\boldsymbol{\beta}}_n)]}{\theta_n^2}\right\}\left[W(e_i+\theta_n)-W(e_i-\theta_n)\right]$$

$$=\left\{\frac{\phi[F_n(e_i+\theta_n;\widehat{\boldsymbol{\beta}}_n)]-2\phi[F_n(e_i;\widehat{\boldsymbol{\beta}}_n)]+\phi[F_n(e_i-\theta_n;\widehat{\boldsymbol{\beta}}_n)]}{\theta_n^2}\right\}\left[\int_{e_i-\theta_n}^{e_i+\theta_n}F_n(z;\widehat{\boldsymbol{\beta}}_n)dz\right]$$

$$=\frac{2}{n\theta_n}\left\{\phi[F_n(e_i+\theta_n;\widehat{\boldsymbol{\beta}}_n)]-2\phi[F_n(e_i;\widehat{\boldsymbol{\beta}}_n)]+\phi[F_n(e_i-\theta_n;\widehat{\boldsymbol{\beta}}_n)]\right\}\left[\sum_{j=1}^{n}I(e_j\le\xi_i)\right],$$

where $\xi_i\in[e_i-\theta_n,e_i+\theta_n]$. Note that we can replace e_i by their order statistics; therefore, the form of $\widehat{C_{jk}^n}$ proposed in equation (3.1) follows.

The following lemma establishes the consistency of $\widehat{C_{jk}^n}$.

Lemma 3.1. *Suppose that* $(H_1)-(H_5)$ *are satisfied. Then* $|M_n(\widehat{C_{jk}^n}-C_{jk})|=o_p(1)$

Proof. Define $\widetilde{C_{jk}^n}=-\int_{-\infty}^{\infty}\frac{d^2}{dz^2}\Big\{\psi_{jk}\big[F(z)\big]\Big\}F_n(z;\widehat{\boldsymbol{\beta}}_n)dz$. Now

$$\widetilde{C_{jk}^n}=\frac{1}{\theta_n^2}\int_{-\infty}^{\infty}\left\{2\psi_{jk}\left[F(z)\right]-\psi_{jk}\left[F(z+\theta_n)\right]-\psi_{jk}\left[F(z-\theta_n)\right]\right\}F_n(z;\widehat{\boldsymbol{\beta}}_n)dz+\mathrm{O}(M_n\theta_n^2)\ .$$

Taking the difference $\widehat{C_{jk}^n}-\widetilde{C_{jk}^n}$ and expanding $\psi_{jk}(F_n)$ about $\psi_{jk}(F)$, we have

$$\begin{aligned}M_n(\widehat{C_{jk}^n}-\widetilde{C_{jk}^n}) \;=\; & \frac{2M_n}{\theta_n^2\sqrt{n}}\int_{-\infty}^{\infty}\Big\{\sqrt{n}[F_n(z;\widehat{\boldsymbol{\beta}}_n)-F(z)]\Big\}\,\psi'_{jk}(\xi_{1,n}(z))F_n(z;\widehat{\boldsymbol{\beta}}_n)dz\\ & -\frac{M_n}{\theta_n^2\sqrt{n}}\int_{-\infty}^{\infty}\Big\{\sqrt{n}[F_n(z+\theta_n;\widehat{\boldsymbol{\beta}}_n)-F(z+\theta_n)]\Big\}\,\psi'_{jk}(\xi_{2,n}(z))F_n(z;\widehat{\boldsymbol{\beta}}_n)dz\\ & -\frac{M_n}{\theta_n^2\sqrt{n}}\int_{-\infty}^{\infty}\Big\{\sqrt{n}[F_n(z-\theta_n;\widehat{\boldsymbol{\beta}}_n)-F(z-\theta_n)]\Big\}\,\psi'_{jk}(\xi_{3,n}(z))F_n(z;\widehat{\boldsymbol{\beta}}_n)dz\\ & +O(M_n\theta_n^2)\ ,\end{aligned}$$

where $\xi_{1,n}(z)$ is between $F_n(z;\widehat{\boldsymbol{\beta}}_n)$ and $F(z)$, $\xi_{2,n}(z)$ is between $F_n(z+\theta_n;\widehat{\boldsymbol{\beta}}_n)$ and $F(z+\theta_n)$, $\xi_{3,n}(z)$ is between $F_n(z-\theta_n;\widehat{\boldsymbol{\beta}}_n)$ and $F(z-\theta_n)$. Since ψ'_{jk} and F_n are bounded with respect to n and $\sup_z|\sqrt{n}(F_n(z;\widehat{\boldsymbol{\beta}}_n) - F(z))| = \mathrm{O}_p(1)$, we have

$$|M_n(\widehat{C^n_{jk}} - \widetilde{C^n_{jk}})| = O_p(M_n/\theta_n^2\sqrt{n}) + O(M_n\theta_n^2). \tag{3.3}$$

On the other hand,

$$\begin{aligned} M_n(C^n_{jk} - \widetilde{C^n_{jk}}) &= -\frac{M_n}{\theta_n^2}\int_{-\infty}^{\infty}(-\theta_n^2)[\psi_{jk}(F)]''(z)F_n(z;\widehat{\boldsymbol{\beta}}_n)dz - \int_{-\infty}^{\infty}[\psi_{jk}(F)]''(z)F(z)dz \\ &\quad - \frac{M_n}{\theta_n^2}\int_{-\infty}^{\infty}\left(-\frac{\theta_n^3}{6}\right)\left\{[\psi_{jk}(F)]'''[\kappa_1(z)] + [\psi_{jk}(F)]'''[\kappa_2(z)]\right\}F_n(z;\widehat{\boldsymbol{\beta}}_n)dz \\ &\quad + O(M_n\theta_n^2) \\ &= M_n\int_{-\infty}^{\infty}\left[F_n(z;\widehat{\boldsymbol{\beta}}_n) - F(z)\right][\psi_{jk}(F)]''(z)dz + O_p(M_n\theta_n) + O(M_n\theta_n^2);, \end{aligned}$$

where $\kappa_1(z) \in (z-\theta_n, z)$ and $\kappa_2(z) \in (z, z+\theta_n)$.

Thus we have

$$M_n(C^n_{jk} - \widetilde{C^n_{jk}}) = \frac{M_n}{\sqrt{n}}\int_{-\infty}^{\infty}\left\{\sqrt{n}\left[F_n(z;\widehat{\boldsymbol{\beta}}_n) - F(z)\right]\right\}[\psi_{jk}(F)]''(z)dz + O_p(M_n\theta_n) + O(M_n\theta_n^2)\ .$$

But

$$\int_{-\infty}^{\infty}[\psi_{jk}(F)]''(z)dz = \int_{-\infty}^{\infty}f'(z)\psi''_{jk}[F(z)]dz + \int_{-\infty}^{\infty}f^2(z)\psi'_{jk}[F(z)]dz)\ .$$

The two integrals on the right are bounded since f is absolutely continuous, $f \in L^2(\mathbb{R})$, and ψ_{jk} has bounded derivatives. Therefore we have

$$|M_n(C^n_{jk} - \widetilde{C^n_{jk}})| = O_p(M_n/\sqrt{n}) + O_p(M_n\theta_n) + O(M_n\theta_n^2). \tag{3.4}$$

Equations (3.3) and (3.4) imply that

$$|M_n(C^n_{jk} - \widehat{C^n_{jk}})| = O_p(M_n/\theta_n^2\sqrt{n}) + O_p(M_n/\sqrt{n}) + O(M_n\theta_n^2) = o_p(1)\ . \tag{3.5}$$

□

Remark 3.2. In view of Remark 2.1, this means that there is a constant $L > 0$ such that

$$|\widehat{C^n_{jk}} - C_{jk}| \leq L\frac{2^{2j}}{M_n} , \quad \text{for all } k \in \mathbb{N} .$$

Now define the wavelet estimator of h_F as

$$\widehat{h^n_F}(t) = \sum_{j=0}^{j_1} \sum_k \widehat{C^n_{jk}} \psi_{jk}(t) ,$$

where j_1 is some chosen resolution level in $\mathbb{N} \cup \{0\}$. Note that since we are using compactly supported wavelets, the sum over k contains only a finite number of terms for a given value of t (see Remark 10.1 on p. 127 of Härdle et al. (1998)).

Theorem 3.3. *Under* $(H_1) - (H_5)$*, we have* $E\|h_F - \widehat{h^n_F}\|_2^2 = o(1)$*.*

Proof. Let

$$h_F(t) = \sum_{j=0}^{j_1} \sum_k \widehat{C_{jk}} \psi_{jk}(t) + \sum_{j>j_1} \sum_k C_{jk}\psi_{jk}(t) ,$$

where the convergence is absolute in $L^2(0, 1)$. Thus

$$E\|h_F - \widehat{h^n_F}\|_2^2 \leq 2E \left(\int_0^1 \left| \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C^n_{jk}})\psi_{jk}(t) \right|^2 dt \right) + 2E \left(\int_0^1 \left| \sum_{j>j_1} \sum_k C_{jk}\psi_{jk}(t) \right|^2 dt \right) .$$

The second term on the right is $o(1)$ by absolute convergence in $L^2(0, 1)$.

For the first term, we have

$$\int_0^1 \left| \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C^n_{jk}})\psi_{jk}(t) \right|^2 dt \leq \int_0^1 \sum_{j=0}^{j_1} (j_1 + 1) \left(\sum_k |C_{jk} - \widehat{C^n_{jk}}| \, |\psi_{jk}(t)| \right)^2 dt .$$

But there is a positive constant L such that $|C_{jk} - \widehat{C_{jk}^n}| \leq L\frac{2^{2j}}{M_n}$ (see Remark 3.2). Thus

$$\int_0^1 \left| \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C_{jk}^n}) \psi_{jk}(t) \right|^2 dt \leq L^2 (j_1 + 1) \sum_{j=0}^{j_1} \frac{2^{4j}}{M_n^2} \int_0^1 \left(\sum_k |\psi_{jk}(t)| \right)^2 dt \ .$$

Since, by Proposition 8.3 of Härdle et al. (1998), the integral on the right hand side is uniformly bounded in j, there is a constant $C > 0$ such that

$$\int_0^1 \left| \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C_{jk}^n}) \psi_{jk}(t) \right|^2 dt \leq \frac{C}{M_n^2} (j_1 + 1) \sum_{j=0}^{j_1} 2^{4j} \leq \frac{L}{M_n^2} (j_1 + 1) 2^{4j_1+1}$$

Choosing j_1 such that $(j_1 + 1) 2^{4j_1+1} < M_n^2$ completes the proof. □

4 Estimation of Slope Parameter

Define

$$\mathbf{U}_n(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i h_F \left[\frac{R(y_i - \mathbf{x}_i' \boldsymbol{\beta})}{n+1} \right] \ .$$

Let $\widehat{\mathbf{U}}_n(\boldsymbol{\beta})$ denote the same expression with h_F replaced by $\widehat{h_F^n}$. Let $\mathbf{X}$ be the $n \times p$ matrix with $\mathbf{x}_i'$ as its ith row. Without loss of generality, we will assume that the design matrix $\mathbf{X}$ is centered; i.e, $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$.

Theorem 4.1. *Under* (H_1) - (H_5),

$$\widehat{\mathbf{U}}_n(\mathbf{0}) \sim AN(\mathbf{0}, \Sigma) \ ,$$

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

Proof. Heiler and Willers (1988) have shown that $\mathbf{U}_n(\mathbf{0}) \sim AN(\mathbf{0}, \Sigma)$. We will have $\widehat{\mathbf{U}}_n(\mathbf{0}) \sim AN(\mathbf{0}, \Sigma)$, if it can be shown that $\widehat{\mathbf{U}}_n(\mathbf{0}) - \mathbf{U}_n(\mathbf{0}) = o_p(1)$. Our approach follows that of Naranjo and McKean (1997) closely with minor modifications to suit wavelets.

It is enough to show that $\widehat{\mathbf{U}}_n(\mathbf{0}) - \mathbf{U}_n(\mathbf{0}) = o_p(1)$ elementwise; so, assume that $\mathbf{U}_n \equiv U_n$ is scalar. Suppose j_1 is such that $2^{4j_1+1} < M_n$. Note that $2^{4j_1+1} < M_n$ implies that $(j_1+1)2^{4j_1+1} < M_n^2$ as required by Theorem 3.3. Let t be a threshold such that $C_{jk} = C_{jk} I(\max_k |C_{jk}| < t)$. We have

$$\begin{aligned} \widehat{U}_n(0) - U_n(0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \left\{ \widehat{h_F^n} \left[\frac{R(y_i)}{n+1} \right] - h_F \left[\frac{R(y_i)}{n+1} \right) \right\} \\ &= \sum_{j=0}^{j_1} \sum_k (C_{jk} - \widehat{C_{jk}^n}) \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \psi_{jk} \left[\frac{R(y_i)}{n+1} \right] \end{aligned}$$

Since $|C_{jk} - \widehat{C_{jk}^n}| = o_p(2^{2j}/M_n)$, it suffices to show that

$$\sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \sum_k \psi_{jk} \left[\frac{R(y_i)}{n+1} \right] = O_p(2^{2j}) \ .$$

This follows from Chebychev's inequality if

$$E \left\{ \sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \sum_k \psi_{jk} \left[\frac{R(y_i)}{n+1} \right] \right\}^2 = O(2^{4j}) \ .$$

To that end, letting $K_{ij} = \sum_k \psi_{jk} \left[\frac{R(y_i)}{n+1} \right]$, we have

$$E \left\{ \sum_{i=1}^{n} \frac{x_i}{\sqrt{n}} \sum_k \psi_{jk} \left[\frac{R(y_i)}{n+1} \right] \right\}^2 = \sum_i (x_i^2/n) E(K_{ij}^2) + \sum_{r \neq s} (x_r x_s / n) E(K_{rj} K_{sj}) \ .$$

But by Theorem 9.6 of Härdle et al. (1998), we have $K_{ij} = O(2^{2j})$ and because $(\sum_i x_i)^2 = 0$, we have by (H_3) that $\sum_i (x_i^2/n) = \sum_{r \neq s} (x_r x_s/n) = O(1)$. The proof is complete. □

Given an initial estimator $\widehat{\boldsymbol{\beta}}_n$, define the one-step estimator as

$$\widehat{\boldsymbol{\beta}}_F^* = \widehat{\boldsymbol{\beta}}_n + \tau\sqrt{n}(\mathbf{X}'\mathbf{X})^{-1}\widehat{\mathbf{U}}(\widehat{\boldsymbol{\beta}}_n) \ , \tag{4.1}$$

where $\tau^{-1} = \int_0^1 |h_F(t)|^2 dt$. The estimators $\widehat{\boldsymbol{\beta}}_F^*$ and $\widehat{\boldsymbol{\beta}}_F$ have the same asymptotic distribution as given in the following theorem. The proof is direct and will not be given here for the sake of brevity.

Theorem 4.2. *If* (H_1)- (H_5) *are satisfied, then* $\widehat{\boldsymbol{\beta}}_F^* \sim AN(0, \tau^2\Sigma^{-1})$.

For practical applications of Theorem 4.2, one needs a consistent estimator of τ^{-1}. Koul et al. (1987) have given a consistent estimator of τ^{-1} for the case where the score function is known. Their estimator is based on a kernel density estimator of the density of the errors based on the residuals of the model. The following theorem gives a consistent estimator of τ^{-1} for the case of estimated scores.

Theorem 4.3. *Define* $(\widehat{\tau_F^n})^{-1} = \int_0^1 |\widehat{h_F^n}(t)|^2 dt$. *Then, under* (H_1) - (H_5),

$$(\widehat{\tau_F^n})^{-1} - \tau^{-1} \xrightarrow{\mathcal{P}} 0 \ .$$

Proof. Note that

$$\left| \int_0^1 |\widehat{h_F^n}(t)|^2 - |h_F(t)|^2 dt \right| \leq \int_0^1 \left| |\widehat{h_F^n}(t)|^2 - |h_F(t)|^2 \right| dt \leq \int_0^1 \left| \widehat{h_F^n}(t) - h_F(t) \right|^2 dt \ .$$

Thus

$$P\left(\left| (\widehat{\tau_F^n})^{-1} - \tau^{-1} \right| > \epsilon \right) \ \leq \ P\left(\int_0^1 \left| \widehat{h_F^n}(t) - h_F(t) \right|^2 dt > \epsilon \right) \ ,$$

which is bounded by $\epsilon^{-1} E\|h_F - \widehat{h_F^n}\|_2^2$ by Markov's inequality. The desired result follows from Theorem 3.3. □

One may estimate $(\widehat{\tau_F^n})^{-1}$ using numerical integration methods (such as Gaussian quadrature) using some grid on $(0,1)$ since the $\widehat{h_F^n}(t)$ can be computed for any given $t \in (0,1)$.

As one application, consider testing the general linear hypothesis

$$H_0 : \mathbf{M}\boldsymbol{\beta} = \mathbf{0} \qquad \text{versus} \qquad H_1 : \mathbf{M}\boldsymbol{\beta} \neq \mathbf{0} \;,$$

where $\mathbf{M}$ is a $q \times p$ matrix of full row rank forming linear constraints. Under H_0, the quantity

$$B_{\mathbf{M}} = \frac{\left(\mathbf{M}\widehat{\boldsymbol{\beta}}_F^*\right)' [\mathbf{M}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{M}']^{-1} \left(\mathbf{M}\widehat{\boldsymbol{\beta}}_F^*\right)}{q(\widehat{\tau_F^n})^2} \;.$$

is asymptotically $\chi^2(q)$ by Theorem 4.2, Theorem 4.3, and Slutsky's Lemma. Thus a level-α Wald test rejects H_0 if $B_{\mathbf{M}}$ exceeds the $(1-\alpha)$ quantile of the $\chi^2(q)$ distribution.

5 Discussion

In this paper, we developed an asymptotically efficient rank estimator based on score functions estimated using wavelets. A consistent estimator for τ is given for the asymptotic variance of the rank estimator. This can be used in constructing Wald tests of general linear hypotheses.

In our treatment we assumed that the errors are independent and identically distributed with cdf F that has pdf f. Estimating the score function to maximize efficiency is generally a very difficult problem for dependent error models. In simple dependent data problems, however, this may be tractable using some of the methodology developed in this paper. For example, consider the model given in (1.1) with $\varepsilon_1, \ldots, \varepsilon_n$ exchangeable and having the same marginal distribution G. Assuming $\mathbf{x}$ is centered, the asymptotic relative efficiency of the rank estimator versus the least squares estimator is given by

$$\text{ARE} = \frac{\sigma^2(1-\rho)}{\tau_\varphi^2(1-\rho_\varphi)} ,$$

where $\rho = \text{Corr}(\varepsilon_1, \varepsilon_2)$, $\sigma^2 = \text{Var}(\varepsilon_1)$, $\rho_\varphi = \text{Corr}[\varphi(G(\varepsilon_1)), \varphi(G(\varepsilon_2))]$, and $\tau_\varphi^{-1} = \int_0^1 \varphi(u) \left\{ -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \right\} du$. We would like to find φ that maximizes the ARE. This amounts to minimizing τ_φ and maximizing ρ_φ. However, the function that minimizes τ_φ does not necessarily maximize ρ_φ. Analytically finding φ that would simultaneously do both is a difficult problem in calculus of variations. The wavelet method developed in this paper provides an approximation the minimizer of τ_φ. The maximizer of ρ_φ can also be found using the techniques in this paper since the wavelet basis of $L^2(\mathbb{R}^2)$ can be given as a product of wavelet basis in $L^2(\mathbb{R})$. Thus we have two approximations of φ. The compromise is to use the estimated score in one in the estimation of the other. This could be iterated until the difference between the two approximations is below a specified level of tolerance. Either one or the average of the two approximations can be taken as the final approximation of φ.

References

Andersson, L., Hall, N., Jawerth, B., and Peters, G. (1994). Wavelets on closed subsets of the real line. In *Recent advances in wavelet analysis*, volume 3 of *Wavelet Anal. Appl.*, pages 1–61. Academic Press, Boston, MA.

Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.*, **2**:63–74.

Cohen, A., Daubechies, I., and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, **1**(1):54–81.

Dionne, L. (1981). Efficient nonparametric estimators of parameters in the general linear hypothesis. *Ann. Statist.*, **9**(2):457–464.

Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Heiler, S. and Willers, R. (1988). Asymptotic normality of R-estimates in the linear model. *Statistics*, **19**(2):173–184.

Hettmansperger, T. P. and McKean, J. W. (1998). *Robust nonparametric statistical methods*, volume 5 of *Kendall's Library of Statistics*. Edward Arnold, London.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.*, **43**(5):1449–1458.

Jureckova, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Statist.*, **42**(4):1328–1338.

Koenker, R., Basset, G. (1987). Regression Quantiles. *Econometrica.*, **46**, 33–50.

Koul, H. L. (1992). *Weighted empiricals and linear models*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 21. Institute of Mathematical Statistics, Hayward, CA.

Koul, H. L., Sievers, G. L., and McKean, J. (1987). An estimator of the scale parameter for the rank analysis of linear models under general score functions. *Scand. J. Statist.*, **14**(2):131–141.

Meyer, Y. (1990). *Ondelettes et opérateurs. II*. Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris. Opérateurs de Calderón-Zygmund. [Calderón-Zygmund operators].

Meyer, Y. (1991). Ondelettes sur l'intervalle. *Rev. Mat. Iberoamericana*, 7(2):115–133.

Naranjo, J. D. and McKean, J. W. (1997). Rank regression with estimated scores. *Statist. Probab. Lett.*, **33**(2):209–216.

van Eeden, C. (1970). Efficiency-robust estimation of location. *Ann. Math. Statist.*, 41:172–181.

Zygmund, A. (1945). Smooth functions. *Duke Math. J.*, 12:47–76.

Zygmund, A. (2002). *Trigonometric series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition. With a foreword by Robert A. Fefferman.