Aligned rank tests for the linear model with heteroscedastic errors

Michael G. Akritas

The Pennsylvania State University, University Park, USA

Willem Albers

Twente University of Technology, Enschede, The Netherlands

Received 3 June 1991; revised manuscript received 18 August 1992

Abstract: We consider the problem of testing subhypotheses in a heteroscedastic linear regression model. The proposed test statistics are based on the ranks of scaled residuals obtained under the null hypothesis. Any estimator that is $n^{1/2}$ -consistent under the null hypothesis can be used to form the residuals. The error variances are estimated through a parametric model. This extends the theory of aligned rank tests to the heteroscedastic linear model. A real data set is used to illustrate the procedure.

AMS Subject Classification: Primary 62J05, 62G10; secondary 62G35.

Key words: Linear regression; subhypotheses; rank transformation; Pitman efficacy; empirical processes.

1. Introduction

Let $(y_i, x_i, \varepsilon_i)$, $i = 1, \dots, n$, be independent identically distributed (i.i.d.) $(p+2)$ -dimensional random vectors such that x_i is p -dimensional, ε_i and x_i are independent and

$$y_i = \beta' x_i + \sigma_i \varepsilon_i = \beta'_1 x_{1i} + \beta'_2 x_{2i} + \sigma_i \varepsilon_i, \quad i = 1, \dots, n. \quad (1.1)$$

Here $\beta' = (\beta'_1, \beta'_2)$ is an unknown p -dimensional parameter with β_1 q -dimensional, $y_i, x'_i = (x'_{1i}, x'_{2i})$ are observable with x_{1i} q -dimensional, the errors ε_i have symmetric marginal distribution F (see the Appendix for exact assumptions), and the σ_i 's are scaling constants which express possible heteroscedasticity. We will assume that the scaling constants are of the form

$$\log(\sigma_i) = \theta' z_i \quad (1.2)$$

Correspondence to: Prof. Michael Akritas, The Pennsylvania State University, Department of Statistics, 219 Pond Laboratory, University Park, PA 16802, USA.

where θ is an $r \times 1$ vector of unknown coefficients and z_i are observable random vectors of dimension r , independent of the ε_i . This is one of the common ways for modeling the variance function (Carroll and Ruppert, 1988). It should be noted that the particular model for the scaling constants is not essential to the proofs, even though different models may require somewhat different assumptions. For instance, the model $\sigma_i = |\beta'x_i|^\theta$ would require that the covariates x_i have finite support and the σ_i 's be strictly bounded away from zero.

Model (1.1) has been considered extensively over the past twenty years. The main focus of this research has been parameter estimation and variance function estimation. See Carroll and Ruppert (1988) for an excellent account of this work. Parameter estimation, however, is only the first step in the analysis of data arising from a linear model. For instance, a classical least squares analysis often focuses upon the analysis of variance, which tests simultaneous hypotheses on large subsets of the parameters. Given the attention the heteroscedastic model has received, it is surprising to note that the testing problem has been essentially overlooked. It is our purpose to provide a class of tests that are easy to apply and thus can serve as a useful alternative two Wald-type tests.

In this paper we are interested in testing the hypothesis

$$H_0: \beta_2 = 0. \quad (1.3)$$

We will consider tests statistics that are based on the ranks of the scaled (reduced model) residuals. This class of statistics will be called *weighted aligned rank statistics*. Aligned rank tests were first proposed by Hodges and Lehmann (1962) in a two-way layout context. Puri and Sen (1971; 1985) derive aligned rank tests for a number of factorial designs. The theory of aligned rank tests for regression is developed by Chiang and Puri (1984); the same problem was also considered by Adichie (1978) but he imposed restrictive assumptions on the design matrix. Other robust tests in the homoscedastic linear model are the M-tests of Schrader and Hettmansperger (1980), Sen (1982) and Ronchetti (1982a, b). (See also Hampel et al., 1986, Chapter 7.) None of these has been extended to the heteroscedastic regression model.

In the next section we outline the methodology and derive the asymptotic distribution of the aligned rank test statistic under the null hypothesis. The Pitman efficacy is derived in Section 3. Section 4 illustrates the test statistic with a real data set. Finally, some technical derivations are given in the Appendix. All the derivations will be presented in terms of logistic scores for the sake of simplicity. However, the present method can accommodate other score functions too.

2. The test statistic

Consider model (1.1) for the observations (y_i, x_i) and let $X = (x_{ij})$ denote the matrix with rows x_i' ; it should be clear from the context when a double subscripted x refers to an element of X or to a partition of x_i (see relation (1.1)). A similar sub-

script notation applies for the elements of the $n \times p$ matrices $D = (D_1, D_2)$ and $\hat{D} = (\hat{D}_1, \hat{D}_2)$ introduced below. In matrix form the model (1.1) can be written as

$$\Sigma^{-1}Y = \Sigma^{-1}X_1\beta_1 + \Sigma^{-1}X_2\beta_2 + \varepsilon$$

where Σ is the diagonal matrix with diagonal elements σ_i , $X = (X_1, X_2)$ with X_1 being $n \times q$, $Y = (y_1, \dots, y_n)'$, and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$. Note that the model for the scaled (y_i, x_i) is homoscedastic. Thus if $\hat{\beta}_1$ is, under H_0 , a $n^{1/2}$ -consistent estimator, the aligned rank test for H_0 is based on the convergence (cf. Akritas, 1991)

$$\tilde{W}'D_2(D_2'D_2)^{-1}D_2'\tilde{W} \rightarrow \frac{1}{12}\chi_{p-q}^2 \quad (2.1)$$

where $\tilde{W}$ is n^{-1} times the vector of ranks of the $(y_i - \hat{\beta}_1'x_{1i})/\sigma_i$, and $D_2 = (I - P)\Sigma^{-1}X_2$, where P is the matrix that projects on the space spanned by the columns of $D_1 = \Sigma^{-1}X_1$ and the unit vector $\mathbf{1} = (1, \dots, 1)'$. However, since the σ_i are unknown (and thus $\tilde{W}$, $D = (D_1, D_2)$ are also unknown), (2.1) cannot be used to test H_0 . In what follows we will show that

$$W^{**'}\hat{D}_2(\hat{D}_2'\hat{D}_2)^{-1}\hat{D}_2'W^{**} \rightarrow \frac{1}{12}\chi_{p-q}^2, \quad (2.2)$$

where W^{**} is n^{-1} times the vector of ranks of the aligned and scaled data $W_i = (y_i - \hat{\beta}_1'x_{1i})/\hat{\sigma}_i$, $\hat{\sigma}_i = \exp(\hat{\theta}'z_i)$, and $\hat{D}_2 = (I - \hat{P})\hat{\Sigma}^{-1}X_2$, $\hat{\Sigma}$ being the diagonal matrix with diagonal elements $\hat{\sigma}_i$ and $\hat{P}$ the matrix that projects on the space spanned by the columns of $\hat{D}_1 = \hat{\Sigma}^{-1}X_1$ and the unit vector $\mathbf{1}$.

Let d_{ij} ($\hat{d}_{ij}$) denote the elements of the matrix D ($\hat{D}$) and note that for $j = 1, \dots, q$, $d_{ij} = x_{ij}/\sigma_i$ ($\hat{d}_{ij} = x_{ij}/\hat{\sigma}_i$). We will also use the notation $\hat{d}_{ij}^* = x_{ij}/\hat{\sigma}_i$, $d_{ij}^* = x_{ij}/\sigma_i$. Then, the model for the aligned and scaled data is

$$\begin{aligned} W_i &= (y_i - \hat{\beta}_1'x_{1i})/\hat{\sigma}_i \\ &= A_i + \beta_2'\hat{d}_{2i}^* + \varepsilon_i\sigma_i/\hat{\sigma}_i, \end{aligned} \quad (2.3)$$

where $A_i = (\beta_1 - \hat{\beta}_1)'\hat{d}_{1i}$. To obtain a suitable representation of the ranks of the W_i 's, let $\Delta' = (\Delta'_1, \Delta'_2)$, where Δ_1 is q -dimensional and Δ_2 is r -dimensional, and define

$$F_n(x; \Delta) = n^{-1} \sum_{i=1}^n I\left(\frac{y_i - \beta_1'x_{1i} - \Delta_1'x_{1i}}{\sigma_i} \leq x\alpha_i(\Delta)\right), \quad (2.4)$$

where $\alpha_i(\Delta) = \exp(\Delta_2'z_i)$. Note that, under H_0 , $F_n(x; 0) = F_n(x)$ is the empirical distribution function of the errors ε_i . Also note that at $\hat{\Delta} = (\hat{\beta}_1 - \beta_1, \hat{\theta} - \theta)$, $\alpha_i(\hat{\Delta}) = \exp[(\hat{\theta} - \theta)'z_i] = \hat{\sigma}_i/\sigma_i$, so that $F_n(x; \hat{\Delta})$ is the empirical distribution function of the W_i 's. It follows that any statistic that is based on the ranks of the W_i 's can be expressed in terms of the $F_n(W_i; \hat{\Delta})$'s. The transformation $W_i \rightarrow F_n(W_i; \hat{\Delta})$ will be referred to as the *rank transformation*. Also let

$$F(x; \Delta) = n^{-1} \sum_{i=1}^n F(x\alpha_i(\Delta) + \Delta_1'd_{1i}) \quad (2.5)$$

so that $F(x; 0) = F(x)$.

Lemma 2.1. Let $\hat{\Delta} = (\hat{\beta}_1 - \beta_1, \hat{\theta} - \theta)$. Then, under H_0 ,

$$n^{1/2} \|F_n(\cdot; \hat{\Delta}) - F(\cdot)\| = O_p(1),$$

where $\|\cdot\|$ denotes sup-norm.

The proof is similar to the proof of Lemma 3.1 of Akritas (1991).

The present methodology constructs first a test statistic based on the *transformed data* $F(W_1), \dots, F(W_n)$. Note that by Lemma 2.1, the transformation $W_i \rightarrow F(W_i)$ is the asymptotic form of the rank transformation. The transformed data are then replaced by the *rank transformed data* $F_n(W_i; \hat{\Delta})$, $i = 1, \dots, n$, in the test statistic and it is shown that the two statistics are asymptotically equivalent.

Under $H_0: \beta_2 = 0$, the model for the transformed data is

$$\begin{aligned} F(W_i) &= F(A_i + \varepsilon_i \sigma_i / \hat{\sigma}_i) = F(\varepsilon_i \sigma_i / \hat{\sigma}_i) + A_i f(\varepsilon_i \sigma_i / \hat{\sigma}_i + \tilde{A}_i) \\ &= F(\varepsilon_i) + [F(\varepsilon_i \sigma_i / \hat{\sigma}_i) - F(\varepsilon_i)] + A_i f(\varepsilon_i \sigma_i / \hat{\sigma}_i + \tilde{A}_i) \\ &= 0.5 + [F(\varepsilon_i \sigma_i / \hat{\sigma}_i) - F(\varepsilon_i)] + A_i f(\varepsilon_i \sigma_i / \hat{\sigma}_i + \tilde{A}_i) + e_i \end{aligned} \quad (2.6)$$

where $\tilde{A}_i$ is between zero and A_i , and $e_i = [F(\varepsilon_i) - 0.5]$. Thus $\text{Var}(e_i) = 1/12$, for all i . We want to show that a quadratic form in the $F(W_i)$'s has asymptotically a chi-squared distribution. To do this we will show that the second and third terms on the right hand side of (2.6) do not affect the asymptotic distribution of the quadratic form (Lemma 2.2). Indeed the usual asymptotic theory for the linear model implies that

$$V' D_2 (D_2' D_2)^{-1} D_2' V \rightarrow \frac{1}{12} \chi_{p-q}^2 \quad (2.7)$$

in distribution, where $V = (V_1, \dots, V_n)'$ with $V_i = 0.5 + e_i$. Relation (2.7) and the following lemma yield the desired result.

Lemma 2.2. Let $W_i^* = F(W_i)$ and $W^* = (W_1^*, \dots, W_n^*)'$. Then under the assumptions stated in the Appendix,

$$V' D_2 (D_2' D_2)^{-1} D_2' V - W^{*'} D_2 (D_2' D_2)^{-1} D_2' W^* \rightarrow 0 \quad (2.8)$$

in probability under H_0 .

The proof is given in the Appendix.

Next we show that the transformed data can be replaced in the quadratic form by the rank transformed data without altering its asymptotic distribution.

Lemma 2.3. Let $W_i^{**} = F_n(W_i; \hat{\Delta})$ and $W^{**} = (W_1^{**}, \dots, W_n^{**})'$, where $\hat{\Delta}$ is defined in Lemma 2.1. Then, under the assumptions stated in the Appendix,

$$W^{**'} D_2 (D_2' D_2)^{-1} D_2' W^{**} - W^{*'} D_2 (D_2' D_2)^{-1} D_2' W^* \rightarrow 0 \quad (2.9)$$

in probability under $H_0: \beta_2 = 0$.

The proof is given in the Appendix.

Finally we show that we can replace the matrix D_2 by the matrix $\hat{D}_2$ in the quadratic form without altering its asymptotic distribution.

Theorem 2.1. *Under the assumptions stated in the Appendix the convergence in (2.2) is true.*

The proof is given in the Appendix.

3. Pitman efficacy

In this section we will consider the sequence of alternatives

$$H_n: \beta_{2n} = n^{-1/2} h, \quad \text{for some } h \in \mathbb{R}^{p-q} \quad (3.1)$$

and will derive the asymptotic distribution of the statistic in (2.2) in terms of the quantity

$$\gamma^2 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n (h' d_{2i}^*)^2. \quad (3.2)$$

From (2.3) it follows that under H_n the model for the transformed aligned and scaled data is (for convenience we write β_2 instead of β_{2n} in what follows)

$$\begin{aligned} F(W_i) &= F(A_i + \beta_2' \hat{d}_{2i}^* + \varepsilon_i \sigma_i \hat{\sigma}_i^{-1}) \\ &= F(\beta_2' \hat{d}_{2i}^* + \varepsilon_i \sigma_i \hat{\sigma}_i^{-1}) + A_i f(\beta_2' \hat{d}_{2i}^* + \varepsilon_i \sigma_i \hat{\sigma}_i^{-1} + \tilde{A}_i) \\ &= F(\beta_2' d_{2i}^* + \varepsilon_i) + [F(\beta_2' \hat{d}_{2i}^* + \varepsilon_i \sigma_i \hat{\sigma}_i^{-1}) - F(\beta_2' d_{2i}^* + \varepsilon_i)] \\ &\quad + A_i f(\beta_2' \hat{d}_{2i}^* + \varepsilon_i \sigma_i \hat{\sigma}_i^{-1} + \tilde{A}_i). \end{aligned} \quad (3.3)$$

From Lemma 3.2 in Akritas (1991) it follows that the mean value and variance of the 'errors' $e_i^* = F(\beta_2' d_{2i}^* + \varepsilon_i)$ satisfy

$$\begin{cases} \sup\{|\text{Var}(e_i^*) - 1/12|; i = 1, \dots, n\} = o(1) \\ E(e_i^*) = 0.5 + \beta_2' d_{2i}^* E[f(\varepsilon_i + \delta_{2i})], \end{cases} \quad (3.4)$$

where δ_{2i} is between zero and $\beta_2' d_{2i}^*$. Adding and subtracting the $E(e_i^*)$ from the right hand side of (3.3) gives

$$\begin{aligned} F(W_i) &= 0.5 + \beta_2' d_{2i}^* E f(\varepsilon_i + \delta_{2i}) + [F(\beta_2' \hat{d}_{2i}^* + \varepsilon_i \sigma_i \hat{\sigma}_i^{-1}) - F(\beta_2' d_{2i}^* + \varepsilon_i)] \\ &\quad + A_i f(\beta_2' \hat{d}_{2i}^* + \varepsilon_i \sigma_i \hat{\sigma}_i^{-1} + \tilde{A}_i) + e_i, \end{aligned} \quad (3.5)$$

where $e_i = e_i^* - E(e_i^*)$. First we will show that the third and fourth term on the right hand side of (3.5) can be ignored in the asymptotics. Let

$$V_i = 0.5 + \beta'_2 d_{2i}^* E f(\varepsilon_i + \delta_{2i}) + e_i, \quad (3.6)$$

and $V = (V_1, \dots, V_n)'$. Note that the notation V_i is also used in (2.7); this is justified by the fact that the two expressions coincide under the null hypothesis.

Lemma 3.1. *Let $W_i^* = F(W_i)$ be given by (3.5) and $W^* = (W_1^*, \dots, W_n^*)'$. Then*

$$V' D_2 (D_2' D_2)^{-1} D_2' V - W^{*'} D_2 (D_2' D_2)^{-1} D_2' W^* \rightarrow 0$$

in probability under H_n .

Proof. Arguing as in the proof of Lemma 2.2 it suffices to show that for $j_2 = q+1, \dots, p$ and conditionally on D_2

$$\begin{aligned} S_{j_2} &= n^{-1/2} \sum_{i=1}^n d_{ij_2} (W_i^* - V_i) \\ &= n^{-1/2} \sum_{i=1}^n d_{ij_2} [F(\beta'_2 \hat{d}_{2i}^* + \varepsilon_i \sigma_i / \hat{\sigma}_i) - F(\beta'_2 d_{2i}^* + \varepsilon_i)] \\ &\quad + n^{-1/2} \sum_{i=1}^n d_{ij_2} A_i f(\beta'_2 \hat{d}_{2i}^* + \varepsilon_i \sigma_i / \hat{\sigma}_i + \tilde{A}_i) \end{aligned} \quad (3.7)$$

converges to zero in probability. These terms can be shown to converge to zero by arguments similar to those used to show that the two terms on the right hand side of (A.3) converge to zero. This completes the proof.

By a similar argument it can be shown that δ_{2i} in the expression of V_i can be ignored in the asymptotics. Indeed it can be shown that

$$V' D_2 (D_2' D_2)^{-1} D_2' V - V^{*'} D_2 (D_2' D_2)^{-1} D_2' V^* \rightarrow 0, \quad \text{in probability,} \quad (3.8)$$

where $V^* = (V_1^*, \dots, V_n^*)'$ with $V_i^* = 0.5 + \beta'_2 d_{2i} E f(\varepsilon_i) + e_i$. Proposition A.2 in Akritas (1991) implies

$$V^{*'} D_2 (D_2' D_2)^{-1} D_2' V^* \rightarrow \frac{1}{12} \chi_{p-q}^2 (12 \gamma^2 [E f(\varepsilon)]^2). \quad (3.9)$$

We are now ready to formulate the main result of this section.

Theorem 3.1. *Under the assumptions stated in the Appendix,*

$$W^{*'} \hat{D}_2 (\hat{D}_2' \hat{D}_2)^{-1} \hat{D}_2' W^{**} \rightarrow \frac{1}{12} \chi_{p-q}^2 (12 \gamma^2 [E f(\varepsilon)]^2)$$

in distribution under H_n .

Proof. By Lemmas 2.3, 3.1, relations (3.8), (3.9) and a standard contiguity argu-

ment, it follows that

$$W^{**'} D_2 (D_2' D_2)^{-1} D_2' W^{**} \rightarrow \frac{1}{12} \chi_{p-q}^2 (12 \gamma^2 [Ef(\varepsilon)]^2)$$

in distribution under H_n . The theorem then follows from the fact that the left hand side of (A.22) converges to zero in probability under H_0 , and thus under H_n .

4. A numerical illustration

To illustrate the use of the test statistic from Section 2 we shall apply it to a real data set. We have chosen the salary survey data example from Chatterjee and Hadi (1988) (Example 4.3 on p. 87), which relates the monthly salary of a random sample of size 31 of employees in a given company to five factors thought to determine salary differentials.

Chatterjee and Hadi fit a linear model to these data and study plots of the residuals versus the fitted values and the regressors. It turns out (see p. 90 and Figures 4.5, 4.6 and 4.8) that these results suggest heteroscedasticity with respect to x_1 , x_2 and x_4 . As x_2 indicates sex a natural testing problem which suggests itself is that of H_0 : no sex difference. In other words, H_0 : $\beta_2 = 0$, where β_2 is the coefficient of x_2 . (In the notation of Section 2, we have $\beta'_1 = (\beta_1, \beta_3, \beta_4, \beta_5)$ and $\beta'_2 = \beta_2$.)

Following (1.2), we model the variance σ_i by

$$\sigma_i / \sigma = \exp(\theta_1 x_{i1} + \theta_2 x_{i2} + \theta_4 x_{i4}). \quad (4.1)$$

To obtain estimates for the θ 's, and hence for the σ_i , we use the pseudo-likelihood method of Carroll and Ruppert (1988, Chapter 3). In this method, the likelihood L is not maximized simultaneously with respect to the β 's and θ 's, but only with respect to the θ 's, while fixing the β 's at the value of a preliminary estimate, obtained e.g. through application of ordinary least squares. The resulting $\hat{\theta}$'s lead through (4.1) to a weighted data set, which in its turn can be used to obtain an updated estimate for the β 's, and through reapplication of the pseudo-likelihood method, to an updated estimate for the θ 's as well. Obviously, this process can be repeated as often as we like, but for our application we stop after this second step. Note in this connection that Carroll and Ruppert state that one step already produces a consistent estimate, but that for practical application a second iteration step is advisable (see p. 73). Another point to be noted in passing is that throughout our paper the alignment is based on residuals under the null hypothesis. As here H_0 : $\beta_2 = 0$, this means that x_2 plays no role in the (weighted) least squares analyses which produce the preliminary and updated estimates for the β 's. But in estimating the θ 's, the variable x_2 does play a role, as Chatterjee and Hadi's analysis suggests heteroscedasticity with respect to x_2 as well.

After rescaling x_1 , x_2 and x_4 to have common median 1.0, the results for $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_4)$ in the present example are $(-0.917, 0.333, 0.011)$ after the first step and

$(-0.955, 0.385, 0.033)$ after the second and final step. Notice that these outcomes are compatible with the conclusions from the residuals analyses in Chatterjee and Hadi. They observe that the variation increases in x_2 and decreases in x_1 and x_4 . Hence $\hat{\theta}_1$ and $\hat{\theta}_2$ have the right sign. On the other hand, $\hat{\theta}_4$ is so close to zero that it mainly suggests that heteroscedasticity with respect to x_4 is negligible.

Graphical inspection of the residuals gives no cause to distrust the assumption of symmetry. Moreover, 15 of the 31 residuals are positive, while the signed rank sum for the positive (negative) observations equals 243 (253). Gastwirth (1971) suggested a simple test for symmetry based on the number of observations exceeding the sample mean. Here this number is 14, which is evidently far from significant.

From the above we can now readily compute the scaled residuals under H_0 , as well as the corresponding vector of ranks $(R_1, \dots, R_n)$. It only remains to evaluate the matrix $\bar{D}_2$ in order to be able to compute the value of our weighted aligned rank statistic from (2.2). As $p - q = 1$ here, this statistic boils down to T^2 , with

$$T = n^{-1} \sum_{i=1}^n \bar{d}_2(i) R_i, \quad (4.2)$$

where $\bar{d}_2(i)$ are the elements of the vector $\bar{D}_2$, standardized such that $\sum_{i=1}^n \bar{d}_2^2(i) = 1$. Straightforward calculation yields that for the present data T attains the value 0.547. As $12T^2$ is asymptotically χ_1^2 , this leads to an approximate P -value of 5.8%.

To conclude this section, we compare the results above to those of the similar, but much simpler, analysis under homoscedasticity. Hence, we now ignore possible variance effects and just obtain unweighted aligned observations by computing the residuals under the null hypothesis. It remains to compute their ranks $(R_1, \dots, R_n)$ and the coefficients $\bar{d}_2(i)$, this time derived from unweighted x_{ij} , to be able to compute T from (4.2) for this situation. Here we find a value 0.403, leading to an approximate P -value of 16%.

Appendix: proofs

The proofs will be derived under the following assumptions.

A1. D is of full rank and

- (i) $n^{-1}D'D$ converges almost surely to a positive definite matrix,
- (ii) The largest diagonal element of $D(D'D)^{-1}D'$ goes to zero almost surely as $n \rightarrow \infty$,

A2. (i) $\sum_{i=1}^n \hat{d}_{ij} = \sum_{i=1}^n d_{ij} = 0$ for all $j = q+1, \dots, p$,

(ii) $\sum_{i=1}^n \hat{d}_{ij_1} \hat{d}_{ij_2} = \sum_{i=1}^n d_{ij_1} d_{ij_2} = 0$ for all $j_1 = 1, \dots, q$ and $j_2 = q+1, \dots, p$,

A3. (i) $n^{-1} \sum_{i=1}^n \|x_i\| = O_p(1),$

(ii) $n^{-1} \sum_{i=1}^n (x_{ij_1} x_{ij_2})^2 = O_p(1),$

(iii) $\sup \{n^{-1/2} \|x_i\|; i=1, \dots, n\} \rightarrow 0, \text{ almost surely,}$

A4. (i) The error distribution function F has finite first moment and a bounded density f ($|f| \leq C_1$) that satisfies $E[\varepsilon f(\varepsilon)] = 0$. The derivative of f exists and is bounded ($|f'| \leq C_2$).

(ii) The distribution of the random vector z_i has bounded support.

(iii) The errors ε_i are independent from x_i and z_i .

(A4(i) can be weakened to include the double exponential distribution.)

A5. The estimators $\hat{\beta}_1, \hat{\theta}$ satisfy

$$n^{1/2}(\hat{\beta}_1 - \beta_1) = O_p(1), \quad n^{1/2}(\hat{\theta} - \theta) = O_p(1).$$

Remark A.1. (i) The relations in assumption A2 follow from the definition of the matrices $D_2, \hat{D}_2$. It is for the sake of convenient reference that these relations are stated as assumptions.

(ii) Relation (1.2) implies that assumption A4(ii) is equivalent to the assumption that the σ_i are bounded away from zero and infinity. The assumption that the σ_i are bounded away from zero is common in the theory of the heteroscedastic linear model.

(iii) The construction of estimators $\hat{\beta}_1, \hat{\theta}$ as described in assumption A5 is considered in Carroll and Ruppert (1988).

We begin with two preliminary results that will be needed in the following proofs.

Lemma A.1. (i) Under assumptions A4(ii) and A5,

$$\sup \{n^{1/2} |\sigma_i \hat{\sigma}_i^{-1} - 1|; i=1, \dots, n\} = O_p(1).$$

(ii) Under assumptions A1(1), A3, A4(ii) and A5,

$$\sup \{|d_{ij} - \hat{d}_{ij}|; i=1, \dots, n\} = o_p(1) \text{ for all } j=1, \dots, p.$$

Proof. (i) From (1.2) we have

$$\begin{aligned} \sup_i \{|\sigma_i \sigma_i^{-1} - 1|\} &= \sup_i \{|\exp[(\theta - \hat{\theta})' z_i] - 1|\} \\ &= \sup_i \{|(\theta - \hat{\theta})' z_i C_i|\} \\ &\leq \|\theta - \hat{\theta}\| \sup_i \{\|z_i\|\} \sup_i \{C_i\}, \end{aligned}$$

where C_i is between 1 and $\sigma_i \hat{\sigma}_i^{-1}$. By assumptions A4(ii) and A5, $\|\hat{\theta} - \theta\| \sup_i \{\|z_i\|\} = O_p(n^{-1/2})$. By a similar argument, $\sup_i \{C_i\} = O_p(1)$, and thus the proof is complete.

(ii) For $j = 1, \dots, q$ we have that $\sup_i |d_{ij} - \hat{d}_{ij}| = \sup_i |d_{ij}(1 - \sigma_i \hat{\sigma}_i^{-1})| \leq \sup_i \|n^{-1/2} d_{ij}\| \sup_i n^{1/2} |1 - \sigma_i \hat{\sigma}_i^{-1}| = o_p(1)$, by assumptions A3(ii), A4(iii), A5 and part (i) of this lemma. For the rest it suffices to show that

$$m(\hat{D}_2 - D_2) = o_p(1),$$

where for any matrix $A = (a_{ij})$, $m(A) = \sup_{i,j} |a_{ij}|$. From their definition,

$$\hat{D}_2 - D_2 = (\hat{\Sigma}^{-1} - \Sigma^{-1})X_2 + \hat{P}(\Sigma^{-1} - \hat{\Sigma}^{-1})X_2 + (P - \hat{P})\Sigma^{-1}X_2.$$

Thus

$$m(\hat{D}_2 - D_2) \leq m((\hat{\Sigma}^{-1} - \Sigma^{-1})X_2) + m(\hat{P}(\Sigma^{-1} - \hat{\Sigma}^{-1})X_2) + m((P - \hat{P})\Sigma^{-1}X_2). \quad (\text{A.1})$$

From the previous analysis of $\sup_i |\hat{d}_{ij} - d_{ij}|$ for $j=1, \dots, q$ it follows that $m((\hat{\Sigma}^{-1} - \Sigma^{-1})X_2) = o_p(1)$. It remains to show that the other two terms on the right hand side of (A.1) are $o_p(1)$. The matrix involved in the second term on the right hand side of (A.1) is

$$n^{-1/2} \hat{D}_1^* (n^{-1} \hat{D}_1^* \hat{D}_1^*)^{-1} n^{-1/2} \hat{D}_1^* (\Sigma^{-1} - \hat{\Sigma}^{-1})X_2,$$

where $\hat{D}_1^* = (\mathbf{1} \mid \hat{D}_1)$. Using the fact that $m(AB) \leq km(A)m(B)$, where k is the common dimension of the matrices A and B it follows that

$$m(\hat{P}(\Sigma^{-1} - \hat{\Sigma}^{-1})X_2) \leq q^2 m(n^{-1/2} \hat{D}_1^*) m((n^{-1} \hat{D}_1^* \hat{D}_1^*)^{-1}) m(n^{-1/2} \hat{D}_1^* (\Sigma^{-1} - \hat{\Sigma}^{-1})X_2).$$

From assumptions A3(iii) and part (i) of this lemma it follows that $m(n^{-1/2} \hat{D}_1^*) = o_p(1)$. By assumption A1(i) and part (i) of this lemma it follows that $m((n^{-1} \hat{D}_1^* \hat{D}_1^*)^{-1}) = O_p(1)$. Finally from assumptions A3(ii) and part (i) of this lemma it follows that $m(n^{-1/2} \hat{D}_1^* (\Sigma^{-1} - \hat{\Sigma}^{-1})X_2) = o_p(1)$. The above imply that the second term on the right hand side of (A.1) is $o_p(1)$. The matrix involved in the third term on the right hand side of (A.1) is

$$\begin{aligned} (\hat{P} - P)\Sigma^{-1}X_2 &= (\hat{D}_1^* - D_1^*)(D_1^* D_1^*)^{-1} D_1^* \Sigma^{-1}X_2 \\ &\quad + \hat{D}_1^* [(\hat{D}_1^* \hat{D}_1^*)^{-1} - (D_1^* D_1^*)^{-1}] D_1^* \Sigma^{-1}X_2 \\ &\quad + \hat{D}_1^* (\hat{D}_1^* \hat{D}_1^*)^{-1} (\hat{D}_1^* - D_1^*) \Sigma^{-1}X_2, \end{aligned}$$

where $D_1^* = (\mathbf{1} \mid D_1)$. Now,

$$\begin{aligned} m(\hat{D}_1^* (\hat{D}_1^* \hat{D}_1^*)^{-1} (\hat{D}_1^* - D_1^*) \Sigma^{-1}X_2) \\ \leq q^2 m(n^{-1/2} \hat{D}_1^*) m((n^{-1} \hat{D}_1^* \hat{D}_1^*)^{-1}) m(A_1) = o_p(1), \end{aligned}$$

where A_1 is the $(1+q) \times (p-q)$ matrix whose first row is the zero vector and the rest of the elements are $n^{-1/2} \sum_i x_{ij} \sigma_i^{-1} (\hat{d}_{ij'} - d_{ij'})$, $j' = 1, \dots, q$, $j = q+1, \dots, p$. Indeed, $m(n^{-1/2} \hat{D}_1^*) = o_p(1)$ and $m((n^{-1} \hat{D}_1^* \hat{D}_1^*)^{-1}) = O_p(1)$, as explained above, and

$|n^{-1/2} \sum_i x_{ij'} \sigma_i^{-1} (\hat{d}_{ij'} - d_{ij'})| \leq \sup \{n^{1/2} |\sigma_i \hat{\sigma}_i^{-1} - 1| \} n^{-1} \sum_i |x_{ij'} x_{ij'} \sigma_i^{-2}| = O_p(1)$, by part (i) of this lemma and assumption A4(ii) (see also Remark A1(ii)). By a similar analysis it follows that $m((\hat{D}_1^* - D_1^*)(D_1^{*'} D_1^*)^{-1} D_1^{*'} \Sigma^{-1} X_2) = o_p(1)$. Finally,

$$\begin{aligned} & m(\hat{D}_1^*[(\hat{D}_1^{*'} \hat{D}_1^*)^{-1} - (D_1^{*'} D_1^*)^{-1}] D_1^{*'} \Sigma^{-1} X_2) \\ & \leq q^2 m(n^{-1/2} \hat{D}_1^*) m((n^{-1} \hat{D}_1^{*'} \hat{D}_1^*)^{-1} - (n^{-1} D_1^{*'} D_1^*)^{-1}) m(A_2) = o_p(1), \end{aligned}$$

where A_2 is the $(q+1) \times (p-q)$ matrix whose first row has elements $n^{-1/2} \sum_i x_{ij'} \sigma_i^{-1}$ and the rest of the elements are $n^{-1/2} \sum_i (d_{ij'} x_{ij'} \sigma_i^{-1})$. Indeed $m(n^{-1/2} \hat{D}_1^*) = o_p(1)$, as explained above; from assumption A3(ii) it follows that $m(A_2) = O_p(n^{1/2})$; finally it is easy to see that $m(n^{-1} \hat{D}_1^{*'} \hat{D}_1^* - n^{-1} D_1^{*'} D_1^*) = O_p(n^{-1/2})$ which implies that $m((n^{-1} \hat{D}_1^{*'} \hat{D}_1^*)^{-1} - (n^{-1} D_1^{*'} D_1^*)^{-1}) = O_p(n^{-1/2})$. This completes the proof of the lemma.

Proof of Lemma 2.2. In the arguments that follow we will condition on the matrix D_2 . The fact that for each given D_2 relation (2.8) holds implies that (2.8) is also true unconditionally. Set $P_2 = D_2(D_2' D_2)^{-1} D_2'$. Then the left side of (2.8) equals $V' P_2(V - W^*) + W^{*'} P_2(V - W^*)$. We will show that each of these terms tends to zero in probability. For the first term we need to show

$$(n^{-1/2} V' D_2)(n^{-1} D_2' D_2)^{-1} [n^{-1/2} D_2'(V - W^*)] \rightarrow 0 \quad \text{in probability.} \quad (\text{A.2})$$

In view of (2.7) and assumption A1(i), (A.2) will hold if

$$S_{j'} = n^{-1/2} \sum_{i=1}^n d_{ij'} (W_i^* - V_i) \rightarrow 0, \quad j' = q+1, \dots, p.$$

By (2.6) and the definitions of V_i , W_i^* , it follows that

$$S_{j'} = n^{-1/2} \sum_{i=1}^n d_{ij'} \{ [F(\varepsilon_i \sigma_i / \hat{\sigma}_i) - F(\varepsilon_i)] + A_i f(\varepsilon_i \sigma_i / \hat{\sigma}_i + \tilde{A}_i) \}. \quad (\text{A.3})$$

By Lemma A.1(i) and assumption A5, the first term on the right hand side of (A.3) will be shown to converge to zero in probability if we show that

$$\sup \left\{ n^{-1/2} \sum_{i=1}^n d_{ij'} [F(\varepsilon_i (1 + z_i' t n^{-1/2})) - F(\varepsilon_i)]; \|t\| < K \right\} \rightarrow 0 \quad (\text{A.4})$$

for any constant K . To show (A.4) we will show that the above process indexed by t converges weakly to the zero process. Before proceeding, it is worth to note that the random variables $F(\varepsilon_i (1 + z_i' t n^{-1/2})) - F(\varepsilon_i)$, $i = 1, \dots, n$, are not i.i.d. because we have conditioned on the $d_{ij'}$'s and, in general, the z_i 's depend on the $d_{ij'}$'s.

Remark A.2. In the case (which is not particularly relevant in practice) that the z_i 's are independent from the x_i 's (and thus from the $d_{ij'}$'s), relation (A.4) and all the results of this paper hold without the assumption that $E[\varepsilon f(\varepsilon)] = 0$ (see calculations involving the expected value below). In particular, the results of this paper would hold for asymmetric error distributions.

To show weak convergence we will show that the conditions in Theorem 10.6 of Pollard (1990) are satisfied. First we need to show that the expected value of $n^{-1/2} \sum_{i=1}^n d_{ij} [F(\varepsilon_i(1 + z_i' t n^{-1/2})) - F(\varepsilon_i)]$ converges to zero uniformly in t . Indeed, this expected value is

$$\begin{aligned}
 & n^{-1/2} \sum_{i=1}^n d_{ij} E[F(\varepsilon_i(1 + z_i' t n^{-1/2})) - F(\varepsilon_i)] \\
 &= n^{-1/2} \sum_{i=1}^n d_{ij} E[z_i' t n^{-1/2} \varepsilon f(\varepsilon + T)], \\
 & \quad \text{for some } T \text{ between zero and } \varepsilon z_i' t n^{-1/2}, \\
 &= n^{-1/2} \sum_{i=1}^n d_{ij} E\{z_i' t n^{-1/2} [\varepsilon f(\varepsilon) + \varepsilon(f(\varepsilon + T) - f(\varepsilon))]\} \\
 &= n^{-1/2} \sum_{i=1}^n d_{ij} E[z_i' t n^{-1/2} \varepsilon(f(\varepsilon + T) - f(\varepsilon))], \\
 & \quad \text{by assumptions A4(i) and A4(iii),}
 \end{aligned}$$

which is easily seen to converge to zero by the dominated convergence theorem and assumptions A3(i), A4(i) and A4(ii); by the fact that t is bounded it is easily seen that the convergence is uniform in t . From now on we consider the process in question to be centered. To show manageability note first that the pseudo-dimension of $\mathcal{F}_{n\omega} = \{(f_{n1}(\omega, t), \dots, f_{nn}(\omega, t)); \|t\| < K\} \subseteq \mathbb{R}^n$, where $f_{ni}(\omega, t) = n^{-1/2} d_{ij} [F(\varepsilon_i(1 + z_i' t n^{-1/2})) - F(\varepsilon_i)]$ is less than or equal to r ; next the envelope functions of $\mathcal{F}_{n\omega}$ are uniformly bounded by virtue of assumptions A3(iii) and A4(ii). This implies that manageability is satisfied with $\lambda(x) = A(1/x)^{4r}$, for some constant A (see Pollard, 1990, Definition 7.9). The rest of the conditions of Theorem 10.6 of Pollard (1990) are easily seen to be satisfied with $H(s, t) = 0$ and $\varrho(s, t) = 0$. Thus relation (A.4) is true.

The second term on the right hand side of (A.3) is

$$n^{-1/2} \sum_{i=1}^n d_{ij} A_i \{f(\varepsilon_i) + [(\hat{\sigma}_i \sigma_i^{-1} - 1) \varepsilon_i + \tilde{A}_i] f'(\varepsilon_i + B_i^*)\}.$$

By Lemma A.1 and assumption A4, the second of the above two terms is an order of magnitude smaller than the first and can easily be shown to converge to zero in probability. The first term is in absolute value

$$\begin{aligned}
 & \left| n^{-1/2} \sum_{i=1}^n d_{ij} (\beta_1 - \hat{\beta}_1)' \hat{d}_{1i} f(\varepsilon_i) \right| \\
 &= \left| (\beta_1 - \hat{\beta}_1)' n^{-1/2} \sum_{i=1}^n d_{ij} [d_{1i} + (\hat{d}_{1i} - d_{1i})] f(\varepsilon_i) \right| \\
 &\leq n^{1/2} \|(\beta_1 - \hat{\beta}_1)\| \left\{ \left\| n^{-1} \sum_{i=1}^n d_{ij} d_{1i} f(\varepsilon_i) \right\| \right\}
 \end{aligned}$$

$$+ \sup_i \{ |\sigma_i \hat{\sigma}_i^{-1} - 1| \} \left\| n^{-1} \sum_{i=1}^n d_{ij'} d_{1i} f(\varepsilon_i) \right\|$$

and both terms on the right hand side of the above inequality converge to zero by assumptions A3(ii), A4, A5, and Lemma A.1(i). This completes the proof of the lemma.

Proof of Lemma 2.3. In the arguments that follow we will condition on the matrix D_2 . The fact that for each fixed D_2 relation (2.9) holds implies that (2.9) is also true unconditionally. In view of (2.7) and Lemma 2.2 it suffices to show

$$S_{2j_2}(\hat{\Delta}) = n^{-1/2} \sum_{i=1}^n d_{ij_2} [F_n(W_i; \hat{\Delta}) - F(W_i)] \rightarrow 0, \quad j_2 = q+1, \dots, p, \quad (\text{A.5})$$

in probability under H_0 . By the fact that $\hat{\beta}$, $\hat{\theta}$ are $n^{1/2}$ -consistent estimators of the corresponding quantities the above will follow from

$$\sup \{ |S_{2j_2}(\Delta)|; \|n^{1/2}\Delta\| \leq K \} \rightarrow 0, \quad \text{in probability,} \quad (\text{A.6})$$

for all $K > 0$. Write

$$\begin{aligned} \sup_{\Delta} |S_{2j_2}(\Delta)| &\leq \sup_{\Delta} \left| n^{-1/2} \sum_{i=1}^n d_{ij_2} [F_n(W_i; \Delta) - F(W_i; \Delta)] \right| \\ &\quad + \sup_{\Delta} \left| n^{-1/2} \sum_{i=1}^n d_{ij_2} [F(W_i; \Delta) - F(W_i)] \right| \end{aligned} \quad (\text{A.7})$$

where $F(\cdot; \Delta)$ is given in (2.3). Writing $F(W_i; \Delta) - F(W_i) = n^{-1} \sum_{i'=1}^n [F(W_i \alpha_{i'}(\Delta) + \Delta'_1 d_{1i'}) - F(W_i)]$ and using a second order Taylor expansion, the second term on the right hand side of (A.7) becomes

$$\begin{aligned} \sup_{\Delta} &\left| n^{-1/2} \sum_{i=1}^n d_{ij_2} \left\{ n^{-1} \sum_{i'=1}^n \{ [W_i(\alpha_{i'}(\Delta) - 1) + \Delta'_1 d_{1i'}] f(W_i) \right. \right. \\ &\quad \left. \left. + 0.5 [W_i(\alpha_{i'}(\Delta) - 1) + \Delta'_1 d_{1i'}]^2 f'(W_i + \delta_{ii'}) \} \right\} \right| \\ &\leq \sup_{\Delta} \left| n^{-1/2} \sum_{i=1}^n d_{ij_2} W_i f(W_i) \left\{ n^{-1} \sum_{i'=1}^n \Delta'_2 z_{i'} \right\} \right| \\ &\quad + \sup_{\Delta} \left| n^{-1/2} \sum_{i=1}^n d_{ij_2} W_i f(W_i) \left\{ n^{-1} \sum_{i'=1}^n \Delta'_2 z_{i'} (C_{i'} - 1) \right\} \right| \\ &\quad + \sup_{\Delta} \left| n^{-1/2} \sum_{i=1}^n d_{ij_2} f(W_i) \left\{ n^{-1} \sum_{i'=1}^n \Delta'_1 d_{1i'} \right\} \right| \\ &\quad + \sup_{\Delta} \left| n^{-1/2} \sum_{i=1}^n d_{ij_2} \left\{ n^{-1} \sum_{i'=1}^n [W_i(\alpha_{i'}(\Delta) - 1) + \Delta'_1 d_{1i'}]^2 f'(W_i + \delta_{ii'}) \right\} \right| \end{aligned} \quad (\text{A.8})$$

where the first two terms on the right hand side of (A.8) resulted from a one term Taylor expansion $\alpha_i(\Delta) - 1 = \Delta'_2 z_i C_i$ where C_i is between $\alpha_i(\Delta)$ and 1. The second and fourth terms on the right hand side of (A.8) are of smaller order of magnitude and can easily be shown to converge to zero. By assumption A3(i) the third term will converge to zero provided that

$$n^{-1} \sum_{i=1}^n d_{ij_2} f(W_i) = o_p(1). \quad (\text{A.9})$$

Recalling that under H_0 $W_i = (\beta'_1 x_{1i} + \varepsilon_i - \hat{\beta}'_1 x_{1i}) \hat{\sigma}_i^{-1}$, it follows that

$$W_i - \varepsilon_i = \varepsilon_i (\hat{\sigma}_i^{-1} \sigma_i - 1) - (\hat{\sigma}_i^{-1} \sigma_i - 1) (\hat{\beta}_1 - \beta_1)' d_{1i} + (\hat{\beta}_1 - \beta_1)' d_{1i}.$$

Thus by assumptions A3(i), (ii), A4, A5 and Lemma A.1(i),

$$\begin{aligned} & \left| n^{-1} \sum_i d_{ij} [f(W_i) - f(\varepsilon_i)] \right| \\ & \leq C_2 n^{-1} \sum_i |d_{ij}(W_i - \varepsilon_i)| = o_p(1). \end{aligned} \quad (\text{A.10})$$

By assumption A2(i),

$$n^{-1} \sum_{i=1}^n d_{ij} f(\varepsilon_i) = o_p(1). \quad (\text{A.11})$$

Relations (A.10), (A.11) show (A.9). By a similar argument the first term on the right hand side of (A.8) converges to zero. Thus we have shown that the second term on the right hand side of (A.7) converges to zero. It remains to show that the first term on the right hand side of (A.7) converges to zero in probability. To do this we will show

$$\sup_{\Delta} \sup_i n^{1/2} |[F_n(W_i; \Delta) - F(W_i; \Delta)] - [F_n(W_i) - F(W_i)]| \rightarrow 0, \quad (\text{A.12})$$

and

$$\sup_i n^{1/2} |[F_n(W_i) - F(W_i)] - [F_n(\hat{\varepsilon}_i) - F(\hat{\varepsilon}_i)]| \rightarrow 0, \quad (\text{A.13})$$

in probability, where $\hat{\varepsilon}_i = (y_i - \hat{\beta}'_1 x_{1i}) \hat{\sigma}_i^{-1} = \varepsilon_i \sigma_i \hat{\sigma}_i^{-1}$, under H_0 . Relations (A.12) and (A.13) imply that the first term on the right hand side of (A.7) is less than or equal to

$$\left| n^{-1/2} \sum_{i=1}^n d_{ij_2} [F_n(\hat{\varepsilon}_i) - F(\hat{\varepsilon}_i)] \right| + o_p(1). \quad (\text{A.14})$$

Let now B^0 denote the Brownian Bridge such that, on a possibly different probability space $\|n^{1/2}[F_n(\cdot) - F(\cdot)] - B^0(F(\cdot))\| \rightarrow 0$ almost surely, where $\|\cdot\|$ stands for the sup-norm (Skorokhod construction). Thus

$$\begin{aligned} & \sup_i n^{1/2} |[F_n(\hat{\varepsilon}_i) - F(\hat{\varepsilon}_i)] - [F_n(\varepsilon_i) - F(\varepsilon_i)]| \\ & \leq 2 \|n^{1/2}[F_n(\cdot) - F(\cdot)] - B^0(F(\cdot))\| + \sup_i |B^0(F(\hat{\varepsilon}_i)) - B^0(F(\varepsilon_i))| \rightarrow 0, \end{aligned}$$

in probability by Lemma A.1 and the uniform continuity of the Brownian Bridge. Thus, by assumptions A3(i) and A4(ii),

$$\left| n^{-1/2} \sum_{i=1}^n d_{ij_2} [F_n(\hat{\varepsilon}_i) - F(\hat{\varepsilon}_i)] - [F_n(\varepsilon_i) - F(\varepsilon_i)] \right| = o_p(1). \quad (\text{A.15})$$

Next,

$$\left| n^{-1/2} \sum_{i=1}^n d_{ij_2} [F_n(\varepsilon_i) - F(\varepsilon_i)] \right| = o_p(1) \quad (\text{A.16})$$

by Proposition A.1 in Akritas (1990). Relations (A.15) and (A.16) imply that the expression in (A.14) is $o_p(1)$ and this proves that the first term on the right hand side of (A.7) is $o_p(1)$. It remains to show relations (A.12), (A.13). Relation (A.13) follows by Skorokhod construction argument similar to the one used following (A.14). To show (A.12) we will first show that for each fixed $t \in \mathbb{R}^{q+r}$, $\|t\| \leq K$,

$$T(\Delta) = \sup_i n^{1/2} |[F_n(W_i; \Delta) - F(W_i; \Delta)] - [F_n(W_i) - F(W_i)]| \rightarrow 0, \quad (\text{A.17})$$

in probability, where $\Delta = n^{-1/2}t$. Let $\mathcal{L}_n(\cdot) = \mathcal{L}_{1n}(\cdot) - \mathcal{L}_{2n}(\cdot)$ where $\mathcal{L}_{1n}(\cdot) = n^{1/2}[F_n(\cdot; \Delta) - F(\cdot; \Delta)]$, $\mathcal{L}_{2n}(\cdot) = n^{1/2}[F_n(\cdot) - F(\cdot)]$. Thus $T(\Delta) \leq \|\mathcal{L}_n(\cdot)\|$, and (A.17) will follow from $\|\mathcal{L}_n(\cdot)\| \rightarrow 0$, in probability. To show this consider first the finite dimensional distributions of $\mathcal{L}_n(\cdot)$. For each fixed x , and under H_0 ,

$$\begin{aligned} \mathcal{L}_n(x) &= n^{1/2}[F_n(x; \Delta) - F(x; \Delta)] - n^{1/2}[F_n(x) - F(x)] \\ &= n^{-1/2} \sum_{i=1}^n \{ [I(\varepsilon_i \leq x\alpha_i(\Delta) + \Delta'_1 d_{1i}) - I(\varepsilon_i \leq x)] \\ &\quad - [F(x\alpha_i(\Delta) + \Delta'_1 d_{1i}) - F(x)] \} \\ &= \sum_{i=1}^n B_i \end{aligned}$$

where the independent variables B_i satisfy $E(B_i) = 0$, $\text{Var}(B_i) = n^{-1}p_i(1-p_i)$ where $p_i = |F(x\alpha_i(\Delta) + \Delta'_1 d_{1i}) - F(x)|$, and $P(|B_i| \leq 2n^{-1/2}) = 1$, for all i . By Bernstein's inequality (Uspensky, 1937)

$$P\left(\left|\sum_{i=1}^n B_i\right| > nc\right) \leq 2 \exp\left\{-\frac{(nc)^2}{2 \sum_{i=1}^n \text{Var}(B_i) + (2/3)(2n^{-1/2})nc}\right\}.$$

Using this with $c = \delta/n$, it follows that $P(|\sum_{i=1}^n B_i| > \delta) \rightarrow 0$ for all $\delta > 0$. It follows that all finite dimensional distributions of the process $\mathcal{L}_n(\cdot)$ converge to those of the zero process. From covariance calculations similar to those of Shorack (1973) it follows that $\tilde{\mathcal{L}}_{1n}(F^{-1}(\cdot))$ $[\tilde{\mathcal{L}}_{2n}(F^{-1}(\cdot))]$ is relatively compact, where $\tilde{\mathcal{L}}_{1n}$ $[\tilde{\mathcal{L}}_{2n}]$ equals $\mathcal{L}_{1n}$ $[\mathcal{L}_{2n}]$ at the jump points and is linear between these points. It follows that $\tilde{\mathcal{L}}_n = \tilde{\mathcal{L}}_{1n} - \tilde{\mathcal{L}}_{2n}$ is relatively compact and hence it converges weakly to the zero process. This implies $\|\tilde{\mathcal{L}}_n(\cdot)\| \rightarrow 0$ in probability and so $\|\mathcal{L}_n(\cdot)\| \rightarrow 0$ in probability which shows (A.17).

Next, using an argument similar to that in Bickel (1975) or Boldin (1983), decompose the cube $C^q = \{s_1: s_1 \in \mathbb{R}^q \text{ and } |s_1| < ([1/\delta] + 1)\delta M\}$, where $|s_1|$ = maximum of the absolute values of the coordinates of s_1 , as the union of cubes with vertices on the grid of points $(j_1 \delta M, \dots, j_q \delta M)$, where $j_i = 0, \pm 1, \dots, \pm [1/\delta] + 1$. In a similar manner decompose the cube $C^r = \{s_2: s_2 \in \mathbb{R}^r \text{ and } |s_2| < ([1/\delta] + 1)\delta M\}$ and let $C =$

$C^q \times C^r$. (Note that for each K in (A.6) there exists M such that $\{\Delta: \|n^{1/2}\Delta\| \leq K\} \subset \{\Delta = n^{-1/2}t: t \in C\}$.) If $s_1 \in C^q$, let $P^q(s_1)$ be the lowest vertex of the cube containing s_1 and set $\Delta^q(s_1) = n^{-1/2}P^q(s_1)$. Similarly define $\Delta^r(s_2) = n^{-1/2}P^r(s_2)$ for any $s_2 \in C^r$, and set $\Delta(t)' = (\Delta^q(s_1)', \Delta^r(s_2)')$ for any $t' = (s_1', s_2') \in C$. For fixed δ (A.17) implies

$$\sup\{T(\Delta(t)); t \in C\} \rightarrow 0 \quad \text{in probability.} \quad (\text{A.18})$$

Next, let C_1^q be any cube of the partition of C^q let P_1^q be its lowest vertex and $\Delta_1^q = n^{-1/2}P_1^q$. Also let C_1^r be any cube of the partition of C^r , P_1^r be its lowest vertex and $\Delta_1^r = n^{-1/2}P_1^r$. Finally let $\Delta_1' = (\Delta_1^q, \Delta_1^r)$ and $C_1 = C_1^q \times C_1^r$. We will show that

$$\sup\{|T(\Delta_2) - T(\Delta_1)|; \Delta_2 = n^{-1/2}t, t \in C_1\} \leq o_p(1) - A\delta, \quad (\text{A.19})$$

where A does not depend on δ . Since for each fixed δ there are finite many C_1 's (A.19) holds uniformly in C_1 . Since

$$\begin{aligned} & \sup\{T(\Delta); \|n^{1/2}\Delta\| \leq K\} \\ & \leq \sup\{T(\Delta(t)); t \in C\} + \max_{C_1} \sup\{|T(\Delta_2) - T(\Delta_1)|; \Delta_2 = n^{-1/2}t, t \in C_1\}, \end{aligned}$$

relations (A.18) and (A.19) imply $\sup\{T(\Delta); \|n^{1/2}\Delta\| \leq K\} \leq o_p(1) + A\delta$. For a suitable choice of δ , this implies that for each $\delta_1, \delta_2 > 0$, $P(\sup\{T(\Delta); \|n^{1/2}\Delta\| \leq K\} > \delta_2) < \delta_1$ for all n large enough and this shows (A.12). It remains to show (A.19). The left hand side of (A.19) is less than or equal to

$$\begin{aligned} & \sup_{\Delta_2} \{n^{1/2} \| [F_n(\cdot; \Delta_2) - F(\cdot; \Delta_2)] - [F_n(\cdot; \Delta_1) - F(\cdot; \Delta_1)] \| \} \\ & \leq n^{1/2} \| [\sup_{\Delta_2} F_n(\cdot; \Delta_2) - \inf_{\Delta_2} F(\cdot; \Delta_2)] - [F_n(\cdot; \Delta_1) - F(\cdot; \Delta_1)] \| \\ & \quad + n^{1/2} \| [\sup_{\Delta_2} F(\cdot; \Delta_2) - \inf_{\Delta_2} F_n(\cdot; \Delta_2)] + [F_n(\cdot; \Delta_1) - F(\cdot; \Delta_1)] \|. \end{aligned} \quad (\text{A.20})$$

For the argument that follows it is convenient to separate the positive and negative arguments in the above sup-norms. Let POS_1 (POS_2) stand for the first (second) term on the right hand side of (A.20) when the sup-norm applies only to positive arguments. Similarly define NEG_1 , NEG_2 . Thus the right hand side of (A.20) is

$$\leq \text{POS}_1 + \text{POS}_2 + \text{NEG}_1 + \text{NEG}_2.$$

Recall that each Δ is a $(q+r)$ -dimensional column vector and that $\alpha_i(\Delta)$ depends only on Δ_2 , the last r coordinates of Δ . It should be clear from the context when Δ_2 denotes the last r coordinates of Δ (i.e., $\Delta' = (\Delta_1', \Delta_2')$) or a $(q+r)$ -dimensional vector (so $\Delta_2 = (\Delta_{11}', \Delta_{12}')$); similarly Δ_1 will be used to denote both the first q components of Δ and a $(q+r)$ -dimensional vector. Let $\Delta'_{iL} = (\Delta'_{1iL}, \Delta'_{2iL})$ be such that $\Delta'_{2iL} z_i = \inf\{\Delta_2' z_i; \Delta_2 = n^{-1/2}s_2, s_2 \in C_1^r\}$, $\Delta'_{1iL} d_{1i} = \inf\{\Delta_1' d_{1i}; \Delta_1 = n^{-1/2}s_1, s_1 \in C_1^q\}$. Also let $\Delta'_{iU} = (\Delta'_{1iU}, \Delta'_{2iU})$ be such that $\Delta'_{2iU} z_i = \sup\{\Delta_2' z_i; \Delta_2 = n^{-1/2}s_2, s_2 \in C_1^r\}$, $\Delta'_{1iU} d_{1i} = \sup\{\Delta_1' d_{1i}; \Delta_1 = n^{-1/2}s_1, s_1 \in C_1^q\}$. Thus we have

$$\text{POS}_1 = n^{1/2} \| [\sup_{\Delta_2} F_n(\cdot; \Delta_2) - \inf_{\Delta_2} F(\cdot; \Delta_2)] - [F_n(\cdot; \Delta_1) - F(\cdot; \Delta_1)] \|_0^\infty$$

$$\begin{aligned}
&\leq n^{1/2} \left\| n^{-1} \sum_{i=1}^n \left[I \left(\frac{\varepsilon_i - \Delta'_{1iU} d_{1i}}{\alpha_i(\Delta_{1U})} \leq \cdot \right) - F(\cdot; \alpha_i(\Delta_{1U}) + \Delta'_{1iU} d_{1i}) \right] \right. \\
&\quad \left. - [F_n(\cdot; \Delta_1) - F(\cdot; \Delta_1)] \right\|_0^\infty \\
&\quad + n^{1/2} \left\| n^{-1} \sum_{i=1}^n [F(\cdot; \alpha_i(\Delta_{1U}) + \Delta'_{1iU} d_{1i}) \right. \\
&\quad \left. - F(\cdot; \alpha_i(\Delta_{1L}) + \Delta'_{1iL} d_{1i})] \right\|_0^\infty. \tag{A.21}
\end{aligned}$$

Arguing as in the proof of (A.17) the first term on the right hand side of (A.21) is seen to be $o_p(1)$. Using a Taylor expansion and assumptions A3, A4 and A5, the second term is seen to be less than $A\delta$ where A does not depend on δ . Similarly POS_2 , NEG_1 , NEG_2 can be shown to be less than $o_p(1) + A\delta$. This shows (A.19) and completes the proof of the lemma.

Proof of Theorem 2.1. Write

$$\begin{aligned}
&W^{**'} D_2 (D_2' D_2)^{-1} D_2' W^{**} - W^{**'} \hat{D}_2 (\hat{D}_2' \hat{D}_2)^{-1} \hat{D}_2' W^{**} \\
&= W^{**'} (D_2 - \hat{D}_2) (D_2' D_2)^{-1} D_2' W^{**} + W^{**'} D_2 (D_2' D_2)^{-1} (D_2 - \hat{D}_2)' W^{**} \\
&\quad - W^{**'} \hat{D}_2 [(D_2' D_2)^{-1} - (\hat{D}_2' \hat{D}_2)^{-1}] \hat{D}_2' W^{**}. \tag{A.22}
\end{aligned}$$

We will show that

$$n^{-1/2} (\hat{D}_2 - D_2)' W^{**} \rightarrow 0 \quad \text{in probability under } H_0. \tag{A.23}$$

In view of Lemma 2.3 this implies that the first and second terms on the right hand side of (A.22) converge to zero in probability. By Lemma 2.3 again, (A.23) also implies that $n^{-1/2} \hat{D}_2' W^{**}$ remains bounded in probability and thus the third term on the right hand side of (A.22) converges to zero in probability. It remains to show (A.23). This will follow from

$$n^{-1/2} \sum_{i=1}^n (\hat{d}_{ij} - d_{ij}) W_i^{**} \rightarrow 0 \quad \text{in probability, } j = q+1, \dots, p. \tag{A.24}$$

First note that

$$\begin{aligned}
&\left| n^{-1/2} \sum_{i=1}^n (\hat{d}_{ij} - d_{ij}) (W_i^{**} - W_i^*) \right| \\
&\leq \sup_i |\hat{d}_{ij} - d_{ij}| n^{1/2} \|F_n(\cdot; \hat{\Delta}) - F(\cdot)\| \rightarrow 0 \tag{A.25}
\end{aligned}$$

in probability by Lemmas 2.1 and A.1(ii). Similarly, by (2.6) and the definition of V_i (with $\varepsilon_i^* = \varepsilon_i \sigma_i \hat{\sigma}_i^{-1} + \tilde{A}_i$, and ε_i^{**} between ε_i and $\varepsilon_i \sigma_i \hat{\sigma}_i^{-1}$),

$$\begin{aligned}
&n^{-1/2} \sum_{i=1}^n (\hat{d}_{ij} - d_{ij}) (W_i^* - V_i) \\
&\leq \left| n^{-1/2} \sum_{i=1}^n (\hat{d}_{ij} - d_{ij}) (\beta_1 - \hat{\beta}_1)' \hat{d}_{1i} f(\varepsilon_i^*) \right|
\end{aligned}$$

$$\begin{aligned}
& + \left| n^{-1/2} \sum_{i=1}^n (\hat{d}_{ij} - d_{ij}) \varepsilon_i (1 - \sigma_i \sigma_i^{-1}) f(\varepsilon_i^{**}) \right| \\
& \leq \|n^{1/2}(\hat{\beta}_1 - \beta_1)\| \sup_i |\hat{d}_{ij} - d_{ij}| n^{-1} \sum_i \|\hat{d}_{1i} f(\varepsilon_i^*)\| \\
& + \sup_i \{ |n^{1/2}(1 - \sigma_i \sigma_i^{-1})| \} \sup_i \{ |\hat{d}_{ij} - d_{ij}| \} n^{-1} \sum_{i=1}^n |\varepsilon_i f(\varepsilon_i^{**})| \rightarrow 0
\end{aligned} \tag{A.26}$$

in probability, by Lemma A.1 and assumptions A3(i), A4(i). Next from the decomposition of $\hat{D}_2 - D_2$ given prior to (A.1) it follows that for $j = q+1, \dots, p$

$$\begin{aligned}
& \hat{d}_{ij} - d_{ij} \\
& = x_{ij} \sigma_i^{-1} (1 - \sigma_i \hat{\sigma}_i^{-1}) + \hat{P}_i (x_{1j} \sigma_1^{-1} (1 - \sigma_1 \hat{\sigma}_1^{-1}), \dots, x_{nj} \sigma_n^{-1} (1 - \sigma_n \hat{\sigma}_n^{-1}))' \\
& + (P_i - \hat{P}_i) (x_{1j} \sigma_1^{-1}, \dots, x_{nj} \sigma_n^{-1})',
\end{aligned}$$

where P_i ($\hat{P}_i$) denotes the i -th row of the matrix P ($\hat{P}$). Thus if $V^c = (V_1 - 0.5, \dots, V_n - 0.5)$ denotes the centered vector V ,

$$\begin{aligned}
& \left| n^{-1/2} \sum_{i=1}^n (\hat{d}_{ij} - d_{ij}) V_i \right| \\
& = \left| n^{-1/2} \sum_{i=1}^n (\hat{d}_{ij} - d_{ij}) (V_i - 0.5) \right| \\
& \leq \left| n^{-1/2} \sum_{i=1}^n x_{ij} \sigma_i^{-1} (\sigma_i \hat{\sigma}_i^{-1} - 1) (V_i - 0.5) \right| \\
& + |n^{-1/2} (x_{1j} \sigma_1^{-1} (1 - \sigma_1 \hat{\sigma}_1^{-1}), \dots, x_{nj} \sigma_n^{-1} (1 - \sigma_n \hat{\sigma}_n^{-1})) \hat{P} V^c| \\
& + |n^{-1/2} (x_{1j} \sigma_1^{-1}, \dots, x_{nj} \sigma_n^{-1}) (P - \hat{P}) V^c|.
\end{aligned} \tag{A.27}$$

For the first term on the right hand side of (A.27) we have

$$\begin{aligned}
& \left| n^{-1/2} \sum_{i=1}^n x_{ij} \sigma_i^{-1} (\sigma_i \hat{\sigma}_i^{-1} - 1) (V_i - 0.5) \right| \\
& = \left| n^{-1/2} \sum_{i=1}^n d_{ij} (V_i - 0.5) \{ (\theta - \hat{\theta})' z_i + 0.5 [(\theta - \hat{\theta})' z_i]^2 C_i \} \right| = o_p(1)
\end{aligned}$$

by assumptions A3(i), A4(ii) and A5. The other two terms on the right hand side of (A.27) can also be shown to converge to zero by straightforward but tedious arguments. Relations (A.25), (A.26) and (A.27) imply (A.24) and complete the proof of the theorem.

Acknowledgement

A successful comment by a referee resulted in the weakening of the assumptions on the error distribution.

References

- Adichie, J.N. (1978). Rank tests of sub-hypotheses in the general linear regression. *Ann. Statist.* **6**, 1012–1026.
- Akritas, M.G. (1991). An alternative derivation of aligned rank tests for regression. *J. Statist. Plann. Inference* **27**, 171–186.
- Bickel, P.J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70**, 428–434.
- Boldin, M.V. (1983). Estimation of the distribution of noise in an autoregressive scheme. *Th. Prob. Appl.* **27**, 866–871.
- Carroll, R.J. and D. Ruppert (1988). *Transformation and Weighting in Regression*. Chapman and Hall, New York.
- Chatterjee, S. and A.S. Hadi (1988). *Sensitivity Analysis in Regression*. Wiley, New York.
- Chiang, C.-Y. and M.L. Puri (1984). Rank procedures for testing sub-hypotheses in linear regression. *Ann. Inst. Statist. Math.* **36**, Part A. 35–50.
- Gastwirth, J.L. (1971). On the sign test for symmetry. *J. Amer. Statist. Assoc.* **66**, 821–823.
- Hampel, F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986). *Robust Statistics: The approach based on influence functions*. Wiley, New York.
- Hodges, J.L., Jr. and E.L. Lehmann (1962). Rank methods for combination of independent experiments in analysis of variance. *Ann. Math. Statist.* **33**, 482–497.
- Pollard, D. (1990). *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.
- Puri, M.L. and P.K. Sen (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.
- Puri, M.L. and P.K. Sen (1985). *Nonparametric Methods in General Linear Models*. Wiley, New York.
- Ronchetti, E.M. (1982a). Robust alternatives to the F-test for the linear model. In: W. Grossmann, C. Pflug and W. Wertz, Eds., *Probability and Statistical Inference*. Reidel, Dordrecht, pp. 329–342.
- Ronchetti, E.M. (1982b). Robust testing in linear models: The infinitesimal approach. Ph.D. Thesis. ETH, Zurich.
- Schrader, R.M. and T.P. Hettmansperger (1980). Robust analysis of variance based upon a likelihood ration criterion. *Biometrika* **67**, 93–101.
- Sen, P.K. (1982). On M tests in linear models. *Biometrika* **69**, 245–248.
- Shorack, G.R. (1973). Convergence of reduced empirical and quantile processes with application to functions of order statistics in the non-i.i.d. case. *Ann. Statist.* **1**, 454–471.
- Uspensky, J.V. (1937). *Introduction to Mathematical Probability*. McGraw-Hill, New York.