Robust Rank Regression for Limited Dependent Variable

Aurobindo Ghosh*

January 31, 2008

JEL Codes: C14, C24
Keywords: Rank Regression, Survivorship bias, Linear rank statistics, Pairwise difference estimator.

Abstract

Survivorship is known to push the performance of mutual funds (both public and private) and their money managers upwards. One possible solution is to impose parametric assumptions. However, many financial data have unknown and possibly heteroscedastic and leptokurtic distribution which severely affects the robustness of the results using traditional parametric models.We propose an estimator of the truncated regression model which is an extension of the well known Hodges and Lehmann (1963) type generalized difference estimator based on an extension of the linear rank statistic. Rank based inference methods has been proposed as early as Adichie (1967) and has been studied by several other authors thereafter. This has also been used to find out the treatment effect in fixed effect models, and then by Honore and Powell (1994) to propose their pairwise difference estimators which is a sequence of minimizers of second order U-processes. We extend their result into a Hodges and Lehmann type estimator in multivariate location as studied by Choudhuri (1992) to a linear regression framework where one or more of the variables might be truncated at variable but observable points. We apply our results to evaluate the performance of funds and managers using Morningstar Principia and other databases.

*School of Economics at Singapore Management University, 90 Stamford Road, #05036, Singapore 178903. Phone: +65 6828-0863. Email: aurobindo@smu.edu.sg. This is a very preliminary and incomplete version of work in progress. Please do not quote without author's permission. Comments are welcome.

1 Introduction

A substantial body of work on rank tests and rank estimates, more popularly known as R-estimates, are mainly aimed at reducing the influence of gross errors in the estimation of the standard multivariate regression model. These estimates designed to be robust against distributions with fatter tails than normal (leptokurtic) are often of the form

$$\mathbf{Y} = [\ \mathbf{1}\ \ \mathbf{X}\ ] \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \mathbf{e}, \tag{1}$$

where $\mathbf{Y}$ is a $N \times 1$ vector, $\mathbf{1}$ is a $N \times 1$ vector of 1's, $\mathbf{X}$ is a $N \times p$ matrix of known regression constants, α is the scalar intercept parameter, β is a $p \times 1$ vector of unknown regression coefficients and $\mathbf{e}$ is a $N \times 1$ vector of *iid* errors with the distribution $F \in \Omega_0$, which is the class of all distributions with median 0. In a rank based inference procedure the intercept term is not identifiable, hence in order to make the columns of $\mathbf{X}$ orthogonal to the intercept term, we consider it in the deviation form

$$\mathbf{Y} = [\ \mathbf{1}\ \ \mathbf{X}_c\ ] \begin{pmatrix} \alpha^* \\ \beta \end{pmatrix} + \mathbf{e} \tag{2}$$

where $\mathbf{X}_c = \mathbf{X} - \mathbf{1}\left(\overline{x}_1, \overline{x}_2, \overline{x}_3, ..., \overline{x}_p\right)$, $\alpha^* = \alpha + \bar{x}'\beta$ and $\bar{\mathbf{x}}' = \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_p\right)$, in particular we have the subspaces spanned by $\mathbf{X}_c$ and $\mathbf{1}$ are orthogonal. We are only going to concentrate on the estimate of the regression slope coefficient β.

The main motivation for this paper is the regularity in empirical finance about performance evaluation of money (or fund) managers that suffers from a problem of survivorship bias as we only observe the survivors. This truncation (not censoring) is also complicated by different levels or benchmarks associated with different types of money managers, for example, the so-called growth and value or other style or industry based fund managers. Without going into the various classification of styles the fact of the matter is that these managers suffer from unobserved heterogeneity in their performance due to their ability ("hot hands") and it has been quite a challenging task to separate out ability from noise. So we can see the need for an estimate of their performance given that the signal of ability to noise ratio is more than sufficiently high. Truncation of this form causes a problem of *missingness* which is neither missing completely at random(MCAR) nor missing at random (MAR), so traditional missing data techniques like imputation using the E-M algorithm might not work. However, that is a possible future direction of research which we are not going to pursue here.

In this paper we are going to explore some of the existing literature in rank based estimates of the coefficients in a multiple regression model and obtain estimators by minimizing a criteria function better known as a class of minimum distance estimators. We are then going to look at a class of truncated regression estimator based on pairwise differences and finally a multivariate extension of the famous Hodges-Lehmann estimator of location and its efficient Bahadur-type representation. We propose an estimator which promises to be an extension of the types of estimators discussed earlier, and that reduces to the multivariate m^{th} order Hodges-Lehmann location estimator as discussed by Chaudhuri (1992). Finally, we are going to look at some motivation for exploring these types of estimators with some empirical examples using Morningstar Principia database before drawing our conclusions and suggested directions of future research.

2 Background and Motivation

2.1 Estimators based on Minimizing Dispersion of Errors

Let us consider an even function $D : \mathbf{R}^N \to \mathbf{R}^+$ which is invariant to translation, this can possibly be used as a measure of dispersion of the residuals and can be minimized to produce an estimate of the regression slope coefficients β based on ranks. This will be an extension of the Hodges and Lehmann (1963) location estimator equivalent in the linear regression setting.

Consider a non-decreasing sequence of score functions $a_1 \leq a_2 \leq ... \leq a_N$ which are non-constant with the added assumption of symmetry in the sense $a_k + a_{N-k+1} = 0$, so that we have $\sum_{i=1}^{n} a_k = 0$. Then we can define a measure of dispersion of $\mathbf{Z}'= (Z_1, Z_2, ..., Z_N)$ by

$$D(\mathbf{Z}) = \sum_{i=1}^{N} a[\mathbf{R}(Z_i)]Z_i \tag{3}$$

where $\mathbf{R}(Z_i)$ is the rank of Z_i among $Z_1, Z_2, Z_3, ..., Z_N$. It is useful to note that this measure of dispersion is not so severely affected by gross errors like the L^2 counterpart variance, as it is a piecewise linear function of $\mathbf{Z}$, which will be explored in more details later in this exposition.

A rank estimate (R-estimate) of β is given by

$$\hat{\beta} = \underset{\beta \in \mathbf{R}^p}{\arg\min} D(\mathbf{Y} - \mathbf{X}\beta) \tag{4}$$
$$= \underset{\beta \in \mathbf{R}^p}{\arg\min} \sum_{i=1}^{N} a[\mathbf{R}(Y_i - \mathbf{x}_i'\beta)](Y_i - \mathbf{x}_i'\beta)$$

where $\mathbf{X}' = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_N)$. It can be shown that $D(\mathbf{Y} - \mathbf{X}\beta)$ is a valid measure of dispersion being a non-negative,continuous and convex function of β (Jaeckel, 1972). He also claims that if the design matrix $\mathbf{X}_c$ is of full rank, then $D(.)$ attains a minimum, and this occurs for a bounded set of β. Also the structure of D suggests that there exists a partition of the space of β such that D is a linear function in each of those polygonal segments, and hence, $D(.)$ is really a piecewise linear function of the errors. Also the partial derivative of D exists almost everywhere as

$$\frac{\partial}{\partial \beta_j} D(\mathbf{Y} - \mathbf{X}\beta) = -\sum_{i=1}^{N} a\left[R\left(Y_i - \mathbf{x}_i'\beta\right)\right]\left(x_{ij} - \bar{x}_j\right)\left[= -\mathbf{S}_j\left(\mathbf{Y} - \mathbf{X}\beta\right)\right] \tag{5}$$

for $j = 1, ..., p$. Hence we see that a function of the regression rank statistic is given by the partial derivative of the proposed measure of dispersion and we can solve an equivalent set of normal equations by simply solving the first order conditions given by

$$\mathbf{S}\left(\mathbf{Y} - \mathbf{X}\beta\right) \doteq 0 \tag{6}$$

which is the negative of the gradient of the measure of dispersion $D(.)$. This also suggests that this is an extension of the Hodges and Lehmann location estimator in the linear model framework as we have $E\mathbf{S}\left(\mathbf{Y} - \mathbf{X}\beta\right) = 0$, for the true parameter β. For example, if we consider the Wilcoxon scores, we have $a(i) = \phi\left(\frac{i}{N+1}\right)$, where $\phi(u) = \sqrt{12}\left(u - \frac{1}{2}\right)$, this is normalized and standardized Wilcoxon scores with mean 0 and variance 1.

Let us now focus on the limiting distribution of the rank estimate we obtained by solving the optimization problem or solving the first order condition. Now, we work backwards from the linear approximation of the negative of the gradient, $-\nabla D\left(\mathbf{Y} - \mathbf{X}\beta\right) = \mathbf{S}\left(\mathbf{Y} - \mathbf{X}\beta\right)$ to get a quadratic approximation of the measure of dispersion $D\left(\mathbf{Y} - \mathbf{X}\beta\right)$. Our plan here is to find out

$$\tilde{\beta} = \underset{\beta \in \mathbf{R}^p}{\arg\min} D(\mathbf{Y} - \mathbf{X}\beta). \tag{7}$$

In can be shown that $\sqrt{n}\left(\tilde{\beta}-\beta_0\right)$ converges to a limiting normal distribution and hence, it can be shown that $\hat{\beta}$ has the same asymptotic distribution as $\tilde{\beta}$.

We start with the following linearity for a vector of rank statistics $\mathbf{S}(\mathbf{Y}-\mathbf{X}\beta)$, using the Wilcoxon scores where β_0 is the true parameter of interest.

$$\frac{1}{\sqrt{N}}\mathbf{S}\left(\mathbf{Y}-\mathbf{X}\beta\right)=\frac{1}{\sqrt{N}}\mathbf{S}\left(\mathbf{Y}-\mathbf{X}\beta\right)-\sqrt{12}\int f^2\left(x\right)dx\frac{1}{N}\mathbf{X}_c'\mathbf{X}_cN^{1/2}\left(\beta-\beta_0\right) + o_p\left(1\right), \tag{8}$$

however, this can be generalized to more general scores namely $\int_0^1\phi\left(u\right)\phi_f\left(u\right)du$ as was done by Jurećkova(1971) where $\int_0^1\phi\left(u\right)du=0$ and $\int_0^1\phi^2\left(u\right)du=1$ or $\phi\left(u\right)$ is a score generating function and $\phi_f\left(u\right)=\frac{f'\left(F^{-1}(u)\right)}{f(F^{-1}(u))}$. Then we have some results under the following assumptions.

2.1.1 Assumptions

1. $\phi\left(u\right)=\sqrt{12}\left(u-\frac{1}{2}\right)$, $a\left(i\right)=\sqrt{12}\left[\frac{i}{N+1}-\frac{1}{2}\right]$.
2. $\left[\begin{array}{cc}1 & \mathbf{X}\end{array}\right]$has full column rank, $p+1$.
3. $N^{-1}\left[\begin{array}{cc}1 & \mathbf{X}\end{array}\right]'\left[\begin{array}{cc}1 & \mathbf{X}\end{array}\right]$ converges to a p.d. matrix. Hence, $N^{-1}\mathbf{X}_c'\mathbf{X}_c\rightarrow$ $\mathbf{\Sigma}$ is a positive definite matrix.
4. $\mathbf{F}\in\mathbf{\Omega}_0$ and $f(.)$ has finite Fisher Information,this implies $\int_{-\infty}^{\infty}f^2\left(x\right)dx<\infty$.

Under these assumptions we have the linearity result earlier hold, i.e.,

$$\frac{1}{\sqrt{N}}\mathbf{S}\left(\mathbf{Y}-\mathbf{X}\beta\right)=\frac{1}{\sqrt{N}}\mathbf{S}\left(\mathbf{Y}-\mathbf{X}\beta_0\right)-\sqrt{12}\int f^2\left(x\right)dx\frac{1}{N}\mathbf{X}_c'\mathbf{X}_cN^{\frac{1}{2}} \times\left(\beta-\beta_0\right)+o_p\left(1\right) \tag{9}$$

this provides a linear approximation to $\bigtriangledown D\left(\mathbf{Y}-\mathbf{X}\beta\right)$, this in turn provides a quadratic approximation to the measure of dispersion $D\left(\mathbf{Y}-\mathbf{X}\beta\right)$, by

$$Q\left(\mathbf{Y}-\mathbf{X}\beta\right)=D\left(\mathbf{Y}-\mathbf{X}\beta\right)-\left(\beta-\beta_0\right)'\mathbf{S}\left(\mathbf{Y}-\mathbf{X}\beta\right)+ \frac{1}{2}\sqrt{12}\int f^2\left(x\right)dxN\left(\beta-\beta_0\right)'\Sigma\left(\beta-\beta_0\right) \tag{10}$$

this function of β, $Q\left(\mathbf{Y}-\mathbf{X}\beta\right)$ coincides with $D\left(\mathbf{Y}-\mathbf{X}\beta\right)$ when $\beta=\beta_0$.

Theorem 1 *For any $B > 0$ and $\varepsilon > 0$, under the given assumptions we have ,*

$$\mathbf{P}\left\{\sup_{\sqrt{N}||\beta-\beta_0||\leq B}|Q(\mathbf{Y}-\mathbf{X}\beta)-D(\mathbf{Y}-\mathbf{X}\beta)|\geq\varepsilon\right\}\longrightarrow 0$$

Proof. A more general version of the proof for general score functions is given in Jaeckel (1972). ■

Theorem 2 *Consider now the $\tilde{\beta}$ defined in the following sense*

$$\tilde{\beta}=\beta_0+\frac{1}{\sqrt{12}\int f^2(x)\,dx}\mathbf{\Sigma}^{-1}\frac{1}{N}\mathbf{S}(\mathbf{Y}-\mathbf{X}\beta_0)$$

Let $\hat{\beta}$ be any point that minimizes $D(\mathbf{Y}-\mathbf{X}\beta)$, then if $\beta_0$is the true parameter ,under the assumptions given

$$\sqrt{N}\left(\tilde{\beta}-\beta_o\right)\rightarrow\mathbf{Z}\sim\mathbf{MVN}\left(\mathbf{0},\frac{1}{12\left(\int f2(x)\,dx\right)}\Sigma^{-1}\right)$$

Proof. A more general version of the proof for general score functions is given in Jaeckel(1972). ■

Jaeckel(1972) proved that the set of minimizers of $D(\mathbf{Y}-\mathbf{X}\beta)$, say, $\mathbf{B}_N$ is bounded. Now we can also say that $\mathbf{B}_N$, is closed as a continuous function $D(.)$ is a constant in $\mathbf{B}_N$.He also showed that for moderate data sets the cardinality of a set like $\mathbf{B}_N$ is quite small.

An extension of this would be to obtain *generalized variance* of a multivariate estimator as the determinant of its variance-covariance matrix (see for example, Wilks). The ARE of two asymptotically MVN estimators is the $\frac{1}{p}^{th}$ root of the ratio of their generalized variance, we can obtain the ARE of the rank estimator $\hat{\beta}$ relative to the least squares estimator β^* is

$$e\left(\hat{\beta},\beta^*\right)=\left\{\frac{|\sigma^2\mathbf{\Sigma}^{-1}|}{\left|\frac{1}{12\left(\int f^2(x)dx\right)^2}\mathbf{\Sigma}^{-1}\right|}\right\}^{\frac{1}{p}}=12\sigma^2\left(\int f^2(x)\,dx\right)^2 \tag{11}$$

which is the same as the Wilcoxon method compared with least squares.

Computationally, it is pretty challenging to compute these estimators mainly because of lack of smoothness. However since $D(.)$ is still a convex surface we can find out the minimum, using a method like the steepest descent or other interior point methods. We can think about a consistent one step

R-estimator form the linear approximation. If we use a consistent estimator of

$$\tau = \frac{1}{\sqrt{12}\int f^2(x)\,dx}\,(= say, \hat{\tau}) \tag{12}$$

we can define the estimator in the following iterative sequence

$$\hat{\beta}_1 = \hat{\beta}_0 + \hat{\tau}\left(\mathbf{X}_c'\mathbf{X}_c\right)^{-1}\mathbf{S}\left(\mathbf{Y} - \mathbf{X}\hat{\beta}_0\right) \tag{13}$$

For the one sample case,we can assume that the underlying distribution $F \in \Omega_s$, then τ can be consistently estimated using the width of the Wilcoxon confidence intervals. The main problem in the above estimation is to consistently estimate $\hat{\tau}$,which in turn boils down to estimating $\gamma = \int f^2(x)\,dx = \int f(x)\,dF(x)$,we can use a kernel density estimator. So as is usual the biggest challenge comes from the bandwidth selection once we know what kernel $w(.)$to select. It can be shown that a consistent estimator of γ is given by

$$\gamma^* = \frac{1}{NK} + \frac{1}{N(N-1)h_N}\sum_{i\neq}\sum_{j} w\left(\frac{Y_i - Y_j}{h_N}\right) \tag{14}$$

where K is a constant and h_N is the optimal bandwidth which is of the order $o\left(n^{-1/2}\right)$

While Jaeckel (1972) and related papers like Adichie (1967), Jurećkova (1971) proposed a more general framework for working with rank based estimators, this can also be related to a more specific problem like one of looking at a limited dependent variable models like truncated regression model as a special case. However in case there is no problem of truncation or censoring the dependent variable method like this would naturally lead to an extension of the Hodges-Lehmann type estimators based on the ranks of pairwise differences or averages. A natural direction to look at now to find some justification in looking at estimators based on minimizing different measures of distance or dispersion.

2.2 Minimum Distance Estimation

If we have a symmetric (wlog around 0) error distribution and assume that $(\mathbf{X}'\mathbf{X})^{-1}$ exists for all $n \geq p$. Consider the following function $\mathbf{V} : \mathbb{R}^1 \times \mathbb{R}^p \longmapsto \mathbb{R}^p$, defined as (we remove the subscript c from X, although it is understood)

$$\mathbf{V}(y, \mathbf{t}) = \mathbf{A}\sum_{i}\mathbf{x}_i\left\{I\left(Y_i - \mathbf{x}_i'\mathbf{t} \leq y\right) - I\left(-Y_i + \mathbf{x}_i'\mathbf{t} < y\right)\right\}, \tag{15}$$

where $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-\frac{1}{2}}$and $I(E)$ is an indicator of event E.

Consider now a measure of distance or dispersion given by

$$\mathbf{M}(\mathbf{t}) := \int \mathbf{V}'(y,\mathbf{t})\,\mathbf{V}(y,\mathbf{t})\,dH(y) \tag{16}$$

where $H(y)$ is a non-decreasing right continuous function. We can immediately see that given the symmetry of ε around 0, $E\mathbf{V}(y,\beta) = 0$. Hence, we can define a class of the so called Minimum Distance estimators β^{+} for different $H(.)$ by the minimizing $M(t)$. It is worthwhile to note that, in $\mathbf{V}(y,\mathbf{t})$,

$$E\{I(Y_i - \mathbf{x}_i'\mathbf{t} \le y) - I(-Y_i + \mathbf{x}_i'\mathbf{t} < y)\} = F(y) - 1 - F(-y) \tag{17}$$

vanishes if the error terms ε_i have a symmetric distribution function $F(.)$. This also appeared before while estimating the coefficients in the minimum dispersion type estimates Jaeckel (1972).

If the symmetry assumption is not given for ε we do not get the moment condition right away as in the previous case. However, if $\left\{(\mathbf{X}-\bar{\mathbf{X}})'(\mathbf{X}-\bar{\mathbf{X}})\right\}^{-1}$ exists when $n \ge p$ and given $\bar{\mathbf{x}}' := (\bar{x}_1, ..., \bar{x}_p)$ and

$$\mathbf{U}(y,\mathbf{t}) := \mathbf{A}_1 \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})\, I(Y_i \le y + \mathbf{x}_i'\mathbf{t}),$$

where $\mathbf{A}_1 = \left\{(\mathbf{X}-\bar{\mathbf{X}})'(\mathbf{X}-\bar{\mathbf{X}})\right\}^{-\frac{1}{2}}$ then $E(\mathbf{U}(y,\beta)) = 0$ given that the errors are identically distributed. We can also see that if

$$\mathbf{Q}(\mathbf{t}) := \int_{-\infty}^{\infty} \mathbf{U}'(y,\mathbf{t})\,\mathbf{U}(y,\mathbf{t})\,dy \tag{18}$$

then we can define an estimator of β, say $\hat{\beta}$ which minimizes $\mathbf{Q}(\mathbf{t})$. It is worthwhile to note that this is an extension of the Hodges and Lehmann (1963) estimator for $H(y) \equiv y$.

If we compare the moment conditions for the minimization of the rank estimator procedure and the minimum distance estimator, we can interpret $I(Y_i \le y + \mathbf{x}_i'\mathbf{t})$ is being used as a *generalized score* as it satisfies the criterion that it is nondecreasing in argument y, also if you consider the symmetric error case then we also have them to be 'symmetric' around 0.

Now it is easy to see that the estimators obtained by minimizing the distance measures are translation invariant. Also it can be shown by simple algebraic manipulation that for the case of symmetric *iid* errors we have the problem reduced to

$$\beta^{+}=\frac{1}{2}\underset{\mathbf{t}\in\mathbb{R}^{p}}{\arg\min}\sum_{k=1}^{p}\sum_{i}\sum_{j}d_{ik}d_{jk}\left\{\begin{array}{c}2\left|H\left(Y_{i}-\mathbf{x}_{i}'\mathbf{t}\right)-H\left(-Y_{j}+\mathbf{x}_{j}'\mathbf{t}\right)\right|\\-\left|H\left(-Y_{i}+\mathbf{x}_{i}'\mathbf{t}\right)-H\left(-Y_{j}+\mathbf{x}_{j}'\mathbf{t}\right)\right|\\-\left|H\left(Y_{i}-\mathbf{x}_{i}'\mathbf{t}\right)-H\left(Y_{j}-\mathbf{x}_{j}'\mathbf{t}\right)\right|\end{array}\right\}\tag{19}$$

where $\mathbf{A}=\left(a_{(1)},a_{(2)},...,a_{(p)}\right),d_{ik}=\mathbf{x}_{i}'a_{(k)}$ and we have used the fact that $2\max\left(a,b\right)\equiv a+b+\left|a-b\right|$ for real a,b. Also we have

$$\beta^{+}=\underset{\mathbf{t}\in\mathbb{R}^{p}}{\arg\min}\sum_{k=1}^{p}\sum_{i}\sum_{j}d_{ik}d_{jk}\left\{\begin{array}{c}\left|H\left(Y_{i}-\mathbf{x}_{i}'\mathbf{t}\right)-H\left(-Y_{j}+\mathbf{x}_{j}'\mathbf{t}\right)\right|-\\\left|H\left(Y_{i}-\mathbf{x}_{i}'\mathbf{t}\right)-H\left(-Y_{j}+\mathbf{x}_{j}'\mathbf{t}\right)\right|\end{array}\right\}\tag{20}$$

where we have an added assumption on the structure on $H\left(.\right)$ namely,

$$\left|H\left(a\right)-H\left(b\right)\right|=\left|H\left(-a\right)-H\left(-b\right)\right|\forall a,b\in\mathbb{R}.\tag{21}$$

Further if we have $H\left(.\right)$ is smooth enough,we can obtain β^{+} solves $\dot{M}\left(\mathbf{t}\right)=0$.If we consider $p=1,x_{i1}=1,H\left(y\right)\equiv y$ we get the famous Hodges and Lehmann(1963) one sample estimator for location which is the median of Walsh averages. Also in a multi-sample setup with different sample sizes we obtain a vector of H-L estimator for the group.

In case of the second case where we only assumed that the errors ε are identically distributed,using $\sum_{i}\left(\mathbf{x}_{i}-\bar{\mathbf{x}}\right)=\mathbf{0}$ we have

$$\mathbf{Q}\left(\mathbf{t}\right)=-\frac{1}{2}\sum_{k=1}^{p}\sum_{i}\sum_{j}d_{ik}d_{jk}\left|Y_{i}-\mathbf{x}_{i}'\mathbf{t}-Y_{j}+\mathbf{x}_{j}'\mathbf{t}\right|\tag{22}$$

where $\mathbf{A}_{1}=\left(a_{1(1)},a_{1(2)},...,a_{1(p)}\right),\ d_{ik}=\left(\mathbf{x}_{i}-\bar{\mathbf{x}}\right)'a_{1(k)}$ Here again by using a similar set of values we can get the two sample H-L location estimator.

This suggests that even if you only started with identically distributed error terms ε_{i}, for the distance measure you consider for setting up the optimization problem,you don't really need to know the individual observations of the errors but only the difference between a pair of them depending on your selection of $H\left(.\right)$. This is a very crucial finding on the part of the paper as you can look at different applications like measures of fixed treatment effects by looking at the difference of observations undergoing the treatment (similar to Honoré ,1992), or the pairwise difference estimator we are going to discuss shortly.

Koul(1985) also claims that under certain regularity conditions on the distribution function F_{i} and the respective densities f_{i} of the error term ε_{i} as well as the function $H\left(y\right)$ (Theorem 1, pp.4) we have

$$\mathbf{A}^{-1}\left(\beta^{+}-\beta\right)=B^{-1}\mathbf{S}+o_{p}\left(1\right)\tag{23}$$

where

$$\begin{aligned}
\mu_i(y) &:= F_i(y) - 1 + F_i(-y), y \in \mathbb{R}, \qquad (24)\\
\sigma_i^2(y) &:= F_i(y) - 1 + F_i(-y) - \mu_i^2(y), y \geq 0; 1 \leq i \leq n.\\
\sigma^2(y) &:= \sum ||\mathbf{c}_i||^2 \sigma_i^2(y), y > 0;\\
\gamma(y) &:= \sum ||\mathbf{c}_i||^2 f_i(y), y \in \mathbb{R}; \mathbf{c}_i \equiv \mathbf{A}\mathbf{x}_i.\\
\Lambda(y) &:= diag(f_1(y), f_2(y), ..., f_n(y));\\
\Lambda^+(y) &:= \Lambda(y) + \Lambda(-y),\\
K(y) &:= \mathbf{A}\mathbf{X}'\Lambda^+(y)\mathbf{X}\mathbf{A}, y \in \mathbb{R};\\
B &:= \int K'(y) K(y) dH(y),\\
\mathbf{S} &: = -\int K(y) \{\mathbf{W}(y) + \mathbf{b}(y)\} dH(y),\\
\mathbf{b}(y) &: = \sum_i \mathbf{c}_i \mu_i(y),\\
\mathbf{W}(y) &:= \mathbf{V}(y, \beta) - \mathbf{b}(y), y \in \mathbb{R}.
\end{aligned}$$

Moreover, given that if all the error terms have the same distribution F with density f , and that f is square integrable with H, as well as $0 < \int_0^\infty (1-F)\, dH < \infty$, and the integrals of f and f^2 are both continuous at 0 an some other regularity conditions if F is symmetric around 0, then

$$\mathbf{A}^{-1}\left(\beta^+ - \beta\right) \Longrightarrow N\left(\mathbf{0}, \tau^2 \mathbf{I}_{p\times p}\right) \qquad (25)$$

where $\tau^2 := \left\{2\int f^2 dH\right\}^{-2} \int\int [F(\max\{x,y\}) - F(x)F(y)]\, d\psi(x)\, d\psi(y)$.

Under similar analogous conditions of the first theorem we have given $\mathbf{X}_c = \mathbf{X} - \bar{\mathbf{X}}$, we have(Theorem 2, pp.5) that if F_i be the distribution function of ε_i with density f_i, and if $H(y) \equiv y$, then replacing the $\mathbf{X}'s$ by $\mathbf{X}_c = \mathbf{X} - \bar{\mathbf{X}}$ and using $d_1 := \mathbf{A}_1(\mathbf{x}_i - \bar{\mathbf{x}})$ instead of $\mathbf{c}$,

$$\mathbf{A}_1^{-1}\left(\hat{\beta} - \beta\right) = B_1^{-1}\mathbf{S}_1 + o_p(1) \qquad (26)$$

Also under the added restrictions of square integrability of f and a finite integral for $F(y)(1-F(y))$ we have

$$\mathbf{A}_1^{-1}\left(\hat{\beta} - \beta\right) \Longrightarrow N\left(\mathbf{0}, \tau_1^2 \mathbf{I}_{p\times p}\right) \qquad (27)$$

where $\tau_1^2 := (12)^{-1}\left\{\int f^2(y)\, dy\right\}^{-2}$ .This is nothing but the asymptotic variance term of Wilcoxon rank estimator of β. He also goes on to establish the ARE of these estimators as compared to least squares as well as robustness properties like bounded influence functions.

2.3 Pairwise Difference Estimators of censored and truncated Regression Models

Consider first the semiparametric censored regression model given by

$$y_i = \max\{0, x_i'\beta + \varepsilon_i\} \Rightarrow y_i - x_i'\beta = \max\{-x_i'\beta, \varepsilon_i\} \tag{28}$$

where ε_i had an unknown distribution independent of x_i. Likelihood based methods for a model like this are in general inconsistent if in particular the distribution of the error terms is misspecified. The new class of estimators suggested here are closely associated with the R-estimators discussed earlier. As our focus is mainly on the truncated regression model let us try to motivate the idea using the following truncated regression model

$$\begin{aligned} y_i^* &= x_i^{*\prime}\beta + \varepsilon_i \\ y_i &= y_i^* \text{ if } y_i^* > t_i^* \end{aligned} \tag{29}$$

where t_i^* can a variable but observed point of truncation can also be referred to as we have a data (y_i, x_i, t_i) from a conditional distribution of (y_i^*, x_i^*, t_i^*). So we can write the residual as $\varepsilon_i = y_i^* - x_i^{*\prime}\beta$. Now let us for the sake of exposition assume $t_i = 0 \forall i = 1, ..., n$. Then we have

$$\varepsilon_i = y_i - x_i'\beta \text{ if } y_i^* > 0 \Rightarrow \varepsilon_i = y_i - x_i'\beta \text{ if } \varepsilon_i > -x_i'\beta \tag{30}$$

Now consider the following pair of observations y_i and y_j, from the above formula the corresponding error terms after it has been "trimmed" in the feasible (observable range) is given by

$$\varepsilon_i = (y_i - x_i'\beta) \text{ if } \varepsilon_i > \max\left\{-x_i'\beta, -x_j'\beta\right\} \tag{31}$$

so what we see is that we only need to make sure that the error term $\varepsilon_i > -x_j'\beta$.However , as we see this still doesn't make it symmetric or iid even if we know x_i and x_j. This goal can be achieved if we consider the difference between 2 such adjusted or trimmed error terms. This could be a worthwhile exercise as we have seen in the case of the minimum distance estimator where we only know that the errors are identically distributed we really don't need to know each individual error term, so long as we know the weighted difference .So,although the motivation for this exercise is totally different for the author,we have the following result

$$\begin{aligned} \varepsilon_i - \varepsilon_j &= \left(y_i - x_i'\beta - y_j - x_j'\beta\right) \text{ if } \varepsilon_i, \varepsilon_j > \max\left\{-x_i'\beta, -x_j'\beta\right\} \\ &\Rightarrow \varepsilon_i - \varepsilon_j = \left(y_i - x_i'\beta - y_j + x_j'\beta\right) I\left(\varepsilon_i > -x_j'\beta, \varepsilon_j > -x_i'\beta\right) \\ &\Rightarrow \varepsilon_i - \varepsilon_j = \left(y_i - y_j - (x_i - x_j)'\beta\right) I\left(\varepsilon_i > -x_j'\beta, \varepsilon_j > -x_i'\beta\right). \end{aligned} \tag{32}$$

The above expression should be symmetrically distributed around 0 given x_i and x_j.The corresponding estimator is given by

$$\begin{aligned}\nu_{ij}(b) &= (y_i - y_j - (x_i - x_j)'b)\, I\left(y_i - x_i'b > -x_j'b, y_j - x_j'b > -x_i'b\right) \\ &= (y_i - y_j - (x_i - x_j)'b)\, I\left(-y_j < (x_i - x_j)'\, b < y_i\right)\end{aligned} \quad (33)$$

which is symmetrically distributed around 0 given the truncated regression model. So if we consider any odd function ξ of $\nu_{ij}(.)$ we can have the moment condition given by $E\left[I\left(-y_j < (x_i - x_j)'\, b < y_i\right)\xi\left((y_i - y_j - (x_i - x_j)'b)\right)\right] = 0$, this is true as we can only define functions on the residual in the feasible or visible range. In order to make the above moment condition into a valid first order condition, we need to define the problem as a problem to minimize $E\left[t\left(y_i, y_j, (x_i - x_j)'b\right)|x_i, x_j\right]$ as a function of $(x_i - x_j)'b$.If we consider a function $\Xi(.)$ with a right continuous first derivative $\xi(.)$ and other characteristics of a distance or dispersion estimator then we have replacing $(x_i - x_j)'b$ by δ,

$$t(y_i, y_j, \delta) = \begin{cases} \Xi(y_1) & \text{if} \quad \delta \leq -y_2 \\ \Xi(y_1 - y_2 - \delta) & \text{if} \quad -y_2 \leq \delta \leq y_1 \\ \Xi(-y_2) & \text{if} \quad \delta \geq y_1 \end{cases} \quad (34)$$

so we can estimate β by minimizing the sample counterpart as the first order condition is satisfied for $(x_i - x_j)'b = (x_i - x_j)'\beta$ as seen from the previous results they

$$T_n(b) = \binom{n}{2}^{-1} \sum_{i<j} t\left(y_i, y_j, (x_i - x_j)'b\right) \quad (35)$$

Honoré and Powell (1994) claim that an added assumption of log-concavity of the errors ε is sufficient to ensure that the solution to the FOC is the global minimizer. If we consider a loss function or a dispersion measure given by $\Xi(d) = |d|$,the moment condition is almost identical to the Mann-Whitney type truncated regression estimator proposed by Bhattacharya,Chernoff and Yang (1983). The method they followed in this paper is to come up with the moment condition first and then look at the optimization problem first instead of the other way round , and use this moment condition as the first order condition.

If there is no truncation or censoring the procedure outlined here will result in the following criterion function to minimize

$$W_n(b) = \binom{n}{2}^{-1} \sum_{i<j} \Xi\left(y_i - y_j - (x_i - x_j)'b\right) \quad (36)$$

which will give the traditional two sample location estimator proposed by Hodges and Lehmann(1963)([6]) if $\Xi(d) = |d|$ .The first order moment condition which arises in this case is $E(\xi(y_i - y_j - (x_i - x_j)'\beta)) = 0$ is similar to the conditions of the rank regression estimators. Trivially, $\Xi(d) = d^2$ gives the least squares estimator.

The procedure outlined here is one of minimizing second order U-Statistics processes for which there are well established large sample theory available, however there is one important shortcoming of the procedure here which is one of the regularity conditions in the literature namely, uniform boundedness of the kernel of the U-statistics. However ,this paper follows the relaxed regularity conditions proposed by Robert Sherman.

Let us consider an estimator $\hat{\theta}$ which minimize the m^{th} order U-statistic for a sample $\{z_i\}$ of iid variables given by

$$U_n(\theta) = \binom{n}{m}^{-1} \sum_{c \in C\{i_1, i_2, \ldots, i_m\}} p(z_{i_1}, z_{i_2}, \ldots, z_{i_m}, \theta) \tag{37}$$

as U-statistics can be seen as an extension of the more popular M-estimators, although most of the large sample results that are valid for the class of M-estimators goes through for U-Statistics as well. Like under the assumptions of a compact parameter space Θ, measurable and continuous kernel $p(.)$ and that the absolute value of kernel $p(.)$ is dominated by an integrable function on the sample space, it can be shown (Theorem1 pp.248) that $U_n(\theta) - E[U_n(\theta)]$ converges to zero almost surely, uniformly in θ. So under the above assumptions the estimator $\hat{\theta} = \arg\min_{\theta\in\Theta} U_n(\theta)$ is strongly consistent for the true parameter θ_0 which minimizes $E[U_n(\theta)]$. In fact the apparent strong assumption of compactness of Θ can be relaxed if $p(.)$ is convex. It can also be shown under certain regularity conditions on the differentiability of $p(.)$ that if the normalized sub-gradient of a m^{th} order U-statistic, given by $Q_n(\theta) \equiv \sqrt{n}\binom{n}{m}^{-1} \sum_c q(z_{i_1}, z_{i_2}, \ldots, z_{i_m}, \theta)$, where each component of the kernel $q(.)$ is a linear combination of left and right partial derivatives of $p(.)$, then

$$Q_n\left(\hat{\theta}\right) = o_p(1) \tag{38}$$

which is an approximate moment condition. The asymptotic normality of these estimators can also be shown following Huber(1967).The result says that if we have θ_0 is an interior point where the kernel $q(.,\theta)$ is well behaved,the unconditional expectation $\lambda(\theta)$ of $q(.,\theta)$ vanishes at the true value and is differentiable at θ_0 with $H_0 \equiv \partial\lambda(\theta_0)/\partial\theta'$ is non-singular. The expected values of both u-statistic and its square is bounded and finally the kernel $q(., \theta_0)$ is square integrable, then we have

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) = -H_0^{-1}\frac{m}{\sqrt{n}}\sum_{i=1}^{n} r(z_i, \theta_0) + o_p(1) \tag{39}$$

and $\hat{\theta}$ is asymptotically normal,

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \Rightarrow N\left(0, H_0^{-1}V_0H_0^{-1}\right), \tag{40}$$

where $V_0 \equiv m^2E\left[r(z_i,\theta_0)\, r(z_i,\theta_0)'\right]$. Estimation would require consistent estimators of the variance covariance matrix, for this we will need consistent estimators for both H_0 as well as V_0. Now from the previous results we have $\tilde{H}_n \equiv \partial Q_n\left(\hat{\theta}\right)/\partial\theta'$, so the l^{th} column is

$$\hat{H}_{nl} \equiv \left(2\hat{h}\right)^{-1}\left[Q_n\left(\hat{\theta}+\hat{h}e_l\right) - Q_n\left(\hat{\theta}-\hat{h}e_l\right)\right] \tag{41}$$

where $\hat{h}$ is a possibly stochastic sequence of bandwidths and e_l is the l^{th} basis vector. Now if we define

$$\hat{r}(z_i,\theta) \equiv \binom{n-1}{m-1}^{-1}\sum_{c'} q(z_i, z_{i_2}, ..., z_{i_m}) \tag{42}$$

where we consider all observations excepting the i^{th} one. Now we can use the average $\frac{1}{n}\sum_{i=1}^{n}\hat{r}\left(z_i,\hat{\theta}\right)$ as an estimator of the conditional average of $r(z_i,\theta)$.

Hence ,they claim that a consistent estimator of V_0 is

$$\hat{V}_n \equiv \frac{m^2}{n}\sum_{i=1}^{n}\hat{r}\left(z_i,\hat{\theta}\right)\hat{r}\left(z_i,\hat{\theta}\right)' \tag{43}$$

Now if we focus on the asymptotic properties of the truncated regression estimator which can be defined as

$$\arg\min_b \sum_{i<j}\left(t\left(y_i, y_j, (x_i - x_j)' b\right) - t\left(y_i, y_j, (x_i - x_j)'\beta\right)\right) \tag{44}$$

Let us assume that expectations exists for $x, \xi(y)$ and $x\xi(y)$.We also assume that no linear subspace of $\mathbb{R}^K$ which contains $(x_1 - x_2)'$ with certainty and last but not the least the error terms ε_i are iid and independent of x_i and have a continuous distribution function F and a continuous density f which is bounded from above.

If the first two conditions hold and the added assumption that the errors ε are iid and the logarithm of its density is strictly concave, then the expectation of the above objective function exists and is uniquely minimized at $b = \beta$ Under certain other regularity conditions it can be shown that the estimator over a compact parameter space is asymptotically normal for interior β

$$\sqrt{n}\left(\hat{\beta} - \beta_0\right) \Rightarrow N\left(0, H_0^{-1}V_0H_0^{-1}\right) \tag{45}$$

As we showed before if

$$\hat{r}_i \equiv \frac{1}{n-1}\sum_{j\neq i} -\xi\left(y_i - y_j - (x_i - x_j)'b\right) I\left(-y_j < (x_i - x_j)'\,\hat{\beta} < y_i\right)$$

then $\hat{V}_n \equiv \frac{4}{n}\sum_{i=1}^{n}\hat{r}_i\hat{r}_i'$ is a consistent estimator for V_0.The numerical derivative estimator has the l^{th} column given by

$$\hat{H}_{nl} \equiv \left(2\hat{h}\right)^{-1}\binom{n}{2}^{-1}\sum_{i<j}\begin{pmatrix} -I\left(-y_j < (x_i - x_j)'\,(\hat{\beta} + \hat{h}e_l) < y_i\right) \\ \times\xi\left(y_i - y_j - (x_i - x_j)'(\hat{\beta} + \hat{h}e_l)\right) \\ +I\left(-y_j < (x_i - x_j)'\,(\hat{\beta} - \hat{h}e_l) < y_i\right)\times \\ \xi\left(y_i - y_j - (x_i - x_j)'(\hat{\beta} - \hat{h}e_l)\right) \end{pmatrix} \tag{46}$$

$$\times\,(x_i - x_j)$$

2.4 Multivariate Location Estimation with R-Estimates

If we have two positive integers m and n such that $1 \leq m \leq n$, if $A_n^{(m)}$ be the collection of all subsets of size m from the set $\{1, 2, ..., n\}$, and $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n \in \mathbb{R}^d$ and for any $\alpha \in A_n^{(m)}$, define $\bar{\mathbf{X}}_\alpha = \frac{1}{m}\sum_{i\in\alpha}\mathbf{X}_i$, then the mth order Hodges and Lehmann estimate could be $\hat{\theta}_n^{(m)}$ defined as

$$\sum_{\alpha\in A_n^{(m)}}\left|\bar{\mathbf{X}}_\alpha - \hat{\theta}_n^{(m)}\right| = \min_{\theta\in\mathbb{R}^d}\sum_{\alpha\in A_n^{(m)}}\left|\bar{\mathbf{X}}_\alpha - \theta\right| \tag{47}$$

Unless we have points $\bar{\mathbf{X}}_\alpha$ forming a single line this estimator will be unique. For $m = 1$ this gives an estimate of a multivariate median and for $m = 2$ we have the standard one sample H-L estimate in d-dimension.

If further we define the unit vector in the direction of $x \in \mathbb{R}^d$ as $U(x) = |x|^{-1}x$ if $x \neq 0; = 0$ if $x = 0$. Also denote the $d \times d$ Hessian matrix of x , $P(x) = |x|^{-1}\left(I_d - |x|^{-2}xx'\right)$ if $x \neq 0; = 0$ if $x = 0$. If $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_m \in \mathbb{R}^d$

be a collection of iid random variables and define $\theta^{(m)}$ be the median of the sampling distribution of $\bar{\mathbf{X}}_m$ if

$$\begin{aligned}\theta^{(m)} &= \arg\min_{\theta\in\mathbb{R}^d} g(\theta) \\ &= \arg\min_{\theta\in\mathbb{R}^d} E\left\{\left|\bar{\mathbf{X}}_m - \theta\right| - \left|\bar{\mathbf{X}}_m\right|\right\}\end{aligned} \tag{48}$$

which is unique if $\mathbf{X}_i's$ are absolutely continuous with respect to the Lebesgue measure Also we have $E\left\{U\left(\left|\bar{\mathbf{X}}_m - \theta^{(m)}\right|\right)\right\} = 0$, so we can without loss of generality shift location $\theta^{(m)}$ to 0.Also define,$D_1^{(m)} = E\left\{P\left(\bar{X}_m\right)\right\}$, $U^{(m)}(\mathbf{X}_1) = E\left\{U\left(\bar{\mathbf{X}}_m|\mathbf{X}_1\right)\right\}$ and $D_2^{(m)} = E\left\{U^{(m)}(\mathbf{X}_1)\,U^{(m)}(\mathbf{X}_1)'\right\}$. Now given the added assumption that besides being iid random variables $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_m \in \mathbb{R}^d$ also have an absolutely continuous density function f which is bounded in every bounded subset of $\mathbb{R}^d$ we have

Theorem 3 *Under the given assumption above,*$D_1^{(m)}$ *is a p.d. matrix and a Bahadur type representation for the* m^{th} *order Hodges-Lehmann estimate is given by:*

$$\hat{\theta}_n^{(m)} = \frac{m!\,(n-m)!}{n!}\left[D_1^{(m)}\right]^{-1}\sum_{\alpha\in A_n^{(m)}} U\left(\bar{\mathbf{X}}_\alpha\right) + \mathbf{R}_n,$$

where as $n\to\infty$*, the remainder term* $\boldsymbol{R}_n$ *is almost surely* $O\left(\frac{\log n}{n}\right)$ *if* $d\geq 3$*. When* $d=2$*,* $\boldsymbol{R}_n$ *is almost surely* $o\left(\left[\frac{\log n}{n}\right]^{\omega}\right)$ *as n tends to* ∞ *for any constant* ω *such that* $o<\omega<1$*.*

Proof. See Chaudhuri(1992). ■

Corollary 4 *If the assumptions above are satisfied, then* $D_2^{(m)}$ *is a positive definite matrix ,and as n tends to* ∞, $\sqrt{n}\hat{\theta}_n^{(m)}$ *converges weakly to a d-dimensional normal random vector with mean zero and the dispersion matrix given by*

$$m^2\left[D_1^{(m)}\right]^{-1}\left[D_2^{(m)}\right]\left[D_1^{(m)}\right]^{-1}.$$

3 Generalized Minimum Distance and 'm-rank' Regression for Truncated Regression Models with Heterogeneity in errors

We suggest two estimators for the semiparametric truncated linear regression model

$$y_i^* = x_i^{*\prime}\beta + \varepsilon_i, \tag{49}$$
$$y_i = y_i^* \text{ if } y_i^* > t_i^*$$

where t_i^* can a variable but observed point of truncation we can consider without loss of generality that the truncation points t_i^* are 0 if they are observed as we can look at a transformed latent variable $y_i^* - t_i^*$ instead of just y_i^*, we also have to make the same adjustment to y_i.

Firstly one which is related to minimum distance estimation. The framework we discussed above from Koul's paper can be easily extended by using a weight or a projection matrix to transform the distance measure to the space with limited dependent variable. So, criterion function is

$$\mathbf{M}(\mathbf{t}) := \int \mathbf{V}'(y, \mathbf{t})\, \mathbf{W}^{-1} \mathbf{V}(y, \mathbf{t})\, dH(y) \tag{50}$$

where $H(y)$ is a non-decreasing right continuous function.

This function reduces to the minimum distance criterion function when $\mathbf{W} = \mathbf{I}_{n\times n}$,we can construct a suitable weight function when we have a truncated regression model using it as a projection operator onto the space of errors ε_i conditional on $\varepsilon_i > -x_i'\beta$. The weight functions which possibly could depend on the parameter estimate as in the case of a truncated regression model discussed earlier could be estimated iteratively using a Hodges-Lehmann type estimator as the initial value.

Our objective is to obtain a general distance function which can be minimized to obtain a set of approximate first order conditions which are then solved to obtain an estimate of regression coefficient in a general semiparametric regression framework. This should be equipped enough to cover cases like a truncated regression model and converge to a standard set of regression coefficients where is no truncation or censoring. Consider the following population regression model

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1} & \mathbf{X}_c \end{bmatrix} \begin{pmatrix} \alpha^* \\ \beta \end{pmatrix} + \mathbf{e}$$

as described in the previous section where columns of $\mathbf{X}_c$ are orthogonal to the constant term. We are only interested in the estimation of β as it is not possible to identify the intercept coefficient in the current framework. Consider the function

$$D^{(m)}(\mathbf{Z}) = \sum_{c\in\mathcal{C}^m} a\left(R\left(h^{(m)}\left(Z_{(c)}\right)\right)\right) h^{(m)}\left(Z_{(c)}\right) \tag{51}$$

where $\mathcal{C}^m$ is a set of m from the set $\{1,2,...,N\}$, $h^{(m)}(.)$ is a real valued function from $\mathbb{R}^m \rightarrow \mathbb{R}$ .This is a valid dispersion function under certain regularity conditions on the function $h^{(m)}$. It is worth noting here the function $h^m(.)$ might actually include a lebesgue measure $H(.)$ As a special case we can see that if we have $m = 1$ and $h^{(1)}(Z_i) = Z_i$ then we get back the traditional rank estimate of dispersion discussed earlier in the report. Also if $m = 2$, the dispersion measure takes the form

$$D^{(2)}(\mathbf{Z}) = \sum_{i>j}^{N} a\left(R\left(Z_i - Z_j\right)\right)\left(Z_i - Z_j\right) \tag{52}$$

which turns out to be the criterion function of the pairwise difference estimators we have discussed earlier. With a suitable choice of the score function $a(.)$ satisfying the criterion of symmetry in the sense $a(k)+a(N-k+1) = 0$ and it is a non-decreasing non-constant function, we can extend this to a whole class of criterion functions which could be minimized to obtain robust R-estimators of the regression coefficients β.

Let us now discuss the case where $m = 3$. Define

$$h^{(3)}\left(Z_{\{i,j,k\}}\right) = Z_i - \frac{Z_j + Z_k}{2} \tag{53}$$

,this gives the measure of dispersion as

$$D^{(3)}(\mathbf{Y} - \mathbf{X}\beta) = \sum_{i>j>k}^{N} a\left(R\left(Y_i - \mathbf{x}_i'\beta - \frac{Y_j - \mathbf{x}_j'\beta - Y_k - \mathbf{x}_k'\beta}{2}\right)\right) \times \left(Y_i - \mathbf{x}_i'\beta - \frac{Y_j - \mathbf{x}_j'\beta - Y_k - \mathbf{x}_k'\beta}{2}\right) \tag{54}$$

this is a valid measure of dispersion being a non-negative,continuous and convex function of β .This follows from the proof in Jaeckel(1972,[10]) if you consider the set of $\left(Y_i - \mathbf{x}_i'\beta - \frac{Y_j - \mathbf{x}_j'\beta - Y_k - \mathbf{x}_k'\beta}{2}\right)$ as $Z_i's$.Hence this should have a minimum which can be obtained by solved by solving the first order conditions.

So as we have the following the following expression as the derivative with respect to the j^{ith} component β_j

$$\frac{\partial}{\partial \beta_l} D^{(3)}(\mathbf{Y}-\mathbf{X}\beta) = \sum_{i>j>k}^{N} a\left(R\left(Y_i - \mathbf{x}_i'\beta - \frac{Y_j - \mathbf{x}_j'\beta - Y_k - \mathbf{x}_k'\beta}{2}\right)\right) \tag{55}$$

$$\times\left(-\mathbf{x}_{il}' - \frac{-\mathbf{x}_{jl}' - \mathbf{x}_{kl}'}{2}\right)$$

$$= -\frac{1}{2}\sum_{i>j>k}^{N} a\left(R\left(Y_i - \mathbf{x}_i'\beta - \frac{Y_j - \mathbf{x}_j'\beta - Y_k - \mathbf{x}_k'\beta}{2}\right)\right)$$

$$\times(2x_{il} - x_{jl} - x_{kl})$$

$$= -\frac{1}{2}\sum_{i>j>k}^{N} a\left(R\left(Y_i - \mathbf{x}_i'\beta - \frac{Y_j - \mathbf{x}_j'\beta - Y_k - \mathbf{x}_k'\beta}{2}\right)\right)$$

$$\times((2x_{il} - \bar{x}_i) - (x_{jl} - \bar{x}_j) - (x_{kl} - \bar{x}_k))$$

hence we can claim that the first order are(given that D is differentiable almost everywhere)

$$-\nabla D(\mathbf{Y}-\mathbf{X}\beta) = \mathbf{S}(\mathbf{Y}-\mathbf{X}\beta) \doteq \mathbf{0} \tag{56}$$

This estimator will have the same distribution asymptotic distributions of multivariate normal under certain regularity conditions. Moreover this can also be viewed as an extension of the regression coefficient estimator of as would be defined by the m^{th} order Hodges-Lehmann location estimator as proposed by Chaudhuri(1992).Hence we should be able to find a Bahadur type representation of the coefficient estimates $\hat{\beta}$ and hence from the corollary this should weakly converge to a p-dimensional normal distribution under certain regularity conditions.

We are proposing an estimator where we use a more robust version of the rank of the difference of the residuals, we can call it the $m - rank$ is the following. If we have a possible heteroscedasticity in the error terms this robust $m - rank$ can substantially reduce it for higher values of m. This also has a trimming effect on the extreme values of these differences particularly the ones on the lower end if we have a non-decreasing score function $a(.)$. It should also improve the breakdown performance of the suggested estimator.

So it is possible at least theoretically to obtain an estimator for limited dependent variable models like truncated(or censored regression models) using methods like a modified version of the method of R-estimates which reduces to the simple problem of a Hodges-Lehmann location estimation when the $\mathbf{X} = \mathbf{1}$. The asymptotic properties of these estimators could be obtained by a slight modification of the existing asymptotics in the case of minimum distance estimation and more precisely the literature on estimates based on regression rank statistics. The estimators proposed here have criteria functions as extensions of U-statistics.

4 Summary and Directions

It is a very interesting problem to look deeper into the properties of the proposed estimator and its comparison in with different methods under varied conditions on the error distribution and functions $h^{(m)}(.)$ as well as more general score functions $a(.)$. The asymptotic properties of the estimator are yet to be proven concretely and also some cases where there will not be enough regularity conditions like a twice continuously differentiable distribution function $F(.)$ for the error. Also it might be worthwhile to look into the geometric interpretation of a m^{th} order Hodges-Lehmann estimator for regression coefficients particularly as $m \to \infty$.

The suggested estimator also has some promise in more general problems with missing data not truncation as the errors could be modified using the function $h^{(m)}(.)$ and also the choice of scores. It is my conjecture that general missing data particularly MCAR and MAR could be relatively easily handled using the m-rank estimates particularly as these should be higher breakdown point than traditional R-estimates, although this a topic to investigate.

Another possible field where this could be used in the study of treatment effects particularly fixed effects as has been explored by Honoré (1992) which talks about differences of treatment effects for similar subjects which removes the group effect and can predict individual effect more accurately. One possible challenge ahead is to suggest a consistent estimate of the variance - covariance matrix of the proposed estimator,although the rank estimators gives some good insight into this problem.

Our objective for this report was trying to find an extension of the traditional R-estimates for regression coefficients particularly to estimate models with truncated data. The pairwise difference techniques proposed by Honoré and Powell(1994) ,which is equivalent to the extension of the Mann-Whitney-Wilcoxon type method proposed by Bhattacharya,Chernoff and Yang(1983) in case of absolute errors are a special case of the proposed estimator. However this work is very preliminary and requires much closer attention and Monte Carlo comparison with other existing methods to actually figure out its merits and shortcomings. We have also looked at a class of minimum distance or minimum chi-square distance estimators to delve into this matter. It is possible to estimate at least theoretically the actual ability variable which is a latent variable in the problem of estimating performance of fund managers as well as efficacy of drugs or evaluation of social programs with a suitable choice of the function $h^{(m)}(.)$ and of course m. However we have not discussed any such procedure or selection criterion of h or m in this report. Theoretical proofs have been omitted to keep the report more heuristic as this is still work in progress.

References

[1] ADICHIE,J.N.(1967).Estimates of regression parameters based on rank test.*Annals of Mathematical Statistics.***38:**894-904

[2] BAHADUR,R.R.(1966).A note on quantiles in large samples.*Annals of Mathematical Statistics.***37:**577-580

[3] BHATTACHARYA,P.K.,H.CHERNOFF AND S.S.YANG(1983).Nonparametric estimation of the slope of a truncated regression.*Annals of Statistics.***11:**505-514.

[4] CHAUDHURI,P.(1992).Multivariate location estimation using extension of R-estimates through U-statistics type approach.*Annals of Statistics.***20:**897-916.

[5] GREGORY,G.G.(1990).Generalized Hodges-Lehmann estimators for the analysis of variance.*Journal of Multivariate Analysis.***33:**294-309.

[6] HODGES,J.L.,JR. AND E.L.LEHMANN(1963)Estimates of location based on rank tests.*Annals of Mathematical Statistics.***34:**598-611.

[7] HONORÉ,B.E.(1992)Trimmed LAD and least Squares estimation of truncated and censored regression models.with fixed effects.*Econometrica.***60:3:**533-565.

[8] HONORÉ,B.E. AND J.POWELL(1994)Pairwise difference estimators of truncated and censored regression models. *Journal of Econometrics.***64:**241-278.

[9] HETTMANSPERGER,T.P.(1985) *Statistical Inference based on ranks.*John Wiley and Sons.New York.

[10] JAECKEL,L.A.(1972).Estimating regression coefficients by minimizing the dispersion of the residuals.*Annals of Mathematical Statistics.***43:**1449-1458

[11] JURECKOVA,J.(1971).Nonparametric estimate of regression coefficients.*Annals of Mathematical Statistics.***42:**1328-1338.

[12] KOUL,H.L.(1985A).Minimum distance estimation in multiple linear regression model.*Sankhyā.***47:1.Series A:**57-74.

[13] KOUL,H.L.(1985B).Minimum distance estimation in multiple linear regression with unknown error distribution.*Statistics and Probability Letters* .**3.** 1-8.

[14] ROSENBAUM,PAUL(1993).Hodges-Lehmann point estimates of treatment effect in observational studies.*Journal of the American Statistical Association.***88:424.**1250-1253.