10-2009

A semi-parametric two-stage projection type estimator of multivalued treatment effects

Aurobindo GHOSH
Singapore Management University, AUROBINDO@SMU.EDU.SG

January 15, 2007

A Semi-parametric Two-Stage Projection Type estimator of Continuous Treatment Effects

Abstract

One of the most well documented regularities in evaluation literature like returns to schooling (or funding for programs) is that several factors come together to confound the measurement of its effect. First, in observational studies the true return is often individual specific, and so it is almost impossible to use a traditional treatment effect models with randomly assigned treatment and control groups. This endogeneity in the model further exacerbates our inability to conduct such trials. Second, the problem is not a classical treatment effect measurement problem where we have discrete or more often binary treatments. Hence, techniques like measuring the Local Average Treatment Effect (LATE) cannot be implemented as it is not very well defined for the continuous treatment case. Third, a traditional 2SLS approach might be misleading because of the non-Gaussian nature of response distribution, in particular, if different quantiles of response have different effects. However, their technique is also not defined for continuous treatments, and cannot measure if different distributions of the treatment might have different effects on the response variable. In this paper, we propose the effects of different multi-valued continuous treatment variable after conditioning for other covariates.

JEL Classification: C13, C14, C21,

Keywords: Treatment Effect, Instrumental variable, Projection type estimator, Quantile Regression, Exogeneity, Monotonicity, returns-to-schooling.

Acknowledgement 1 *This is a preliminary draft of the paper, please do not quote without author's permission. Thanks to the Center for Academic Computing at Singapore Management University for providing research assistance. This project has been funded by SMU Office of Research Grant No.06-C208-SMU-007. Usual disclaimers apply.*

Aurobindo Ghosh
SMU, 90 Stamford Road, Singapore 178903. Tel +65 68280863. Fax: +65 6828 0833. aurobindo@smu.edu.sg

1 BACKGROUND AND MOTIVATION

Seminal papers in the statistics literature to identify causal relationship like Neymen (1923) which tried to look at the treatment effect had very few choices but to randomize the group of possible recipients of the treatment into the experimental and the control groups and then compare the differences in the means or the distributions of these two populations and test for a significant treatment effect. The recent interest in this field like the Rubin Causal Model (RCM) as coined by Holland grew in the context of social experiments and observational studies where randomization of treatment assignment into treatment and control groups is not a viable alternative. Moreover the focus has shifted from the strength of the relationship between the endogenous regressor and the instrument to the validity of the instrument itself.

Angrist, Imbens and Rubin (1996, henceforth AIR) looked at the problem of identifying the treatment effect for a population where the treatment received was not random by clearly mentioning the assumptions required for obtaining a valid instrument. They also showed that the IV estimand can be given a precise and straightforward causal interpretation in a potential outcomes framework despite non-ignorability of treatment received which means that the treatment received is not independent or in a weaker sense *ignorable* (Rubin,1974) of the person treated with some extension to multiple treatment effects as opposed to the binary treatment effects are discussed in Angrist and Imbens (1995) and Angrist, Graddy and Imbens (1995). However they only talk about the two-stage least squares estimation.

Under the usual assumptions like the exclusion restriction which makes sure that the only effect on the outcome (potential outcome) of the treatment assignment is through the treatment received, the treatment effect could only be identified for a sub-population of *compliers* whose treatment could be changed by the treatment assignment mechanism. This is what they call the *Local Average Treatment Effect* (LATE) as opposed to an *average treatment effect* (ATE) for the whole population as proposed by Robins (see Imbens, 2004). However, the population treatment effect cannot be identified even if the monotonicity assumption holds, only some bounds (Manski, 1990; Balke and Pearl, 1997) for the average treatment effect could be obtained .

The major criticism of this genre of papers is that the LATE effect is for a subgroup which could not be identified (Heckman, 1996; Moffitt, 1996), although we can find the treatment effect with the so called RCM model. Moreover, even monotonicity alone will not

be sufficient to identify more complicated treatment assignments like the bioequivalence trials which compares a standard treatment with a new treatment with an option of not having any treatment at all. One important interpretation of the instrumental variable (IV) procedure is the relation with the propensity score of Rosenbaum and Rubin (1983) which could be used as a instrument itself in the usual two-stage least square procedure.

Rosenbaum's suggestion of a more robust estimator for treatment effects might be one of the reasons for the demand for the quantile treatment effects estimator (Rosenbaum, 2002; Hodges and Lehmann, 1963). The two-stage Least Absolute Deviation (LAD) estimator which is a precursor to the robust treatment effects estimators is also asymptotic normality in the general case. (Amemiya, 1982; Powell 1983).

Amemiya (1982) proposed the LAD equivalent of a 2SLS estimator that minimizes

$$\sum_{t=1}^{T} \left| q y_t + (1-q)\, P_t^{'} y_t - P_t^{\prime} Z \alpha \right| \tag{1}$$

where q is predetermined, and $P_t = X\left(X^{\prime}X\right)^{-1} X^{\prime}$ is the projection matrix. This combines the idea of a two-stage least squares where pure LAD estimation is obtained for $q = 1$. These estimators are more appealing in cases of fatter tailed or leptokurtic distributions, and displayed strong consistency under different mixtures of normal.

Mean treatment effect is restricted to a simple "location shift", so any quantile specific regression is redundant. However, observational studies suffer from non-randomness or *non-ignorability* in treatment assignment, so comparison between *potential outcomes* of treatment and control groups are often misleading. If we focus on only a subgroup of the whole population (viz. the compliers), the subjects whose treatment could be changed by the instrument, we can identify the treatment effects or LATE effect for the subgroup (Imbens and Angrist, 1994; Angrist et. al, 1996).

Our objective in this paper is to extend the results of Abadie, Angrist and Imbens (2002) beyond binary to a more general framework of multiple treatments and treatment assignments. We want to look at the quantile treatment effects for a group of subjects who are assigned a treatment through a "randomized" mechanism. As in Angrist and Imbens (1994), the average treatment effect is not identified for the entire population (we can at most get upto a bound, say, Robins-Manski bound, see Manski, 1990), however the treatment effect is identified for a subset of the whole population. In particular, those

whose treatment could be changed by their assignment i.e. where there is at least partial compliance to the treatment assignment – the compliers. However, so long as the treatment received are in "one direction" in response to the assignment the effect could be identified. The estimator suggested is identified and consistent under certain regularity conditions, it can also be proved that the estimator would have asymptotic normality under more stringent conditions.

We also discuss some computational aspects of the linear programming problem of computational complexity of a well known interior point algorithm used for this purpose (Koenker and D'Orey, 1987). The real problem with implementing a standard Linear Programming (LP) algorithm here is that, the problem is not really LP as in case we have a negative weight function κ(to be defined later), we have to include a non-linear constraint of the error terms u_i, such that, $u_i^+ u_i^- = 0$ to ensure not having an unbounded solution for the program (Abadie et. al., 2002). They use a modified version of the Barrodale and Roberts (BR) algorithm, also implemented, by Koenker and D'Orey (1987) for solving the LP formulation of the quantile regression. However, the nature for our suggested estimator do not ensure that we have a convex function to minimize, so chance of a global minimum is very slim, if any. The way Abadie et. al. (2002) gets around the problem is to use a modified version of the BR algorithm and start at different starting points to check for consistency, after starting with the quantile regression solutions as the initial seed. The main advantage of using an interior point algorithm (like the Frisch–Newton algorithm, based on the logarithmic barrier) over the traditional BR algorithm is the time of computation, as the former can become notoriously slow once we approach the true value. The other reason for using an interior point algorithm like the affine scaling Primal-Dual algorithm is the fact that our problem might not be convex because of the non-linearity in the constraint .So the primal-dual algorithm which uses Newton steps to improve the objective function as well as the barrier method to ensure feasibility simultaneously is designed more for a non-linear programming type problems so it would be better equiped to handle non-linearities. This and related issues including the problem setup is given in Appendix B. Finally, the speed of convergence is a very important issue if we are working with large datasets particularly a panel data with a few thousand observation over a few years. The empirical example we talk about is taken from a panel with 838 individuals with 16 years of data.

2 MULTI-VALUED TREATMENT EFFECT

2.1 Model for Treatment and Assignment

Consider the following model

$$Y = \gamma_0 + X_1\gamma_1 + \rho S + \varepsilon \tag{2}$$

where Y is the dependent variable or response variable, X_1 is the set of included exogenous variables, S is the treatment intensity and γ_0, γ_1 and ρ are constants while ε is a random variable with mean 0 and constant variance that is uncorrelated with X_1. We also consider another augmented equation for the treatment received as

$$S = \delta_0 + X_1\delta_1 + X_2\delta_2 + \eta. \tag{3}$$

The problem is even if the true response is linear, OLS will give biased estimates of ρ, our parameter of interest as the random variables ε and η might be correlated due to some unobserved variables common to both, for example, ability affects both schooling and earnings.

For the general IV formulation the estimate is given by $(Z'W)^{-1} Z'y$ where Z is the matrix of instruments and W is the matrix of regressors including X_1 and S. For a consistent estimator a necessary condition is that Z must be uncorrelated with the regression errors and $plim(\frac{Z'W}{n})$ is non-singular. An alternative formulation is the 2SLS case. In the first stage we start of by regressing S on X_1 and all possible candidate instruments not included in (2), say, X_2. Then using the predicted value of S, or the conditional expectation of S given the covariates and the instruments in equation (2),

$$Y = \gamma_0 + X_1\gamma_1 + \rho\widehat{S} + \nu \text{ where } \nu = \left\{\varepsilon - \rho\left(S - \widehat{S}\right)\right\}. \tag{4}$$

Let us now consider the case where we have a rational number of levels of treatment and we express them as integers $j = 0, 1, \ldots J$, and Y_j is the response given treatment level j, we assume that for each individual there exists Y_j for all j, although we can observe only one of them. We also assume the standard *Stable Unit Treatment Value Assumption* (SUTVA) as formulated by Rubin (1974), which states that the potential outcomes of a

single individual are independent of the potential outcomes and treatment status of all other individuals.

We are interested in the distribution of $Y_j - Y_{j-1}$, which is the return to the j^{th} unit of treatment. The estimates can be interpreted as a *causal* effect when its probability limit approaches some weighted average of $E[Y_j - Y_{j-1}]$ for all j in some sub-population of interest. However, as S is not randomly assigned we cannot simply subtract the average response of subjects receiving $(j-1)^{th}$ level of treatment from the average of subjects receiving j^{th} level of treatment.

We also consider $S_Z \in \{0, 1, \ldots, J\}$ to be the level of treatment given the value of the instrument Z. Here again we assume that S_Z exists for all values of Z although only one is observed. So for each individual we observe the triplet $\{Z, S, Y\}$.

We make the following assumptions for the estimation.

Condition 1 A1(INDEP): *The random variables S_Z, $Z = 0, 1, ..., k$ and Y_j, $j = 0, 1, \ldots, J$ are jointly independent of Z.*

This assumption implies that among other things the value of the instrument has no direct impact on the response Y. However, this is not sufficient for the identification of the causal effect. The other assumption we need is monotonicity.

Condition 2 A2(MONOTONICITY): *With probability 1, either $S_1 - S_0 \geq 0$ (or $S_1 - S_0 \leq 0$) for each person when Z is binary.*

One implication of assumption (2) that the conditional cumulative distribution function (CDF) of S given $Z = k$, and the CDF of S given $Z = k - 1$ never cross *with probability*.1. This implies

$$F_S(j|Z = k - 1) \geq F_S(j|Z = k)$$

where F_S is the CDF of S. Without loss of generality, we consider just the case where $S_1 - S_0 \geq 0$.

In order to incorporate the above result in a multi-valued instrument we can consider K mutually orthogonal binary instruments indicated by $d_k = I\{Z = k\}$, the indicator function for $Z = k$. We can now redefine the average treatment effect as $\beta_{k,k-1}$ in place of β changing 1 to k and 0 to $k - 1$. We also order the points of support of Z to conform with the ranks of the treatment intensity. Then we have the following theorem.

Theorem 1 *(Theorem 3, Angrist and Imbens,1995). Let $g[X]$ be a design matrix constructed from the indicator variables for each value of X. If we consider the $2SLS$ estimator constructed using $g(X)$ and a full set of interactions between $g(X)$ and Z as instruments for the regression of Y on rows of $g[X]$ and S. The resulting estimate is*

$$\beta_x = \frac{E\{Y.(E[S|X,Z] - E[S|X])\}}{E\{S.(E[S|X,Z] - E[S|X])\}} = \frac{E\{\beta(X)\Theta(X)\}}{E\{\Theta(X)\}} \tag{5}$$

where $\Theta(X) = E\{E[S|X,Z](E[S|X,Z] - E[S|X])|X\}$ *and*

$$\beta(X) = \frac{E\{Y.(E[S|X,Z] - E[S|X])|X\}}{E\{S.(E[S|X,Z] - E[S|X])|X\}}.$$

As a corollary to the above theorem we propose the following result.

Proposition 2 *(LATE in case of multivalued treatment and instrument):Under assumptions* A1 *and* A2 *besides SUTVA, given that we have the triplet (Z,S,Y) for each individual and conditional on the covariates X we have the following IV estimator*

$$\beta_x = \frac{E\{Y.(E[S|X,Z] - E[S|X])\}}{E\{S.(E[S|X,Z] - E[S|X])\}} = \frac{Cov\{Y, E[S|X,Z]\}}{Cov\{S, E[S|X,Z]\}}. \tag{6}$$

That is to say that the average treatment effect β_x is given by the ratio of the covariance of Y with the conditional expectation of treatment S given X and Z, and the covariance of treatment S and the conditional expectation of S given X and Z.

Proof. The first equality follows from the Theorem 1 (Theorem 3 in Angrist and Imbens ,1995). Then we can re-write as

$$\begin{aligned}
\beta_x &= \frac{E\{Y.(E[S|X,Z] - E[S|X])\}}{E\{S.(E[S|X,Z] - E[S|X])\}} \\
&= \frac{E\{(Y - E(Y)).(E[S|X,Z] - E[S|X])\}}{E\{(S - E(S)).(E[S|X,Z] - E[S|X])\}} \\
&= \frac{Cov\{Y, E[S|X,Z]\}}{Cov\{S, E[S|X,Z]\}}
\end{aligned}$$

as both $E(Y)$ and $E(S)$ are finite constants. This is the proof of the proposition. ■

It might be worth mentioning that the second equality in the Theorem 1 follows from dividing both the numerator and the denominator by the conditional variance of S given X and Z. So the average treatment effect is a variance weighted treatment effect .

However our main interest is not in the average treatment effect of the entire population, but in a subpopulation, namely, the compliers. These are the subjects whose treatment status could be changed by the assignment mechanism. Investigators such as policymakers are often interested in the distribution of the treatment effects beyond a "location shift," particularly the non-constant effect of a policy on the tails of a distribution like income. However, it is sufficient to identify the marginal distributions of $Y_i(1)$ and $Y_i(0)$ for the compliers in case of binary treatments. Under SUTVA, it is irrelevant for an individual to know the population distribution of all types of individuals, all that counts are the individuals with similar characteristics. So the distribution of the treatment effect $Y_i(1) - Y_i(0)$ can be obtained if one knows the marginal distributions of $Y_i(1)$ and $Y_i(0)$ only. It is however, non-trivial to get the marginal distribution of the compliers mainly because, compliers are not identifiable from the population units. We can however identify some *non-compliers* namely, the *always-takers*, who volunteers for treatment even if she is not chosen by the treatment assignment mechanism; and the *never-takers*, who never receives the treatment even if the treatment assignment assigns a treatment. In the binary case, for example, an always taker is identified if given $Z = 0$ we have $S = 1$. Similarly , a never taker is identified if given $Z = 1$ we have $S = 0$ for an individual. So anyone, who is neither an always taker nor a never taker must be a complier given monotonicity.

Now let us assume that ϕ_n, ϕ_a and ϕ_c be the population proportions of never-takers, always takers and compliers. Under monotonicity if we have $\phi_a = P(S_{obs,i} = 1|Z_{obs,i} = 0)$ and ϕ_n=$P(S_{obs,i} = 0|Z_{obs,i} = 1)$, then

$$\phi_c = 1 - \phi_n - \phi_c \tag{7}$$

that identifies the proportion of compliers.

Let the relevant probability density functions (PDF) of $Y_i(1)$ and $Y_i(0)$ be $g_{c1}(y)$ and $g_{c0}(y)$, respectively. If $f_{zs}(y)$ denotes the directly estimable PDF when $Z_{obs,i} = z$ and $S_{obs,i} = s$, then $g_n(y) = f_{10}(y)$ and $g_a(y) = f_{01}(y)$, hence applying Bayes' rule

$$f_{00}(y) = \frac{\phi_c}{\phi_c + \phi_n} g_{c0}(y) + \frac{\phi_n}{\phi_c + \phi_n} g_n(y); \tag{8}$$

$$f_{11}(y) = \frac{\phi_c}{\phi_c + \phi_a} g_{c1}(y) + \frac{\phi_a}{\phi_c + \phi_a} g_a(y). \tag{9}$$

From equations (8) and (9) we can back out the PDFs $g_{c0}(.)$ and $g_{c1}(.)$ as we know $g_n(y)$ and $g_a(y)$. Therefore, we can identify the distributions of the treatment effect as we know the two marginal distributions for the compliers. However one major problem in backing out the marginal densities of the response variables for the treated and the untreated is that the resulting densities might turn out to be negative in some cases from the data, particularly if the proportion of compliers is small in the population. If the non-negativity constraint is imposed then we might get totally different inference.

2.2 Assumptions for estimation

Now let us assume $S_Z = 0, 1 \dots J$ and $Z = 0, 1, \dots K$.

Condition 3 *The following assumptions should hold almost surely over the support of treatments*

1. *(INDEPENDENCE)* (Y_{ZS}, S_Z) *jointly indep of* Z *given* X.
2. *(EXCLUSION)* $P(Y_{ZS} = Y_S|X, Z = z) = 1$ *almost everywhere.*
3. *(NON-TRIVIAL ASSIGNMENT)* $Var(Z|X) \neq 0$
4. *(FIRST STAGE)* $Cov(S, Z) \neq 0$
5. *(MONOTONICITY)* $S_j - S_{j-1} \geq 0$

The above assumptions are based on multi-valued but discrete treatment intensities and instruments. Note that, assumptions 1 and 2 are standard for IV estimation. Assumption 3 guarantees conditional distribution of Z is non-degenerate. For binary treatment, $P(Z = 1|X) = E(Z|X) \in (0, 1)$ is an equivalent assumption. Similarly, for binary treatment and assignment assumptions 4 and 5 are equivalent to $E(S_1|X) \neq E(S_0|X)$ and $P(S_1 \geq S_0|X) = 1$, respectively (Abadie et al., 2002). We will now extend these assumptions to a continuous treatment case by writing the monotonicity assumption in terms of derivatives or expected derivatives.

As we discussed earlier if we use the derivative of Y and D with respect to Z, a plausible alternative to LATE (Heckman,1997) gives

$$\frac{E\left(\frac{\partial Y}{\partial Z}|X\right)}{E\left(\frac{\partial D}{\partial Z}|X\right)} = E\left[\frac{\partial Y}{\partial D}, E\left\{\frac{\partial D}{\partial Z}|X\right\}|X\right] > 0. \tag{10}$$

In order to match pairs we could also possibly take a *propensity score* type measure like Robins, to compare observation with similar covariates up to the propensity score. In this case if the subject is aware of the individual specific effect the instrumental variable estimator will be inconsistent. If, however, the mean effect of the *treatment on the treated* is an unknown constant, instrumental variable will still give us a consistent estimator. Unfortunately, as counterfactual states are not observed and compliers cannot be identified, it would not be possible to find out the effect of the treatment on an individual who is a complier. If we have independence of treatment and potential outcomes or at least ignorable treatment assignment, any difference in status could be assigned to treatment alone. Here we do have independence of instrument and potential outcomes but treatment itself is not ignorable. However following Imbens and Angrist (1994), IV estimation can identify the compliers whose treatment received could change with instruments.

Lemma 1 *Given* Condition 3 *and conditional on X, the conditional expectation of the treatment status S on Z and X, is ignorable for compliers, symbolically,*

$$Y_S \; II \; S|X, \; S_k \geq j > S_{k-1} \quad \textit{for all values of } j, k. \tag{11}$$

In other words, for the compliers the treatment assignment is as good as randomly assigned.

Proof. From assumptions 1 and 2 in Condition 3 we have $(Y_S, S_Z) \; II \; Z$ given X, so we have $Y_S \; II \; Z$ given X and $S_k \geq j > S_{k-1}$ for all j, k. So, we have $Y_S \; II \; S$ as S is a monotonic function of Z given X. This is because we can write S as a linear combination of orthogonal components $I\{Z = k\}$. ■

However as mentioned before compliers as individuals are not identified in the population, so we need a mechanism to isolate the compliers in order to compute any moment restricted to the subpopulation of compliers. Although the compliers are not identified we can identify some non-compliers given monotonicity as given in equation (7).

The probability of finding a complier is the expected value of the "indicator function" for compliers in the population,

$$\begin{aligned}
E\left[I\left\{Complier|X\right\}\right] &= P\left\{Complier|X\right\} \\
&= 1 - P\left\{always\ ta\ker|X\right\} - P\left\{never\ ta\ker|X\right\} \\
&= 1 - P\left\{S=1|Z=0,X\right\} - P\left\{S=0|Z=1,X\right\} \\
&= 1 - \frac{P\left\{S=1,Z=0|X\right\}}{P\left\{Z=0|X\right\}} - \frac{P\left\{S=0,Z=1|X\right\}}{P\left\{Z=1|X\right\}} \\
&= 1 - E\left[\frac{I\left\{S=1,Z=0\right\}}{P\left\{Z=0|X\right\}}|X\right] - E\left[\frac{I\left\{S=0,Z=1\right\}}{P\left\{Z=1|X\right\}}|X\right] \\
&= E\left[1 - \frac{I\left\{S=1,Z=0\right\}}{P\left\{Z=0|X\right\}} - \frac{I\left\{S=0,Z=1\right\}}{P\left\{Z=1|X\right\}}|X\right] \\
&\quad \left(= E\left[1 - I\left\{S=1|Z=0\right\} - I\left\{S=0|Z=1\right\}|X\right]\right) \qquad (12)
\end{aligned}$$

This shows two things. First, in a conditional expected sense, the expression inside equation (12)

$$\begin{aligned}
& 1 - \frac{I\left\{S=0,Z=1\right\}}{P\left\{Z=1|X\right\}} - \frac{I\left\{S=1,Z=0\right\}}{P\left\{Z=0|X\right\}} \\
= \ & 1 - \frac{I\left\{S=0\right\}I\left\{Z=1\right\}}{P\left\{Z=1|X\right\}} - \frac{I\left\{S=1\right\}I\left\{Z=0\right\}}{1-P\left\{Z=1|X\right\}} \\
\equiv \ & \kappa\left(Z,S,X\right). \qquad (13)
\end{aligned}$$

is a consistent unbiased estimator of the probability that individual i is a complier. Second, this gives an estimator of the projection or conditional expectation of S on to the space of compliers (given individual i is a complier). In the above derivation we have dropped the subscripts "obs,i" of Z, S and X for ease of exposition. This is the same $\kappa\left(Z,S,X\right)$ which appears in Abadie et. al. (2002) in case of a binary treatment and binary instrument.

It is very straightforward to check that if individual i is a complier $\kappa = 1$. However, if individual i is a non-complier $\kappa < 0$ as $P[Z=1|X] \in (0,1)$. Given that Z is not a degenerate random variable, we are actually shrinking the value of the original function when we restrict it to a subset of the whole population, namely the compliers. So, the individuals who are non-compliers have a negative weight on the estimand.

In case we have multivalued treatment $S = 0, 1 \ldots, J$ and a multivalued treatment assignment $Z = 0, 1, ..., K$, we want to find out a consistent estimator the probability $P(Complier|X)$. We can also find out the expectation of any real measurable function $\psi(Y, D, X)$ in place of $I\{Complier\}$

$$\begin{aligned}
E\left[I\left\{Complier\right\}|X\right] &= P\left\{Complier|X\right\} = P\left[\bigcup_k \bigcup_j \left\{S_k \geq j > S_{k-1}|X\right\}\right] \\
&= \sum_{k=1}^{K}\sum_{j=0}^{J} P\left\{S_k \geq j > S_{k-1}|X\right\} \\
&= \sum_{k=1}^{K}\sum_{j=0}^{J} \left[P\left\{S_k \geq j|X\right\} - P\left\{S_{k-1} \geq j|X\right\}\right] \\
&= \sum_{k=1}^{K}\sum_{j=0}^{J} \left[1 - P\left\{S < j|Z = k\right\} - P\left\{S \geq j|Z = k-1\right\}|X\right] \\
&= \sum_{k=1}^{K}\sum_{j=0}^{J} \left[1 - \frac{P\left\{S < j, Z = k\right\}}{P\left[Z = k|X\right]} - \frac{P\left\{S \geq j, Z = k-1\right\}}{P\left(Z = k-1|X\right)}|X\right] \\
&= E\left[\sum_{k=1}^{K}\sum_{j=0}^{J} \left[1 - \frac{I\left\{S < j, Z = k\right\}}{P\left[Z = k|X\right]} - \frac{I\left\{S \geq j, Z = k-1\right\}}{P\left(Z = k-1|X\right)}|X\right]\right] \\
&= E\left[\sum_{k=1}^{K}\sum_{j=0}^{J} \left[1 - \frac{I\left\{S < j\right\}I\left\{Z = k\right\}}{P\left[Z = k|X\right]} - \frac{I\left\{S \geq j\right\}I\left\{Z = k-1\right\}}{P\left(Z = k-1|X\right)}|X\right]\right] \\
&= E\left[\sum_{k=1}^{K}\sum_{j=0}^{J} \left[\kappa^{jk}|X\right]\right], \qquad (14)
\end{aligned}$$

where $\kappa^{jk} = 1 - \frac{I\{S<j\}I\{Z=k\}}{P[Z=k|X]} - \frac{I\{S\geq j\}I\{Z=k-1\}}{P(Z=k-1|X)}$, and $\sum_k \sum_j \kappa^{jk}$ is an unbiased estimator of $P\left\{Complier|X\right\}$. Given the conditional cumulative distribution function $F_S(.|Z)$ we

can also simplify

$$\begin{aligned} P\{Complier|X\} &= \sum_{k=1}^{K}\sum_{j=0}^{J}\left(P\{S\geq j|Z=k\}-P\{S\geq j|Z=k-1\}|X\right) \\ &= \sum_{k=1}^{K}\sum_{j=0}^{J}\left(P\{S<j|Z=k-1\}-P\{S<j|Z=k\}|X\right) \\ &= \sum_{k=1}^{K}\sum_{j=0}^{J}\left(F_S\left(j|Z=k-1\right)-F_S\left(j|Z=k\right)|X\right). \end{aligned} \tag{15}$$

This implies that the difference of two conditional CDFs for the data, i.e., the difference of the two empirical distribution functions (EDF) is unbiased for the probability of compliers given the covariates X. Each of the EDF's is identified by an argument similar to Angrist et. al (1996). So the value of κ could be negative.

Lemma 2 *(Extension to Abadie, 1997) Let $\psi(Y,S,X)$ be any measurable real function of (Y,S,X), the given Assumption 3 we have*

$$\begin{aligned} &E\left[\psi\left(Y,S,X\right)|S_k\geq j>S_{k-1} \text{ for all } j=0,1,...,J;k=1,2,...,K\right] \\ =\ &\frac{E\left[\sum_{k=1}^{K}\sum_{j=0}^{J}\left[\kappa^{jk}.\psi(Y,S,X)|X\right]\right]}{P\left[S_k\geq j>S_{k-1} \text{ for all } j=0,1,...,J;k=1,2,...,K\right]} \end{aligned} \tag{16}$$

where $\kappa^{jk}=1-\frac{I\{S<j\}I\{Z=k\}}{P[Z=k|X]}-\frac{I\{S\geq j\}I\{Z=k-1\}}{P(Z=k-1|X)}$, where $I\{A\}$ is the indicator function of event A, $S_Z\in\{0,1,2,...,J\}$, $Z\in\{1,2,...,K\}$ and Y_{ZS} can take any real value.

Proof. The proof strongly relies on monotonicity and the exclusion restriction and also ignorability from Lemma 1. We have to show that

$$\begin{aligned} &E\left[\psi\left(Y,S,X\right)|S_k\geq j>S_{k-1} \text{ for all } j,k\right] \\ =\ &\frac{E\left[\sum_{k=1}^{K}\sum_{j=0}^{J}\left[\kappa^{jk}.\psi(Y,S,X)\right]\right]}{P\left[S_k\geq j>S_{k-1} \text{ for all } j,k\right]}. \end{aligned}$$

Following the derivation of the probability of the complier in equation (14), we can find the expectation of the function $\psi(Y,S,X)$

$$E\left[\psi(Y,S,X)\,|S_k \geq j > S_{k-1} \text{ for all } j,k|X\right] = \frac{E\left[\sum_{k=1}^{K}\sum_{j=0}^{J}\left[\kappa^{jk}.\psi(Y,S,X)\right]|X\right]}{P\left[S_k \geq j > S_{k-1} \text{ for all } j,k|X\right]}$$

$$\Rightarrow E\left[\psi(Y,S,X)\,|S_k \geq j > S_{k-1} \text{ for all } j,k|X\right] f(X) = \frac{E\left[\sum_{k=1}^{K}\sum_{j=0}^{J}\left[\kappa^{jk}.\psi(Y,S,X)\right]|X\right] f(X)}{P\left[S_k \geq j > S_{k-1} \text{ for all } j,k\right]}.$$

Now integrating both sides with respect to X we get lemma. ■

The main implication of the lemma is that given any characteristics of the subpopulation of compliers we can identify the characteristics or moment conditions for the whole population using the appropriate weight.

3 QUANTILE TREATMENT EFFECT

Like the usual 2SLS or IV estimation, the IV estimators of the quantile functions collapse to the conditional quantile regression estimates for perfect compliance. To make the problem identifiable the treatment effects are assumed to be linear and additive in covariates. Using ignorability from Lemma 1 and the exclusion restriction we have the θ^{th} quantile treatment effect is

$$Q_\theta\left(Y|X,S,S_k \geq j > S_{k-1},\ \text{for all } j,k\right) = \alpha_\theta S + X'\beta_\theta. \tag{17}$$

As only the marginal distributions are identified for compliers α_θ is the difference in the quantile treatment effects of the θ^{th} quantiles of Y's of treatments j and $j-1$.

Hence, the solution,

$$\begin{aligned}(\alpha_\theta,\beta_\theta) &\equiv \arg\min_{(\alpha,\beta)} E\left[\rho_\theta\left(Y-\alpha S-X'\beta\right)|S_k \geq j > S_{k-1},\ \text{for all } k \text{ and } j\right] \\ &\equiv \arg\min_{(\alpha,\beta)} E\left[\sum_k\sum_j \kappa^{jk}.\rho_\theta\left(Y-\alpha S-X'\beta\right)\right].\end{aligned} \tag{18}$$

For a random sample of size n, $\{y, d, x, z\}$, $i = 1, \ldots, n$ this reduces to

$$\left(\widehat{\alpha}_\theta, \widehat{\beta}_\theta\right) \equiv \arg\min_{(\alpha,\beta)} \sum_{i=1}^{n} \sum_{k} \sum_{j} \kappa_i^{jk} . \rho_\theta \left(y_i - \alpha s_i - x_i'\beta\right),$$

that can be implemented using weighted quantile regression formula after evaluating $P\left[Z = k|X\right]$, and plugging in to get κ_i^{jk}'s (Powell, 1994).

This might be a good place to ponder on what exactly this estimate is, particularly, in the context of its precursors like the *2SLAD* estimate of Amemiya (1982). In Amemiya the 2SLAD estimate was given by minimizing equation (1) or

$$\alpha_q = \arg\min_{\alpha} \sum_{t=1}^{T} \left| q y_t + (1-q) P_t^{'} y_t - P_t' Z\alpha \right|. \tag{19}$$

Taking $q = 0$, we get pure 2SLAD estimator reduces to

$$\alpha_q = \arg\min_{\alpha} \sum_{t=1}^{T} \left| P_t^{'} y_t - P_t' Z\alpha \right| = \arg\min_{\alpha} \sum_{t=1}^{T} \left| P_t' \left(y_t - Z\alpha\right) \right| \tag{20}$$

where P_t' is the t^{th} row of projection matrix or $X(X'X)^{-1}X$, X being the matrix of exogenous variables or candidate instruments. In the case of binary treatment and assignment case discussed in there is only one instrument for the treatment received (Abadie et. al., 2002). So the problem in binary treatment effect can be recast as in equation (20) where

$$P_t' = 1 - I\left\{S = 1|Z = 0\right\} - I\left\{S = 0|Z = 1\right\}$$

such that

$$\begin{aligned} E\left[P_t'|X\right] &= E\left[1 - \frac{I\left\{S = 1, Z = 0\right\}}{P\left\{Z = 0|X\right\}} - \frac{I\left\{S = 0, Z = 1\right\}}{P\left\{Z = 1|X\right\}}|X\right] \\ &= E\left[1 - \frac{Z.(1-S)}{P\left\{Z = 1|X\right\}} - \frac{(1-Z).S}{1 - P\left\{Z = 1|X\right\}}|X\right]. \end{aligned} \tag{21}$$

However, this becomes more involved if we go to a continuous treatment or to more general IV or 2SLAD type estimation.

4 ASYMPTOTIC DISTRIBUTION THEORY RESULTS

This section follow Abadie et. al (2002) closely with appropriate extensions.

Condition 4 *(IDENTIFICATION ASSUMPTION): There exists unique* $\alpha \in \Lambda$ *and* $\beta \in \Theta$ *such that the* θ^{th} *quantile of* Y_S *conditional on* X *and* $S_k \geq j > S_{k-1}$ *for all* $j = 0, 1, ..., J$ *and* $k = 1, 2, ..., K$ *is unique and equals* $\alpha S + X'\beta$.

Proposition 3 *(IDENTIFICATION) Given Assumptions 3 and 4,*

$$\arg \min_{(a \times b) \in (\Lambda, \Theta)} E \left[\begin{array}{c} \sum_{k=1}^{K} \sum_{j=0}^{J} \left(1 - \frac{I\{S<j\} I\{Z=k\}}{P[Z=k|X]} - \frac{(1-I\{S<j\}) I\{Z=k-1\}}{P(Z=k-1|X)} \right) \\ \times (Y - aS - X'b) (\theta - I\{Y - aS - X'b < 0\}) \end{array} \right] \tag{22}$$

is unique and equals (α, β).
Proof. *See Appendix A.* ■

Condition 5 *(CONSISTENCY ASSUMPTION):*

1. *Let* $W = (Y, D, X', Z)'$ *are* iid.
2. *For estimation of the expected value of the projection of* Z *on* X, *let* $\gamma_j \in \Gamma$ *be a subset of* R^L, *and* $P(Z = k|X) = P(X; \gamma_k)$ *which must strictly lie between* 0 *and* 1, *and is continuous in* $g \in \Gamma$.
3. *There exists a consistent first stage estimator* $\hat{\gamma}_k$ *of* γ_k.
4. $E|Y|$ *and* $E||X||$ *are finite.*
5. Λ *and* Θ *are compact (just convex would have been sufficient if the function* κ *was convex, which is not the case).*
6. *The function* $I\{Y - aD - X'b < 0\}$ *is continuous at each* (a, b) *in* $\Lambda \times \Theta$ *w. p.*1.

Proposition 4 *(CONSISTENCY): Given Assumptions (3)-(5) hold, then the conditional quantile regression estimators*

$$\begin{aligned} \left(\hat{\alpha}, \hat{\beta}\right) \equiv & \arg \min_{a \in \Lambda, b \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \sum_{j} \sum_{k} \left(1 - \frac{I\{S < j\} I\{Z = k\}}{P\left[x_i; \hat{\gamma}_k\right]} - \frac{(1 - I\{S < j\}) I\{Z = k-1\}}{P\left[x_i; \hat{\gamma}_{k-1}\right]} \right) \\ & \times \rho_\theta \left(y_i - a s_i - x_i' b\right) \end{aligned}$$

is consistent for (α, β).

Proof. See Appendix B. ■

Under certain other regularity conditions like the generalized method of moment estimator for the first stage estimator■ γ, and the absolute continuity of the distribution function. Abadie et. al.(2002) shows that estimator is asymptotically normal for binary treatment effect.

5 EMPIRICAL EXAMPLE

I wish to do an example of the returns to schooling example I gave as an example, however I haven't had any significant luck with the data, so far. If I can't I would talk about an artificial example in the line of Rosenbaum (1997)'s comment to Angrist et. al. (1997) and compare the different models.

The example we talk about is more like a motivation for the need of an instrumental variable estimator for quantile regression type problems. The data is taken from the Panel Data of Income Dynamics, 1969-1984, although we only use the data on individuals from 1969. The variables are YEAR=1969, ln of Wage (WAGE), EDUC(in years of education), EXP(Experience in years), TEN(Tenure in months) and an id number for the observation. This was used by Solomon Polachek and Bong Joon Yoon in "Panel Estimates for two tier Earnings Forecast" in Journal of Applied Econometrics v11.2, 1996.

6 SUMMARY AND OBJECTIVES

We tried to illustrate the effect of non-ignorable treatment on quantiles of the potential outcomes distribution when the treatment received and the assignment mechanism are continuous.The Quantile Treatment Effect (QTE) estimator takes into account the endogeneity problem inherent in a model and gives consistent estimators.

One major concern of the exercise was convergence to the QTE estimators and if they exist at all was not guaranteed from the algorithm used. Traditional 2SLS tends to restrict the effect of covariates to be the same for all groups unlike QTE, this might reduce the sampling variation in 2SLS estimates which might lead to bias.

As mentioned before some behavioral assumptions are extremely essential for the estimator to be consistent as discussed in Heckman (1997).

Some of the assumptions, for consistency like the compact parameter space could have been removed if we consider a convex hull of the κ function, this might be a useful exercise to pursue.

7 APPENDIX (Proof of Lemma and Propositions)

7.1 Appendix A

Proof. Proof of Proposition 3 (IDENTIFICATION): From Assumption 4 we have that,

$$E\left[\left(Y-h\left(S,X\right)\right)\left(\theta-I\left\{Y-h\left(,X\right)\right\}\right)|S_k\geq j>S_{k-1},\text{ for all } j,k\right]$$

has a minimum when $h(S,X)$ is the θ^{th} quantile of the conditional distribution of Y given X and that this is uniquely equal to $\alpha S+X'\beta$.Thus we have

$$\begin{aligned}(\alpha,\beta) &\equiv \arg\min_{a\in\Lambda,b\in\Theta} E\left[\left(Y-aS-X'b\right)\times\left(\theta-I\left\{Y-aS-X'b\right\}\right)|S_k\geq j>S_{k-1}\text{ for all } j,k\right]\\ &\Rightarrow (\alpha,\beta)=\arg\min_{a\in\Lambda,b\in\Theta} E\left[\sum_j\sum_k \kappa^{jk}\left(Y-aS-X'b\right)\left(\theta-I\left\{Y-aS-X'b<0\right\}\right)\right] \quad (23)\end{aligned}$$

using lemma 1. ■

Proof. Proof of Proposition 5 (CONSISTENCY):

By proposition 3, we have shown that the equation (22) is uniquely minimized at (α,β) over a compact space (Λ,Θ). ■Consider the following function

$$f\left(W_i,l,G\right)=\sum_{k=1}^{K}\sum_{j=0}^{J}\kappa_i^{jk}\left(g_k\right)\left(\theta-I\left\{Y_i-aS_i-X_i'b<0\right\}\right)\left(Y_i-aS_i-X_i'b\right) \quad (24)$$

where $W_i = (S_i, X_i')\,, l = (a, b) \in (\Lambda, \Theta)$ and $G = (g_0, g_1, g_{2,...,}g_K) \in \Gamma^K$. Then

$$\begin{aligned} & \sup_{l\in(\Lambda,\Theta)} \left\| \frac{1}{n}\sum_{i=1}^{n} f\left(w_i, l, \hat{G}\right) - E\left[f\left(W, l, G\right)\right] \right\| \\ \leq & \sup_{l\in(\Lambda,\Theta)} \left\| \frac{1}{n}\sum_{i=1}^{n} f\left(w_i, l, \hat{G}\right) - E\left[f\left(W, l, G\right)\right] \right\| + \sup_{l\in(\Lambda,\Theta)} \left\| E\left[f\left(w_i, l, \hat{G}\right)\right] - E\left[f\left(W, l, G\right)\right] \right\| \\ = & R_1 + R_2, \text{ say.} \end{aligned} \tag{25}$$

We can further show using standard notations of teh check function $\rho_\theta(y) = (\theta - I\{y < 0\})\, y$ and $\hat{\kappa}_i^{jk} = \kappa_i^{jk}(\hat{g}_k)$

$$\begin{aligned} R_1 = & \sup_{l\in(\Lambda,\Theta)} \left\| \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{j=0}^{J} \hat{\kappa}_i^{jk}\rho_\theta\left(Y_i - aS_i - X_i'b\right) - \sum_k\sum_j E\left[\kappa_i^{jk}\rho_\theta\left(Y_i - aS_i - X_i'b\right)\right] \right\| \\ \leq & \sup_{l\in(\Lambda,\Theta)} \sum_{k=1}^{K}\sum_{j=0}^{J} \left\| \frac{1}{n}\sum_{i=1}^{n} \left(\hat{\kappa}_i^{jk}\rho_\theta\left(Y_i - aS_i - X_i'b\right) - E\left[\kappa_i^{jk}\rho_\theta\left(Y_i - aS_i - X_i'b\right)\right]\right) \right\| \end{aligned} \tag{26}$$

From Assumptions 3-5, we have that $f(.)$ is a continuous function in (W, l, G). From the Assumption 5 part 2., each κ_i^{jk} is bounded by some real $\bar{K}$ while $|\theta - I\{Y_i - aS_i - X_i'b < 0\}|$ is bounded by 1. As we are maximizing over compact space $\Lambda \times \Theta$, we have a finite value for all $\|l\| \leq \bar{l}$ for all $l \leq \Lambda \times \Theta$. Hence, $||f(.)|| \leq \bar{K}\left(|Y| + \bar{l}\left(\|X\| + 1\right)\right)$, the bound is finite from assumptions. Using Lemma 2.4 from McFadden and Newey (1994) we have that $E[f(W, l, G)]$ is continuous at each $(l, G) \in \left(\Lambda, \Theta, \Gamma^K\right)$. Then,

$$R_1 \leq \sup_{l\in(\Lambda,\Theta)} \sum_{k=1}^{K}\sum_{j=0}^{J} \left\| \frac{1}{n}\sum_{i=1}^{n} \left(\hat{\kappa}_i^{jk}\rho_\theta\left(Y_i - aS_i - X_i'b\right) - E\left[\kappa_i^{jk}\rho_\theta\left(Y_i - aS_i - X_i'b\right)\right]\right) \right\| \xrightarrow{p} 0, \tag{27}$$

as each of its components goes to zero.

Similarly, as $E[f(W, l, G)]$ is a continuous function of (l, G)

$$E\left[f\left(W, l, \hat{\gamma}\right)\right] \xrightarrow{p} E\left[f\left(W, l, \gamma\right)\right]$$

uniformly in $l \in (\Lambda \times \Theta)$. Hence, from Theorem 2.1 in Newey and Mcfadden's (1994),

$$R_2 = \sup_{l \in (\Lambda, \Theta)} \left\| E\left[f\left(w_i, l, \hat{G} \right) \right] - E\left[f\left(W, l, G \right) \right] \right\| \xrightarrow{p} 0. \tag{28}$$

From equations (27) and (28), applying to equation (25) we have a sufficient condition for the convergence of $\left(\hat{\alpha}, \hat{\beta} \right)$ to (α, β), that shows the consistency of the estimators. ■

7.2 Appendix B

Let us consider the problem of quantile regression in this context as given by Koenker and Bassett (1978), we can write the problem as the following

$$\min_{\boldsymbol{\beta} \in B} \sum_{i=1}^{n} \rho_\theta \left(y_i - x_i' \boldsymbol{\beta} \right)$$

where $\rho_\theta (r) = r \left(1 - I \left\{ r < 0 \right\} \right)$ can be explained as a linear programming problem in standard form $\min c'x$ s.t.$Ax = y, x \geq 0$, in this case

$$\begin{aligned} & \min_{\boldsymbol{\beta} \in B} \left\{ \sum_i \theta u_i^+ + (1 - \theta) u_i^- | y_i = x_i' \boldsymbol{\beta} + u_i^+ - u_i^- \text{ for all } i = 1, ..., n; \left(u_i^+, u_i^- \right) \in R^{2n+} \right\} \\ \equiv & \min_{\boldsymbol{\beta} \in B} \left\{ \sum_i \theta e' u^+ + (1 - \theta) e' u^- | y = \mathbf{X}' \boldsymbol{\beta} + u^+ - u^-, \left(u^+, u^- \right) \in R^{2n+} \right\}. \end{aligned}$$

where $e' = (1, 1, .., 1, ..., 1)$ We can write this in the standard LP form where

$$c' = \left(0, \theta e', (1 - \theta) e' \right), x = \left(\boldsymbol{\beta}', u^{+\prime}, u^{-\prime} \right), A = (\mathbf{X}, \mathbf{I}, -\mathbf{I}), b = y.$$

Further, in this special case, we have $\boldsymbol{\beta}' = \left(\alpha, \beta' \right)'$ and $\mathbf{X} = (d, X)$, hence the problem becomes a linear program with

$$c' = \left(0, 0, 0, 0, \theta e', (1 - \theta) e' \right), x = \left(\alpha^+, \beta^{+\prime}, \alpha^-, \beta^{-\prime}, u^{+\prime}, u^{-\prime} \right), A = (d, X, -d, -X, I, -I), b = y.$$

This problem could easily be solved using a Barrodale-Roberts (BR) type algorithm

as was suggested by Koenker and D'Orey (1987). However, here the problem is a little different as we have to replace the $e' = (1, 1, \ldots 1)'$ by $\kappa = (\kappa_1, \kappa_2, ..., \kappa_n)$ where some of the κ's might be negative. To keep the solution of LP bounded, and to keep the definitions of u^+ and u^- consistent with the intended definition, they include a new constraint namely, $u_i^+ . u_i^- = 0$ which is trivially true in the LP.

We propose a Primal-Dual Interior point algorithm for the estimator developed by Karmarkar among others (Arbel, 1993), the basic idea is to use the following primal and dual problems

$$\min_x c'x - \mu \left(\sum_{i=1}^{n} \ln (x_i) \right) \text{ s.t.} Ax = b \text{ and } \mu > 0. \tag{29}$$

$$\min_x b'y + \mu \left(\sum_{i=1}^{n} \ln (z_i) \right) \text{ s.t. } A'y + z = c, \ \mu > 0. \tag{30}$$

where we optimize both together maintaining the feasibility requirement for both the Primal and Dual problems we stop by a minimum relative duality gap criterion.

However, we need to include an extra constraint to take into account the problem of unbounded solutions by including the constraint $x_2' x_3 = 0$ where $x' = (x_1', x_2', x_3')$. One way of doing that would be to include that as an extra constraint in the primal problem.

References

ABADIE, A., J.D. ANGRIST AND G. IMBENS, (2002),"Instrumental Variable Estimates of the effect of subsidized training on the quantiles of Trainee earnings," *Econometrica* 70, 91-117.

AHN, H. AND J.L. POWELL,(1993),"Semiparametric Estimation of Censored Selection Models with Non-parametric Selection Mechanism," Journal of Econometrics, 58, 3-27.+

ANDREWS,D.W.K (1994),"Empirical Process Methods in Econometrics," In Chapter 37, *Handbook of Econometrics*, IV, Elsevier Science Publisher: Amsterdam.+

AMEMIYA,T.(1982),"Two-Sage Least Absolute Deviations Estimators," *Econometrica*, 50, 689-711.

ANGRIST, J.D, K. GRADDY AND G. IMBENS,(1995), "Nonparametric Demand Analysis With an Application to the Demand of Fresh Fish," *Technical Working Paper 178*, NBER, MIT.

ANGRIST, J.D. AND W.N. EVANS (1997) ,"Children and Their Parents' Labor Supply:Evidence from Exogenous Variation in Family Size," *American Economic Review* (forthcoming)+

ANGRIST, J.D. AND G.IMBENS, (1994),"Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

ANGRIST, J.D. AND G.IMBENS, (1995),"Two-Stage Least Squares Estimation of Average Causal Effects with Variable Intensity," *Journal of the American Statistical Association*, 90, 431-442.+

ANGRIST,J.D.,G.W. IMBENS AND D.B. RUBIN, (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*,91,444-472.

ARBEL, A., (1993), *Exploring Interior Point Linear Programming*, MIT Press: Cambridge, MA.

BALKE, A. AND J.PEARL, 1997,"Bounds on Treatment Effects From Studies with Imperfect Compliance," J*ournal of the American Statistical Association*, 92,1171-1176.

BASSETT, G. AND R. KOENKER, (1982), "An Empirical Quantile Function for Linear Models with iid Errors," *Journal of the American Statistical Association*, 77, 407-415.+

HODGES, J. AND E. LEHMANN, (1963),"Estimates of Location Based on Rank Tests," *Annals of Mathematical Statistics*,34, 598-611.

HECKMAN, J.J., (1996), *Comments* to AIR (1996), *Journal of the American Statistical Association*, 91, 459-462.

HECKMAN, J.J., (1997),"Instrumental Variable: A Study of Implicit Behaioral Assumptions Used in Making Program Evaluations," *The Journal of Human Resources*,32 No. 3, 441-462.

HUBER, P.J., (1981) , *Robust Statistics*, Wiley: New York.+

IMBENS, G., (2004),"Non-parametric Estimation of Average Treatment Effects under Exogeneity: A Review," *The Review of Economics and Statistics*, 2004, 86(1), 4–29.

KOENKER, R.(2005), *Quantile Regression*, Cambridge University Press: Cambridge, MA.+

KOENKER, R. AND G. BASSETT (1978),"Regression Quantiles," *Econometrica* 46, 33-50.

KOENKER , R. AND V. D'OREY, (1987),"Computing Regression Quantile," Journal of the Royal Statistical Society, *Applied Statistics*, 36, 383-393.

KRASKER, W.S. AND R.E. WELSCH,(1985),"Resistant Estimation for Simultaneous Equation Models using Weighted Instrumental Variables," *Econometrica*, 53,1475-1488.+

KRISHNAKUMAR, J. AND E. RONCHETTI, (1997),"Robust Estimators for Simultaneous Equations Models," *Journal of Econometrics*, 78, 295-314.+

LIPSITZ,S.R.,G. FITZMAURICE,G. MOLENBERGHS AND L.P.ZHAO, (1997), "Quantile Regression Methods for Longitudinal Data with Drop-outs: Application to CD4 Cell Counts of Patients Infected with the Human Immunodeficiency Virus," *Journal of the Royal Statistical Society, Applied Statistics*, 46, 463-476.+

MANSKI, C.F., (1990),"Nonparametric Bounds on Treatment Effects," *American Economic Review*, Papers and Proceedings, 80, 319-323.

MOFFITT, R. A., (1996), Comments to AIR(1996), Journal of the American Statistical Association, 91, 462-465.

NEYMAN, J., (1923). "On the Application of Probability Theory to Agricultural Experiments: Essay on Principle. Section 9," *Roczniki Nauk Rolniczych Tom* X, 1-51 (*Annals of Agricultural Sciences*). Translated and Edited by D.M. Dabrowska and T.P. Speed for *Statistical Science*, 1990, 465-480.

NEWEY ,W.K. AND D. MCFADDEN, (1994),"Large Sample Estimation and Hypothesis Testing," Chapter 36, *Handbook of Econometrics, IV*. Elsevier Science Publisher: Amsterdam.

NEWEY, W. K. AND J.L. POWELL, (1993),"Efficiency Bounds for some Semiparametric Selection Models," *Journal of Econometrics*, 58, 169-184.+

POWELL, J.L., (1983),"The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econom*terica, 51,1569-1575.

POWELL, J.L., (1994),"Estimation of Semiparametric Models," In Chapter 41, *Handbook of Econometrics,IV*. Elsevier Science Publisher: Amsterdam.

POWELL, J.L., J.H. STOCK AND T. M. STOKER, (1989),"Semiparametric Estimation of Index Coefficients," *Econometrica,* 57, 1403-1430.+

ROBINS, J.M. AND S. GREENLAND, (1996), "Comments to Angrist et. al. (1996)", *Journal of the American Statistical Association*, 91, 456-458.+

ROSENBAUM, P. R., (2002), *Observational Studies*, Second edition, Springer-Verlag: New York.

ROSENBAUM, P.R. AND D.B. RUBIN, (1983),"The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, 41-55.

RUBIN, D.B., (1974),"Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688-701.

RUPPERT, D. AND R. CARROLL, (1980),"Trimmed Least Squares estimation in a Linear Model," *Journal of the American Statistical Association*, 75, 828-838.+

VENABLES, W.N. AND B.D. RIPLEY, (1997), *Modern Applied Statistics with S-Plus*, 2nd Edition, Springer-Verlag: Berlin.+