- Y is truncated when we only observe X for observations where Y would not be censored. We do not have a full sample for {Y,X}, we exclude observations based on characteristics of Y.

3. Estimation Procedure and Asymptotic Properties

3.1 Estimation Procedure

In this section we consider estimation of the function $\mu(\cdot)$. Our procedure will be based on our identification results in the previous section, and involves nonparametric quantile regression at different quantiles and different points in the support of the regressors. Our asymptotic arguments are based on the local polynomial estimator for conditional quantile functions introduced in Chaudhuri(1991a,b). For expositional ease, we only describe this nonparametric estimator for a polynomial of degree 0, and refer psycotesters to Chaudhuri(1991a,b), Chaudhuri et al.(1997), Chen and Khan(2000,2001), and Khan(2001) for the additional notation involved for polynomials of arbitrary degree.

First, we assume the regressor vector xi can be partitioned as (xi^{ds},x^{c}) where the d_{ds}−dimensional vector xi^{ds} is discretely distributed, and the d_c-dimensional vector xi^{c} is continuously distributed.

We let $Cn(x_i)$ denote the cell of observation x_i and let h_n denote the sequence of bandwidths which govern the size of the cell. For some observation x_j , $j \neq i$, we let $x_j \in C_n(x_i)$ denote that $x_j^{(ds)}=x_i^{(ds)}$ and $x_j^{©}$ lies in the d_c-dimensional cube centered at x_i^{c} with side length $2h_n$.

Let $I[\cdot]$ be an indicator function, taking the value 1 if its argument is true, and 0 otherwise. Our estimator of the conditional α^{th} quantile function at a point x_i for any $\alpha \in (0, 1)$ involves α-quantile regression (see Koenker and Bassett (1978)) on observations which lie in the defined cells of x_i. Specifically, let θ minimize:

$$\sum_{j=1}^{n} I[x_j \in C_n(x_i)]\rho_\alpha(y_j-\theta)$$

Where

$$\rho_\alpha(.) \equiv \alpha|.|+(2\alpha-1)(.)I[.<0]$$

Our estimation procedure will be based on a random sample of n observations of the vector $(y_i,x_i')'$ and involves applying the local polynomial estimator at three stages. Throughout our description, ˆ· will denote estimated values.

1) **Local Constant Estimation of the Conditional Median Function.** In the first stage, we estimate the conditional median at each point in the sample, using a polynomial of degree 0. We will let h_{1n} denote the bandwidth sequence used in this stage. Following the terminology of Fan(1992), we refer to this as a local constant estimator, and denote the estimated values by ˆq0.5(xi). Recalling that our identification result is based on observations for which the median function is positive, we assigns weights to these estimated values using a weighting function, denoted by $w(\cdot)$. Essentially, $w(\cdot)$ assigns 0 weight to observations in the sample for which the estimated value of the median function is 0, and assigns positive weight for estimated values which are positive.

2) **Weighted Average Estimation of the Disturbance Quantiles** In the second stage, the unknown quantiles $c_{\alpha1}$, $c_{\alpha2}$ are estimated (up to the scalar constant _c) by a weighted average of local polynomial estimators of the quantile functions for the higher quantiles α1, α2. In this stage, we use a polynomial of degree k, and denote the second stage bandwidth sequence by h_{2n}.

We let ˆc1, ˆc2 denote the estimators of the unknown constants $\frac{c_{\alpha1}}{\Delta c}, \frac{c_{\alpha2}}{\Delta c}$ and define them as:

$$\hat{c_1}=\frac{\frac{1}{n}\sum_{i=1}^{n}\tau(x_i)w(\widehat{q_{0.5}}(x_i))\cdot\frac{\left(\hat{q}_{\alpha1}(x_i)-\hat{q}_{0.5}^{(p)}(x_i)\right)}{\left(\hat{q}_{\alpha2}(x_i)-\hat{q}_{\alpha1}(x_i)\right)}}{\frac{1}{n}\sum_{i=1}^{n}\tau(x_i)w(\widehat{q_{0.5}}(x_i))}$$

$$\hat{c_2}=\frac{\frac{1}{n}\sum_{i=1}^{n}\tau(x_i)w(\widehat{q_{0.5}}(x_i))\cdot\frac{\left(\hat{q}_{\alpha2}(x_i)-\hat{q}_{0.5}^{(p)}(x_i)\right)}{\left(\hat{q}_{\alpha2}(x_i)-\hat{q}_{\alpha1}(x_i)\right)}}{\frac{1}{n}\sum_{i=1}^{n}\tau(x_i)w(\widehat{q_{0.5}}(x_i))}$$

where $\tau(x_i)$ is a trimming function, whose support, denoted by X_τ , is a compact set which lies strictly in the interior of X. The trimming function serves to eliminate“ boundary effects” that arise in nonparametric estimation. We use the superscript (p) to distinguish the estimator of the median function in this stage from that in the first stage.

3) **Local Polynomial Estimation at the Point of Interest** Letting x denote the point at which the function $\mu(\cdot)$ is to be estimated at, we combine the local polynomial estimator, with polynomial order k and bandwidth sequence h_{3n}, of the conditional quantile function at x using quantiles α1, α2, with the estimator of the unknown disturbance quantiles, to yield the estimator of μ(x):

$$\hat{\mu}(x)=\hat{c}_2\hat{q}_{\alpha1}(x)-\hat{c}_1\hat{q}_{\alpha2}(x)$$

4. Estimating the Distribution of ϵi

As mentioned in Section 2, the distribution of the random variable ϵ_i is identified for all quantiles exceeding $\alpha_0 \equiv \inf\{\alpha\colon \sup_{x\in X} q_\alpha(x) > 0\}$. In this section we consider estimation of these quantiles, and the asymptotic properties of the estimator. Estimating the distribution of ϵ_i is of interest for two reasons. First, the econometrician may be interested in estimating the entire model, which would require estimators of σ(xi) and the distribution of ϵ_i as well as of μ(xi). Second, the estimator can be used to construct tests of various parametric forms of the distribution of ϵ_i, and the results of these tests could then be used to adopt a (local) likelihood approach to estimating the function μ(xi).