æ

Using Modes to Identify and Estimate Truncated Regression Models.

Bo E. Honoré*

Northwestern University

October, 1988.

Revised: January, 1989.

Abstract.

In this paper we investigate the identifiability of the truncated regression model. We show that in a model with arbitrary heteroskedasticity neither mean independence nor median independence is sufficient identification, but that mode independence is. The discussion of identification suggests estimators for the model. We prove consistency of the estimators under general heteroskedasticity. A small scale Monte Carlo study is performed to investigate the small sample properties of the estimators.

*Mailing Address: Bo E Honoré, Department of Economics, Northwestern University, 2003 Sheridan Road, Evanston, Illinois 60208.

1. Introduction.

The truncated regression model is a regression model where the sample has been restricted to the observations for which the dependent variable is positive. Besides being of interest in its own right, a study of the truncated regression model is interesting because it may give new insights into related models.

The model most closely related to the truncated regression model is the censored regression model. In the censored regression model, the value of the dependent variable is not observed if it is negative, but it is observed that a negative value of the dependent variable is associated with that particular value of the regressors. The censored regression model can therefore be thought of as being a combination of the discrete choice model and the truncated regression model. The discrete choice model gives the probability that the dependent variable is positive, and the truncated regression model gives the distribution of the dependent variable conditional on it being positive. A study of the truncated regression model will therefore also cast light on the censored regression model, as do studies of the discrete choice model.

The topic of Section 2 of this paper is the identifiability of the truncated regression model. The main conclusion of Section 2 is that the natural identifying condition for the truncated regression model is that the mode of the error term is zero conditional on the regressors. Mean independence, median independence, or even symmetry do not suffice. In Section 3, we show how the identification results in Section 2 suggest estimators for the truncated regression model, and we prove that the estimators are consistent under very general conditions.

Semiparametric estimators of truncated regression models have previously been suggested by Newey (1986), Lee (1988a,b) and Powell (1986). The estimator presented in Section 3 of this paper is consistent under more general circumstances. Newey (1986) assumes that the errors are independent of the regressors, while Powell (1986) assumes that the distribution of the error term conditional on the regressors is symmetric. Lee (1988) assumes that the error distribution is either symmetric conditional on the regressors or independent of the regressors. One of the estimators suggested in section 3 is very much related to the estimator suggested by Lee. Section 4 presents the results of a small scale Monte Carlo study of the performance of the estimators presented in this paper compared to the performance of some alternative estimators. Section 5 concludes the paper.

2. Identification.

The truncated regression model is defined by observations of[1] (Y, X) generated by

$$Y = X\beta + U$$

conditional on $Y > 0$. In other words, it is a linear regression model with a sampling plan that samples (Y, X) from the conditional distribution of (Y, X) given $Y > 0$. The parameter β will be considered the parameter of interest, and we denote alternative elements of the parameter space by b.

The topic of this section is to investigate natural identifying conditions to be put on the conditional distribution of U given X. We will denote the distribution of X and the conditional distribution of U given X by F_X and $F_{U|X}$ respectively. The model is then fully characterized by β, F_X, and $F_{U|X}$. $\mathcal{X} \subset R^P$ will denote the support for X. Throughout this paper, we will assume that the distribution of U given X is absolutely continuous.

The motivation for our study of identification is the same as the motivation in Manski (1988). A study of identification "can expose the foundations" of a model. Most frequently, identification is dealt with as part of a consistency proof. This means that the conditions for identification are often not separated from the conditions needed for consistency of a particular estimator. Another reason for studying identification is that such a study often suggests estimators in situations where it is hard to think of "natural estimators". The truncated regression model with heteroskedastic, nonsymmetric errors is an example of this. The identification analysis in this section immediately suggests estimators, the consistency of which are proved in the next section.

It is clear that data from a truncated regression model can be constructed from data from a censored regression model, and that data from a censored regression model can in turn be generated from data from a linear regression model. This implies that conditions sufficient for identification of the truncated regression model will also be sufficient for identification of the censored and linear regression models, and that conditions necessary for identification of the censored and linear regression models will also be necessary for identification of the truncated regression model.

[1] Unless we explicitly state otherwise, we will assume that one of the elements of X is a constant.

It is therefore natural to commence by investigating whether some of the identification results for censored and linear regression models also deliver identification for the truncated regression model.

The most frequently used assumptions for identification of linear and censored regression models are mean independence, $\mathrm{E}[U|X] = 0$, and median independence,[2] $\mathrm{P}(U < 0) = \frac{1}{2}$. The following theorem shows that neither mean independence, nor median independence, nor both, give identification of the truncated regression model.

THEOREM 1. *For any specification* $(\beta, F_X, F_{U|X})$ *define*

$$\tilde{\beta} = 0$$

$$\tilde{F}_X(x) = \tilde{\mathrm{P}}(X \leq x) = \mathrm{P}(X \leq x \mid Y > 0)$$

$$\tilde{F}_{U|X}(u|x) = \tilde{\mathrm{P}}(U \leq u \mid X = x) = \begin{cases} \frac{1}{2} + \frac{1}{2}\mathrm{P}(Y \leq u \mid X = x, Y > 0) & \text{for } u > 0; \\ \frac{1}{2} & \text{for } u = 0; \\ \frac{1}{2} - \frac{1}{2}\mathrm{P}(Y < -u \mid X = x, Y > 0) & \text{for } u < 0. \end{cases}$$

Then $(\tilde{\beta}, \tilde{F}_X, \tilde{F}_{U|X})$ *gives the same distribution of* (Y, X) *conditional on* $Y > 0$ *as does* $(\beta, F_X, F_{U|X})$.

Proof: Let $\tilde{P}$ denote probabilities under $(\tilde{\beta}, \tilde{F}_X, \tilde{F}_{U|X})$. To prove the theorem, it must be established that

$$\tilde{\mathrm{P}}(X \leq x \mid Y > 0) = \mathrm{P}(X \leq x \mid Y > 0) \tag{2.1}$$

and

$$\tilde{\mathrm{P}}(Y \leq y \mid Y > 0, X = x) = \mathrm{P}(Y \leq y \mid Y > 0, X = x) \tag{2.2}$$

First notice that

$$\tilde{\mathrm{P}}(Y > 0 \mid X = x) = \tilde{\mathrm{P}}(X\tilde{\beta} + U > 0 \mid X = x) = \tilde{\mathrm{P}}(U > 0 \mid X = x) = \frac{1}{2}$$

[2] Median independence is often generalized to quantile independence. In this paper, we will ignore this generalization.

so that the event $\{Y > 0\}$ is independent of X under the specification $(\tilde{\beta}, \tilde{F}_X, \tilde{F}_{U|X})$. This implies that

$$\tilde{\mathrm{P}}(X \le x \mid Y > 0) = \tilde{\mathrm{P}}(X \le x) = \mathrm{P}(X \le x \mid Y > 0)$$

so we have proved (2.1). Also

$$\begin{aligned}
\tilde{\mathrm{P}}(Y \le y \mid X = x, Y > 0) &= \tilde{\mathrm{P}}(X\tilde{\beta} + U \le y \mid X = x, X\tilde{\beta} + U > 0) \\
&= \tilde{\mathrm{P}}(U \le y \mid X = x, U > 0) \\
&= \frac{\tilde{\mathrm{P}}(0 < U \le y \mid X = x)}{\tilde{\mathrm{P}}(U > 0 \mid X = x)} \\
&= \frac{\left(\frac{1}{2} + \frac{1}{2}\mathrm{P}(Y \le y \mid X = x, Y > 0)\right) - \frac{1}{2}}{\frac{1}{2}} \\
&= \mathrm{P}(Y \le y \mid X = x, Y > 0)
\end{aligned}$$

This proves (2.2). ■

Theorem 1 shows that for any specification of $(\beta, F_X, F_{U|X})$ there exists another specification, $(\tilde{\beta}, \tilde{F}_X, \tilde{F}_{U|X})$, such that:

a. $\tilde{\beta} = 0$.

b. $\tilde{F}_{U|X}$ is symmetric around zero, and hence has median equal to 0 conditional on X as well as $\mathrm{E}[U|X] = 0$, provided that the expectation exists.

c. The observed distribution of X, i.e., the distribution of X conditional on $Y > 0$, is the same under the two specifications.

The parameter β is therefore not identified relative to $b = 0$ under the assumption that the conditional distribution of U given X, and assumptions about the conditional distribution of X given $Y > 0$ do not help identification.[3]

Powell (1986) points out that symmetry is not sufficient for consistency of his estimator. Theorem 1 shows that this problem is not specific to his estimator but rather is an implication of the truncated regression model. To prove identification, Powell assumes the distribution of U given X is symmetric and unimodal and that $\mathrm{E}[X'X \mid X\beta > 0]$ is positive definite.[4] The following

[3] It would be interesting to investigate to what extent β is identified relative to other $b \in R^P$ under the assumption of symmetry, for example.

[4] Powell's assumption on the regressors is formulated a little more generally, as he allows the regressors not to be identically distributed.

theorem shows that it is the unimodality rather than the symmetry that is the key to identifiability.

THEOREM 2. *Suppose that*

i. *The distribution of U given X is absolutely continuous and unimodal with* $\mathrm{mode}(U|X) = 0$,

ii. $\mathrm{P}(X\beta > 0) > 0$, *and*

iii. *The distribution of X conditional on $X\beta > 0$ is not concentrated in a linear subspace of R^P.*

Then β is identified.

Proof: Under the conditions, there exists a compact set, $A \subset \mathcal{X}$, such that $\mathrm{P}(X \in A) > 0$ and $\mathrm{mode}[Y|X = x] > 0$ for all $x \in A$. Then

$$\begin{aligned}\mathrm{E}\big[(X'X) \mid \mathrm{mode}(Y|X) > 0,\ X \in A\big]^{-1}\mathrm{E}\big[X'\,\mathrm{mode}(Y|X) \mid \mathrm{mode}(Y|X) > 0,\ X \in A\big]& \\ = \mathrm{E}\big[(X'X) \mid X\beta > 0,\ X \in A\big]^{-1}\mathrm{E}\big[X'X\beta \mid X\beta > 0,\ X \in A\big] = \beta&\end{aligned}$$

which shows that β is identified. ■

The conditioning on $X \in A$ in the proof of Theorem 2 guarantees that the expectations exist. Except for that the set A plays no role.

To use the proof of Theorem 2 for construction of an estimator, we would have to estimate the mode of Y given X for each observation of X. This is in principle not impossible: we could for example use a kernel estimator to estimate the joint distribution of Y and X and then for each $x \in \mathcal{X}$ we would find the value of y that maximizes the joint density. The problem with this approach is that estimation of the density requires some kind of grouping of the data (in R^{P+1}) and the asymptotic theory requires that the groups depend on sample size in such a way that all the X's in the same group become arbitrarily close to each other as the sample size goes to infinity. In other words, the asymptotic distribution probably is not a good approximation to the finite sample distribution if the X's are too far apart. On the other hand, we can only expect to find a "precise" estimate of the value of y that maximizes the joint distribution of Y and X for given X if the estimate of the joint distribution is based on a lot of observations, i.e., if a lot of the X's are grouped together.

We will therefore present an identification result which is less general than Theorem 2, but which allows the X's to be grouped in fixed groups. This means that to estimate a sample analog, it is not necessary to let the distance between the X's in the group decrease as the sample size increases.

The basic observation that underlies the identification result is that, conditional on X, the mode of $Y - \max\{X\beta, 0\}$ equals 0. This implies that $\text{mode}\big[Y - \max\{X\beta, 0\} \bigm| X \in \Gamma\big] = 0$ where Γ is any subset of $\mathcal{X}$ such that $\mathrm{P}(X \in \Gamma) > 0$. The idea of the following identification result is to choose a collection of subsets of $\mathcal{X}$, $\{\Gamma_j : j = 1, \ldots J\}$ such that

$$\text{mode}\big[Y - \max\{X\beta, 0\} \bigm| X \in \Gamma_j\big] = 0 \qquad \text{for} \quad j = 1, \ldots, J$$

suffices to identify β.

THEOREM 3. *Suppose that the distribution of U given X is absolutely continuous and unimodal with* $\text{mode}(U|X) = 0$*. Let $\{\Gamma_j : j = 1, \ldots J\}$ be a collection of subsets of $\mathcal{X}$ such that there exists a subcollection of connected sets $\{\Gamma_{j_p} : p = 1, \ldots, P\}$ such that $\mathrm{P}(X \in \Gamma_{j_p}) > 0$ for $p = 1, \ldots, P$ and such that for any $(x_1, \ldots, x_P) \in \Gamma_{j_1} \times, \ldots \times \Gamma_{j_P}$, $x_1, \ldots, x_P$ are linearly independent. Assume that $\mathrm{P}(X\beta > 0 | X \in \Gamma_{j_p}) = 1$ for $p = 1, \ldots, P$. Then β is the only element in R^P that satisfies*

$$\text{mode}\big[Y - \max\{X\beta, 0\} \bigm| X \in \Gamma_j\big] = 0 \qquad \text{for} \quad j = 1, \ldots, J. \tag{2.3}$$

Proof: β clearly satisfies (2.3) by the discussion preceding the theorem, so we have only to show that there exists no other b that satisfies (2.3).

Let b be a candidate. First notice that if $xb < x\beta$ for all $x \in \Gamma_{j_p}$ then

$$\begin{aligned} &\text{mode}\big[Y - \max\{Xb, 0\} \bigm| X \in \Gamma_{j_p}\big] \\ &\qquad = \text{mode}\big[Y - \max\{X\beta, 0\} + (\max\{X\beta, 0\} - \max\{Xb, 0\}) \bigm| X \in \Gamma_{j_p}\big] > 0 \end{aligned}$$

Likewise if $xb > x\beta$ for all $x \in \Gamma_{j_p}$ then

$$\begin{aligned} &\text{mode}\big[Y - \max\{Xb, 0\} \bigm| X \in \Gamma_{j_p}\big] \\ &\qquad = \text{mode}\big[Y - \max\{X\beta, 0\} + (\max\{X\beta, 0\} - \max\{Xb, 0\}) \bigm| X \in \Gamma_{j_p}\big] < 0 \end{aligned}$$

So for b to satisfy the condition, there cannot exist a set Γj_p such that either $xb > x\beta \quad \forall x \in \Gamma_{j_p}$ or $xb < x\beta \quad \forall x \in \Gamma_{j_p}$. Since the sets γ_{j_p} are all connected, this means that for $p = 1, \ldots, P$ there exists $x_{j_p} \in \Gamma_{j_p}$ such that

$$x_{j_p} b = x_{j_p} \beta \qquad p = 1, \ldots, P$$

or

$$x_{j_p}(b-\beta)=0 \qquad p=1,\ldots,P$$

But as x_{j_p} are linearly independent, this implies that $b=\beta$. ■

The conditions in Theorem 3 require some explanation. Suppose that X has a multinomial distribution that takes the values $x_1,\ldots,x_J$. We can then define $\Gamma_j=\{x_j\}$. The conditions in Theorem 3 are then exactly equal to the conditions of Theorem 2. This is the motivation for the conditions in Theorem 3. The added complexity in the formulations of Theorem 3 comes from the grouping of the data into a finite set of groups.

The purpose of Theorem 3 is to use the identification result in Theorem 2 to construct an estimator. Another way to construct estimators from Theorem 2 uses the intuition from an alternative proof of Theorem 2:

Alternative proof of Theorem 2: Let $f_{Y|X}$ be the density of Y given X in the truncated sample. Denote the density at η of $Y-\max\{0,X\beta\}$ by $f(\eta;b)$. Then for $b\neq\beta$,

$$\begin{aligned} f(0,b) &= \mathrm{E}_X\left[f_{Y|X}\left(\max\{0,Xb\}\right)\right] \\ &= \mathrm{E}_X\left[f_{Y|X}\left(\max\{0,X\beta\}\right)\right]-\mathrm{E}_X\left[f_{Y|X}\left(\max\{0,X\beta\}\right)-f_{Y|X}\left(\max\{0,Xb\}\right)\right] \\ &= f(0;\beta)-\mathrm{E}_X\left[f_{Y|X}\left(\max\{0,X\beta\}\right)-f_{Y|X}\left(\max\{0,Xb\}\right)\right] \\ &< f(0;\beta) \end{aligned}$$

where the inequality follows from the fact that the mode of Y given X is $\max\{0,X\beta\}$. This completes the alternative proof of Theorem 2. ■

In the next section we will see how this proof can be used to construct an estimator for β.

3. Consistent Estimation.

In this section we will define an estimator which can be interpreted as the solution of a population analog to the identification condition presented in Theorem 3. There we saw that, subject to regularity conditions, $b=\beta$ is the only value for b for which

$$\operatorname{mode}\left[Y-\max\{Xb,0\}\;\middle|\;X\in\Gamma_j\right]=0 \qquad \text{for} \quad j=1,\ldots J. \tag{3.1}$$

The general idea behind the first estimator defined in this section is to construct an estimator of the left hand side of (3.1) and then estimate β by the value of b that makes this as close to zero as possible for $j = 1, \ldots, J$.

To formally define the estimator, we need to define some notation. Assume that we have a sample of n i.i.d observations of (Y, X) from (2.1). Let n_j be the number of observations for which $X \in \Gamma_j$. We can then label the observations according to whether $X \in \Gamma_j$. We write[5] (Y_i^j, X_i^j) for $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$.

Define $V_i^j(b) = Y_i^j - \max\{X_i^j b, 0\}$. We can then rewrite (3.1) as

$$\operatorname{mode}\big[V^j(b)\big] = 0 \qquad \text{for} \quad j = 1, \ldots J \tag{3.2}$$

Let $f^j(\cdot; b)$ be the density of $V^j(b)$. We will implement (3.2) by estimating $f^j(\cdot, b)$ by a kernel estimator.[6] Let $K : R \mapsto R$ be a function that satisfies

$$\int_{-\infty}^{\infty} K(y)\, dy = 1 \tag{3.3}$$

$$C_1 = \sup_x K(x) < \infty \tag{3.4}$$

$$\sup_x \big| K(x + y) - K(x) \big| \leq C_2 |y| \tag{3.5}$$

for some finite $C_2 > 0$. Also let h_n be a sequence of real numbers such that

$$h_n \to 0 \qquad \text{as} \qquad n \to \infty \tag{3.6}$$

$$\sum_{n=1}^{\infty} h_n^{2(P+1)} \exp\big(-n h_n^2\big) < \infty \tag{3.7}$$

We then define the kernel estimator of $f^j(\eta; b)$ by

$$\hat{f}_n^j(\eta; b) = \frac{1}{n_j h_{n_j}} \sum_{i=1}^{n_j} K\Big(\frac{\eta - V_i^j}{h_{n_j}}\Big)$$

We now turn to the estimation of the mode of $V^j(b)$. We would like to define the mode of $V^j(b)$ as the value of η that maximizes $f^j(\cdot; b)$ and the estimated mode of $V^j(b)$ as the value that maximizes

---

[5] Notice that some of the observations might appear more than once with this labelling, and that not all observations have to appear.

[6] See Prakasa–Rao (1983) for a general discussion of kernel estimators and other density estimators.

$\hat{f}_n^j(\cdot;b)$, but the distribution of $V^j(b)$ is not necessarily unimodal, so the maximizers of $f^j(\cdot;b)$ and $\hat{f}_n^j(\cdot;b)$ may not be unique. Also, since we are only interested in whether or not the mode of $V^j(b)$ is zero, we can restrict ourselves to maximizing $f^j(\cdot;b)$ and $\hat{f}_n^j(\cdot;b)$ over $\eta \in [-1,1]$. This will make the proof of consistency easier, and it will also be easier to implement. With this in mind we define

$$m^j(b) = \{\mu : f^j(\mu;b) = \sup_{\eta\in[-1,1]} f^j(\eta;b)\}$$

and

$$\hat{m}_n^j(b) = \{\mu : \hat{f}_n^j(\mu;b) = \sup_{\eta\in[-1,1]} \hat{f}_n^j(\eta;b)\}$$

This suggests estimating β by the value of b that makes the estimated value of $\text{mode}\big[V^j(b)\big]$ as close to zero as possible in all the J groups. This can, for example, be done by minimizing the sum of the absolute values of the estimates of $\text{mode}\big[V^j(b)\big]$. This is the basic idea behind the estimator, but in order to formally define the estimator we must first deal with some technical complications.

Consider the function

$$Q_n(b) = \sum_{j=1}^{J} \inf\{|\mu| : \mu \in \hat{m}_n^j(b)\}$$

We would then like to define $\hat{b}_n$ as a value that satisfies

$$Q_n(\hat{b}_n) = \min_{b\in\mathbf{B}} Q_n(b)$$

Unfortunately this is not possible, as Q_n will not in general be a continuous function, so $\inf_{b\in\mathbf{B}} Q_n(b)$ may not be attained. Instead we can define a sequence $\hat{b}_n^\ell$ such that

$$\lim_{\ell\to\infty} Q_n(\hat{b}_n^\ell) = \inf_{b\in\mathbf{B}} Q_n(b).$$

The sequence $\{\hat{b}_n^\ell \mid \ell = 1,\ldots\}$ has a convergent subsequence. We can therefore define $\hat{b}_n$ as the limit of any convergent subsequence of $\{\hat{b}_n^\ell \mid \ell = 1,\ldots\}$.

Because this estimator is not necessarily measurable, we will prove that it is consistent in the sense that for any open neighborhood, $N(\beta)$, of β there exists a measurable set A and an integer n_1 such that $\mathrm{P}(A) = 1$ and $A \subset \{\hat{b}_n \in N(\beta)\}$ for all $n \geq n_1$.

Theorem 4 below gives the consistency result. In order to prove the consistency result, we need to make an assumption about the continuity of the density of $Y - \max\{Xb, 0\}$.

ASSUMPTION 1. *Let* $\mathbf{B}$ *denote the parameter space and let* Γ_j *be a subset of* $\mathcal{X}$. *Let* $b \in \mathbf{B}$ *be given. Define*

$$c_b(\Gamma_j) = \sup_{x \in \Gamma_j} - \max\{xb, 0\}$$

The interval $[c_b(\Gamma_j), \infty)$ *will then be a subset of the support of the distribution of* $Y - \max\{Xb, 0\}$ *conditional on* X *for* $X \in \Gamma_j$. *Furthermore, let* $\tilde{\mathbf{B}}$ *be any open subset of* $\mathbf{B}$, *and define*

$$c_{\tilde{\mathbf{B}}}(\Gamma_j) = \sup_{b \in \tilde{\mathbf{B}}} c_b(\Gamma_j)$$

It is then assumed that

i. For given b, *the density of* $Y - \max\{Xb, 0\}$ *conditional on* $X \in \Gamma_j$, $f^j(\eta; b)$, *is continuous over* $(c_b(\Gamma_j), \infty)$.

and

ii. For $\eta \in (c_{\tilde{\mathbf{B}}}(\Gamma_j), \infty)$, $f^j(\eta, b)$ *is continuous in* $b \in \tilde{\mathbf{B}}$.

THEOREM 4. *Assume that* $\{(Y_i, X_i)\}_{i=1}^n$ *is a sample of* n *i.i.d. observations from the truncated regression model given by (2.1). Assume that the distribution of* U *given* X *is absolutely continuous with support equal to* R *and unimodal with* $\mathrm{mode}(U|X) = 0$. *Further, assume that the parameter space* $\mathbf{B}$ *is compact.*

Let $\{\Gamma_j \mid j = 1, \ldots J\}$ *be a collection of subsets of* $\mathcal{X}$ *each satisfying Assumption 1. Assume that there exists a subcollection of connected sets* $\{\Gamma_{j_p} \mid p = 1, \ldots, P\}$ *such that*

i. $\mathrm{P}(X \in \Gamma_{j_p}) > 0$ *for* $p = 1, \ldots, P$.

ii. For any $(x_1, \ldots, x_P) \in \Gamma_{j_1} \times, \ldots \times \Gamma_{j_P}$, $x_1, \ldots, x_P$ *are linearly independent.*

iii. $\mathrm{P}(X\beta > 0 | X \in \Gamma_{j_p}) = 1$ *for* $p = 1, \ldots, P$.

Assume that Assumption 1 is satisfied for all sets Γ_j.

Then $\hat{b}_n$ *is consistent in the sense discussed above if we restrict* $K(\eta)$ *to be increasing for* $\eta < 0$ *and decreasing for* $\eta > 0$.

Proof: Recall that

$$Q_n(b) = \sum_{j=1}^{J} \inf\{|\hat{m}_n^j(b)|\}.$$

It then suffices to prove that for all open neighborhoods of β, $N(\beta)$, there exists a $\nu > 0$ such that

$$\mathrm{P}\Big(\lim_{n\to\infty} \sup_{b\in\mathbf{B}\setminus N(\beta)} Q_n(b) - Q_n(\beta) \geq \nu\Big) = 1.$$

Without loss of generality assume that $j_p = p$ for $p = 1, \ldots, P$. In this proof we will let ω denote points in the underlying probability space, and we will write $Q_n(b,\omega)$, $m_n^j(b,\omega)$, etc. whenever we need to emphasize that the functions are random. We will use A (with subscripts) to denote events.

Let an open neighborhood, $N(\beta)$, of β be given. Then, by the same argument as in the proof of Theorem 3, for any $b \in \mathbf{B} \setminus N(\beta)$ there exists a p such that either $xb > x\beta\ \forall x \in \Gamma_p$ or $xb < x\beta\ \forall x \in \Gamma_p$. Let

$$k(b) = \max_{1\leq p\leq P} \inf_{x\in\Gamma_p} \big|xb - x\beta\big|$$

Then $k(b) > 0$ and k is a continuous function of b. As $\mathbf{B}$ is compact there therefore exists an $\varepsilon > 0$ such that

$$k(b) > \varepsilon \qquad \forall b \in \mathbf{B}.$$

Without loss of generality assume that $\varepsilon < 1$. We can then divide $\mathbf{B} \setminus N(\beta)$ into $2P$ compact subsets, $\mathbf{B}_{11}$, $\mathbf{B}_{12}, \ldots, \mathbf{B}_{1P}$, $\mathbf{B}_{21}$, $\mathbf{B}_{22}, \ldots, \mathbf{B}_{2P}$, such that

$$b \in \mathbf{B}_{1p} \Rightarrow xb - x\beta > \varepsilon \qquad \forall x \in \Gamma_p$$

$$b \in \mathbf{B}_{2p} \Rightarrow xb - x\beta < -\varepsilon \qquad \forall x \in \Gamma_p$$

Let $\nu = \frac{\varepsilon}{10}$. We will then show

(a) There exists an event A_{01}, such that $\mathrm{P}(A_{01}) = 0$, and for all $\omega \notin A_{01}$, there exists an n_0 such that for $n \geq n_0$, $Q_n(\beta,\omega) < \nu$,

and

(b) There exists an event A_{02}, such that $\mathrm{P}(A_{02}) = 0$, and for all $\omega \notin A_{02}$, there exists an n_0 such that for $n \geq n_0$, $Q_n(b,\omega) > 2\nu \quad \forall b \in \mathbf{B} \setminus N(\beta)$.

First note that there exists an event A_1 with $\mathrm{P}(A_1) = 0$, such that for $\omega \notin A_1$, $n_j(\omega) \to \infty$ as $n \to \infty$.

To prove (a), we will show that for any j ($j = 1, \ldots, J$), there is an n_1 such that for $n \geq n_1$, $\inf\{|m_n^j(\beta)|\} \leq \frac{\nu}{J}$ on a set of probability 1. Let j be given. Notice that there exists a $\delta > 0$ such that

$$f^j(0;\beta) > f^j(\eta;\beta) + \delta \qquad \forall |\eta| \geq \frac{\nu}{2J}.$$

From Prakasa–Rao (1983) Lemma 2.1.2, we know that there exists an event A_{2j} with $\mathrm{P}(A_{2j}) = 0$ and for $\omega \notin A_{2j} \cup A_1$

$$\sup_{\eta} \bigl| \hat{f}_n^j(\eta, \beta) - E\bigl[\hat{f}_n^j(\eta, \beta)\bigr] \bigr| \to 0 \qquad \text{as } n \to \infty$$

So for $\omega \notin A_{2j} \cup A_1$, we can find an n_2 such that for $n \geq n_2$

$$\bigl| \hat{f}_n^j(\eta; \beta) - E\bigl[\hat{f}_n^j(\eta; \beta)\bigr] \bigr| < \frac{\delta}{6} \qquad \forall \eta \tag{3.8}$$

By Assumption 1, $f^j(\eta; \beta)$ is continuous for $\eta \geq 0$. We can therefore find a $\eta_0 \in (0, \frac{\nu}{2J})$ and an open neighborhood, $N(\eta_0)$ of η_0 such that

$$f^j(\eta; \beta) > f^j(0; \beta) - \frac{\delta}{6} \qquad \forall \eta \in N(\eta_0)$$

Now

$$\begin{aligned}
E\bigl[\hat{f}_n^j(\eta_0; \beta)\bigr] &= \frac{1}{h_{n_j}} \int_{-\infty}^{\infty} K\Bigl(\frac{\eta_0 - y}{h_{n_j}}\Bigr)\ F^j(dy; \beta) \\
&> \frac{1}{h_{n_j}} \int_{N(\eta_0)} K\Bigl(\frac{\eta_0 - y}{h_{n_j}}\Bigr) f^j(y; \beta)\ dy \\
&> \bigl(f^j(0; \beta) - \delta/6\bigr) \frac{1}{h_{n_j}} \int_{N(\eta_0)} K\Bigl(\frac{\eta_0 - y}{h_{n_j}}\Bigr)\ dy \\
&= \bigl(f^j(0; \beta) - \delta/6\bigr) \int_{\frac{1}{h_{n_j}}(\eta_0 - N(\eta_0))} K(y)\ dy \\
&> f^j(0; \beta) - \delta/3
\end{aligned} \tag{3.9}$$

where the last inequality is true for n greater than some n_3 depending on K and $N(\eta_0)$.

Equations (3.8) and (3.9) imply that for $\omega \notin A_1 \cup A_{2j}$

$$\hat{f}_n^j(\eta_0; \beta) > f^j(0, \beta) - \delta/2 \qquad \text{for} \quad n \geq \max\{n_2, n_3\} \tag{3.10}$$

Now consider an $|\eta_1| > \frac{\nu}{J}$. First notice that

$$\begin{aligned}
E\bigl[\hat{f}_n^j(\eta_1; \beta)\bigr] &= \frac{1}{h_{n_j}} \int_{-\infty}^{\infty} K\Bigl(\frac{\eta_1 - y}{h_{n_j}}\Bigr)\ F^j(dy; \beta) \\
&= \frac{1}{h_{n_j}} \Bigl[\int_{|y| \leq \nu/2J} K\Bigl(\frac{\eta_1 - y}{h_{n_j}}\Bigr)\ F^j(dy; \beta) + \int_{|y| > \nu/2J} K\Bigl(\frac{\eta_1 - y}{h_{n_j}}\Bigr)\ F^j(dy; \beta) \Bigr]
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{h_{n_j}}\Big[\int_{|y|\leq \nu/2J} K\Big(\frac{\eta_1 - y}{h_{n_j}}\Big) f^j(0;\beta)\, dy + \int_{|y|>\nu/2J} K\Big(\frac{\eta_1 - y}{h_{n_j}}\Big)\big(f^j(0;\beta) - \delta\big)\, dy\Big] \\
&= \frac{1}{h_{n_j}}\big(f^j(0;\beta) - \delta\big)\int_{-\infty}^{\infty} K\Big(\frac{\eta_1 - y}{h_{n_j}}\Big)\, dy + \frac{1}{h_{n_j}}\delta \int_{|y|\leq \nu/2J} K\Big(\frac{\eta_1 - y}{h_{n_j}}\Big)\, dy \\
&= f^j(0;\beta) - \delta + \delta \int_{\frac{\eta_1 - \nu/2J}{h_{n_j}}}^{\frac{\eta_1 + \nu/2J}{h_{n_j}}} K(y)\, dy \\
&\leq f^j(0;\beta) - \delta + \delta \int_{\nu/2Jh_{n_j}}^{\infty} K(y)\, dy + \delta \int_{-\infty}^{-\nu/2Jh_{n_j}} K(y)\, dy \\
&\leq f^j(0;\beta) - \frac{5\delta}{6} \qquad\qquad (3.11)
\end{aligned}
$$

where the last inequality is true for n greater than some n_4 not depending on η_1, and the next to last inequality follows from $|\eta_1| > \nu/J$.

Equations (3.8) and (3.11) imply that for $\omega \notin A_1 \cup A_{2j}$

$$
\hat{f}_n^j(\eta_1;\beta) < f^j(0;\beta) - \frac{2\delta}{3} \qquad \text{for} \quad n \geq \max\{n_2, n_4\} \tag{3.12}
$$

In conclusion, equations (3.10) and (3.12) give that for $\omega \notin A_1 \cup A_{2j}$, there exists an $\eta_0 < \nu/J$ and an $n_0 = \max\{n_2, n_3, n_4\}$ such that for $n \geq n_0$

$$
\hat{f}_n^j(\eta_1;\beta) < \hat{f}_n^j(\eta_0;\beta) - \delta/6 \qquad \forall |\eta_1| > \frac{\nu}{J}. \tag{3.13}
$$

This implies that $\hat{m}_n^j(\beta)$ cannot contain any elements greater than ν/J in absolute value. As $\mathrm{P}\big(A_1 \cup \bigcup_{j=1}^J A_{2j}\big) = 0$, we have thus proved (a) above.

We now turn to (b).

Let b be any element of $\mathbf{B} \setminus N(\beta)$. For some p, b belongs to either $\mathbf{B}_{1p}$ or $\mathbf{B}_{2p}$. We will now show that there exists an event A_{3p} with $\mathrm{P}(A_{3p}) = 0$ such that for $\omega \notin A_{3p}$ there is an n_5 such that for $n \geq n_5$, $\inf\{|\hat{m}_n^p(b)|\} \geq 2\nu$.

Recall that $\varepsilon = 10\nu$. Define

$$
\delta_1 = \min\Big\{\inf_{b\in\mathbf{B_{1p}}} f^p(-6\nu;b) - f^p(-4\nu;b), \inf_{b\in\mathbf{B_{2p}}} f^p(6\nu;b) - f^p(4\nu;b)\Big\}
$$

$$
\delta_2 = \max\Big\{\sup_{b\in\mathbf{B_{1p}}} f^p(-10\nu;b), \sup_{b\in\mathbf{B_{2p}}} f^p(10\nu;b)\Big\}.
$$

Assume that $b \in \mathbf{B}_{1p}$. Then $xb > x\beta + \varepsilon\ \forall x \in \Gamma_p$, and since $x\beta > 0 \quad \forall x \in \Gamma_p$, we have $-\max\{xb, 0\} < -\max\{x\beta, 0\} - \varepsilon\ \forall x \in \Gamma_p$. This implies that $c_b(\Gamma_p) \leq -\varepsilon$ and that for $X \in \Gamma_p$ the conditional mode of $Y - \max\{Xb, 0\}$ given X is less than $-\varepsilon$. $f^p(\eta, b)$ is therefore non-increasing in η for $\eta > -\varepsilon$.

Let n_1 be big enough that for $n \geq n_1$:

$$\int_{-2\nu/h_{n_p}}^{2\nu/h_{n_p}} K(y)\, dy > 1 - \delta_2^{-1}\delta_1/6$$

and

$$\int_{2\nu/h_{n_p}}^{\infty} K(y)\, dy < \delta_2^{-1}\delta_1/6$$

For $\eta \in [-2\nu, 2\nu]$ given, we then have

$$\begin{aligned}
& E\big[\hat{f}_n^p(-8\nu; b)\big] - E\big[\hat{f}_n^p(\eta, b)\big] \\
= \; & \frac{1}{h_{n_p}}\Big[\int_{-\infty}^{\infty} K\Big(\frac{-8\nu - y}{h_{n_p}}\Big)\, F^p(dy; b) - \int_{-\infty}^{\infty} K\Big(\frac{\eta - y}{h_{n_p}}\Big)\, F^p(dy; b)\Big] \\
= \; & \frac{1}{h_{n_p}}\Big[\int_{-\infty}^{-10\nu} K\Big(\frac{-8\nu - y}{h_{n_p}}\Big)\, F^p(dy; b) + \int_{-10\nu}^{-6\nu} K\Big(\frac{-8\nu - y}{h_{n_p}}\Big)\, F^p(dy; b) \\
& + \int_{-6\nu}^{\infty} K\Big(\frac{-8\nu - y}{h_{n_p}}\Big)\, F^p(dy; b) - \int_{-\infty}^{-10\nu} K\Big(\frac{\eta - y}{h_{n_p}}\Big)\, F^p(dy; b) \\
& - \int_{-10\nu}^{-4\nu} K\Big(\frac{\eta - y}{h_{n_p}}\Big)\, F^p(dy; b) - \int_{-4\nu}^{\infty} K\Big(\frac{\eta - y}{h_{n_p}}\Big)\, F^p(dy; b)\Big] \\
> \; & \frac{1}{h_{n_p}}\Big[\int_{-\infty}^{-10\nu} \Big(K\Big(\frac{-8\nu - y}{h_{n_p}}\Big) - K\Big(\frac{\eta - y}{h_{n_p}}\Big)\Big)\, F^p(dy; b) + f^p(-6\nu; b)\int_{-10\nu}^{-6\nu} K\Big(\frac{-8\nu - y}{h_{n_p}}\Big)\, dy \\
& - f^p(-10\nu; b)\int_{-10\nu}^{-4\nu} K\Big(\frac{\eta - y}{h_{n_p}}\Big)\, dy - f^p(-4\nu; b)\int_{-4\nu}^{\infty} K\Big(\frac{\eta - y}{h_{n_p}}\Big)\, dy\Big] \\
> \; & \frac{1}{h_{n_p}}\Big[f^p(-6\nu; b)\int_{-10\nu}^{-6\nu} K\Big(\frac{-8\nu - y}{h_{n_p}}\Big)\, dy \\
& - f^p(-10\nu; b)\int_{-10\nu}^{-4\nu} K\Big(\frac{\eta - y}{h_{n_p}}\Big)\, dy - f^p(-4\nu; b)\int_{-4\nu}^{\infty} K\Big(\frac{\eta - y}{h_{n_p}}\Big)\, dy\Big] \\
= \; & f^p(-6\nu; b)\int_{-2\nu/h_{n_p}}^{2\nu/h_{n_p}} K\big(y\big)\, dy \\
& - f^p(-10\nu; b)\int_{(\eta+4\nu)/h_{n_p}}^{(\eta+10\nu)/h_{n_p}} K\big(y\big)\, dy - f^p(-4\nu; b)\int_{-\infty}^{(\eta+4\nu)/h_{n_p}} K\big(y\big)\, dy \\
\geq \; & f^p(-6\nu; b)(1 - \delta_2^{-1}\delta_1/6) - f^p(-10\nu; b)\delta_2^{-1}\delta_1/6 - f^p(-4\nu; b)
\end{aligned}$$

$$> \quad f^p(-6\nu;b) - f^p(-4\nu;b) - 2\delta_2^{-1}\delta_1 f^p(-10\nu;b)/6$$

$$> \quad 2\delta_1/3 \tag{3.14}$$

A similar calculation can be used to show that for $b \in \mathbf{B}_{2p}$,

$$E\big[\hat{f}_n^p(8\nu;b)\big] - E\big[\hat{f}_n^p(\eta;b)\big] > 2\delta_1/3 \tag{3.15}$$

for all $\eta \in [-2\nu, 2\nu]$ and for $n \geq n_6$ for some n_6 not depending on η.

We will now show that there exists an event A_{3p} with $\mathrm{P}(A_{3p}) = 0$, such that for $\omega \notin A_{3p} \cup A_1$

$$\hat{f}_n^p(\eta;b) - E\big[\hat{f}_n^p(\eta;b)\big] \to 0 \qquad \text{as} \quad n \to \infty \tag{3.15a}$$

uniformly in $b \in \mathbf{B}_p = \mathbf{B}_{1p} \cup \mathbf{B}_{2p}$ and $\eta \in [-1,1]$.

The following calculation closely follows Prakasa–Rao (1983) pages 186–188. Ichimura (1988) uses the same approach.

Let $\|\cdot\|$ denote the sup-norm of a vector. When we talk about balls we will mean with respect to this norm. Notice that $\big|\max\{0, xb_1\} - \max\{0, xb_2\}\big| \leq \big|xb_1 - xb_2\big| = \big|x(b_1 - b_2)\big| \leq P\big\|x\big\|\big\|b_1 - b_2\big\|$ Also, since $\mathbf{B}_p$ is compact there exists a closed ball, S, in R^{P+1} such that $[-1,1] \times \mathbf{B}_p \subset S$.

Let $C_3 = \max\big\{2, 2P\sup_{x\in\Gamma_p}\|x\|\sup_{b\in\mathbf{B}_p}\|b\|\big\}$.

For any number $c > 0$, let $k = C_3/c$. Then there exists a sequence of k^{P+1} elements (η_ℓ, b_ℓ), $\ell = 1, \ldots, k^{P+1}$, in S, such that for any element (η, b) in S there exists a (η_ℓ, b_ℓ) such that $|\eta - \eta_\ell| \leq c$ and $\big|\max\{0, xb\} - \max\{0, xb_\ell\}\big| \leq c$ for all $x \in \Gamma_p$.

Denote $[-1,1] \times \mathbf{B}_p$ by Θ_p.

Let $\gamma > 0$ be arbitrary and let

$$c_n = \gamma h_{n_p}^2 / 24 C_2$$

$$k_n = C_3 / c_n$$

(k_n is really the smallest integer greater than or equal to C_3/c_n), and let (η_ℓ, b_ℓ), $\ell = 1, \ldots, k_n^{P+1}$, in S, such that for any element (η, b) in S there exists a (η_ℓ, b_ℓ) such that $|\eta - \eta_\ell| \leq c_n$ and $\big|\max\{0, xb\} - \max\{0, xb_\ell\}\big| \leq c_n$ for all $x \in \Gamma_p$. Also let $\Theta_{p\ell} = \{(\eta, b) : |\eta - \eta_\ell| \leq c_n,\ \big|\max\{0, xb\} - \max\{0, xb_\ell\}\big| \leq c_n \text{ for all } x \in \Gamma_p\}$.

Then

$$
\begin{aligned}
&\mathrm{P}\Big(\sup_{(\eta,b)\in\Theta_p}\big|\hat f_n^p(\eta;b)-E\big[\hat f_n^p(\eta;b)\big]\big|\geq\gamma\Big)\\
\leq\quad&\sum_{\ell=1}^{k_n^{P+1}}\mathrm{P}\Big(\sup_{(\eta,b)\in\Theta_{p\ell}}\big|\hat f_n^p(\eta;b)-E\big[\hat f_n^p(\eta;b)\big]\big|\geq\gamma\Big)\\
\leq\quad& k_n^{P+1}\sup_{1\leq\ell\leq k_n^{P+1}}\mathrm{P}\Big(\sup_{(\eta,b)\in\Theta_{p\ell}}\big|\hat f_n^p(\eta;b)-E\big[\hat f_n^p(\eta;b)\big]\big|\geq\gamma\Big)\\
\leq\quad& k_n^{P+1}\sup_{1\leq\ell\leq k_n^{P+1}}\Big[\mathrm{P}\Big(\sup_{(\eta,b)\in\Theta_{p\ell}}\big|\hat f_n^p(\eta;b)-\hat f_n^p(\eta_\ell;b_\ell)\big|\geq\gamma/3\Big)\\
&+\mathrm{P}\Big(\sup_{(\eta,b)\in\Theta_{p\ell}}\big|E\big[\hat f_n^p(\eta;b)\big]-E\big[\hat f_n^p(\eta_\ell;b_\ell)\big]\big|\geq\gamma/3\Big)\\
&+\mathrm{P}\Big(\big|\hat f_n^p(\eta_\ell;b_\ell)-E\big[\hat f_n^p(\eta_\ell;b_\ell)\big]\big|\geq\gamma/3\Big)\Big]
\end{aligned}
$$

Now for $(\eta,b)\in\Theta_{p\ell}$:

$$
\begin{aligned}
\big|\hat f_n^p(\eta,b)-\hat f_n^p(\eta_\ell,b_\ell)\big| &\leq \frac{1}{n_p h_{n_p}}\sum_{i=1}^{n_p}\Big|K\Big(\frac{\eta-Y_i^p-\max\{X_i^p b,0\}}{h_{n_p}}\Big)-K\Big(\frac{\eta_\ell-Y_i^p-\max\{X_i^p b_\ell,0\}}{h_{n_p}}\Big)\Big|\\
&\leq\frac{2}{h_{n_p}^2}C_2c_n\\
&=\gamma/4
\end{aligned}
$$

and therefore

$$
\big|E\big[\hat f_n^p(\eta,b)\big]-E\big[\hat f_n^p(\eta_\ell,b_\ell)\big]\big|\leq\gamma/4
$$

It follows that

$$
\begin{aligned}
&\mathrm{P}\Big(\sup_{(\eta,b)\in\Theta_p}\big|\hat f_n^p(\eta;b)-E\big[\hat f_n^p(\eta;b)\big]\big|\geq\gamma\Big)\\
&\leq k_n^{P+1}\sup_{1\leq\ell\leq k_n^{P+1}}\mathrm{P}\Big(\big|\hat f_n^p(\eta_\ell;b_\ell)-E\big[\hat f_n^p(\eta_\ell;b_\ell)\big]\big|\geq\gamma/3\Big)\\
&=k_n^{P+1}\sup_{1\leq\ell\leq k_n^{P+1}}\mathrm{P}\Big(\Big|\sum_{i=1}^{n}\Big\{K\Big(\frac{\eta_\ell-Y_i^p-\max\{X_i^p b_\ell,0\}}{h_{n_p}}\Big)\\
&\quad-E\Big[K\Big(\frac{\eta_\ell-Y_i^p-\max\{X_i^p b_\ell,0\}}{h_{n_p}}\Big)\Big]\Big\}\Big/ nh_{n_p}\Big|\geq\gamma/3\Big)
\end{aligned}
$$

If we define

$$Z_i = \Big\{K\Big(\frac{\eta_\ell - Y_i^p - \max\{X_i^p b_\ell, 0\}}{h_{n_p}}\Big) - E\Big[K\Big(\frac{\eta_\ell - Y_i^p - \max\{X_i^p b_\ell, 0\}}{h_{n_p}}\Big)\Big]\Big\}\Big/ h_{n_p}$$

then $0 \le |Z_i| \le 2C_1/h_{n_p}$, and Hoeffdings Inequality (Prakasa–Rao, Lemma 3.1.1) gives

$$\mathrm{P}\Big(\Big|\frac{1}{n_p}\sum_{i=1}^{n_p} Z_i\Big| \ge \gamma/3\Big) \le 2\exp\Big(\frac{-n_p h_{n_p}^2 \gamma^2}{18C_1^2}\Big).$$

Therefore

$$\mathrm{P}\Big(\sup_{(\eta,b)\in\Theta}\big|\hat{f}_n^p(\eta;b) - E\big[\hat{f}_n^p(\eta;b)\big]\big| \ge \gamma\Big) \le 2\Big(1 + \frac{C_3^2}{6\gamma_1 h_{n_p}^2}\Big)^{P+1}\exp\Big(\frac{-n_p h_{n_p}^2\gamma^2}{18C_1^2}\Big)$$

and hence

$$\sum_{n_p=1}^{\infty}\mathrm{P}\Big(\sup_{(\eta,b)\in\Theta}\big|\hat{f}_n^p(\eta;b) - E\big[\hat{f}_n^p(\eta;b)\big]\big| \ge \gamma\Big) < \infty,$$

so by the Borel–Cantelli lemma there exists an event A_{3p} with $\mathrm{P}(A_{3p}) = 0$ such that for $\omega \notin A_{3p}$

$$\sup_{(\eta,b)\in\Theta_p}\big|\hat{f}_n^p(\eta;b) - E\big[\hat{f}_n^p(\eta;b)\big]\big| \to 0. \tag{3.16}$$

This implies that for $\omega \notin A_{3p}$, there is an n_6, such that for $n \ge n_6$:

$$\big|\hat{f}_n^p(\eta;b) - E\big[\hat{f}_n^p(\eta;b)\big]\big| < \delta_1/3$$

for all $b \in \mathbf{B}_p$ and $\eta \in [-1,1]$. Combining this with (3.14), we find that for $\omega \notin A_{3p} \cup A_1$ and $b \in \mathbf{B}_{1p}$, there is an $n_7 = \max\{n_5, n_6\}$ such that for $n \ge n_7$

$$\hat{f}_n^p(-8\nu;b) > \hat{f}_n^p(\eta;b) + \delta_1/3 \qquad \forall\eta \in [-2\nu, 2\nu]. \tag{3.17}$$

Likewise, from (3.16) and (3.15) we find that for $\omega \notin A_{3p} \cup A_1$ and $b \in \mathbf{B}_{2p}$, there is an n_8 such that for $n \ge n_8$

$$\hat{f}_n^p(8\nu;b) > \hat{f}_n^p(\eta;b) + \delta_1/3 \qquad \forall\eta \in [-2\nu, 2\nu]. \tag{3.18}$$

Equations (3.17) and (3.18) imply that for $n \ge \max\{n_7, n_8\}$, $\inf\{|\hat{m}_n^p(b)|\} > 2\nu$ for all $b \in \mathbf{B}_p$. This implies (b). ■

It is natural to try to find the asymptotic distribution of the mode estimator. This seems to be very hard as the objective function is not only not necessarily continuous, but is also very hard to characterize.

We now turn to an estimator that is based on the alternative proof of the identification result of Theorem 2.

The idea of the alternative proof of Theorem 2 was that if we consider the density of $Y - \max\{0, Xb\}$ evaluated at 0, then this is maximized (as a function of b) at $b = \beta$. The idea behind the next estimator is to estimate the density of $Y - \max\{0, Xb\}$ at 0 by a kernel estimator and then search for the value of b that maximizes this. In other words, we define the estimator to be the maximizer of

$$\hat{f}_n(0; b) = \frac{1}{nh_n} \sum_{i=1}^{n} K\Big(\frac{y_i - \max\{0, x_i b\}}{h_n}\Big) \tag{3.19}$$

where K and h_n satisfy the conditions of (3.3)–(3.7).

This estimator is very close to the estimator proposed by Lee (1988). Maximization of (3.19) with uniform kernel, $\max\{0, x_i b\}$ replaced by $\max\{h_n, x_i b\}$ and h_n fixed (as opposed to a function of sample size), gives his estimator.

We can now state the consistency result for the estimator defined by maximization of (3.19).

THEOREM 5. *Assume that $\{(Y_i, X_i)\}_{i=1}^{n}$ is a sample of n i.i.d. observations from the truncated regression model given by (2.1). Assume that the parameter space* $\mathbf{B}$ *is compact and that $K(\eta) = 0$ for $\eta < 0$.*

Further, assume that the distribution of U given X is absolutely continuous with support equal to R and unimodal with $\mathrm{mode}(U|X) = 0$ *and that X is continuously distributed with compact support $\mathcal{X}$. Also assume that the conditional density of U given X is uniformly bounded, $f_{U|X} \leq f_{\max}$, and uniformly unimodal in a neighborhood of zero, i.e., there exists some ν_0 and some strictly increasing function, g, such that $f_{U|X}(0) - f_{U|X}(\eta) \geq g(|\eta|)$ for $|\eta| \leq \nu_0$. Also assume conditions (ii) and (iii) of Theorem 2.*

Then the estimator defined by maximization of (3.19) is consistent in the sense discussed before the statement of Theorem 4.

Proof: Let $f_{U|X}$, $\tilde{f}_{U|X}$, f_X and $\tilde{f}_X$ denote the density of U given X, the density of U given X and $Y > 0$, the density of X, and the density of X given $Y > 0$, respectively. It then follows that

$$\tilde{f}_{U|X}(\eta)\tilde{f}_X(x) = \frac{f_{U|X}(\eta) f_X(x)}{\mathrm{P}(Y > 0)} \qquad \text{for } \eta > -x\beta$$

Let f_{V_b} be the density of $Y - \max\{0, Xb\}$ conditional on $Y > 0$ and on X. Also let $Z = \max\{0, Xb\} - \max\{0, X\beta\}$ and

$$\breve{f}_{U|X}(\eta) = 1_{\{\eta > -\max\{0, X\beta\}\}} f_{U|X}\big(\eta + \max\{0, -X\beta\}\big).$$

Notice that this implies that $\breve{f}_{U|X}$ is maximized at $\eta = 0$ and is decreasing on R_+. We then have

$$\begin{aligned}
&E\big[\hat{f}_n(0;\beta)\big] - E\big[\hat{f}_n(0,b)\big] \\
= \;& \frac{1}{h_{n_p}} E_{X|Y>0}\Big[\int_{-\infty}^{\infty} K\Big(\frac{\eta}{h_n}\Big)\, f_{V_\beta}(\eta)\, d\eta - \int_{-\infty}^{\infty} K\Big(\frac{\eta}{h_n}\Big)\, f_{V_b}(\eta)\, d\eta\Big] \\
= \;& \frac{1}{h_{n_p}} E_{X|Y>0}\Big[\int_{0}^{\infty} K\Big(\frac{\eta}{h_n}\Big)\, \tilde{f}_{U|X}(\eta + \max\{0, X\beta\} - X\beta)\, d\eta \\
& - \int_{0}^{\infty} K\Big(\frac{\eta}{h_n}\Big)\, \tilde{f}_{U|X}(\eta + \max\{0, X\beta\} - X\beta + Z)\, d\eta\Big] \\
= \;& \frac{1}{\mathrm{P}(Y>0)} \frac{1}{h_{n_p}} E_X\Big[\int_{0}^{\infty} K\Big(\frac{\eta}{h_n}\Big)\, \big(1_{\{\eta > -\max\{0,X\beta\}\}} f_{U|X}(\eta + \max\{0, X\beta\} - X\beta) \\
& - 1_{\{\eta + Z > -\max\{0,X\beta\}\}} f_{U|X}(\eta + \max\{0, X\beta\} - X\beta + Z)\big)\, d\eta\Big] \\
= \;& \frac{1}{\mathrm{P}(Y>0)} E_X\Big[\int_{0}^{\infty} K\Big(\eta\Big)\, \big(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\, d\eta\Big].
\end{aligned}$$

By assumption there exists P compact and connected subsets of $\mathcal{X}$, $\Gamma_1, \Gamma_2, \ldots, \Gamma_P$, such that $\mathrm{P}(X\beta > 0 | X \in \Gamma_p) = 1$ for $p = 1, \ldots, P$.

Let $N(\beta)$ be given. Then there is an $\varepsilon > 0$ and a division of $\mathbf{B} \setminus N(\beta)$ into $2P$ compact subsets, $\mathbf{B}_{11}$, $\mathbf{B}_{12}, \ldots, \mathbf{B}_{1P}$, $\mathbf{B}_{21}$, $\mathbf{B}_{22}, \ldots, \mathbf{B}_{2P}$, such that

$$b \in \mathbf{B}_{1p} \Rightarrow xb - x\beta > \varepsilon \qquad \forall x \in \Gamma_p$$

$$b \in \mathbf{B}_{2p} \Rightarrow xb - x\beta < -\varepsilon \qquad \forall x \in \Gamma_p$$

Let $\gamma = \min_p \mathrm{P}(X \in \Gamma_p)$. Then $\gamma > 0$.

Let $b \in \mathbf{B}_{1p} \cup \mathbf{B}_{2p}$. We can then write

$$\begin{aligned}
&E\big[\hat{f}_n(0;\beta)\big] - E\big[\hat{f}_n(0,b)\big] \\
= \;& \frac{1}{\mathrm{P}(Y>0)} E_X\Big[\int_{0}^{\infty} K(\eta)\, \breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\, d\eta \Big| X \in \Gamma_p\Big]\ \mathrm{P}(X \in \Gamma_p) \\
& + \frac{1}{\mathrm{P}(Y>0)} E_X\Big[\int_{0}^{\infty} K(\eta)\, \big(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\, d\eta \Big| X \in \Gamma_p^c\Big]\ \mathrm{P}(X \in \Gamma_p^c) \qquad (3.20)
\end{aligned}$$

We will now study the two terms separately. If $b \in \mathbf{B}_{1p}$ then $Xb > X\beta + \varepsilon > \varepsilon$, so $Z > \varepsilon$. Let $\zeta = \min_{x \in \Gamma_p}\{x\beta\} > 0$. If $b \in \mathbf{B}_{2p}$ then $X\beta > Xb + \varepsilon$, and $X\beta > \zeta$, so $Z > \min\{\varepsilon, \zeta\}$. Without loss of generality assume that $\varepsilon < \zeta$.

We can then conclude from Lemma 2 of Appendix A that there exists an n_{1p}, such that for $n \geq n_{1p}$, the first term in (3.20) is greater than $5\mathrm{P}(\Gamma_p) f_{\max} g(|\varepsilon/2|)\big(7f_{\max} + g(|\varepsilon/2|)\big)^{-1}$.

To evaluate the second term of (3.20) notice that Lemma 1 implies that there exists a $\kappa > 0$ such that

$$\mathrm{P}(0 < |Z| < \kappa, X\beta > 2\kappa) < \frac{\mathrm{P}(\Gamma_p)}{7f_{\max} + g(|\varepsilon/2|)} \qquad \text{and} \qquad \mathrm{P}(|X\beta| \leq 2\kappa) < \frac{\mathrm{P}(\Gamma_p)}{7f_{\max} + g(|\varepsilon/2|)}$$

The second term of (3.20) can then be written as

$$\begin{aligned}
& E_X\left[\int_0^\infty K(\eta)\,\big(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\,d\eta \,\Big|\, X \in \Gamma_p^c\right] \mathrm{P}(X \in \Gamma_p^c) \\
= \quad & E_X\left[\int_0^\infty K(\eta)\,\big(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\,d\eta \,\Big|\, X \in \Gamma_p^c, |Z| \geq \kappa\right] \mathrm{P}(X \in \Gamma_p^c, |Z| \geq \kappa) \\
& + E_X\left[\int_0^\infty K(\eta)\,\big(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\,d\eta \,\Big|\, X \in \Gamma_p^c, |Z| = 0\right] \mathrm{P}(X \in \Gamma_p^c, |Z| = 0) \\
& + E_X\left[\int_0^\infty K(\eta)\,\big(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\,d\eta \,\Big|\, X \in \Gamma_p^c, 0 < |Z| < \kappa, X\beta > 2\kappa\right] \\
& \mathrm{P}(X \in \Gamma_p^c, 0 < |Z| < \kappa, X\beta > 2\kappa) \\
& + E_X\left[\int_0^\infty K(\eta)\,\big(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\,d\eta \,\Big|\, X \in \Gamma_p^c, 0 < |Z| < \kappa, -2\kappa \leq X\beta \leq 2\kappa\right] \\
& \mathrm{P}(X \in \Gamma_p^c, 0 < |Z| < \kappa, -2\kappa \leq X\beta \leq 2\kappa) \\
& + E_X\left[\int_0^\infty K(\eta)\,\big(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\big)\,d\eta \,\Big|\, X \in \Gamma_p^c, 0 < |Z| < \kappa, X\beta < -2\kappa\right] \\
& \mathrm{P}(X \in \Gamma_p^c, 0 < |Z| < \kappa, X\beta < -2\kappa)
\end{aligned}$$

By Lemma 2 there exists an n_{2p} (not depending on b) such that for $n \geq n_{2p}$ the first term of this is positive, whereas the second term is zero. The third and fourth terms are each greater than $-f_{\max}\mathrm{P}(\Gamma_p)\big(7f_{\max} + g(|\varepsilon/2|)\big)^{-1}$ by definition of κ. The fifth term is positive because $Z > 0$ and $\breve{f}_{U|X}$ is decreasing on R_+

This implies that for $n \geq \max\{n_{1p}, n_{2p}\}$ and $b \in \mathbf{B}_{1p} \cup \mathbf{B}_{2p}$

$$E\big[\hat{f}_n(0;\beta)\big] - E\big[\hat{f}_n(0,b)\big] \geq \frac{2\mathrm{P}(\Gamma_p) f_{\max} g(|\varepsilon/2|)}{\mathrm{P}(Y > 0)\big(7f_{\max} + g(|\varepsilon/2|)\big)} \geq \frac{2\gamma f_{\max} g(|\varepsilon/2|)}{7f_{\max} + g(|\varepsilon/2|)}$$

By the same argument that led to (3.15a) there exists an event of probability 1, such that that event implies that there exists an n_{3p} such that for $n \geq n_{3p}$

$$\sup_{b \in \mathbf{B}_{1p} \cup \mathbf{B}_{2p}} \{\hat{f}_n(0; b) - E\big[\hat{f}_n(0; b)\big]\} \leq \frac{\gamma f_{\max} g(|\varepsilon/2|)}{7 f_{\max} + g(|\varepsilon/2|)}$$

This implies that for $\geq \max_p \max\{n_{1p}, n_{2p}, n_{3p}\}$

$$\mathrm{P}\big(\lim_{n\to\infty} \hat{f}_n(\beta) - \sup_{b \in \mathbf{B} \setminus N(\beta)} \hat{f}_n(b) \geq \frac{\gamma f_{\max} g(|\varepsilon/2|)}{7 f_{\max} + g(|\varepsilon/2|)}\big) = 1.$$

This concludes the proof of Theorem 5. ■

4. Monte Carlo Evidence.

Having discussed the large sample properties of the estimators in the last section, it is natural to turn to their small sample properties. To partially characterize the small sample performance of the estimators and some of its alternatives, a small scale Monte Carlo study was performed. We consider seven different specifications, differing only in the conditional distribution of the error term conditional on the regressor.

Data from the seven models is generated by generating observations from the linear regression model

$$Y = \beta_0 + \beta_1 X + U$$

with $(\beta_0, \beta_1) = (0, 1)$ and where X has a uniform distribution, $X \sim \mathrm{U}(-\sqrt{3}, \sqrt{3})$. The conditional distribution for U conditional on X differs from experiment to experiment, but in all cases $\mathrm{E}[U] = 0$ and $\mathrm{V}[U] = 1$. Data is generated until there are 500 observations for which $Y > 0$. The observations for which $Y > 0$ will be the datasets used in the estimation. For each specification we will generate 400 datasets.

For each dataset we consider five different estimators of (β_0, β_1). The five estimators are ordinary least squares (OLS), normal maximum likelihood (MLE), the two mode estimators from the previous section (MOD1 and MOD2) and Powell's symmetrically trimmed least squares estimator[7] (STLS).

[7] Powell's estimator minimizes $\sum_{i=1}^{N} \big(y_i - \max\{\frac{1}{2} y_i, x\beta\}\big)^2$

The seven specifications of the conditional distribution of U given X are:

1. $U|X = x \sim \mathrm{N}(0, 1)$.
2. $U|X = x \sim$ LaPlace (double exponential) with mean 0 and variance 1.
3. $U|X = x \sim \mathrm{N}\big(0, \big(1 + (\sqrt{\tfrac{4}{3}} - 1)x\big)^2\big)$.
4. $U|X = x \sim \mathrm{N}\big(0, \big(1 - (\sqrt{\tfrac{4}{3}} - 1)x\big)^2\big)$.
5. $U|X = x \sim \mathrm{N}(0, x^2)$.
6. $U|X = x \sim \chi^2_3$ (normalized to have mean 0 and variance 1).
7. $U|X = x \sim -\chi^2_3$ (normalized to have mean 0 and variance 1).

For all of these specifications, the truncation of the sample excluded approximately 50% of the observations.

The grouping necessary for the first mode estimator is done by grouping the data into 5 groups of equal size according to the value of X.

Decisions about kernels and bandwidths have to be made in order to estimate (β_0, β_1) by the mode estimators. To calculate the first mode estimator we must estimate the mode (and therefore the density) in five subgroups each with 100 observations. We arbitrarily choose $h_n = 0.1$ in the estimation of the mode in each of the groups. To calculate the second mode estimator, we must estimate a density using all 500 observations. It is required that $h_n \to 0$ as $n \to \infty$, so it is natural to use a smaller h_n to estimate the second mode estimator. It is known that for estimation of densities, the optimal rate[8] of h_n is of order $n^{-1/5}$. We therefore choose $h_n = 0.072 \approx 0.1 \times 5^{-1/5}$.

The choice of kernels for the estimators is also arbitrary. The first mode estimator requires that the mode be estimated in each group. We use a LaPlace kernel because it has a spike at zero, and we therefore speculate that it will estimate the mode well. For the second mode estimator, we use a folded normal kernel.[9] This choice is motivated by the familiarity of the normal distribution. It is known that the integrated mean squared error of the density estimate is minimized when the kernel has the form of a second degree polynomial, but almost any reasonable kernel will give almost optimal results (Prakasa–Rao (1983) pp. 63-66). We therefore conjecture that the mode estimators will not be too sensitive to the choice of kernel.

---

[8] Note that it does not necessarily follow that this will imply an optimal rate for the mode estimators considered here.

[9] This will not satisfy the condition (3.5).

The results for specifications 1 to 7 are presented in Table 1 to Table 7.

For each specification we present the lower and upper quantiles of the estimator as well as its trimmed mean using only the observations between the 5^{th} and the 95^{th} percentile. For the estimator of β_1, we also present the trimmed mean squared error using the smallest 90% of the squared errors.[10] We do not present the latter for the estimator of β_0, as the estimators implicitly adopt different location normalizations. The reason for presenting the *trimmed* means and *trimmed* mean squared is that asymptotic properties such as consistency and asymptotic normality do not guarantee that the moments exist for any finite sample size.[11] It therefore seems reasonable to present summary statistics that do not rely on finite moments. In this Monte Carlo study it turned out that for some specifications, very extreme estimates were obtained in a few cases.

Focussing on the slope parameter, the most striking result of the Monte Carlo study is that OLS performs remarkably well in terms of trimmed mean squared error. OLS outperforms MLE except for specification one, when the likelihood is correctly specified, and specification seven. The performance of the symmetrically trimmed least squares estimator is very impressive for models with symmetric errors, but seems to be sensitive to deviations from symmetry, cf. Table 6 and Table 7. As a general statement, the mode estimators perform very poorly, primarily because they have high variance. If only biases are considered, the mode first estimator performs rather well. Its worst (trimmed) bias in the seven specifications is 0.5347, which is better that the worst (trimmed) bias for any of the other estimators.

As noted above, there is reason to believe that the mode estimators will not be too sensitive to the choice of kernel. It could however be very sensitive to the choice of h_n. To investigate this sensitivity, we re–estimate the model given by specification 1, using bandwidths that are half and twice as big as the ones used in Table 1. The results are reported in Table 8 and Table 9. As expected, it seems that a larger bandwidth gives a lower variance, but more bias. This is in particular true for the first mode regressor. For the second mode regressor, the (trimmed) mean squared error does not depend on strongly on h_n.

It is realistic to assume that the sample size is 500 when one is using only one regressor.

---

[10] The conclusions below are not sensitive to whether 90% or 95% were used.

[11] Compactness of the parameter space does imply the existence of the moments, but compactness is not enforced in practice.

In Tables 10, 11 and 12 we therefore present the results for specifications 1, 6 and 7 with 100 observations. The bandwidths used here are $h = 0.138$ and $h_n = 0.1$ for the two estimators, reflecting the smaller sample size.

5. Conclusion.

In this paper we have discussed the identifiability of the truncated regression model in the case where there is heteroskedasticity in the distribution of the errors. Neither mean nor median independence will suffice for identification, but mode independence will. The proof of identification under mode independence suggests an estimator for the case where there is mode independence and general heteroskedasticity. We prove the consistency of this estimator.

The motivation for the discussion of identification in this paper is very much inspired by the analysis in Manski (1988). It is our hope that the discussion in Section 2 has helped *"expose the foundations"* of the truncated regression model.

Having established identification, the Construction consistent estimators is primarily a technical problem. In Manski's (1988) words, *"the gap between identification and consistent estimability is 'smoothness'. Verification that an estimation problem is smooth in an appropriate sense is often a tedious, unenlightening task, ..."* Unfortunately the proofs of consistency in this paper are examples of this. The proofs do not give any new insight about the truncated regression model. It is our hope, however, that the proof strategy will be useful in consistency proofs in other cases where the objective function is defined in terms of a kernel estimator.

Appendix A: Lemmas.

LEMMA 1. *Assume that $\mathcal{X}$ is compact and that X is continuously distributed over $\mathcal{X}$. Further, assume that $\mathbf{B} \subset R^P$ is compact and that $N(\beta)$ is an open neighborhood of β. Then*

$$\forall\alpha > 0\ \exists\kappa > 0\ :\ \mathrm{P}\big(0 < |\max\{0, Xb\} - \max\{0, X\beta\}| < \kappa, X\beta > 2\kappa\big) < \alpha \quad \forall b \in \mathbf{B} \setminus N(\beta)$$

Proof: Notice that

$$\mathrm{P}\big(0 < |\max\{0, Xb\} - \max\{0, X\beta\}| < \kappa, X\beta > 2\kappa\big) \leq \mathrm{P}\big(|X(b-\beta)| < \kappa\big)$$

Since b is bounded away from β, the right hand side can be made arbitrarily close to zero which proves the result. ■

LEMMA 2. *Make the assumptions of Theorem 5. Let $b \in \mathbf{B}$ and $\varepsilon \neq 0$ be given and let Z and $\breve{f}_{U|X}$ be defined as in the proof of Theorem 5. If A is a subset of $\mathcal{X}$ such that either $Z > 2\varepsilon > 0$ for all $X \in \mathcal{X}$ or $Z < 2\varepsilon < 0$ for all $X \in \mathcal{X}$, then there exists an n_1 not depending on b such that for $n > n_1$,*

$$E_X\left[\int_0^\infty K(\eta)\,\left(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\right)\,d\eta\middle|X \in A\right] \geq \frac{5f_{\max}g(|\varepsilon|)}{\left(7f_{\max} + g(|\varepsilon|)\right)}$$

Proof: Without loss of generality assume that $|\varepsilon| \leq \nu_0$. Let M be such that

$$\int_0^M K(\eta)\,d\eta > \frac{f_{\max}}{\left(f_{\max} + g(|\varepsilon|)/7\right)}$$

and let $\bar{f}_U = E[f_{U|X}|X \in A]$. $\bar{f}_U(\eta)$ will then be continuous for $\eta > 0$, so there will exist an $\varepsilon_1 > 0$ such that $\bar{f}_U(\eta) > \bar{f}_U(0) - g(|\varepsilon|)/7$ for all $\eta < \varepsilon_1$ Let n_1 be big enough that for $n \geq n_1$, $Mh_n < \min\{\varepsilon_1|\varepsilon|\}$. Then

$$\begin{aligned}
&E_X\left[\int_0^\infty K(\eta)\,\left(\breve{f}_{U|X}(h_n\eta) - \breve{f}_{U|X}(h_n\eta + Z)\right)\,d\eta\middle|X \in A\right] \\
&\geq E_X\left[\int_0^\infty K(\eta)\,\breve{f}_{U|X}(h_n\eta)\,d\eta - \int_0^M K(\eta)\,\breve{f}_{U|X}(h_n\eta + Z)\,d\eta - \int_M^\infty K(\eta)\,f_{\max}\,d\eta\middle|X \in A\right] \\
&\geq E_X\left[\int_0^\infty K(\eta)\,\breve{f}_{U|X}(h_n\eta)\,d\eta - \int_0^M K(\eta)\,\left(\breve{f}_{U|X}(0) - g(\varepsilon|)\right)\,d\eta - \int_M^\infty K(\eta)\,f_{\max}\,d\eta\middle|X \in A\right] \\
&\geq E_X\left[\int_0^M K(\eta)\,\breve{f}_{U|X}(h_n\eta)\,d\eta - \int_0^M K(\eta)\,\left(\breve{f}_{U|X}(0) - g(\varepsilon|)\right)\,d\eta - \int_M^\infty K(\eta)\,f_{\max}\,d\eta\middle|X \in A\right] \\
&= \int_0^M K(\eta)\,\bar{f}(h_n\eta)\,d\eta - \int_0^M K(\eta)\,\left(\bar{f}(0) - g(\varepsilon|)\right)\,d\eta - f_{\max}\int_M^\infty K(\eta)\,d\eta \\
&\geq \frac{6}{7}g(|\varepsilon|)\int_0^M K(\eta)\,d\eta - f_{\max}\int_M^\infty K(\eta)\,d\eta \\
&\geq \frac{6}{7}g(|\varepsilon|)\frac{f_{\max}}{\left(f_{\max} + g(|\varepsilon|)/7\right)} - f_{\max}\frac{g(|\varepsilon|)/7}{\left(f_{\max} + g(|\varepsilon|)/7\right)} \\
&= \frac{5g(|\varepsilon|)f_{\max}}{\left(7f_{\max} + g(|\varepsilon|)\right)} \qquad \blacksquare
\end{aligned}$$

Appendix B: Additional Information about the Monte Carlo Study.

The random number generator used in the Monte Carlo study is from the NAG library. It is based on the a multiplicative congruential method and its cycle length is 2^{57}.

The optimization is done using the Simplex method from the NAG library. In the calculation of MOD1, the OLS estimates were used as starting values. Upon convergence, the optimization is restarted from its current values.

Local optima presented a problem in the calculation of MOD2. Five different sets of starting values were therefore used: $(\hat{b}_0^{\rm OLS}, \hat{b}_1^{\rm OLS})$, $(\frac{1}{2}\hat{b}_0^{\rm OLS}, \frac{1}{2}\hat{b}_1^{\rm OLS})$, $(\frac{1}{2}\hat{b}_0^{\rm OLS}, 2\hat{b}_1^{\rm OLS})$, $(2\hat{b}_0^{\rm OLS}, \frac{1}{2}\hat{b}_1^{\rm OLS})$, and $(2\hat{b}_0^{\rm OLS}, 2\hat{b}_1^{\rm OLS})$. Upon convergence, the optimization is restarted from the "best" optimum.

The calculation were performed on an AST 286 and on an Everex 386.

Acknowledgements.

The research reported in this paper is supported by the NSF Grant No SES–8809352. I would like to thank the participants of the Northwestern University Econometrics Workshop for valuable comments and Bruce Meyer for many discussions related to this paper.

References.

Ichimura H. (1988): "Estimation of Single Index Models," unpublished.

Lee, M. (1988a): "Mode Regression," unpublished.

Lee, M. (1988b): "Quadratic Kernel Mode Regression," unpublished.

Manski, C. (1988): "Identification of Binary Response Models", *Journal of the Americal Statistical Association*, **83**, pp. 729–738.

Newey, N. K. (1986): "Interval Estimation of the Truncated Regression Model," unpublished.

Powell, J. (1986): "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, **54**, pp. 1435–1460.

Prakasa–Rao, B. L. S. (1983): *Nonparametric Functional Estimation*. Orlando: Academic Press.

æ

Table 1: Results of Simulations. N = 500. 400 Replications. $h = 0.10$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim \mathrm{N}(0,1)$

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.4340	0.3881	0.4789	1.1712	0.0313	1.1014	1.2448
MLE	0.0109	−0.0932	0.1237	0.9964	0.0071	0.9282	1.0570
MOD1	0.3406	0.1375	0.5615	0.7029	0.1567	0.4441	0.9301
MOD2	−0.0115	−0.2407	0.3351	0.9755	0.1623	0.5989	1.2387
STLS	−0.0625	−0.2986	0.2377	1.0514	0.0531	0.8342	1.2343

Table 2: Results of Simulations. N = 500. 400 Replications. $h = 0.10$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim \mathrm{LaPlace}(0,1)$

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.4354	0.3731	0.4961	1.0541	0.0116	0.9720	1.1362
MLE	−0.3881	−0.5829	−0.1825	1.1210	0.0297	0.9862	1.2538
MOD1	0.1104	−0.0217	0.2392	0.9010	0.0331	0.7591	1.0439
MOD2	−0.0280	−0.1704	0.1565	0.9626	0.0536	0.7594	1.1127
STLS	−0.0061	−0.1056	0.0955	1.0063	0.0110	0.9160	1.0992

Table 3: Results of Simulations. N = 500. 400 Replications. $h = 0.10$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim \mathrm{N}\big(0, \big(1 + (\sqrt{\tfrac{4}{3}} - 1)x\big)^2\big)$

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.3704	0.3195	0.4243	1.3163	0.0963	1.2212	1.4048
MLE	−0.4240	−0.5809	−0.2589	1.3629	0.1270	1.2492	1.4604
MOD1	0.3352	0.1676	0.5359	0.7680	0.1815	0.4908	1.0103
MOD2	−0.0024	−0.2782	0.3285	1.0027	0.1398	0.6712	1.3053
STLS	−0.0210	−0.2740	0.3103	0.9970	0.0887	0.7360	1.1987

Table 4: Results of Simulations. N = 500. 400 Replications. $h = 0.10$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim \mathrm{N}\big(0, \big(1 - (\sqrt{\tfrac{4}{3}} - 1)x\big)^2\big)$

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.5169	0.4742	0.5598	0.9966	0.0059	0.9352	1.0660
MLE	0.3264	0.2526	0.4067	0.7000	0.0856	0.6365	0.7555
MOD1	0.4504	0.2068	0.6652	0.6483	0.1808	0.4257	0.9051
MOD2	0.0654	−0.1360	0.3730	0.8851	0.1252	0.5958	1.1248
STLS	−0.0121	−0.2087	0.1994	1.0053	0.0295	0.8510	1.1455

Table 5: Results of Simulations. N = 500. 400 Replications. $h = 0.10$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim \mathrm{N}(0, x^2)$

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.0753	0.0186	0.1356	1.5839	0.3232	1.4838	1.6753
MLE	−3.6020	−4.1863	−3.0136	3.3363	5.1980	3.0214	3.6450
MOD1	−0.2104	−0.4265	0.0325	1.2142	0.1766	0.8194	1.5568
MOD2	−0.6864	−1.2722	−0.1130	1.6633	0.7037	1.0146	2.2534
STLS	−0.1871	−0.3561	0.0377	1.1164	0.0924	0.9174	1.3053

Table 6: Results of Simulations. N = 500. 400 Replications. $h = 0.10$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim \chi_3$ with mean 0 and variance 1.

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.7192	0.6526	0.7941	0.6496	0.1183	0.5580	0.7466
MLE	−2.9301	−3.9415	−1.7807	1.2627	0.1518	0.9337	1.5560
MOD1	0.0785	−0.0731	0.3466	0.4653	0.4453	0.1904	0.6227
MOD2	−0.5154	−0.9374	0.0294	0.8182	0.2550	0.4065	1.1727
STLS	−1.5722	−2.0601	−1.0273	1.7872	0.6622	1.4080	2.1255

Table 7: Results of Simulations. N = 500. 400 Replications. $h = 0.10$.
$\beta_0 = 0$, $\beta_1 = 1$, $\varepsilon \sim -\chi_3$ with mean 0 and variance 1.

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.1780	0.1379	0.2197	1.4306	0.1762	1.3719	1.4888
MLE	0.0304	−0.0499	0.0916	1.1847	0.0372	1.1498	1.2409
MOD1	0.6787	0.5714	0.8066	0.9631	0.0265	0.8378	1.0770
MOD2	0.4723	0.2857	0.6685	1.1285	0.0558	0.9324	1.3169
STLS	0.5847	0.5544	0.6202	0.7458	0.0614	0.7035	0.7839

Table 8: Results of Simulations. N = 500. 400 Replications. $h = 0.05$.
$\beta_0 = 0$, $\beta_1 = 1$, $\varepsilon \sim \mathrm{N}(0, 1)$

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.4340	0.3881	0.4789	1.1712	0.0313	1.1014	1.2448
MLE	0.0109	−0.0932	0.1237	0.9964	0.0071	0.9282	1.0570
MOD1	0.3695	0.1205	0.6210	0.7262	0.2377	0.3960	1.0306
MOD2	0.1006	−0.1131	0.4166	0.9001	0.1761	0.5707	1.1992
STLS	−0.0625	−0.2986	0.2377	1.0514	0.0531	0.8342	1.2343

Table 9: Results of Simulations. N = 500. 400 Replications. $h = 0.20$.
$\beta_0 = 0$, $\beta_1 = 1$, $\varepsilon \sim \mathrm{N}(0, 1)$

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.4340	0.3881	0.4789	1.1712	0.0313	1.1014	1.2448
MLE	0.0109	−0.0932	0.1237	0.9964	0.0071	0.9282	1.0570
MOD1	0.3707	0.1764	0.5566	0.6942	0.1441	0.4710	0.9214
MOD2	−0.1748	−0.4491	0.2296	1.0321	0.1608	0.6496	1.2845
STLS	0.0625	−0.2986	0.2377	1.0514	0.0531	0.8342	1.2343

Table 10: Results of Simulations. N = 100. 400 Replications. $h = 0.138$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim \mathrm{N}(0,1)$

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.4530	0.3561	0.5566	1.1331	0.0485	0.9599	1.2991
MLE	−0.0082	−0.2399	0.2242	0.9964	0.0387	0.8297	1.1406
MOD1	0.4013	0.1844	0.6756	0.6326	0.3516	0.2675	0.9755
MOD2	0.1130	−0.1209	0.4702	0.9064	0.2119	0.5385	1.2295
STLS	−0.1052	−0.4087	0.4173	1.0289	0.2510	0.6346	1.3305

Table 11: Results of Simulations. N = 100. 400 Replications. $h = 0.138$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim \chi_3$ with mean 0 and variance 1.

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.7187	0.5483	0.8754	0.6531	0.1621	0.4510	0.8762
MLE	−3.1764	−4.0906	−1.0168	1.3130	0.4970	0.6794	1.6826
MOD1	0.2019	−0.0804	0.5478	0.5766	0.5270	0.1547	0.9059
MOD2	−0.2550	−0.6629	0.2591	0.7007	0.3260	0.3281	1.0062
STLS	−1.2513	−1.8692	−0.3157	1.5017	0.7928	0.8351	2.0664

Table 12: Results of Simulations. N = 100. 400 Replications. $h = 0.138$.
$\beta_0 = 0,\ \beta_1 = 1,\ \varepsilon \sim -\chi_3$ with mean 0 and variance 1.

	Intercept			Slope			
	TMean	LQ	UQ	TMean	TMSE	LQ	UQ
OLS	0.1995	0.1136	0.2971	1.3898	0.1535	1.2510	1.5314
MLE	0.0521	−0.0688	0.1517	1.1683	0.0417	1.0797	1.2597
MOD1	0.5422	0.3532	0.7542	0.9134	0.0942	0.6753	1.1672
MOD2	0.4231	0.2006	0.6568	1.1017	0.0756	0.8715	1.3433
STLS	0.5338	0.4490	0.6486	0.7850	0.0628	0.6646	0.8714