NONPARAMETRIC CENSORED AND TRUNCATED REGRESSION*

Arthur Lewbel — Boston College

Oliver Linton — London School of Economics

December 19, 2000

Abstract

This paper proposes new estimators of the latent regression function in nonparametric censored and truncated regression models. Our estimators are computationally convenient, consisting only of two nonparametric regressions and a univariate integral. We establish consistency and asymptotic normality for an implementation based on local linear kernel estimators. An extension permits estimation in the presence of a general form of heteroscedasticity.

1 Introduction

Consider the censored regression model $Y_i = \max[c, m(X_i) - e_i]$, where X_i is an observed d vector of regressors X_{ki} for $k = 1, \ldots, d$, and e_i is an unobserved error term that is independent of X_i (writing the model as $m - e$ instead of the more usual $m + e$ simplifies later results). We will also later consider errors having general forms of heteroscedasticity. In the case where $E(e_i) = 0$, the function m equals the regression function of the uncensored population. We assume that the censoring point c is a known constant, which we can take to be zero without loss of generality, by subtracting c from Y_i and from $m(X_i)$.

A common economic example of fixed censoring is where Y_i is observed purchases, which may either be censored from above by rationing, or censored from below by zero if consumers can only buy but not sell the product.

*We would like to thank Yan Li for excellent research assistance, and Joel Horowitz, Andrew Chesher, the associate editor and some referees for helpful comments. This research was supported in part by the National Science Foundation through grants SBR-9514977, SES-9905010, and SBR-9730282.

Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. Phone: (617) 552-3678. E-mail lewbel@bc.edu.

Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: lintono@lse.ac.uk.

Both the function $m(\cdot)$ and the distribution $F(\cdot)$ of the error e are unknown. The errors are not assumed to be symmetric. This paper provides a simple consistent estimator of $m(x) + k$ for some constant k. Under an additional tail condition, we provide a consistent estimator of $m(x)$ when m is the conditional mean function of the uncensored population. Also, we show that $F(\cdot)$ can be estimated given $m(x)$.

The proposed estimator is extended to deal with the truncated regression model, where Y_i is only observed when it is not censored. We also describe extensions to deal with a general form of heteroskedasticity, in which the distribution of e could depend in unknown ways on a subset of elements of x.

For any continuously distributed element x_k of x, let $m_k(x) = \partial m(x)/\partial x_k$. We also provide direct estimators of the derivatives $m_k(x)$ in both the censored and truncated regression models. These derivatives are interpretable as the marginal effect of a change in x on the underlying uncensored population. They can also be used to test or estimate parametric or semiparametric specifications of $m(x)$. For example, $m_k(x)$ is constant if $m(x)$ is linear in x_k, and $m_k(x)$ depends only on x_k if $m(x)$ is additive in a function of x_k. Rate root n converging estimates of a weighted average of $m_k(x)$ can be constructed, and used as estimates of the coefficients in a partly linear specification of $m(x)$.

Parametric and semiparametric estimators of censored or truncated regression models include Amemiya (1973), seminal Heckman (1976), Buckley and James (1979), Koul, Suslara, and Van Ryzin (1981), Powell (1984), (1986a), (1986b), Duncan (1986), Fernandez (1986), Horowitz (1986,1988), Moon (1989), Powell, Stock and Stoker (1989), Nawata (1990), Ritov (1990) Ichimura (1993), Honoré and Powell (1994), Lewbel (1998a, 2000), Buchinsky and Hahn (1998). Unlike the present paper, most of these models either assume $m(x) = \beta' x$ or some other parametric form, or they provide estimates of average derivatives only up to an unknown scale, or they assume that the error distribution is parametric. The fully nonparametric $m(x)$ model we consider is important because of the sensitivity of the parametric and semiparametric estimators to misspecification of functional form.

A small number of estimators exist for nonparametric censored regression models, in most cases focusing on the case where c is a random censoring point independent of X (which is a model adopted in many medical applications). Fan and Gijbels (1994) proposed a nonparametric censored regression estimator based on a local version of Buckley and James (1979). While this estimator is consistent when the censoring point is drawn from a continuous distribution, it is inconsistent in our situation of fixed censoring. This is because it relies on the existence of uncensored observations which are smaller than a given censored observation. This can not happen when censored observations always take the same value (zero in our case). It is not known if nonparametric Buckley James estimators can be constructed that are consistent under fixed censoring.

Other possible nonparametric censored regression estimators are based on quantile regressions. See, e.g., Fan and Gijbels (1996, pp 200-203) for definitions and references, Dabrowska (1995) for combining quantiles, or Chaudhuri (1991) for local polynomial quantile regression, and Chen and Khan (2000). Let $\rho(x)$ denote the proportion of observations that are censored at point $X = x$, and let $\alpha_q = \theta_q(e|X = x)$ denote the q'th conditional quantile of e, which is constant with respect to x when e is independent of X. Then $\theta_q(Y|X = x) = m(x) + \alpha_q$ when $q < 1 - \rho(x)$, and therefore a q'th quantile regression of Y on

X can used to estimate $m(x)$ (up to a constant α_q) but *only* if $q < 1 - \rho(x)$. The difficulty with using quantile methods is that, at each point x, only quantiles q that are less than $\rho(x)$ can be used to estimate $m(x)$. Notice that quantiles at different values of x (such as those where there is little censoring) provide information about α_q but, unlike for parametric models, cannot be used or combined to help estimate $m(x)$. Therefore, extreme quantiles may be required if some values of x result in heavy censoring.

Our estimator converges at the same rate as nonparametric quantiles, and will typically be more efficient in applications where the errors are thin tailed, or where heavy censoring would require the use of extreme quantiles. Our estimator also has standard errors that are easier to compute than quantile standard errors, because the latter depend on estimates of the error density in the tails.

Unlike censored regression, we do not know of any existing estimator, other than the one we propose here, for the nonparametric truncated regression model. However, it is likely that alternative estimators for either censored or truncated models could be obtained by taking existing semiparametric estimators that assume a known functional form for $m(x)$, and replacing that functional form assumption with a polynomial expansion. An advantage of our estimator over these actual and potential alternatives is its simplicity, and its known limiting distribution. Also, we will show that our estimator can be extended to deal with some very general forms of heteroskedasticity.

Our proposed estimators employ a novel technique of first nonparametrically regressing Y on X, then regressing a different function of Y on the fits of this first stage. It is likely that this new methodology will be applicable to other contexts where identification can be based on differential expressions involving index functions or latent variables.

2 The Censored Regression Function and its Derivatives

We will suppose that the following condition holds.

ASSUMPTION A1. Suppose that $Y^* = m(X) - e$ and we observe X and $Y = I(Y^* \geq 0)Y^*$, where I is the indicator function that equals one if its argument is true and zero otherwise. The $d \times 1$ random vector X can contain both discrete and continuously distributed elements; let its support be Ω. The function m is differentiable and has finite derivatives $m_k(x) = \partial m(x)/\partial x_k$ with respect to the elements x_k of x that are continuously distributed, for all $x \in \Omega$. The error e is independent of x, with absolutely continuous distribution function $F(e)$ and Lebesgue density function $f(e)$. Let Ω_e be the support of e.

We assume also that the observed data are independent, identically distributed observations (Y_i, X_i) for $i = 1, \ldots, n$, although our main results, Theorems 1-4, under reasonable conditions hold as stated when $\{Y_i, X_i\}$ is a stationary mixing process with $\{e_i\}$ independent of $\{X_i\}$, as in Robinson (1982).

Define the following functions:

$$\begin{aligned}
\mathfrak{F}_0(m) &= F(m) \\
\mathfrak{F}_\kappa(m) &= \int_{-\infty}^{m} \mathfrak{F}_{\kappa-1}(e)de, \ \kappa = 1, 2, \ldots
\end{aligned}$$

Theorem 1 *Let Assumption A1 hold. For any nonnegative integer* κ, *if* $\mathfrak{F}_\kappa[m(x)]$ *exists and* $\lim_{e\to-\infty} e^\kappa F(e) = 0$, *then*

$$E[Y^\kappa I(Y > 0)|X = x] = \kappa!\mathfrak{F}_\kappa[m(x)]. \tag{1}$$

PROOF. Since the conditional distribution of $Y|X = x$ only depends on x through $m(x)$, we have $E[Y^\kappa I(Y > 0)|X = x] = E[Y^\kappa I(Y > 0)|m(X) = m(x)]$. For $\kappa > 0$

$$\begin{aligned}
\frac{\partial E[Y^\kappa I(Y > 0)|m(X) = m(x)]}{\partial m(x)} &= \frac{\partial \int_{-\infty}^{m(x)}[m(x) - e]^\kappa f(e)de}{\partial m(x)} \\
&= \int_{-\infty}^{m(x)} \kappa[m(x) - e]^{\kappa-1} f(e)de \\
&= \kappa E[Y^{\kappa-1} I(Y > 0)|m(X) = m(x)],
\end{aligned}$$

and $\lim_{e\to-\infty} E[Y^\kappa I(Y > 0)|m(X) = e] = 0$, so $E[Y^\kappa I(Y > 0)|m(X) = e] = \int_{-\infty}^{e} \kappa E[Y^{\kappa-1} I(Y > 0)|m(X) = e]de$. The result can now be proved by induction. For $\kappa = 0$ we have $E[I(Y > 0)|X = x] = \Pr[e < m(x)] = F[m(x)] = \mathfrak{F}_0[m(x)]$, and assuming that the theorem holds for $\kappa - 1$, we have $E[Y^\kappa I(Y > 0)|m(X) = e] = \int_{-\infty}^{e} \kappa E[Y^{\kappa-1} I(Y > 0)|m(X) = e]de = \int_{-\infty}^{e} \kappa(\kappa - 1)!\mathfrak{F}_{\kappa-1}(e)de = \kappa!\mathfrak{F}_\kappa(e)$. ■

Equation (1) has long been known for the special case of $m(x) = \beta' x$ and $\kappa = 1$. See, e.g., Rosett and Nelson (1975), Heckman (1976), McDonald and Moffitt (1980), and Horowitz (1986). Theorem 1 shows that this expression holds for arbitrary m, F, and integers κ, and so can be exploited for nonparametric estimation of $m(x)$.

Define the following functions:

$$\begin{aligned}
r(x) &= E(Y|X = x), \quad r_k(x) = \frac{\partial r(x)}{\partial x_k} \\
s(x) &= E[I(Y > 0)|X = x], \quad s_k(x) = \frac{\partial s(x)}{\partial x_k} \\
q(r) &= E[I(Y > 0)|r(X) = r],
\end{aligned}$$

where x_k is the k'th element of x. The function q is only defined on the support of r, but we continue it beyond the support by setting it constant [and equal to the value at the corresponding end of the support] elsewhere. Under A1, the function $\mathfrak{F}_1$ is invertible on the set $[\inf_{e\in\Omega_e} e, \infty)$ with range $[0, \infty)$. Let $\mathfrak{F}_1^{-1}$ denote the inverse function of $\mathfrak{F}_1$, which is well-defined on $[0, \infty)$. Let λ_0 be any nonnegative constant, and let $\lambda_r = \sup_{x\in\Omega} r(x)$. If $\lambda_0 < r(x)$, then integrals of the form $\int_{r(x)}^{\lambda_0}$ below are to be interpreted as $-\int_{\lambda_0}^{r(x)}$.

Theorem 2 *Let Assumption A1 hold and suppose that* $\lim_{e\to-\infty} eF(e) = 0$. *Then for all* $x \in \Omega$, $r(x) = \mathfrak{F}_1[m(x)]$, $s(x) = F[m(x)]$, *and* $q(r(x)) = F\left(\mathfrak{F}_1^{-1}[r(x)]\right)$. *Also, for all* $x \in \Omega$ *having* $F[m(x)] \neq 0$, *we have*

$$m(x) + k = \lambda_0 - \int_{r(x)}^{\lambda_0} \frac{1}{q(r)}dr, \tag{2}$$

for some location constant $k(\lambda_0)$. Furthermore, for each continuously distributed element X_k of X,

$$m_k(x) = \frac{r_k(x)}{s(x)}. \tag{3}$$

PROOF. The equations for r, s, and q follow from Theorem 1. First, suppose that $\lambda_0 \leq \lambda_r$. Then using the change of variables $r = \mathfrak{F}_1(m)$, $dr = F(m)dm$, and the fact that $q(r) = F\left(\mathfrak{F}_1^{-1}[\mathfrak{F}_1(m)]\right) = F(m)$, we obtain $\lambda_0 - \int_{r(x)}^{\lambda_0}[1/q(r)]dr = \lambda_0 - \int_{\mathfrak{F}_1^{-1}(\mathfrak{F}[m(x)])}^{\mathfrak{F}_1^{-1}(\lambda_0)}[1/F(m)]F(m)dm = \lambda_0 - \int_{m(x)}^{\mathfrak{F}_1^{-1}(\lambda_0)} 1dm = \lambda_0 - \mathfrak{F}_1^{-1}(\lambda_0) + m(x)$, so equation (2) holds with $k = \lambda_0 - \mathfrak{F}_1^{-1}(\lambda_0)$. If $\lambda_0 > \lambda_r$, we write $\int_{r(x)}^{\lambda_0}[1/q(r)]dr = \int_{r(x)}^{\lambda_r}[1/q(r)]dr + \int_{\lambda_r}^{\lambda_0}[1/q(r)]dr = \int_{r(x)}^{\lambda_r}[1/q(r)]dr + k_1 = m(x) + [\lambda_r - \mathfrak{F}_1^{-1}(\lambda_r) + k_1]$ for some constant $k_1(\lambda_0)$ depending on λ_0 and on the constant value of q on the range $[\lambda_r, \lambda_0]$. Finally, $r_k(x) = \partial\mathfrak{F}_1[m(x)]/\partial x_k = F[m(x)]m_k(x) = s(x)m_k(x)$.

■

A general concern in latent variable models is the extent to which identification is based on information in the tails of the data. This applies particularly to estimation of the location or intercept. See, e.g., Andrews and Schafgans (1998). In Theorem 2, the derivatives $m_k(x)$ are identified locally, since $m_k(x) = r_k(x)/s(x)$, and both $r_k(x)$ and $s(x)$ are estimated just using data in the neighborhood of x. Similarly, $m(x)$ itself is identified up to the arbitrary location constant k without tail data, since equation (2) only depends on a range of X values that is large enough to obtain the function $r(X)$ everywhere in the interval from $r(x)$ to λ_0.

Let $\widehat{r}(x)$ be a kernel or other nonparametric regression of Y on X, let $\widehat{s}(x)$ be a nonparametric regression of $I(Y > 0)$ on X, and let $\widehat{q}(r)$ be a nonparametric regression of $I(Y > 0)$ on $\widehat{r}(X)$. Many such nonparametric estimators $\widehat{r}$ and $\widehat{s}$ have been shown to be uniformly consistent. Let λ_0 be some constant, and let $\widehat{m}(x) = \lambda_0 - \int_{\widehat{r}(x)}^{\lambda_0}[1/\widehat{q}(r)]dr$. Then, based on Theorem 2, $\widehat{m}(x)$ is a consistent estimator of $m(x) + k(\lambda_0)$ for some constant k, and $\widehat{r_k}(x)/\widehat{s}(x)$ is a consistent estimator of $m_k(x)$. We will later provide their limiting normal distributions.

Theorem 2 does not require the error distribution to have a finite mean [because the upper tail of the error distribution is unrestricted]. However, if the errors have mean zero, then the following theorem shows identification (using tail data) of the location of $m(x)$.

Theorem 3 *Let Assumption A1 hold and suppose that $\lim_{e\to-\infty} eF(e) = 0$ and that $E(e) = 0$. Let $\lambda_e = \sup_{e\in\Omega_e} e$, and suppose that $\lambda_r \geq \lambda_e$. Then for all $\lambda_0 \geq \lambda_r$ and for all $x \in \Omega$ having $F[m(x)] \neq 0$*

$$m(x) = \lambda_0 - \int_{r(x)}^{\lambda_0} \frac{1}{q(r)}dr. \tag{4}$$

PROOF. By the tail condition, $q(\lambda_r) = 1$ and we have $q(r) = 1$ for all $r \geq \lambda_r$. Therefore, $\lambda_0 - \int_{r(x)}^{\lambda_0}[1/q(r)]dr = \lambda_r - \int_{r(x)}^{\lambda_r}[1/q(r)]dr$. By Theorem 2, $\lambda_r - \int_{r(x)}^{\lambda_r}[1/q(r)]dr = \lambda_r - \mathfrak{F}_1^{-1}(\lambda_r) + m(x)$. Then, using an integration by parts, $E(e) = 0 = \int_{-\infty}^{\lambda_e} ef(e)de = -\int_{-\infty}^{\lambda_e}[F(e) - I(e > 0)]de = -\mathfrak{F}_1(\lambda_e) + \lambda_e$, so $\mathfrak{F}_1(\lambda_e) = \lambda_e$. Finally, since $F(m) = 1$ for all $m \geq \lambda_e$, we have $\mathfrak{F}_1(\lambda_r) = \lambda_r$.

■

The assumption that $\lambda_e \leq \lambda_r$ is equivalent to requiring that the censoring probability for any e is less than 100%, and it implies that $\lambda_0 - \mathfrak{F}_1^{-1}(\lambda_0) = 0$ for any $\lambda_0 \geq \lambda_r$, so that if such a λ_0 is chosen in $\widehat{m}(x) = \lambda_0 - \int_{\widehat{r}(x)}^{\lambda_0}[1/\widehat{q}(r)]dr$, it will converge to $m(x)$. In practice, we may want to replace λ_0 by some estimate of the upper bound like $\widehat{\lambda}_0 = \widehat{\lambda}_r = \max_{i=1,\ldots,n}\widehat{r}(X_i)$ (which converges to λ_r under general conditions) or let $\widehat{\lambda}_0$ be some large fixed number that is known to lie above λ_r. We do not need to consistently estimate λ_r, all that is required is a $\widehat{\lambda}_0$ that is greater than or equal to $\widehat{\lambda}_r$ with probability tending to one.

If the probability of 100% censoring is small but not equal to zero, then $\lambda_r - \mathfrak{F}_1^{-1}(\lambda_r)$ will be small, so the asymptotic location of this $\widehat{m}(x)$ will still yield close to mean zero latent errors. This is illustrated later in a Monte Carlo study

Since $s(x) = q[r(x)]$, an alternative derivative estimator would be $m_k(x) = \widehat{r}_k(x)/\widehat{q}[\widehat{r}(x)]$, which might have different small sample behavior. Note also that, given this expression for $m_k(x)$, our integral expression for $m(x)$ could be derived from $\int[1/q(r)]dr = \int r_k(x)/q[r(x)]dx_k$, using a change of variables from x_k to r for each k.

2.1 The Error Distribution

For any e^*, $E[I(Y > 0)|m(X) = e^*] = F(e^*)$, where F is the distribution function of the errors e. Therefore, given the estimated regression function $\widehat{m}(x)$, the distribution function F can be estimated as a nonparametric regression of $I(Y_i > 0)$ on $\widehat{m}(X_i)$. Lemma 1 in Lewbel (1997) can then be used to directly estimate the variance and other moments of e, and this estimate of F can differentiated to provide an estimate of the density of e. An alternative (less smooth) estimate of F is the Kaplan-Meier estimate based on the residuals $\widehat{e}_i = Y_i - \widehat{m}(X_i)$. Let $\widehat{e}_{(i)}$ be the i^{th} largest residual and let $\delta_{(i)} = 0$ when observation $Y_{(i)}$ is censored, and $\delta_{(i)} = 1$ otherwise. Then let

$$\widehat{F}(e) = 1 - \prod_{i:\widehat{e}_{(i)} \leq e} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}. \tag{5}$$

3 Nonparametric Truncated Regression

This section shows how $m(x)$ and its derivatives $m_k(x)$ can be estimated in a nonparametric truncated regression model. The nonparametric truncated regression model is identical to the nonparametric censored regression model, except that data are only observed when $Y > 0$.

Define the following functions:

$$\begin{aligned} R(x) &= E(Y|X=x, Y>0), \quad R_k(x) = \frac{\partial R(x)}{\partial x_k} \\ T(x) &= E(Y^2/2|X=x, Y>0), \quad T_k(x) = \frac{\partial T(x)}{\partial x_k} \end{aligned}$$

$$
\begin{aligned}
U[R] &= E[(Y^2/2)|R(X) = R, Y > 0], \quad U'(R) = \frac{\partial U(R)}{\partial R} \\
\widetilde{R}(m) &= \mathfrak{F}_1(m)/F(m),
\end{aligned}
$$

where x_k is the k'th element of x. The function U is only defined on the support of R, but we continue it beyond the support by setting it constant [and equal to the value at the corresponding end of the support] elsewhere. In assumption A1* below we assume that the function $\widetilde{R}$ is invertible on the set $(\inf_{e \in \Omega_e} e, \infty)$ with range $(0, \infty)$. Let $\widetilde{R}^{-1}$ denote the inverse function of $\widetilde{R}$, which is well-defined on $(0, \infty)$.Let $\widetilde{R}^{-1}$ denote the inverse function of $\widetilde{R}$, and let $\lambda_R = \sup_x R(x)$.

To save space, we will simply assume the case in which $E(e) = 0$ and $\lambda_e \leq \lambda_R \leq \lambda_0$. If these conditions do not hold, then the estimator will still yield $m(x)$, but with an arbitrary location, exactly as was the case with censored regression.

Assumption A1*. Let Assumption A1 hold, except that what is now observed is $Y = Y^* I(Y^* > 0)$ and $X^* = XI(Y^* > 0)$. The function $\widetilde{R}(e)$ is invertible for all $e > \inf_{e \in \Omega_e} e$, $E(e) = 0$, and $\lim_{e \to -\infty} e^2 F(e) = 0$. Assume that $\lambda_e \leq \lambda_R$.

Theorem 4 *Let Assumption A1* hold. Then for all* $x \in \Omega$, $R(x) = \widetilde{R}[m(x)]$, *and* $U[R(x)] = T(x) = \mathfrak{F}_2[m(x)]/F[m(x)]$. *Also, letting* λ_0 *be any constant such that* $\lambda_0 \geq \lambda_R$, *for all* $x \in \Omega$ *such that* $F[m(x)] \neq 0$,

$$
m(x) = \lambda_0 - \int_{R(x)}^{\lambda_0} \frac{U(R) - RU'(R)}{U(R) - R^2} dR, \tag{6}
$$

and for each continuously distributed element X_k *of* X,

$$
m_k(x) = \frac{R(x)T_k(x) - T(x)R_k(x)}{R(x)^2 - T(x)}. \tag{7}
$$

Proof. For positive k, $E(Y^k/k|X = x) = E(Y^k/k|X = x, Y > 0)F[m(x)] + E(Y^k/k|X = x, Y = 0)(1 - F[m(x)])$. The equations for R, U, and T then follow from Theorem 1. To derive the expression for $m(x)$, apply the change of variables $R = \widetilde{R}(m)$, so the claim is that $m(x)$ equals $\lambda_0 - \int_{\widetilde{R}^{-1}[R(x)]}^{\widetilde{R}^{-1}[\lambda_0]} \left(U[\widetilde{R}(m)] - \widetilde{R}(m)U'[\widetilde{R}(m)]\right) / \left(U[\widetilde{R}(m)] - \widetilde{R}(m)^2\right) [\partial \widetilde{R}(m)/\partial m] dm$. To simplify this expression, observe that $\partial \widetilde{R}(m)/\partial m = [1 - \widetilde{R}(m) f(m)/F(m)]dm$, $U[\widetilde{R}(m)] = \mathfrak{F}_2(m)/F(m)$, and $U'[\widetilde{R}(m)] = (d[\mathfrak{F}_2(m)/F(m)]dm)\, dm/d\widetilde{R}(m) = \left(\widetilde{R}(m) - U[\widetilde{R}(m)] f(m)/F(m)\right) / [1 - \widetilde{R}(m) f(m)/F(m)]$. Substituting each of these expressions into the integral, the claimed expression for $m(x)$, simplifies to $\lambda_0 - \int_{m(x)}^{\widetilde{R}^{-1}(\lambda_0)} 1 dm = \lambda_0 - [\widetilde{R}^{-1}(\lambda_0) - m(x)]$. It was shown in the proof of Theorem 3 that, given $\lambda_e \leq \lambda_R$, we have $\mathfrak{F}_1(\lambda_0) = \lambda_0$ for any $\lambda_0 \geq \lambda_R$. By definition, $F(\lambda_0) = 1$ for any $\lambda_0 \geq \lambda_R$, so $\widetilde{R}(\lambda_0) = \lambda_0$, and therefore $\lambda_0 = \widetilde{R}^{-1}(\lambda_0)$, which completes the derivation of the expression for $m(x)$. Finally, taking derivatives of the derived expressions for $R(x)$ and $T(x)$ gives $R_k(x) = (1 - R(x) f[m(x)]/F[m(x)])\, m_k(x)$ and

$T_k(x) = (R(x) - T(x) f[m(x)]/F[m(x)])\, m_k(x)$, which when substituted into the claimed expression for $m_k(x)$ yields $m_k(x)$. ■

With truncated data, a nonparametric regression of Y on X will equal $\widehat{R}(x)$, an estimator of $R(x)$. Similarly, nonparametrically regressing $Y^2/2$ on X with truncated data will yield an estimator $\widehat{T}(x)$, and we have derivative estimators $\widehat{R}_k(x)$ and $\widehat{T}_k(x)$ for continuously distributed elements x_k of x. Finally, nonparametrically regressing $Y^2/2$ on $\widehat{R}(X)$ with truncated data will yield an estimator $\widehat{U}(R)$, and $\widehat{U}'(R) = \partial \widehat{U}(R)/\partial R$. Given the above theorem, these nonparametric regressions can be substituted into the above expression for $m(x)$ and $m_k(x)$ to yield semiparametric plug-in estimators for these functions.

3.1 The Error Distribution in Truncated Regression

It follows from Theorem 3 that, for any e^*, $E[Y|m(X) = e^*, Y > 0] = \widetilde{R}(e^*)$, and $1/\widetilde{R}(e^*) = F(e^*)/\mathfrak{F}_1(e^*) = \partial \ln \mathfrak{F}_1(e^*)/\partial e^*$, so $\mathfrak{F}_1(e^*) = \exp \int_{-\infty}^{e^*} 1/\widetilde{R}(m)dm$, and $F(e^*) = \partial \mathfrak{F}_1(e^*)/\partial e^* = [1/\widetilde{R}(e^*)] \exp \int_{-\infty}^{e^*} 1/\widetilde{R}(m)dm$. Therefore, given the estimated regression function $\widehat{m}(x)$, the distribution function $F(e)$ for any e can be estimated as $\widehat{F}(e) = [1/\widehat{\widetilde{R}}(e)] \exp \int_{-\infty}^{e} 1/\widehat{\widetilde{R}}(m)dm$, where the estimated function $\widehat{\widetilde{R}}$ is a nonparametric regression of Y_i on $\widehat{m}(X_i)$ using the truncated data, and the integral is evaluated numerically.

4 Estimation

We propose estimators based on local linear regression because of their attractive properties with regard to boundary bias and design adaptiveness [see Fan and Gijbels (1996) for discussion and references]. This is important here because we may be integrating over boundary regions in (2) and (6). We just define the estimators of m and its partial derivatives in the censored regression case. The estimator in the truncated case involves analogous substitutions; we refer the reader to Lewbel and Linton (1999) for further details.

Given generic observations $\{Y_i, X_i\}_{i=1}^n$, we estimate the regression function $g(x) = E(Y_i|X_i = x)$ and its derivatives using the multivariate weighted least squares criterion

$$\sum_{i=1}^{n} [Y_i - b_0 - b_1 \cdot (X_i - x)]^2 \mathcal{K}((X_i - x)/h_n), \tag{8}$$

where $\mathcal{K}(u)$ is a nonnegative kernel function on $\mathbb{R}^d$ and h_n is a bandwidth parameter. Minimizing (8) with respect to the scalar b_0 and the vector $b_1 \in \mathbb{R}^d$ gives an estimate $(\widehat{b}_0(x), \widehat{b}_1(x))$ of $(g(x), \partial g(x)/\partial x)$.

Let $\widehat{r}(x)$ be the nonparametric regression of Y_i on X_i, constructed as in (8), and let $\widehat{q}(r)$ be the one-dimensional nonparametric regression of $I(Y_i > 0)$ on the generated regressor $\widehat{r}(X_i)$ evaluated at $\widehat{r}(X_i) = r$. We then let

$$\widehat{m}(x) = \lambda_0 - \int_{\widehat{r}(x)}^{\lambda_0} \frac{1}{\widehat{q}(r)} dr \tag{9}$$

for some fixed positive λ_0, and to conserve space redefine $m(x) = m(x) + \lambda_0 - \mathfrak{F}_1^{-1}(\lambda_0)$. The univariate integral can be evaluated numerically, and $\widehat{m}(x)$ can be computed very quickly.

If $\lambda_r > \lambda_e$, then the same limiting distribution will be obtained if λ_0 is replaced by $\widehat{\lambda}_r = \max_{i=1,\ldots,n} \widehat{r}(X_i)$ or $\widehat{\lambda}_Y = \max_{i=1,\ldots,n} Y_i$ because (as discussed after Theorem 3) in that case $\widehat{\lambda}_r - \mathfrak{F}_1^{-1}(\widehat{\lambda}_r) = 0$ for all $\widehat{\lambda}_r > \lambda_e$, and the probability that $\widehat{\lambda}_r > \lambda_e$ goes to one at a fast rate. Note that in this case $m(x)$ will be located so as to make $E(e) = 0$.

Let $\widehat{r}_k(x)$ and $\widehat{s}(x)$ be nonparametric estimators of the functions $r_k(x)$ and $s(x)$ as defined above. Specifically, for $\widehat{r}_k(x)$ and $\widehat{s}(x)$ we take $Y_i = Y_i$ and $Y_i = 1(Y_i > 0)$ in (8), respectively, while X_i are the given covariates. We then let

$$\hat{m}_k(x) = \frac{\widehat{r}_k(x)}{\widehat{s}(x)} \tag{10}$$

for $k = 1, \ldots, d$.

We now provide the pointwise distribution theory for $\widehat{m}(x)$; the distribution theory for $\hat{m}_k(x)$ is trivial, and can be found in Lewbel and Linton (1999). We make the following assumptions on the kernel $\mathcal{K}$ and on the data distribution.

ASSUMPTION A2. The support Ω is compact. The functions $\sigma_r^2, \sigma_s^2, f_X$, and s, where $\sigma_r^2(x) = \text{var}(Y|X = x)$ and $\sigma_s^2(x) = \text{var}[1(Y > 0)|X = x]$, while f_X is the Lebesgue density of X, are continuous on Ω, and $\inf_{x\in\Omega} f_X(x) > 0$ and $\inf_{x\in\Omega} s(x) > 0$. The conditional distribution $G(y|u)$ of Y given $X = u$ is continuous at the point $u = x$. $E\left[|Y|^t\right] < \infty$ for some $t > 2$. The regression functions r and s are three times continuously differentiable on Ω.

ASSUMPTION A3. The kernel $\mathcal{K}$ is symmetric about zero, bounded, and has compact connected support ($\mathcal{K}(u) = 0$ for $||u|| > A_0$ some A_0), and is differentiable in all its arguments. Let $||\mathcal{K}||^2 = \int \mathcal{K}^2(u)du$. The bandwidths satisfy $h_n \to 0$ and $\limsup_{n\to\infty} nh_n^{d+4} < \infty$.

Theorem 5 *Suppose that Assumptions A1-A3 hold. Then, there exists a bounded continuous function $b_m(\cdot)$ such that for any x in the interior of Ω,*

$$\sqrt{nh_n^d}\left(\widehat{m}(x) - m(x) - h_n^2 b_m(x)\right) \Longrightarrow N\left(0, \frac{\sigma_r^2(x)}{f_X(x)s^2(x)} \|\mathcal{K}\|^2\right).$$

The asymptotic variance of $\widehat{m}(x)$ increases with the amount of censoring $[1 - s(x)]$ and the variance $\sigma_r^2(x)$ of Y. The corresponding α-quantile estimator has asymptotic variance proportional to $\alpha(1 - \alpha)/f(F^{-1}(\alpha))^2$, where f is the density of the error e. The relative efficiency of these two estimators depends as usual on the tail thickness of the error distribution. The asymptotic variance of our estimator can be consistently estimated from the estimates of $\sigma_r^2(x)$, $s(x)$, and $r(x)$ [note that the bias term is of smaller order than the standard deviation provided $nh_n^{d+4} \to 0$]. In contrast, to estimate the asymptotic variance of the quantile estimator requires estimates of the error density.

The truncated regression estimator, and the estimators of the derivatives $m_k(x)$ are also asymptotically normal. Their distributions are provided in Lewbel and Linton (1999). That working paper also shows that averages of the derivative estimators $\widehat{m}_k(x)$ can converge at rate root n. An application of these average derivative estimates is that they can be used to estimate censored or truncated regression models in which $m(x)$ is specified as partly linear.

5 Monte Carlo Simulation

A Monte Carlo study is employed to check the finite sample behavior of our estimator. The design for the study is $Y = \max[m(X) - e, 0]$, $m(x) = x^3$, with scalar $X \sim$ Uniform$[-1, 1]$ and $e \sim N(0, 0.25)$. Given this design, the amount of censoring as a function of x is given by $1 - \Phi\left(2x^3\right)$, where $\Phi(\cdot)$ is the standard normal c.d.f., so the percent of censoring ranges from 100% at $x = -1$, to 50% at $x = 0$, to 0% at $x = 1$. The sample size is $n = 200$, and the number of Monte Carlo simulations is 1000.

We consider the following censored regression estimator

$$\widehat{m}(x) = \widehat{\lambda}_r - \int_{\widehat{r}(x)}^{\widehat{\lambda}_r} \frac{1}{\widehat{q}(r)} dr,$$

where $\widehat{\lambda}_r = \max_{i=1,\ldots,n} \widehat{r}(X_i)$. The component functions such as $\widehat{r}(x)$ and $\widehat{q}(r)$ are estimated as nonparametric kernel regressions, using normal kernels. The integral in $\widehat{m}(x)$ is evaluated numerically using the trapezoid method. Bandwidths are selected by grid search to minimize simulation based estimates of the integrated squared error, ISE$=\int [\widehat{m}(x) - m(x)]^2 f_X(x)dx$. Average absolute error and average squared error were also evaluated and yielded virtually the same bandwidths, which were $h = 0.2$ for $\widehat{r}(x)$ and $h = 0.05$ for $\widehat{q}(r)$.

Details of this procedure, and GAUSS code for all of the Monte Carlo simulations reported here, are available from the authors on request.

For comparison, the function $m(x)$ is also estimated using quantile regression, as follows. The conditional empirical distribution function is first estimated as

$$\widehat{F}(y|x) = \frac{\sum_{i=1}^{n} \phi\left(\frac{x-X_i}{h_1}\right) \Phi\left(\frac{y-Y_i}{h_2}\right)}{\sum_{i=1}^{n} \phi\left(\frac{x-X_i}{h_1}\right)},$$

where $\phi(\cdot)$ is the standard normal density function. Then $\widehat{F}(y|x)$ is numerically inverted and the q-quantile estimate is

$$\widehat{m}_q(x) = \widehat{F}_q^{-1}(y|x) - \alpha_q,$$

where α_q is the q-th quantile of the error term. The true α_q is used here, to make the location of the quantile estimates comparable to the $E(e) = 0$ location of our estimator. The optimal bandwidth for the quantile regression estimator $\widehat{m}_q(x)$ is obtained using the same procedure as for $\widehat{m}(x)$.

Figure 1 shows the results for the censored regression estimator $\widehat{m}(x)$, and Figure 2 shows the median regression estimator $\widehat{m}_q(x)$ for $q = 0.5$. On these figures the solid line is the true $m(x)$, while dotted lines show the mean, median, 5% and 95% quantiles of the estimates of $m(x)$, across the 1000 monte carlo simulations. The difference between the solid line and the mean or median dotted lines provides a measure of bias of the estimator, while the 5% and 95% lines provide a measure of spread of the estimates, and may be interpreted as simulation based estimates of confidence bands.

An interesting feature of this design is that it formally violates our assumption regarding location estimation, since $\lambda_r = \sup_x r(x) = 1$ while $\lambda = \sup e = \infty$. Therefore, in this design the "bias" in location (relative to locating $m(x)$ so as to make the errors have mean zero) is $1 - \mathfrak{F}_1^{-1}(1)$, where the function $\mathfrak{F}_1(e^*)$ equals the integral from $-\infty$ to e^* of the distribution function of a normal having mean zero, variance one fourth. However, since $\Pr(e > 1)$ is tiny, the magnitude of the location bias seen in Figure 1 is correspondingly small.

Comparing figures 1 and 2 shows that for positive x, where the amount of censoring is less than 50%, both our estimator $\widehat{m}(x)$ and the nonparametric median regression $\widehat{m}_{.5}(x)$ perform about equally well. However, for negative x, our estimator continues to perform well, with confidence bands only mildly enlarged by the greater degree of censoring in that region. Median regression is of course inconsistent in that region; consistent quantile estimation in the negative x region requires more extreme quantiles. Experiments (not reported) using lower quantiles, e.g., $q = 0.25$, increase the range of x values for which $\widehat{m}_q(x)$ is consistent, but also correspondingly widen the estimator's confidence bands. Use of different quantiles also changes the location of quantile estimates (through α_q). Efficiency of the quantile estimators might be increased by combining estimates from multiple quantiles. Our estimator does not require arbitrary selection of one or more quantiles, remains consistent everywhere inside of the support of x, and can have location determined by $E(e) = 0$.

Limited experiments (not reported) with different bandwidths were also performed. Doubling the bandwidths flattens $\widehat{m}(x)$, causing increased bias, primarily in the tails of the data. Halving the bandwidths has little effect on the average or median values of $\widehat{m}(x)$ across the simulations, but increases the variance of the estimates and hence widens the confidence bands.

Similar results to those reported are obtained when comparing our derivative estimator $\widehat{m}_k(x)$ to nonparametric quantile derivatives. See Lewbel and Linton (1999) for details.

6 Extensions and Conclusions

We have provided estimators for the nonparametric censored and truncated regression models with fixed censoring. Our estimators are computationally convenient, consisting only of two nonparametric regressions and a univariate integral. Specifically, we employ a novel method of first nonparametrically regressing Y on X, next regressing a different function of y on the fits of this first stage, and finally integrating the result. This new methodology exploits derivative relationships between conditional expectations of y and of functions of y, and so might be extended to other contexts in which differential expressions involving index functions or latent variables can be obtained.

Our estimator could be used if (instead of a fixed censoring point) the censoring point is a random variable C_i that is known for all observations, by redefining Y_i and $m(X_i)$ as $Y_i - C_i$ and $m(X_i) - C_i$, and then redefining X_i to include C_i. Our estimator would then permit the variable C_i to affect Y_i like any other regressor in X_i, in addition to determining the point of censoring.

We provided limiting distributions assuming all the elements of x are continuous, but the estimator can very easily handle inclusion of discrete regressors as well. The first stage nonparametric regression $\widehat{r}(x)$ would simply include both types of regressors , either by doing a separate local linear regression for each discrete cell, or by smoothing over cells as in, e.g., Racine and Li (2000). The rest of the estimation would then proceed exactly as before.

Our estimators can be extended to allow for very general forms of heteroskedasticity. Let z be any subset of the elements of x. Instead of homoskedasticity, assume now that the error distribution depends in arbitrary, unknown ways on the subset of regressors z. For example, in a demand model the latent errors are usually interpreted as unobserved preference attributes, and so are typically assumed to not depend on prices. In that application z might equal all of the elements of x except prices.

Assume that $F(e|x) = F(e|z)$, $E(e|z) = 0$, and $\text{supp}(e|z) = \text{supp}(e) \subseteq \text{supp}[m(x)|z]$. Let $\mathfrak{F}_1(m|z) = \int_{-\infty}^{m} F(e|z)de$. Assume the function $\mathfrak{F}_1$ is invertible on its first element, and define the function $\mathfrak{F}_1^{-1}$ by $\mathfrak{F}_1^{-1}[\mathfrak{F}_1(m|z), z] = m$. As before, let $r(x) = E(y|x)$, and now define $q[r(x), z] = E[I(Y > 0)|r(x), z]$. Then by Theorem 1, but now conditioning on z,

$$r(x) = \mathfrak{F}_1[m(x)|z] \quad ; \quad q[r(x), z] = F\left(\mathfrak{F}_1^{-1}[r(x), z]|z\right).$$

Similarly, following the steps of Theorem 2 while conditioning on z shows that, for all $x \in \Omega$ having $F[m(x)|z] \neq 0$,

$$m(x) + k = \lambda_0 - \int_{r(x)}^{\lambda_0} \frac{1}{q[r, z]} dr \tag{11}$$

for suitable λ_0 and $k(\lambda_0)$. The estimator based on this equation is identical to the homoskedastic estimator, except that $\widehat{q}$ will be a nonparametric regression on $\widehat{r}$ and on z. An analogous derivation can be applied to the truncated regression estimator. We do not know of any other estimators for censored or truncated regression that permit this general form of heteroskedasticity, even in a semiparametric context where $m(x)$ is finitely parameterized.

A Appendix

SKETCH PROOF OF THEOREM 5. Write $\widehat{q}(s) = \widehat{q}(s; \widehat{r}_1, \ldots, \widehat{r}_n)$, where $\widehat{r}_j = \widehat{r}(X_j)$. Rearranging terms, and making a mean value expansion we obtain

$$\begin{aligned}\widehat{m}(x) - m(x) = \frac{1}{q(r(x))}(\widehat{r}(x) - r(x)) + \int_{r(x)}^{\lambda_0} \frac{(\widehat{q}(s) - q(s))}{q^2(s)} ds \quad & - \frac{\widehat{q}'(\overline{r}(x))}{2\widehat{q}^2(\overline{r}(x))}(\widehat{r}(x) - r(x))^2 \\ - \int_{\widehat{r}(x)}^{\lambda_0} \frac{(\widehat{q}(s) - q(s))^2}{\widehat{q}(s) q^2(s)} ds \quad & + \frac{\widehat{q}(\overline{\overline{r}}(x)) - q(\overline{\overline{r}}(x))}{\widehat{q}(\overline{\overline{r}}(x)) q(\overline{\overline{r}}(x))}(\widehat{r}(x) - r(x)),\end{aligned}$$

where $\overline{r}(x)$ and $\overline{\overline{r}}(x)$ are intermediate values between $r(x)$ and $\widehat{r}(x)$. Let $\delta_n = \max\{1/\sqrt{nh_n^d}, h_n^2\}$. Under our conditions, for some $\epsilon > 0 : \inf_{r(x)-\epsilon\leq s\leq\lambda_0+\epsilon} q(s) > 0$, $\sup_{r(x)-\epsilon\leq s\leq\lambda_0+\epsilon} |\widehat{q}'(s)| = O_p(1)$, and $(\sup_{r(x)-\epsilon\leq s\leq\lambda_0+\epsilon} |\widehat{q}(s) - q(s)|)^2 = o_p(\delta_n)$ [see Masry (1996a,b)]. Therefore, we have

$$\widehat{m}(x) - m(x) = \frac{\widehat{r}(x) - r(x)}{s(x)} + \int_{r(x)}^{\lambda_0} \frac{(\widehat{q}(s) - q(s))}{q^2(s)} ds + o_p(\delta_n).$$

More detailed arguments, given in Lewbel and Linton (1999), show that

$$\int_{r(x)}^{\lambda_0} \frac{(\widehat{q}(s) - q(s))}{q^2(s)} ds = h_n^2 b(x) + o_p(\delta_n)$$

for some bounded continuous function $b(x)$. Finally, we apply Theorems in Masry (1996a,b) to $\widehat{r}(x) - r(x)$ to obtain the result. ■

References

[1] AHN, H. (1995): "Nonparametric two-stage estimation of conditional choice probabilities in a binary choice model under uncertainty," *Journal of Econometrics* **67**, 337-378.

[2] AMEMIYA, T. (1973), "Regression Analysis When the Dependent Variable is Truncated Normal," *Econometrica*, 41, 997–1016.

[3] AMEMIYA, T. (1985) *Advanced Econometrics*. Harvard University Press.

[4] ANDREWS, D. W. K., AND M. M. A. SCHAFGANS, (1998), "Semiparametric Estimation of the Intercept of a Sample Selection Model," *Review of Economic Studies*, 65, 497-516.

[5] BUCHINSKY, M. AND J. HAHN, "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica,* 66, 653-671.

[6] BUCKLEY, J. AND I. JAMES, (1979), "Linear Regression With Censored Data," *Biometrika*, 66, 429–436.

[7] CHAUDHURI, P. (1991). "Nonparametric estimates of regression quantiles and their local Bahadur representation," *Annals of Statistics* **19**, 760-777.

[8] CHAUDHURI, P., K. DOKSUM, AND A. SAMAROV, (1997), "On Average Derivative Quantile Regression,". *Annals of Statistics* **25**, 715-744.

[9] CHEN, S. AND S. KHAN (2000), "Estimation of a Nonparametric Censored Regression Model," unpublished manuscript.

[10] DABROWSKA, D. M. (1995): "Nonparametric regression with censored covariates," *Journal of Multivariate Analysis* 54, 253-283.

[11] DUNCAN, G. M., (1986), "A Semi-parametric Censored Regression Estimator," *Journal of Econometrics*, 32, 5–24.

[12] FAN, J., AND I. GIJBELS (1994): "Censored Regression: Local Linear Approximations and their Applications," *Journal of the American Statistical Association* 89, 560-570.

[13] FAN, J., AND I. GIJBELS (1996), *Local Polynomial Modelling and Its Applications* Chapman and Hall.

[14] FERNANDEZ, L., (1986), "Non-parametric Maximum Likelihood Estimation of Censored Regression Models," *Journal of Econometrics*, 32, 35–57.

[15] HECKMAN, J. J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 15, 475–492.

[16] HONORÉ, B. E. AND J. L. POWELL, (1994), "Pairwise Difference Estimators of Censored and Truncated Regression Models," *Journal of Econometrics*, 64, 241–278.

[17] HOROWITZ, J. L., (1986), "A Distribution Free Least Squares Estimator for Censored Linear Regression Models," *Journal of Econometrics,* 32, 59-84.

[18] HOROWITZ, J. L., (1988), "Semiparametric M-Estimation of Censored Linear Regression Models," *Advances in Econometrics,* 7, 45-83.

[19] ICHIMURA, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS estimation of Single-index Models," *Journal of Econometrics*, 58, 71–120.

[20] VAN KEILEGOM, I., AND M.G. AKRITAS (1999): "Transfer of tail information in censored regression models," Forthcoming in the *Annals of Statistics*.

[21] KOUL, H., V. SUSLARA, AND J. VAN RYZIN (1981), "Regression Analysis With Randomly Right Censored Data," *Annals of Statistics*, 42, 1276–1288.

[22] LEWBEL, A. (1995), "Consistent Nonparametric Tests With An Application to Slutsky Symmetry," *Journal of Econometrics*, 67, 379–401.

[23] LEWBEL, A. (1997), "Semiparametric Estimation of Location and Other Discrete Choice Moments," *Econometric Theory, 13, 32-51.*

[24] LEWBEL, A. (1998a), “Semiparametric Latent Variable Model Estimation With Endogenous or Mismeasured Regressors,” *Econometrica*, 66, 105–121.

[25] LEWBEL, A. (2000), “Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables,” *Journal of Econometrics* 97, 145-177.

[26] LEWBEL, A., AND O.B. LINTON (1999): “Nonparametric Censored and Truncated Regression,” Sticerd working paper no. EM/00/389.

[27] MCDONALD, J. AND R. MOFFITT (1980), “The Uses of Tobit Analysis,” *Review of Economics*, 62, 318–321.

[28] MADDALA, G. S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Econometric Society Monograph No. 3, Cambridge: Cambridge University Press.

[29] MASRY, E. (1996a), “Multivariate local polynomial regression for time series: Uniform strong consistency and rates,” *J. Time Ser. Anal.* 17, 571-599.

[30] MASRY, E., (1996b), “Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processes and their Applications* 65, 81-101.

[31] MOON, C.-G., (1989), “A Monte Carlo Comparison of Semiparametric Tobit Estimators. *Journal of Applied Econometrics,* 4, 361-382.

[32] NAWATA, K. (1990), “Robust Estimation Based on Group-Adjusted Data in Censored Regression Models,” *Journal of Econometrics*, 43, 337–362.

[33] POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989), “Semiparametric Estimation of Index Coefficients,” *Econometrica* 57, 1403–1430.

[34] POWELL, J. L. (1984), “Least Absolute Deviations Estimation for the Censored Regression Model,” *Journal of Econometrics*, 25, 303–325.

[35] POWELL, J. L. (1986a), “Symmetrically Trimmed Least Squares Estimation For Tobit Models,” *Econometrica*, 54, 1435–1460.

[36] POWELL, J. L. (1986b), “Censored Regression Quantiles,” *Journal of Econometrics*, 32, 143–155.

[37] RACINE, J. AND Q. LI (2000), “Nonparametric Estimation of Conditional Distributions with Mixed Categorical and Continuous Data,” Unpublished Manuscript.

[38] RITOV, Y. (1990): “Estimation in a linear regresion model with censored data,” *Annals of Statistics* 18, 303-328.

[39] ROBINSON, P. M. (1982), "On the Asymptotic Properties of Estimators of Models Containing Limited Dependent Variables," *Econometrica*, 50, 27-41.

[40] ROSETT, R. AND F. NELSON (1975), "Estimation of the Two-Limit Probit Regression Model," *Econometrica*, 43, 141–146.

[41] SCHMEE, J., AND HAHN, GERALD J. (1979): "A simple method for regression analysis with censored data" *Technometrics* 21, 417-432.