NONPARAMETRIC CENSORED AND TRUNCATED REGRESSION*

Arthur Lewbel† Boston College

Oliver Linton‡ LSE and Yale University

March 27, 2000

Abstract

The nonparametric censored regression model, with a fixed, known censoring point (normalized to zero), is $y = \max[0, m(x) + e]$, where both the regression function $m(x)$ and the distribution of the error e are unknown. This paper provides consistent estimators of $m(x)$ and its derivatives. The convergence rate is the same as for an uncensored nonparametric regression and its derivatives. We also provide root n estimates of weighted average derivatives of $m(x)$, which equal the coefficients in linear or partly linear specifications for $m(x)$. An extension permits estimation in the presence of a general form of heteroscedasticity. We also extend the estimator to the nonparametric truncated regression model, in which only uncensored data points are observed. The estimators are based on the relationship $\partial E(y^k|x)/\partial m(x) = kE[y^{k-1}I(y > 0)|x]$, which we show holds for positive integers k.

JEL Codes: C14 C24 C13

*We would like to thank Yan Li for excellent research assistance, and Joel Horowitz and Andrew Chesher 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.

Keywords: Semiparametric, Nonparametric, Censored Regression, Truncated Regression, Tobit, Latent Variable.

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 mean zero error that is independent of X_i (writing the model as $m - e$ instead of the more usual $m + e$ simplifies later results). Here, 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 $m(X_i)$.

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.

Both the regression function $m(\cdot)$ and the distribution $F(\cdot)$ of the error e is unknown. The errors are not assumed to be symmetric. This paper provides a simple consistent estimator of $m(x)$, which equals the conditional mean function for the uncensored population. Also, we show that the distribution function of the errors 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 heteroscedasticity, in which the distribution of e could depend in unknown ways on all but one element of x.

For any continuously distributed element x_k of x, let $m_k(x) = \partial m(x)/\partial x_k$. This paper also provides 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 used as estimates of the coefficients in a partly linear specification of $m(x)$.

Parametric and semiparametric estimators of censored 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, 1998b), Buchinsky and Hahn (1998), and Levy (1999). 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). We do not know of any other estimator for the nonparametric truncated regression model.

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, we show that it is inconsistent in our situation of fixed censoring. We do not know if any other nonparametric version of Buckley and James can be constructed that would not, for similar reasons, be inconsistent under fixed censoring.

Other possible nonparametric censored regression estimators are based on quantile methods, e.g., Dabrowska (1995). As we will later demonstrate, the main advantage of our estimator over quantile regression estimators is that consistent quantile estimators require some a priori information about the degree of censoring at each point, and our estimator does not. Also, our estimator can be extended to handle nonparametric truncated regression.

The estimators we propose are functions of nonparametric regressions. While these estimators remain consistent when ordinary kernel regressions are used in these functions, we instead employ local polynomials which have some advantages over ordinary kernels [see, e.g., Fan and Gijbels (1996)] that we will exploit. We show that the uniform convergence rate of the estimators is the same as for an uncensored regression. We also construct root n consistent and asymptotically normal estimators of weighted averages of the derivatives $m_k(x)$, which equal the coefficients in partly linear censored or truncated regression models.

2 The Censored Regression Function and its Derivatives

Let Y^* be an unobserved latent variable with $E|Y^*| < \infty$, and define $m(x) = E(Y^*|X = x)$ and $e = Y^* - m(X)$. The random vector X can contain both discrete and continuously distributed elements. The unknown function m is continuous and differentiable with respect to the continuously distributed elements of X. For each continuously distributed element X_k of X define

$$m_k(x) = \frac{\partial m(x)}{\partial x_k}$$

Assume that the mean zero error e is independent of X, and is continuously distributed with unknown distribution function $F(e)$ and probability density function $f(e)$ (the model will later be extended to let the distribution of e depend on x in some general ways). The observed dependent variable Y equals the latent variable censored at zero, so $Y = I(Y^* \geq 0)Y^*$, where I is the indicator function that equals one if its argument is true and zero otherwise. We assume throughout that our 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:

$$\begin{aligned}
\mathfrak{F}_0(m) &= F(m) \\
\mathfrak{F}_\kappa(m) &= \int_{-\infty}^{m} \mathfrak{F}_{\kappa-1}(e)de, \ \kappa = 1, 2, \ldots \\
\mathfrak{F}(m) &= \mathfrak{F}_1(m).
\end{aligned}$$

Theorem 1 *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. $E[Y^\kappa I(Y > 0)|X = x] = E[Y^\kappa I(Y > 0)|m(X) = m(x)]$. 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:

$$\begin{aligned}
r(x) &= E(Y|X = x), \quad r_k(x) = \frac{\partial r(x)}{\partial x_k} \\
s(x) &= E[I(Y > 0)|X = x], \quad s_k(x) = \frac{\partial s(x)}{\partial x_k} \\
t(x) &= E(Y^2/2|X = x), \quad t_k(x) = \frac{\partial t(x)}{\partial x_k} \\
q[r(x)] &= E[I(Y > 0)|r(X) = r(x)],
\end{aligned}$$

where x_k is the k'th element of x.

Assumption A1. *Suppose that $Y^* = m(X) - e$ and we observe X and $Y = I(Y^* \geq 0)Y^*$. Let Ω be a compact subset of the support of the $d \times 1$ vector x. 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 has mean zero, is continuously distributed, independent of x, with probability distribution function $F(e)$ and probability density function $f(e)$. $\mathfrak{F}_2[m(x)]$ exists for all $x \in \Omega$. The function $\mathfrak{F}$* is *invertible, and $\lim_{e \to -\infty} e^2 F(e) = 0$. Let $\mathfrak{F}^{-1}$ denote the inverse function of $\mathfrak{F}$, let Ω_e denote the support of e, and let $\lambda = \sup_{e \in \Omega_e} e$. Assume that $\lambda < \sup_x m(x)$.*

Theorem 2 *Let Assumption A1 hold. Then for all $x \in \Omega$, $r(x) = \mathfrak{F}[m(x)]$, $s(x) = F[m(x)]$, $t(x) = \mathfrak{F}_2[m(x)]$, and $q[r(x)] = F\left(\mathfrak{F}^{-1}[r(x)]\right)$. Also, for all $x \in \Omega$ having $F[m(x)] \neq 0$,*

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

and 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, t, and q follow from Theorem 1. For $m(x)$, use the change of variables $r = \mathfrak{F}(m)$, $dr = F(m)dm$, and $q(r) = F\left(\mathfrak{F}^{-1}[\mathfrak{F}(m)]\right) = F(m)$ to get $\int_{r(x)}^{\lambda}[1/q(r)]dr = \int_{\mathfrak{F}^{-1}(\mathfrak{F}[m(x)])}^{\mathfrak{F}^{-1}(\lambda)}[1/F(m)]F(m)dm = \int_{m(x)}^{\mathfrak{F}^{-1}(\lambda)} 1dm = \mathfrak{F}^{-1}(\lambda) - m(x)$. Next, using an integration by parts, $E(e) = 0 = \int_{-\infty}^{\lambda} ef(e)de = -\int_{-\infty}^{\lambda}[F(e) - I(e > 0)]de = -\mathfrak{F}(\lambda) + \lambda$, so $\mathfrak{F}^{-1}(\lambda) = \lambda$, which completes the derivation of the expression for $m(x)$. Finally, $r_k(x) = \partial \mathfrak{F}[m(x)]/\partial x_k = F[m(x)]m_k(x) = s(x)m_k(x)$.

■

Note that $m(x) = \lambda_* - \int_{r(x)}^{\lambda_*} \frac{1}{q(r)} dr$ for any $\lambda_* \geq \lambda$. Let $\lambda_r = \sup_x r(x)$. Then under our assumptions $\lambda_r > \lambda$ because $\sup_x r(x) = \sup_x \mathfrak{F}(m(x)) = \mathfrak{F}(\sup_x m(x)) > \mathfrak{F}(\lambda) = \lambda$. 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, let $\widehat{q}(r)$ be a nonparametric regression of $I(Y > 0)$ on $\widehat{r}(X)$, and let $\widehat{\lambda}_r = \max_{i=1,\ldots,n} \widehat{r}(X_i)$. It is a standard result that $\widehat{r}, \widehat{s}$ and $\widehat{\lambda}_r$ are consistent. Therefore, based on the above theorem, we will show that $\widehat{\lambda}_r - \int_{\widehat{r}(x)}^{\widehat{\lambda}_r}[1/\widehat{q}(r)]dr$ (which can be evaluated using numerical integration) and $\widehat{r}_k(x)/\widehat{s}(x)$ are consistent estimator of $m(x)$ and $m_k(x)$, respectively, and we will provide their limit normal distributions.[1]

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.

[1]One could use $\max_{i=1,\ldots,n} Y_i$ instead of $\max_{i=1,\ldots,n} \widehat{r}(X_i)$, but we have found better finite sample performance with our chosen upper bound estimator.

2.1 Identification

A general concern in latent variable models is the extent to which identification is based on assumptions and behavior in the tails of the data. This applies particularly to estimation of the location or intercept. See, e.g., Andrews and Schafgans (1998).

In our estimator, 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 an arbitrary location constant without appeal to tail data, since for any constant ζ, we have $m(x) - \mathfrak{F}^{-1}(\zeta) = -\int_{r(x)}^{\zeta}[1/q(r)]dr$. This entails observing a range of X values that is large enough obtain the function $r(X)$ everywhere in the interval from $r(x)$ to ζ.

Our estimator uses tail information only to identify the location constant of $m(x)$. We define $m(x)$ to equal the expected value of y given x if y were not censored, so the location is the constant required to make $E(e) = 0$. Theorem 1 provides an estimator of $m(x) + \lambda_r - \mathfrak{F}^{-1}(\lambda_r)$. To estimate location, Theorem 1 assumes that $\lambda < \sup_x m(x)$, which means that for any value that e can take on, there exists an observable x that results in an uncensored y. This assumption makes $\lambda_r > \lambda$. and hence $\lambda_r - \mathfrak{F}^{-1}(\lambda_r) = 0$. If this tail assumption is violated, that is, if there exist a range of e values having 100% censoring, then only the location constant of $m(x)$ will be affected. If the probability of 100% censoring is small, then the resulting bias in the location estimate, which equals $\lambda_r - \mathfrak{F}^{-1}(\lambda_r)$, will be small. This is illustrated later in a Monte Carlo study.

2.2 Average Derivatives and Partly Linear Models

Given any weighting function $w(x)$, define the average regression function derivative $\delta_{wk} = E[w(X)m_k(X)]/E[w(X)]$. Since $m_k(x) = r_k(x)/s(x)$, this δ_{wk} can be estimated at rate root n by replacing the expectations with sample averages and substituting in nonparametric regression based estimates of $r_k(x)$ and $s(x)$.

Taking $w(x) = 1$ results in unweighted average derivatives. Taking $w(x)$ to equal $s(x)$ times the density of x yields a particularly simple form for δ_{wk} if kernel regressions are used to estimate $r_k(x)$ and $s(x)$, since then δ_{wk} will equal the Powell, Stock, and Stoker's (1989) weighted average derivative divided by the mean of a kernel regression numerator (see, e.g., Lewbel 1995).

If the latent regression function is linear or partly linear, that is, if for some $j \leq d$, $m(x) = \beta_1 x_1 + ... + \beta_j x_j + \widetilde{m}(x_{j+1}, \dots, x_k)$, then for $1 \leq k \leq j$, $\beta_k = \delta_{wk}$. Root n estimation of the coefficients in uncensored partly linear regression models is described in Robinson (1988), among others. In contrast, what is provided here is estimation of the same parameters when the partly linear model is censored. For small amounts of censoring, Chaudhuri, Doksum and Samarov (1997) might be a useful alternative. As an estimator of β_k, δ_{wk} has the advantage that if $m(x)$ turns out to not be linear or partly linear, δ_{wk} will still equal the usual interpretation of β_k as a measure of the average effect on the latent variable of a marginal change in x_k.

2.3 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. An alternative 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

$$\widehat{F}(e) = 1 - \prod_{i:\widehat{e}_{(i)} \leq e} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}}. \tag{4}$$

2.4 Comparison With Alternative Estimators

Consider first the Buckley and James (1979) censored regression estimator, which consists of transforming the dependent variable so as to make it have the right conditional expectation. This method is usually presented in random censoring models, but for finitely parameterized censored regression functions such estimators may work given fixed censoring as well. If m and F were known, then the ideal Buckley-James transform would be

$$Y_i^{BJ} = \delta_i Y_i + (1 - \delta_i) \frac{\int_{m(X_i)}^{\infty} e dF(e)}{\int_{m(X_i)}^{\infty} dF(e)},$$

where $\delta_i = 0$ when observation Y_i is censored, and $\delta_i = 1$ otherwise. It follows that

$$E(Y_i^{BJ} | X_i = x) = m(x).$$

In practice, both m and F are unknown and have to be replaced by estimators. When $m(x) = \beta' x$ we can use standard semiparametric profiling techniques as in Klein and Spady (1993) to estimate β. Specifically, we can estimate F by the Kaplan-Meier estimator constructed from the residuals $Y_i - \beta' X_i$, where the resulting 'estimator' depends on β. We then find a zero of the resulting score function. See Breiman, Tsur and Zemel (1993) for a simple version. Ritov (1990) provides a rigorous treatment and discussion of more general score functions and efficiency.

It is not known if Buckley-James type estimators can consistently estimate a nonparametric $m(x)$ with fixed censoring. Fan and Gijbels (1994) present a local Buckley-James estimator for nonparametric $m(x)$ that is consistent given random censoring. Fan and Gijbels do not consider what happens to their estimator under fixed censoring (they refer to the case where the censoring density is not continuous as a technicality to be ignored for simplicity). However, it turns out that their estimator is inconsistent under 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]. We suggest an alternative implementation of the Buckley-James algorithm below, which makes use of our consistent estimates of m and F.

Other 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. To demonstrate the advantage of our proposed estimator over quantile regression methods, 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 given our assumption that e is independent of X. Then $\theta_q(Y|X = x) = m(x) + \alpha_q$ if $q < 1 - \rho(x)$, 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 $q < 1 - \rho(x)$.[2]

The problem with using quantile methods to estimate $m(x)$ is that they require a priori knowledge about the amount of censoring at each point x, specifically, only quantiles q that are less than the unknown function $\rho(x)$ 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 $m(x)$. For example, if for a given x, $\rho(x) = 0.6$ (sixty percent censoring), then only quantiles $q < 0.4$ can be used to estimate the function m at that point x. If some other point x^* has less than fifty percent censoring then median regression can be used to estimate $m(x^*) + \alpha_{0.5}$, but that does not help to estimate $m(x)$ for x not in the neighborhood of x^*. The problem is not imprecision, but rather that consistency of the quantile estimator requires either knowing a priori some bound on the amount of censoring $\rho(x)$ at each x, or requires some mechanism, presumably based on an estimate of $\rho(x)$, to choose an appropriate quantile or set of quantiles for estimation. It is not clear how any such quantile selection procedure would work, or how it would affect the limiting distribution of the estimator.

Our estimator of $m(x)$ converges at the same rate as nonparametric quantile estimators. Whether our estimator or nonparametric quantile estimation is more efficient depends on the application. The main advantage of our estimator over quantiles is that ours does not require knowledge about the degree of censoring for consistency.

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:

$$R(x) = E(Y|X = x, Y > 0), \quad R_k(x) = \frac{\partial R(x)}{\partial x_k}$$

[2] We can also write $m(x) = \int_0^1 \theta_q(Y|X = x)dq$, but in general this requires knowledge of all quantiles, and so is not feasible when there is censoring.

$$
\begin{aligned}
T(x) &= E(Y^2/2|X=x, Y>0), \quad T_k(x) = \frac{\partial T(x)}{\partial x_k} \\
U[R(x)] &= E[(Y^2/2)|R(X)=R(x), Y>0], \quad U'(R) = \frac{\partial U(R)}{\partial R} \\
\widetilde{R}(m) &= \mathfrak{F}(m)/F(m),
\end{aligned}
$$

where x_k is the k'th element of x.

Assumption A1*. *Suppose that $Y^* = m(X) - e$ and we observe $Y = Y^* I(Y^* > 0)$ and $X^* = XI(Y^* > 0)$. Let Ω be a compact subset of the support of the $d \times 1$ vector x. 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 has mean zero, is continuously distributed, independent of x, with probability distribution function $F(e)$ and probability density function $f(e)$. $\mathfrak{F}_2[m(x)]$ exists and $F[m(x)] > 0$ for all $x \in \Omega$. The function $\widetilde{R}(m)$* is *invertible, and* $\lim_{e \to -\infty} e^2 F(e) = 0$. *Let $\widetilde{R}^{-1}$ denote the inverse function of $\widetilde{R}$, let Ω_e denote the support of e, and let $\lambda = \sup_{e \in \Omega_e} e$. Assume that $\lambda < \sup_x m(x)$.*

Theorem 3 *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, for all $x \in \Omega$,*

$$
m(x) = \lambda - \int_{R(x)}^{\lambda} \frac{U(R) - RU'(R)}{U(R) - R^2} dR, \tag{5}
$$

and for each continuously distributed element X_k of X,

$$
m_k(x) = \frac{R(x)T_k(x) - T(x)R_k(x)}{R(x)^2 - T(x)}. \tag{6}
$$

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 $\lambda - \int_{\widetilde{R}^{-1}[R(x)]}^{\widetilde{R}^{-1}[\lambda]} \left(U[\widetilde{R}(m)] - \widetilde{R}(m)U'[\widetilde{R}(m)]\right) / \left(U[\widetilde{R}(m)] - \widetilde{R}(m)^2\right) [\partial \widetilde{R}(m)/\partial m] dm$. 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 $\lambda - \int_{m(x)}^{\widetilde{R}^{-1}(\lambda)} 1dm = \lambda - [\widetilde{R}^{-1}(\lambda) - m(x)]$. It was shown in the proof of Theorem 1 that $\mathfrak{F}(\lambda) = \lambda$. By definition, $F(\lambda) = 1$, so $\widetilde{R}(\lambda) = \lambda$, and therefore $\lambda = \widetilde{R}^{-1}(\lambda)$, 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. As discussed earlier, we do not know of any other consistent estimator for these functions in the nonparametric truncated regression model.

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 $1/\widetilde{R}(e^*) = F(e^*)/\mathfrak{F}(e^*) = \partial \ln \mathfrak{F}(e^*)/\partial e^*$, so $\mathfrak{F}(e^*) = \exp \int_{-\infty}^{e^*} 1/\widetilde{R}(m)dm$, and $F(e^*) = \partial\mathfrak{F}(e^*)/\partial e^* = [1/\widetilde{R}(e^*)] \exp \int_{-\infty}^{e^*} 1/\widetilde{R}(m)dm$. 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

For the remainder of the paper we will discuss estimation using local polynomials. We use local polynomials instead of ordinary kernel or sieve estimators 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 in our estimation of the censored regression function and truncated regression function because we may be integrating over boundary regions in (2) and (5).

We shall use the following notation. For functions g and vectors $\mathbf{k} = (k_1, \ldots, k_d)$ and $x = (x_1, \ldots, x_d)$, let

$$\mathbf{k}! = k_1! \times \cdots \times k_d!, \ |\mathbf{k}| = \sum_{i=1}^{d} k_i, \ x^{\mathbf{k}} = x_1^{k_1} \times \cdots \times x_d^{k_d}$$

$$\sum_{0\le|\mathbf{k}|\le p} = \sum_{j=0}^{p} \sum_{k_1=0}^{j} \cdots \sum_{\substack{k_d=0 \\ k_1+\cdots+k_d=j}}^{j}, \ (D^{\mathbf{k}}g)(y) = \frac{\partial^{|k|}g(y)}{\partial y_1^{k_1} \cdots \partial y_d^{k_d}}.$$

To be consistent with our earlier usage of the subscript k, we will also use the special notation $g_k(x) = D^{e_k}g(x)$, where e_k is the k^{th} elementary vector, and $g_{k\ell}(x) = D^{(e_k+e_\ell)}g(x)$. We also stack the first derivatives into a vector so that $Dg(x) = (g_1(x), \ldots, g_d(x))'$.

4.1 Generic Nonparametric Regression Function and Derivatives

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

$$\sum_{i=1}^{n}\left[Y_i - \sum_{0\le|k|\le p} b_{\mathbf{k}}(x)(X_i - x)^{\mathbf{k}}\right]^2 \mathcal{K}((X_i - x)/h_n), \tag{7}$$

where $\mathcal{K}(u)$ is a nonnegative weight function on $\mathbb{R}^d$ and h_n is a bandwidth parameter, while p is an integer with $p \geq 2$. Minimizing (7) with respect to each $b_{\mathbf{k}}$ gives an estimate $\hat{b}_{\mathbf{k}}(x)$ such that $(D^{\mathbf{k}}g)\hat{}(x) = \mathbf{k}!\hat{b}_{\mathbf{k}}(x)$ estimates $(D^{\mathbf{k}}g)(x)$. Let also $\widehat{g}_k(x) = (D^{e_k}g)\widehat{}(x)$ and $\widehat{Dg}(x) = (\widehat{g}_1(x), \ldots, \widehat{g}_d(x))'$.

4.2 The Censored Regression Function

Let $\widehat{r}(x)$ be the nonparametric regression of Y_i on X_i, constructed as in (7). We then let

$$\widehat{m}(x) = \widehat{\lambda}_r - \int_{\widehat{r}(x)}^{\widehat{\lambda}_r} \frac{1}{\widehat{q}(r)} dr, \tag{8}$$

where $\widehat{q}(r)$ is the one-dimensional nonparametric regression of $I(Y_i > 0)$ on the generated regressor $\widehat{r}(X_i)$ evaluated at $\widehat{r}(X_i) = r$, while $\widehat{\lambda}_r = \max_{1\le i\le n} \widehat{r}(X_i)$. The integral can be evaluated numerically. We later show, under regularity conditions, that suitably centered $\widehat{m}(x)$ is asymptotically normal.

4.3 The Censored Regression Function Derivatives

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 (7), respectively, while X_i are the given covariates. We then let

$$\hat{m}_k(x) = \frac{\widehat{r}_k(x)}{\widehat{s}(x)}, \quad k = 1, \ldots, d. \tag{9}$$

4.4 Censored Regression Weighted Average Derivatives

Given any weighting function $w(x)$, the weighted average regression function derivative $\delta_{wk} = E[w(X)m_k(X)]/E[w(X)]$ is estimated by

$$\widehat{\delta}_{wk} = \frac{\sum_{i=1}^n w(x_i)\hat{m}_k(x_i)}{\sum_{i=1}^n w(x_i)}$$

Alternatively, the weighting function $w(x) = \widetilde{w}(x)/s(x)$ can be used, yielding the estimator

$$\widehat{\delta}_{wk} = \frac{\sum_{i=1}^n \widetilde{w}(x_i)\widehat{r}_k(x_i)}{\sum_{i=1}^n \widetilde{w}(x_i)\widehat{s}(x_i)}$$

which can have a simpler limiting distribution.

If the latent regression function has the partly linear form $m(x) = \beta_1 x_1 + ... + \beta_j x_j + \widetilde{m}(x_{j+1}, \ldots, x_k)$ for some $j \leq d$, then for $1 \leq k \leq j$, $\widehat{\beta}_k = \widehat{\delta}_{wk}$. Given regularity conditions, $\widehat{\delta}_{wk}$ is root n consistent and asymptotically normal, as in Powell, Stock, and Stoker (1989) or Härdle and Stoker (1989).

4.5 The Truncated Regression Function and its Derivatives

Let $\widehat{R}(x)$ be the nonparametric regression of Y_i on X_i constructed as in (7), but using only observations having $Y_i > 0$, that is, truncated data. Let $\widehat{U}(R)$ be a one-dimensional nonparametric regression of $Y_i^2/2$ on the generated regressor $\widehat{R}(X_i)$ evaluated at $\widehat{R}(X_i) = R$, again using only observations having $Y_i > 0$, and let $\widehat{U}'(R)$ be an estimator of the first derivative of that regression function. Then

$$\widehat{m}(x) = \widehat{\lambda}_R - \int_{\widehat{R}(x)}^{\widehat{\lambda}_R} \frac{\widehat{U}(s) - s\widehat{U}'(s)}{\widehat{U}(s) - s^2} ds, \tag{10}$$

where $\widehat{\lambda}_R = \max_{1 \leq i \leq n} \widehat{R}(X_i)$. Likewise,

$$\hat{m}_k(x) = \frac{\widehat{R}(x)\widehat{T}_k(x) - \widehat{T}(x)\widehat{R}_k(x)}{\widehat{R}(x)^2 - \widehat{T}(x)}, \quad k = 1, \ldots, d,$$

where $\widehat{T}(x)$ is the nonparametric regression of $Y^2/2$ on X, while $\widehat{R}_k(x)$ and $\widehat{T}_k(x)$ are the derivative estimators for continuously distributed elements x_k of x.

5 Asymptotic Properties

5.1 Assumptions

We first give some general definitions for our local polynomial kernel nonparametric regression estimators. Let

$$N_\ell = \left(\begin{array}{c} \ell + d - 1 \\ d - 1 \end{array} \right)$$

be the number of distinct d-tuples j with $|j| = \ell$. Arrange these N_ℓ d-tuples as a sequence in a lexicographical order (with highest priority to last position so that $(0, \ldots, 0, \ell)$ is the first element in the sequence and $(\ell, 0, \ldots, 0)$ the last element) and let ϕ_ℓ^{-1} denote this one-to-one map. Arrange the distinct values of $(D^{\mathbf{k}})\hat{}(g)$, $0 \leq |\mathbf{k}| \leq p$, as a column vector of dimension $N \times 1$, where $N = \sum_{\ell=0}^{p} N_\ell \times 1$, where the i^{th} element of that vector is obtained by the following relation

$$i = \phi_{|j|}^{-1}(j) + \sum_{k=0}^{|j|-1} N_k. \tag{11}$$

Similarly, arrange the vector $(D^{\mathbf{k}})(g)$. For each j with $0 \leq |j| \leq 2p$, let

$$\mu_j(\mathcal{K}) = \int_{\mathbb{R}^d} u^j \mathcal{K}(u)du, \;\; \nu_j(\mathcal{K}) = \int_{\mathbb{R}^d} u^j \mathcal{K}^2(u)du,$$

and define the $N \times N$ dimensional matrices M and Γ and $N \times 1$ vector B by

$$M = \begin{bmatrix} M_{0,0} & M_{0,1} & \cdots & M_{0,p} \\ M_{1,0} & M_{1,1} & \cdots & M_{1,p} \\ \vdots & & & \vdots \\ M_{p,0} & M_{p,1} & \cdots & M_{p,p} \end{bmatrix}, \; \Gamma = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & \Gamma_{0,p} \\ \Gamma_{1,0} & \Gamma_{1,1} & \cdots & \Gamma_{1,p} \\ \vdots & & & \vdots \\ \Gamma_{p,0} & \Gamma_{p,1} & \cdots & \Gamma_{p,p} \end{bmatrix}, \quad B = \begin{bmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{bmatrix}, \tag{12}$$

where $M_{i,j}$ and $\Gamma_{i,j}$ are $N_i \times N_j$ dimensional matrices whose (ℓ, m) element are, respectively, $\mu_{\phi_i(\ell)+\phi_j(m)}$ and $\nu_{\phi_i(\ell)+\phi_j(m)}$. Note that the elements of the matrices M and Γ are simply multivariate moments of the kernel $\mathcal{K}$ and $\mathcal{K}^2$, respectively. Finally, arrange the N_{p+1} elements of the derivatives $(1/j!)(D^j g)(x)$ for $|j| = p+1$ as a column vector $\mathcal{D}_{p+1}(x; g)$ using the lexicographical order introduced earlier.

For each j with $0 \leq |j| \leq 2p+1$ define the function

$$H_j(u) = u^j \mathcal{K}(u).$$

We make the following assumptions on the kernel $\mathcal{K}$ and on the data distribution. Assumptions A are used for the pointwise result, while assumption B contains the strengthening needed for our uniform convergence result.

Assumption A2

(a) 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).

(b) For all j with $0 \leq |j| \leq 2p+1$, there exists finite C_4 such that

$$|H_j(u) - H_j(v)| \leq C_4 ||u - v||.$$

Assumption A3.

(a) The regression functions r and s are $p+1$-times continuously differentiable.

(b) The conditional distribution $G(y|u)$ of Y given $X = u$ is continuous at the point $u = x$.

(c) $E\left[|Y_1|^2\right] < \infty$.

(d) 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 at the point x, and $f_X(x), s(x) > 0$.

Assumption B

(a) For any k with $|k| = p+1$, there exists finite C_6 such that

$$|(D^k r)(u) - (D^k r)(v)|, |(D^k s)(u) - (D^k s)(v)| \leq C_6 ||u - v||.$$

(b) $E\left[|Y_1|^t\right] < \infty$ for some $t > 2$.

(c) The density function f_X and the regression function s satisfy

$$\inf_{x \in \mathcal{X}} f_X(x) > 0 \quad ; \quad \inf_{x \in \mathcal{X}} s(x) > 0$$

on some compact subset $\mathcal{X}$ of $\mathbb{R}^d$. The functions σ_r^2, σ_s^2, and f_X are continuous on $\mathcal{X}$.

5.2 Distribution of Censored Regression Function Derivatives

We are now ready to give the asymptotic properties of our estimate $\widehat{Dm}(x)$ of $(Dm)(x)$ computed using our estimates $\widehat{Dr}(x)$ and $\widehat{s}(x)$.

Theorem 4 *Suppose that Assumptions A1-A3 hold and that $h_n = O(n^{-1/(d+2p+2)})$. Then, we have*

$$\sqrt{nh_n^{d+2}}\left[\left\{\widehat{Dm}(x) - Dm(x)\right\} - h_n^p \frac{(M^{-1}B\mathcal{D}_{p+1}(x;r))_1}{s(x)}\right] \Longrightarrow N\left[0,\ \frac{\sigma_r^2(x)}{f_X(x)s^2(x)}(M^{-1}\Gamma M^{-1})_{1,1}\right],$$

where $(M^{-1}\Gamma M^{-1})_{1,1}$ and $(M^{-1}B\mathcal{D}_{p+1}(x;r))_1$ are the corresponding [as in (12)] submatrix of $M^{-1}\Gamma M^{-1}$ and subvector of $M^{-1}B\mathcal{D}_{p+1}(x;r)$, respectively.
Suppose in addition that Assumption B holds, and that the bandwidth $h_n \to 0$ slowly enough such that the right hand side of (13) below is $o(1)$. Then, we have with probability one

$$\sup_{x \in \mathcal{X}} |\widehat{Dm}(x) - (Dm)(x)| = O\left\{\left(\frac{\ln n}{nh_n^{d+2}}\right)^{1/2}\right\} + O(h_n^p). \tag{13}$$

The proof of this theorem involves a standard linearization argument and application of Masry (1996a, Theorem 6) and Masry (1996b, Theorem 5), and is omitted.

Remarks A.

1. The optimal bandwidth for estimating the j^{th} first order partial derivative $(D^{e_j}m)(x)$ can be defined as the one which minimizes the sum of the squared bias and "variance" above; it is asymptotically

$$h_n^{opt} = n^{-1/(d+2p+2)}\left[\frac{2p\left(\frac{(M^{-1}B\mathcal{D}_{p+1}(x;r))_1}{s(x)}\right)^2}{(d+2)\frac{\sigma_r^2(x)}{f_X(x)s^2(x)}(M^{-1}\Gamma M^{-1})_{1,1}}\right]^{\frac{1}{2p+d+2}}.$$

The rate of "mean-square convergence" is then $O(n^{-2p/(d+2p+2)})$ which matches the optimal rate given by Stone (1980,1982) in the i.i.d. regression setting.

2. The quantity $s(x)$ measures the amount of censoring: when $s(x) = 1$ there is no censoring, while when $s(x) = 1/2$ there is 50% censoring. Both variance and bias deteriorate as $s(x)$ decreases, but $\widehat{Dm}$ is still consistent for any $s(x) > 0$ in contrast to any given nonparametric quantile estimator.

3. The asymptotic variance can easily be estimated from consistent estimates of $\sigma_r^2(x)$, $f_X(x)$, and $s^2(x)$, thus allowing consistent confidence to be constructed.

5.3 Distribution of The Censored Regression Function Estimator

We present this result for the local linear estimator [i.e., $p = 1$] with product kernels, i.e., we take $\mathcal{K}(u) = \prod_{\ell=1}^{d} K(u_\ell)$. We have the following theorem.

Theorem 5 *Suppose that Assumptions A1-A3 hold and that $r(x)$ has three continuous partial derivatives, and that $h_n \to 0$ and $\limsup_{n\to\infty} nh_n^{d+4} < \infty$. Then, there exists a bounded continuous function $b_m(x)$ such that*

$$\sqrt{nh_n^d}\left(\widehat{m}(x) - m(x) - h_n^2 b_m(x)\right) \Longrightarrow N\left(0, \frac{\sigma_r^2(x)}{f_X(x)q^2(r(x))}\nu_0(\mathcal{K})\right).$$

Note that the bias term is of smaller order provided $nh_n^{d+4} \to 0$. The asymptotic variance reflects the censoring through the function q. The asymptotic variance can be estimated from the estimates of $\sigma_r^2(x)$, $q(r)$, and $r(x)$.

5.4 Distribution of The Truncated Regression Function and Derivative Estimators

Define $\sigma_R^2(x) = \mathrm{var}(Y|X = x, Y > 0)$, $\sigma_T^2(x) = \mathrm{var}(Y^2/2|X = x, Y > 0)$, and $\sigma_{TR}(x) = \mathrm{cov}(Y, Y^2/2|X = x, Y > 0)$.

Theorem 6 *Suppose that Assumptions A1-A3 hold except that $R(x)$ and $T(x)$ have three continuous partial derivatives, and that $h_n \to 0$ and $\limsup_{n\to\infty} nh_n^{d+4} < \infty$. Then there exists some bounded continuous function $b_V^T(x)$ such that*

$$\sqrt{nh_n^{d+2}}\left(\widehat{Dm}(x) - Dm(x) - h_n^2 b_V^T(x)\right) \Longrightarrow N\left(0, v(x)\frac{\nu_2(\mathcal{K})}{\mu_2^2(\mathcal{K})}I_d\right),$$

where

$$v(x) = \frac{R(x)^2\sigma_T^2(x) + T(x)^2\sigma_R^2(x) - 2R(x)T(x)\sigma_{TR}(x)}{(R(x)^2 - T(x))^2 f_X(x)}.$$

Furthermore, there exists a bounded continuous function $b^T(x)$ such that

$$\sqrt{nh_n^d}\left(\widehat{m}(x) - m(x) - h_n^2 b^T(x)\right) \Longrightarrow N\left(0, \left(\frac{U(R(x)) - R(x)U'(R(x))}{U(R(x)) - R(x)^2}\right)^2 \frac{\sigma_R^2(x)}{f_X(x)}\nu_0(\mathcal{K})\right).$$

6 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 $e \sim N(0, \frac{1}{4})$. Given this design, the amount of censoring as a function of x is given by $1 - \Phi(2x^3)$, 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.

As described in the text, the censored regression and censored derivative estimators are

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

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 functions $m(x)$ and $m_k(x)$ are also estimated using quantile regression and quantile derivative estimation, as follows. The conditional empirical distribution function is 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 $\widehat{m}(x)$.

The quantile derivative estimator is obtained by taking the total derivative of

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

which yields

$$\widehat{m}_{qk}(x) = \frac{dy_q}{dx} = \frac{h_2}{h_1} \frac{\sum_{i=1}^{n} \phi'\left(\frac{x-X_i}{h_1}\right) \Phi\left(\frac{y_q-Y_i}{h_2}\right) - \sum_{i=1}^{n} \phi\left(\frac{x-X_i}{h_1}\right) \Phi\left(\frac{y_q-Y_i}{h_2}\right) \sum_{i=1}^{n} \phi'\left(\frac{x-X_i}{h_1}\right)}{\sum_{i=1}^{n} \phi\left(\frac{x-X_i}{h_1}\right) \sum_{i=1}^{n} \phi^2\left(\frac{x-X_i}{h_1}\right)},$$

where $\phi'(\cdot)$ is the derivative of normal density function.

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 $q = .5$. 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 $\widehat{m}(x) \to m(x) + 1 - \mathfrak{F}^{-1}(1)$, where the function $\mathfrak{F}(e^*)$ equals the integral from $-\infty$ to e^* of the distribution function of a normal having mean zero, variance one fourth. The location bias is therefore given by $1 - \mathfrak{F}^{-1}(1)$. 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. In contrast, the median regression is inconsistent in that region, centering on zero. Experiments (not reported) using lower quantiles, e.g., $q = .25$, 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 estimator (through α_q). Our estimator does not require arbitrary selection of a quantile, remains consistent everywhere inside of the support of x, and has location determined by $E(e) = 0$.

Figures 3 and 4 show the same information for the derivative estimators $\widehat{m}_k(x)$ and $\widehat{m}_{.5k}(x)$. The sample size $n = 200$ is rather small for nonparametric derivative estimation, which is reflected in wide confidence bands and flattening in the tails.

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.

7 Extensions and Conclusions

We have provided estimators for the nonparametric censored and truncated regression models with fixed censoring. Our estimator could also be used if 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 therefore permit the variable C_i to affect Y_i like any other regressor in X_i, in addition to determining the censoring point.

We conclude with some extensions.

7.1 Additional Moments

The estimators we provide are based on the conditional means $E(Y^\kappa|X = x)$ for low integers k. Moments of other functions of Y could also be employed. Let $\phi(y)$ be a differentiable function having $\phi(0) = 0$, and let $\phi'(y) = \partial\phi(y)/\partial y$. Theorem 1 can be extended to

$$\frac{\partial E[\phi(Y)I(Y > 0)|m(X) = m(x)]}{\partial m(x)} = E[\phi'(Y)I(Y > 0)|m(X) = m(x)]$$

The conclusions of Theorems 1, 2, and 5 will then hold, replacing the functions r, s, and q, with

$$\begin{aligned} r(x) &= E[\phi(Y)I(Y > 0)|X = x], \\ s(x) &= E[\phi'(Y)I(Y > 0)|X = x] \\ q[r(x)] &= E[\phi'(Y)I(Y > 0)|r(X) = r(x)], \end{aligned}$$

Different choices of the function ϕ might yield more efficient estimators. In particular, by Theorem 5, to maximize efficiency we would want to choose ϕ to minimize $\sigma_r^2(x)/q^2(r(x))$. Alternatively, estimates using different ϕ functions might be combined to increase efficiency, or compared to test the model. For example, letting $\phi(y) = y^\kappa$ for $\kappa > 1$ would yield estimates based on higher moments of y.

7.2 Heteroscedastic errors

Assume $F(e|x) = F[e|w(x)]$ and $E[e|w(x)] = 0$ for some known, vector valued function w. Assume $\text{supp}[e|w(x)] = \text{supp}(e) \subseteq \text{supp}[m(x)|w(x)]$. This allows for very general forms of heteroscedasticity, for example, $w(x)$ could equal the vector of all of the regressors except for one (continuously distributed one), so the errors could depend in an arbitrary, unknown way on all but one of the regressors.

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

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

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

$$m(x) = \lambda - \int_{r(x)}^{\lambda} \frac{1}{q[r, w(x)]} dr \tag{14}$$

The estimator based on this equation is identical to the homoscedastic estimator, except that $\widehat{q}$ will be a nonparametric regression on $\widehat{r}$ and on w.

7.3 A Feasible Buckley-James Transform

For any e^*, let $\widehat{F}(e^*)$ be the nonparametric regression of $I(Y_i > 0)$ on $\widehat{m}(X_i)$ evaluated at the point e^*. We may then define a feasible B-J transform

$$\widehat{Y}_i^{BJ} = \delta_i Y_i + (1 - \delta_i) \frac{\int_{\widehat{m}(X_i)}^{\widehat{\lambda}} e \cdot d\widehat{F}(e)}{\int_{\widehat{m}(X_i)}^{\widehat{\lambda}} d\widehat{F}(e)}, \tag{15}$$

and apply local linear regression to the observations $\{\widehat{Y}_i^{BJ}, X_i\}$. The integration in (15) can be done numerically. This local linear regression is then a revised estimate of m, denoted $\widehat{m}^{BJ}$. This process can be repeated until some convergence criterion is satisfied or one can just take a finite number of steps; since the starting point is a consistent estimate of m, F, asymptotically only one-step should be required. See Rothenberg and Leenders (1965) and Bickel (1975). Given the known advantages of parametric of Buckley-James estimators, $\widehat{m}^{BJ}$ may have better small sample or asymptotic properties than $\widehat{m}$. See van Keilegom and Akritas (1999) for some analysis of the Kaplan-Meier estimator constructed from nonparametric residuals.

A Appendix

We first give some facts and definitions for the generic local linear estimators $\widehat{g}(x), \widehat{g}_k(x)$ of a regression function $g(x)$ [of $Y|X$] and its partial derivative $g_k(x)$, which will be needed in the proof of Theorems 5 and 6. We write $\widehat{g}(x) - g(x) = e_0' M_n^{-1}(x) U_n(x) + e_0' M_n^{-1}(x) B_n(x)$ and $\widehat{g}_k(x) - g_k(x) = h_n^{-1} e_k' M_n^{-1}(x) U_n(x) + h_n^{-1} e_k' M_n^{-1}(x) B_n(x)$, where $e_k = (0, 0, \ldots, 0, 1, 0, \ldots, 0)'$ is the $d+1$ vector with the one in the $k+1$ position. Here, the $(d+1) \times (d+1)$ symmetric matrix $M_n(x)$ is

$$M_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} \begin{bmatrix} \mathcal{K}\left(\frac{x-X_i}{h_n}\right) & \mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_1-X_{1i}}{h_n}\right) & \cdots & \mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_d-X_{di}}{h_n}\right) \\ & \mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_1-X_{1i}}{h_n}\right)^2 & \cdots & \mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_1-X_{1i}}{h_n}\right)\left(\frac{x_d-X_{di}}{h_n}\right) \\ & & \ddots & \vdots \\ & & & \mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_d-X_{di}}{h_n}\right)^2 \end{bmatrix}.$$

The stochastic term $U_n(x)$ is the $d+1\times 1$ vector

$$U_n(x)=\begin{bmatrix}\frac{1}{nh_n^d}\sum_{i=1}^n\mathcal{K}\left(\frac{x-X_i}{h_n}\right)\epsilon_i\\ \frac{1}{nh_n^d}\sum_{i=1}^n\mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_1-X_{1i}}{h_n}\right)\epsilon_i\\ \vdots\\ \frac{1}{nh_n^d}\sum_{i=1}^n\mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_d-X_{di}}{h_n}\right)\epsilon_i\end{bmatrix}\equiv\begin{bmatrix}U_{n0}(x)\\ U_{n1}(x)\\ \vdots\\ U_{nd}(x)\end{bmatrix},$$

where $\epsilon_i=Y_i-g(X_i)$ is the error term that satisfies $E(\epsilon_i|X_i)=0$ a.s.; the bias term is the $d+1\times 1$ vector

$$B_n(x)=\begin{bmatrix}\frac{1}{nh_n^d}\sum_{i=1}^n\mathcal{K}\left(\frac{x-X_i}{h_n}\right)\Delta_i(x)\\ \frac{1}{nh_n^d}\sum_{i=1}^n\mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_1-X_{1i}}{h_n}\right)\Delta_i(x)\\ \vdots\\ \frac{1}{nh_n^d}\sum_{i=1}^n\mathcal{K}\left(\frac{x-X_i}{h_n}\right)\left(\frac{x_d-X_{di}}{h_n}\right)\Delta_i(x)\end{bmatrix}\equiv\begin{bmatrix}B_{n0}(x)\\ B_{n1}(x)\\ \vdots\\ B_{nd}(x)\end{bmatrix},$$

where $\Delta_i(x)=g(X_i)-g(x)-\sum_{k=1}^d g_k(x)(X_{ki}-x_k)$.

Let $B_0(x)=\frac{1}{2}\mu_2(K)\sum_{j=1}^d g_{jj}(x)$ and $B_k(x)=\frac{1}{6}\sum_{j=1}^d\sum_{l=1}^d\sum_{m=1}^d\int\mathcal{K}(u)u_ku_ju_lu_m du\times\{3g_{jl}(x)f_m(x)+g_{jlm}(x)f(x)\}$, $k=1,\ldots,d$, where f is the marginal density of the covariates. Some of our results must allow for x in the boundary region; in this case, the range of integration in the kernel moments depends on x. For example, the matrix M defined in (12) depends on x when x is in the boundary region; however, since $\int_a^b K(u)du>0$ for any $a<b$ contained in the support of the kernel, the resulting matrix is positive definite for all x. In the sequel we have avoided explicitly writing out this complication for notational simplicity.

We have the following results:

$$\sup_x|M_n(x)-f(x)M|=O_p(h_n)+O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right)$$

$$\sup_x\left|B_{n0}(x)-\frac{h_n^2}{2}B_0(x)\right| = o_p(h_n^2) \tag{16}$$

$$\sup_x\left|B_{nk}(x)-h_n^3B_k(x)\right| = o_p(h_n^3), \tag{17}$$

which follow from the results of Masry (1996a).

A.1 Main Result

PROOF OF THEOREM 5. The proof is based on the series of lemmas given below. Write $\widehat{q}(s)=\widehat{q}(s;\widehat{r}_1,\ldots,\widehat{r}_n)$, where $\widehat{r}_j=\widehat{r}(X_j)$, and define also $\widehat{q}(s;r_1,\ldots,r_n)$, where $r_j=r(X_j)$, to be the one-dimensional nonparametric regression of $I(Y_i > 0)$ on the true regressor $r(X_i)$ evaluated at $r(X_j) = s$. We let M_{nr} and M_{nq} denote the matrices M_n defined in the previous section when the regression functions are r and q respectively. In the local linear case, the limiting matrices M are both diagonal. Similarly let U_{nr}, U_{nq}, B_{nr}, and B_{nq} denote the stochastic and bias terms in the corresponding regressions. Then define the regression errors $\varepsilon_i = Y_i - r_i$ and $u_i = \mathbf{1}(Y_i > 0) - q(r_i)$, where $E(\varepsilon_i|X_i) = 0$ and $E(u_i|r_i) = 0$. Let $\mathcal{F}_X$ and $\mathcal{F}_r$ be the sigma algebras generated by X and $r(X)$ respectively. Since $\mathcal{F}_X \supseteq \mathcal{F}_r$ we have $E(\varepsilon_i|r_i) = 0$ by the tower property of conditional expectations, see Billingsley (1986, Theorem 34.3). However, $E(u_i|X_i) \neq 0$. Therefore, we write $u_i = g_u(X_i) + \eta_i$, where $E(\eta_i|X_i) = 0$ by construction. Define also the conditional moments $\sigma_\eta(X_i) = E(\eta_i^2|X_i)$, $\sigma_{\varepsilon\eta}(X_i) = E(\varepsilon_i\eta_i|X_i)$, $\sigma_\varepsilon^2(X_i) = E(\varepsilon_i^2|X_i)$, and $\sigma_u^2(X_i) = E(u_i^2|X_i)$.

Rearranging terms, we have

$$\begin{aligned}\widehat{m}(x) - m(x) &= \widehat{\lambda}_r - \int_{\widehat{r}(x)}^{\widehat{\lambda}_r} \frac{1}{\widehat{q}(s)} ds - \left(\lambda_r - \int_{r(x)}^{\lambda_r} \frac{1}{q(s)} ds\right) \\ &= \left(\widehat{\lambda}_r - \lambda_r\right) - \left(\int_{\widehat{r}(x)}^{\widehat{\lambda}_r} - \int_{r(x)}^{\lambda_r}\right) \frac{1}{q(s)} ds + \int_{\widehat{r}(x)}^{\widehat{\lambda}_r} \left(\frac{\widehat{q}(s) - q(s)}{\widehat{q}(s)q(s)}\right) ds.\end{aligned}$$

By mean value expansions we obtain

$$\begin{aligned}\widehat{m}(x) - m(x) &= \left(1 - \frac{1}{q(\lambda_r)}\right)(\widehat{\lambda}_r - \lambda_r) + \frac{1}{q(r(x))}(\widehat{r}(x) - r(x)) + \int_{r(x)}^{\lambda_r} \frac{(\widehat{q}(s) - q(s))}{q^2(s)} ds && (18)\\ &+ \frac{\widehat{q}'(\overline{\lambda})}{2\widehat{q}^2(\overline{\lambda})}(\widehat{\lambda}_r - \lambda_r)^2 - \frac{\widehat{q}'(\overline{r}(x))}{2\widehat{q}^2(\overline{r}(x))}(\widehat{r}(x) - r(x))^2 - \int_{\widehat{r}(x)}^{\widehat{\lambda}_r} \frac{(\widehat{q}(s) - q(s))^2}{\widehat{q}(s)q^2(s)} ds && (19)\\ &- \frac{\widehat{q}(\overline{\lambda}_r) - q(\overline{\lambda}_r)}{\widehat{q}(\overline{\lambda}_r)q(\overline{\lambda}_r)}(\widehat{\lambda}_r - \lambda_r) + \frac{\widehat{q}(\overline{r}(x)) - q(\overline{r}(x))}{\widehat{q}(\overline{r}(x))q(\overline{r}(x))}(\widehat{r}(x) - r(x)), && (20)\end{aligned}$$

where $\overline{\lambda}$ and $\overline{r}(x)$ are intermediate values [they are not necessarily the same in the two expressions, but we have adopted this for notational convenience]. The terms in (18) are all linear in the estimation error from the two nonparametric regressions, while the terms (19) and (20) are both quadratic in such errors, and can thus be expected to be of smaller order. Since $q(\lambda_r) = 1$, the first term in (18) is zero. The second term is just a constant times the estimation error of $\widehat{r}(x)$, and can be analyzed directly from the results of Masry (1996ab). To analyze the third term we make another Taylor series expansion

$$\begin{aligned}\int_{r(x)}^{\lambda_r} \frac{(\widehat{q}(s) - q(s))}{q^2(s)} ds &= \int_{r(x)}^{\lambda_r} \frac{(\widehat{q}(s; r_1, \ldots, r_n) - q(s))}{q^2(s)} ds + \sum_{j=1}^{n} (\widehat{r}_j - r_j) \int_{r(x)}^{\lambda_r} \frac{\partial \widehat{q}(s; r_1, \ldots, r_n)}{\partial r_j} \frac{ds}{q^2(s)} \\ &+ \frac{1}{2} \sum_{j=1}^{n} \sum_{l=1}^{n} (\widehat{r}_j - r_j)(\widehat{r}_l - r_l) \int_{r(x)}^{\lambda_r} \frac{\partial^2 \widehat{q}(s; \overline{r}_1, \ldots, \overline{r}_n)}{\partial r_j \partial r_l} \frac{ds}{q^2(s)}, && (21)\end{aligned}$$

where $\overline{r}_j$ are intermediate values. Denote (21) by R_{n1}, and the quadratic terms in (19)-(20) by R_{n2}-R_{n6}, and let $\mathcal{R}_n = \sum_{j=1}^{6} R_{nj}$. We have obtained the second order expansion

$$\begin{aligned} \widehat{m}(x) - m(x) &= \frac{1}{q(r(x))}\left(\widehat{r}(x) - r(x)\right) + \int_{r(x)}^{\lambda_r} \frac{\left(\widehat{q}(s; r_1, \ldots, r_n)\right) - q(s))}{q^2(s)} ds \\ &\quad + \sum_{j=1}^{n} (\widehat{r}_j - r_j) \int_{r(x)}^{\lambda_r} \frac{\partial \widehat{q}(s; r_1, \ldots, r_n)}{\partial r_j} \frac{ds}{q^2(s)} + R_n \\ &\equiv \mathcal{A}_n + \mathcal{B}_n + \mathcal{C}_n + \mathcal{R}_n. \end{aligned} \tag{22}$$

Let $\delta_n = \max\{1/\sqrt{nh_n^d}, h_n^2\}$. The proof of our theorem consists of evaluating the magnitudes of the terms $\mathcal{A}_n, \mathcal{B}_n$, and $\mathcal{C}_n$, and then the remainder term $\mathcal{R}_n$.

LEMMA 1. *There exists a bounded continuous function $b_a(x)$ such that*

$$\sqrt{nh_n^d}\left(\mathcal{A}_n - h_n^2 b_a(x)\right) \Longrightarrow N\left(0, \frac{\sigma_r^2(x)}{q^2(r(x)) f_X(x)} \nu_0(\mathcal{K})\right).$$

We next consider the terms $\mathcal{B}_n$ and $\mathcal{C}_n$. For this we need the following decompositions for $\widehat{q}(s; r_1, \ldots, r_n)$ and $\widehat{r}_j$: $\widehat{q}(s; r_1, \ldots, r_n) - q(s) = e_0' M_{nr}^{-1}(s) U_{nr}(s) + e_0' M_{nr}^{-1}(s) B_{nr}(s)$ and $\widehat{r}_j - r_j = e_0' M_{nq}^{-1}(X_j) U_{nq}(X_j) + e_0' M_{nq}^{-1}(X_j) B_{nq}(X_j)$. Note that the matrices $M_{nq}(X_j)$ and $M_{nr}(s)$ are measurable functions of $X_1, \ldots, X_n$. The term $\mathcal{B}_n$ is just an integral of a one dimensional smoother and its variance will be of order n^{-1}, although its bias is $O(h_n^2)$.

LEMMA 2. *As $n \to \infty$ we have*

$$\left| \mathcal{B}_n - h_n^2 \cdot \int_{r(x)}^{\lambda_r} \frac{B_{q0}(s)}{q^2(s)} ds \right| = o_p(h_n^2).$$

We now turn to the term $\mathcal{C}_n$. Note that

$$\begin{aligned} \frac{\partial \widehat{q}(s; r_1, \ldots, r_n)}{\partial r_i} &= e_0' M_{nr}^{-1}(s) \frac{\partial U_{nr}(s)}{\partial r_i} + e_0' M_{nr}^{-1}(s) \frac{\partial B_{nr}(s)}{\partial r_i} \\ &\quad - e_0' M_{nr}^{-1}(s) \frac{\partial M_{nr}(s)}{\partial r_i} M_{nr}^{-1}(s) [U_{nr}(s) + B_{nr}(s)], \end{aligned}$$

where

$$\frac{\partial M_{nr}(s)}{\partial r_i} = \frac{1}{nh_n^2}\begin{bmatrix} K_i'(s) & L_i'(s)+K_i(s) \\ L_i'(s)+K_i(s) & J_i'(s)+2L_i(s) \end{bmatrix} \quad ;$$

$$\frac{\partial U_{nr}(s)}{\partial r_i} = \frac{1}{nh_n^2}\begin{bmatrix} K_i'(s) \\ L_i'(s)+K_i(s) \end{bmatrix} u_i - \frac{1}{nh_n}\begin{bmatrix} K_i(s) \\ L_i(s) \end{bmatrix} q'(r_i) \equiv \frac{\partial U_{nr}^{+}(s)}{\partial r_i} + \frac{\partial U_{nr}^{-}(s)}{\partial r_i}$$

$$\frac{\partial B_{nr}(s)}{\partial r_i} = \frac{1}{nh_n^2}\begin{bmatrix} K_i'(s) \\ L_i'(s)+K_i(s) \end{bmatrix} \Delta_i(s) + \frac{1}{nh_n^2}\begin{bmatrix} K_i(s) \\ L_i(s) \end{bmatrix} \Delta_i'(s)$$

where $K_i(s) = K((s-r_i)/h_n)$, $K_i'(s) = K'((s-r_i)/h_n)$, $L_i(s) = K((s-r_i)/h_n)((s-r_i)/h_n)$, $L_i'(s) = K'((s-r_i)/h_n)((s-r_i)/h_n)$, and $J_i'(s) = K'((s-r_i)/h_n)((s-r_i)/h_n)^2$, while $\Delta_i'(s) = q'(r_i) - q'(s)$. Now, substitute into the definition of $\mathcal{C}_n$ the three terms constituting $\partial\widehat{q}(s; r_1, \ldots, r_n)/\partial r_j$; also write $\partial U_{nr}^{+}(s)/\partial r_i = \partial U_{nr}^{g}(s)/\partial r_i + \partial U_{nr}^{\eta}(s)/\partial r_i$, where: $\partial U_{nr}^{g}(s)/\partial r_i$ is like $\partial U_{nr}^{+}(s)/\partial r_i$ with $g_u(X_i)$ substituting for u_i, and $\partial U_{nr}^{\eta}(s)/\partial r_i$ is like $\partial U_{nr}^{+}(s)/\partial r_i$ with η_i substituting for u_i. With these definitions we can now divide C_n into four pieces, i.e., $\mathcal{C}_n = \mathcal{C}_{n1} + \mathcal{C}_{n2} + \mathcal{C}_{n3} + \mathcal{C}_{n4}$, where:

$$\begin{aligned} \mathcal{C}_{n1} &= \sum_{j=1}^{n} e_0' M_{nq}^{-1}(X_j) U_{nq}(X_j) \int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s) \frac{\partial U_{nr}^{g}(s)}{\partial r_j} \frac{1}{q^2(s)} ds \\ &\quad + \sum_{j=1}^{n} e_0' M_{nq}^{-1}(X_j) B_{nq}(X_j) \int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s) \frac{\partial U_{nr}^{g}(s)}{\partial r_j} \frac{1}{q^2(s)} ds \equiv \mathcal{C}_{n11} + \mathcal{C}_{n12}, \end{aligned}$$

and $\mathcal{C}_{n2} \equiv \mathcal{C}_{n21} + \mathcal{C}_{n22}$ is like $\mathcal{C}_{n1}$ but with $\partial U_{nr}^{\eta}(s)/\partial r_j$ replacing $\partial U_{nr}^{g}(s)/\partial r_j$, while $\mathcal{C}_{n3} \equiv \mathcal{C}_{n31} + \mathcal{C}_{n32}$ is like $\mathcal{C}_{n1}$ but with $\partial B_{nr}(s)/\partial r_j$ replacing $\partial U_{nr}^{g}(s)/\partial r_j$. Finally,

$$\mathcal{C}_{n4} = \sum_{j=1}^{n} (\widehat{r}_j - r_j) \times \int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s) \frac{\partial U_{nr}^{-}(s)}{\partial r_i} \frac{1}{q^2(s)} ds.$$

$$\mathcal{C}_{n5} = \sum_{j=1}^{n} (\widehat{r}_j - r_j) \times \int_{r(x)}^{\lambda_r} \frac{1}{q^2(s)} e_0' M_{nr}^{-1}(s) \frac{\partial M_{nr}(s)}{\partial r_j} M_{nr}^{-1}(s) \left[U_{nr}(s) + B_{nr}(s)\right] ds.$$

The properties of $\mathcal{C}_n$ and $\mathcal{R}_n$ are given in the following lemmas, which are proved below.

Lemma 3. *Then*: (1) $\mathcal{C}_{n11} = o_p(\delta_n)$; (2)

$$\mathcal{C}_{n12} = h_n^2 \left[\frac{E\left(B_{q0}(X) g_u(X) \,|r(X) = \lambda_r\right)}{q^2(\lambda_r)} - \frac{E\left(B_{q0}(X) g_u(X) \,|r(X) = r(x)\right)}{q^2(r(x))}\right] + o_p(\delta_n);$$

(3) $\mathcal{C}_{n21} = o_p(\delta_n)$; (4) $\mathcal{C}_{n22} = o_p(\delta_n)$; (5) $\mathcal{C}_{n31} = o_p(\delta_n)$; (6) $\mathcal{C}_{n32} = o_p(\delta_n)$; (7) $o_p(\delta_n)$

$$\mathcal{C}_{n4} = -h_n^2 \int_{r(x)}^{\lambda_r} E\left(B_{q0}(X) \,|r(X) = s\right) \frac{q'(s)}{q^2(s)} ds + o_p(\delta_n);$$

(8) $\mathcal{C}_{n5} = o_p(\delta_n)$.

Lemma 4. $\mathcal{R}_n = o_p(\delta_n)$.

A.2 Proofs of Lemmas

Denote by E_X and var_X the conditional expectation and variance given $X_1, \ldots, X_n$, respectively; likewise let E_r and var_r denote the conditional expectation and variance given $r_1, \ldots, r_n$, respectively. For any random sequences X_n, Y_n we write $X_n \simeq Y_n$ whenever $X_n = Y_n + o_p(Y_n)$.

Proof of Lemma 1. This follows by Theorem 6 of Masry (1996b), since $q(r(x)) > 0$. ■

Proof of Lemma 2. We first write $\mathcal{B}_n = \mathcal{B}_{n1} + \mathcal{B}_{n2}$, where $\mathcal{B}_{n1} = \int_{r(x)}^{\lambda_r} q^{-2}(s)e_0' M_{nr}^{-1}(s)U_{nr}(s)ds$ and $\mathcal{B}_{n2} = \int_{r(x)}^{\lambda_r} q^{-2}(s)e_0' M_{nr}^{-1}(s)B_{nr}(s)ds$. The term $\mathcal{B}_{n1}$ is, conditionally on $X_1, \ldots, X_n$, a sum of mean zero independent random variables. We have $E_r(\mathcal{B}_{n1}) = 0$, while

$$\text{var}_r(\mathcal{B}_{n1}) = \frac{1}{n^2}\sum_{i=1}^n \sigma_u^2(r_i)\left(\frac{1}{h_n}\int_{r(x)}^{\lambda_r}\frac{1}{q^2(s)}e_0' M_{nr}^{-1}(s)\begin{bmatrix} K_i(s) \\ L_i(s) \end{bmatrix} ds\right)^2.$$

Now note that for any vectors a, b, and real symmetric matrix A, we have $|a'A^{-1}b| \leq (a'A^{-1}a)^{1/2}(b'A^{-1}b)^{1/2}$ $\leq \lambda_{\max}(A^{-1})(a'a)^{1/2}(b'b)^{1/2}$, and $\lambda_{\max}(A^{-1}) = \lambda_{\min}^{-1}(A)$. The matrix $M_{nr}(s)$ is real and symmetric. Therefore,

$$\begin{aligned}\left|\frac{1}{h_n}\int_{r(x)}^{\lambda_r}\frac{1}{q^2(s)}e_0' M_{nr}^{-1}(s)\begin{bmatrix} K_i(s) \\ L_i(s) \end{bmatrix} ds\right| &\leq \frac{1}{h_n}\int_{r(x)}^{\lambda_r}\frac{1}{q^2(s)}\left|e_0' M_{nr}^{-1}(s)\begin{bmatrix} K_i(s) \\ L_i(s) \end{bmatrix}\right| ds \\ &\leq \frac{1}{\inf_s \lambda_{\min}(M_{nr}(s))}\frac{1}{h_n}\int_{r(x)}^{\lambda_r}\frac{\left(|K_i(s)|^2 + |L_i(s)|^2\right)^{1/2}}{q^2(s)}ds.\end{aligned}$$

Furthermore, $\inf_s \lambda_{\min}(M_{nr}(s)) \geq \inf_s \lambda_{\min}(f_r(s)M) - \sup_s |\lambda_{\max}(M_{nr}(s) - f_r(s)M)|$, and $M_{nr}(s)$ converges uniformly to the matrix $Mf_r(s)$, so that $\inf_s \lambda_{\min}(M_{nr}(s)) \geq \inf_s \lambda_{\min}(f_r(s)M) + o_p(1)$. Finally, the matrix M is positive definite, while $\inf_s f_r(s) > 0$ and $\inf_s q(s) > 0$. Therefore, there is some finite positive constant c such that with probability tending to one

$$\left|\frac{1}{h_n}\int_{r(x)}^{\lambda_r}\frac{1}{q^2(s)}e_0' M_{nr}^{-1}(s)\begin{bmatrix} K_i(s) \\ L_i(s) \end{bmatrix} ds\right| \leq c\int_{\frac{r(x)-r_i}{h_n}}^{\frac{\lambda_r - r_i}{h_n}}\left(|K(t)|^2 + |L(t)|^2\right)^{1/2} dt, \tag{23}$$

where have applied the change of variables $s \longmapsto t = (s - r_i)/h_n$ and dominated convergence. In conclusion, $\text{var}_r(\mathcal{B}_{n1}) = O_p(n^{-1})$ and so $\mathcal{B}_{n1} = O_p(n^{-1/2})$.

The term $\mathcal{B}_{n2}$ just depends on $X_1, \ldots, X_n$. We replace $M_{nr}^{-1}(s)$ and $B_{nr0}(s)$ by their probability limits $[f_r^{-1}(s)M^{-1}$ and $h_n^2 B_{r0}(s)]$, and obtain

$$\mathcal{B}_{n2} = h_n^2\int_{r(x)}^{\lambda_r}\frac{B_{r0}(s)}{f_r(s)q^2(s)}ds + o_p(h_n^2).$$

Again, this is justified by dominated convergence and the uniform convergence. ■

Proof of Lemma 3.1. Let

$$\varrho_{ni} = \frac{1}{h_n}\int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s)\begin{bmatrix} K_i'(s) \\ L_i'(s) + K_i(s) \end{bmatrix}\frac{1}{q^2(s)}ds$$

$$Z_{ni} \;=\; \frac{1}{n^2 h_n^{d+1}} \sum_{j=1}^{n} \mathcal{K}\left(\frac{X_j - X_i}{h_n}\right) g_u(X_j) e_0' M_{nq}^{-1}(X_j) v_{ji} \varrho_{nj},$$

where $v_{ji} = (1, (X_{1j} - X_{1i})/h_n, \ldots, (X_{dj} - X_{di})/h_n)'$. We can now write $\mathcal{C}_{n11} = \sum_{i=1}^{n} \varepsilon_i Z_{ni}$, where Z_{ni} depends only on $X_1, \ldots, X_n$. Therefore, conditionally on $X_1, \ldots, X_n$, $\mathcal{C}_{n11}$ is a sum of independent random variables with mean zero and variance $\sum_{i=1}^{n} \sigma_{\varepsilon}^2(X_i) Z_{ni}^2$.

We next bound the terms in Z_{ni}. We have that

$$\left| e_0' M_{nq}^{-1}(X_j) v_{ji} \right| \leq \frac{1}{\min_{1 \leq j \leq n} \lambda_{\min}(M_{nq}(X_j))}$$

on the set where $\mathcal{K}((X_j - X_i)/h_n \neq 0$ using the inequality of the previous lemma. The matrix $M_{nq}(X_j)$ is real and symmetric and has a vanishingly small probability of being singular. Furthermore, $M_{nq}(x)$ converges to $f_X(x)M$ uniformly in x, and so by the continuous mapping theorem $\min_{1 \leq j \leq n} \lambda_{\min}(M_{nq}(X_j))$ converges in probability to $\lambda_{\min}(M) \times \inf_x f_X(x)$, which is bounded away from zero. Therefore, we have found a constant c such that with probability tending to one $\left| e_0' M_{nq}^{-1}(X_j) v_{ji} \right| \leq c$ for all i, j such that $\mathcal{K}((X_j - X_i)/h_n \neq 0$.

Also write

$$\varrho_{ni} = \frac{1}{h_n} \int_{r(x)}^{\lambda_r} \frac{K'(\frac{s - r_i}{h_n})}{f_r(s) q^2(s)} ds - \frac{1}{h_n} \int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s) \left[M_{nr}(s) - f_r(s) M \right] M^{-1} \begin{bmatrix} K_i'(s) \\ L_i'(s) + K_i(s) \end{bmatrix} \frac{ds}{f_r(s) q^2(s)}$$

$$\equiv \varrho_n^0(r_i) + \varrho_{ni}^1.$$

Using integration by parts and change of variables, we have

$$\varrho_n^0(r_i) = \frac{K\left(\frac{\lambda_r - r_i}{h_n}\right)}{f_r(\lambda_r) q^2(\lambda_r)} - \frac{K\left(\frac{r(x) - r_i}{h_n}\right)}{f_r(r(x)) q^2(r(x))} - \int_{r(x)}^{\lambda_r} K\left(\frac{s - r_i}{h_n}\right) \left(\frac{1}{f_r(s) q^2(s)}\right)' ds. \tag{24}$$

Clearly, the first two terms in $\varrho_n^0(r_i)$ are $O_p(h_n)$, while the last term in (24) is also of this order, which can easily be shown by change of variables argument. We will also replace ϱ_{nj}^1 by an upper bound that only depends on r_j and n, thus for some constant c

$$\left| \varrho_{nj}^1 \right| \leq c \int_{r(x)}^{\lambda_r} \left(K'\left(\frac{s - r_j}{h_n}\right)^2 + L'\left(\frac{s - r_j}{h_n}\right)^2 \left(\frac{s - r_j}{h_n}\right)^2 + K\left(\frac{s - r_j}{h_n}\right)^2 \right)^{1/2} ds \equiv \overline{\varrho}_n^1(r_j)$$

with probability tending to one. This uses the fact that $M_{nr}(s)$ converges uniformly to $f_r(s)M$ with rate no worse than h_n and so the elements of $M_{nr}^{-1}(s)[M_{nr}(s) - f_r(s)M]M^{-1}$ are all bounded by some constant times h_n with probability tending to one. Combining these relations and using the triangle inequality, we have for some finite c on a set whose probability tends to one,

$$\sum_{i=1}^{n} \sigma_{\varepsilon}^2(X_i) Z_{ni}^2 \leq c \sum_{i=1}^{n} \sigma_{\varepsilon}^2(X_i) \left[(Z_{ni}^0)^2 + (Z_{ni}^1)^2 \right], \tag{25}$$

where $Z_{ni}^0 = n^{-2}h_n^{-(d+1)}\sum_{j=1}^n |\mathcal{K}((X_j - X_i)/h_n)||g_u(X_j)||\varrho_n^0(r_j)|$, and Z_{ni}^1 is like Z_{ni}^0 but with $\overline{\varrho}_n^1(r_j)$ replacing $|\varrho_n^0(r_j)|$.

We next establish the order in probability of the right hand side of (25). By the Markov inequality, for any $\delta_n > 0$,

$$\Pr\left[\sum_{i=1}^n \sigma_\varepsilon^2(X_i)(Z_{ni}^j)^2 > \delta_n\right] \leq \frac{E\left[\sum_{i=1}^n \sigma_\varepsilon^2(X_i)(Z_{ni}^j)^2\right]}{\delta_n} \leq \frac{\overline{\sigma}_\varepsilon^2 E\left[n(Z_{ni}^j)^2\right]}{\delta_n},$$

where $\overline{\sigma}_\varepsilon^2$ is an upper bound on $\sigma_\varepsilon^2(X_i)$. We have $E[n(Z_{ni}^j)^2] = E^2[\sqrt{n}Z_{ni}^j] + \mathrm{var}[\sqrt{n}Z_{ni}^j]$, $j = 0, 1$. We will just show the working for $j = 0$, because the case $j = 1$ is similar. By the triangle inequality $|Z_{ni}^0|$ is bounded by some constant times $n^{-2}h_n^{-(d+1)}\sum_{j=1}^n |\mathcal{K}((X_j - X_i)/h_n)||K((\lambda_r - r_j)/h_n)|$ plus $n^{-2}h_n^{-(d+1)}\sum_{j=1}^n |\mathcal{K}((X_j - X_i)/h_n)||K((r(x) - r_j)/h_n|$ plus a similar term involving the integral term in (24). We first show that $EZ_{ni}^0 = O(n^{-1})$ for each i. We have

$$\begin{aligned}
E\left[\left|\mathcal{K}\left(\frac{X_j - X_i}{h_n}\right)\right|\left|K\left(\frac{\lambda_r - r_j}{h_n}\right)\right|\right] &= \int \left|\mathcal{K}\left(\frac{X_j - X_i}{h_n}\right)\right|\left|K\left(\frac{\lambda_r - r_j}{h_n}\right)\right| f_X(X_i)f_X(X_j)dX_idX_j \\
&= h_n^d \int |\mathcal{K}(u)| \left|K\left(\frac{\lambda_r - r_j}{h_n}\right)\right| f_X(X_j + hu)f_X(X_j)dudX_j \\
&\leq c \cdot h_n^d \cdot \int E\left[f_X(X_j)\,|r(X) = s\right] \left|K\left(\frac{\lambda_r - s}{h_n}\right)\right| f_r(s)ds \\
&= O(h_n^{d+1}),
\end{aligned}$$

where the second line follows from a change of variables $X_i \longmapsto u = (X_j - X_i)/h_n$, while the third line follows from dominated convergence [using the bound on f_X], and the law of iterated expectations $[\int h(X)f_X(X)dX = Eh(X) = E[E[h(X)|r(X)] = \int E[h(X)|r(X) = s]f_r(s)ds$ for any measurable function h.] We have shown that $E|Z_{ni}^0| = O(1/n)$. Because Z_{ni}^0 is a sum of independent random variables, conditional on X_i, we have

$$E\mathrm{var}_i[Z_{ni}^0] \leq \frac{1}{n^4 h_n^{2(d+1)}}\sum_{j=1}^n E\left[\left|\mathcal{K}\left(\frac{X_j - X_i}{h_n}\right)\right|^2\left|K\left(\frac{\lambda_r - r_j}{h_n}\right)\right|^2\right] = O(\frac{1}{n^3 h_n^{d+1}})$$

by the same arguments as above. Furthermore, $\mathrm{var}[E_i\sqrt{n}Z_{ni}^0] = O(n(h_n^{(d+1)}/nh_n^{(d+1)})^2) = O(1/n)$. Therefore, we have $\mathrm{var}[\sqrt{n}Z_{ni}^0] = E\mathrm{var}_i[\sqrt{n}Z_{ni}^0] + \mathrm{var}[E_i\sqrt{n}Z_{ni}^0] = O(1/n^2h_n^{(d+1)})$. We now conclude that $\sum_{i=1}^n \sigma_\varepsilon^2(X_i)(Z_{ni}^0)^2 = O_p(n^{-1}h_n^{-(d+1)/2})$. In conclusion, $\mathcal{C}_{n11} = O_p(n^{-1}h_n^{-(d+1)/2}) = o_p(\delta_n)$. ■

PROOF OF LEMMA 3.2. Substituting the leading terms of M_{nq}^{-1} and M_{nr}^{-1} and using the representation (24) we have

$$
\begin{aligned}
\mathcal{C}_{n12} &\simeq \frac{1}{nh_n}\sum_{j=1}^{n} B_{nq0}(X_j)g_u(X_j)\varrho_n^0(r_j) \\
&\simeq h_n^2\left[\frac{E\left(B_{q0}(X)g_u(X)\,|r(X)=\lambda_r\right)}{q^2(\lambda_r)}-\frac{E\left(B_{q0}(X)g_u(X)\,|r(X)=r(x)\right)}{q^2(r(x))}\right],
\end{aligned}
$$

by the law of large numbers.

■

Proof of Lemma 3.3. Dividing into $j=i$ and $j\neq i$ terms, we get $\mathcal{C}_{n21}=\mathcal{C}_{n21a}+\mathcal{C}_{n21b}$, where $\mathcal{C}_{n21a}=n^{-2}h_n^{-(d+1)}\mathcal{K}(0)\sum_{j=1}^n\varepsilon_j\eta_je_0'M_{nq}^{-1}(X_j)e_0\varrho_{nj}$ and $\mathcal{C}_{n21b}=n^{-2}h_n^{-(d+1)}\sum\sum_{i\neq j}\mathcal{K}((X_j-X_i)/h_n)\varepsilon_i\eta_je_0'M_{nq}^{-1}(X_j)v_{ji}\varrho_{nj}$. The first term is conditional on $X_1,\ldots,X_n$ a sum of independent random variables. Taking expectations conditional on $X_1,\ldots,X_n$, we find that $E_X(\mathcal{C}_{n21a})=n^{-2}h_n^{-(d+1)}\mathcal{K}(0)\sum_{j=1}^n\sigma_{\varepsilon\eta}(X_j)e_0'M_{nq}^{-1}(X_j)e_0\varrho_{nj}$. This term is bounded by some constant times $n^{-2}h_n^{-(d+1)}\sum_{j=1}^n|\varrho_n^0(r_j)|+|\overline{\varrho}_n^1(r_j)|$, with probability tending to one as $n\to\infty$, which is a sum of independent random variables of order $1/nh_n^d$ in probability. The conditional variance of $\mathcal{C}_{n21a}$ is $n^{-4}h_n^{-2(d+1)}\mathcal{K}(0)^2\sum_{j=1}^n E(\varepsilon_j^2\eta_j^2|X_j)(e_0'M_{nq}^{-1}(X_j)e_0)^2\varrho_{nj}^2$, which is bounded by some constant times $n^{-4}h_n^{-2(d+1)}\sum_{j=1}^n(|\varrho_n^0(r_j)|+|\overline{\varrho}_n^1(r_j)|)^2$ with probability tending to one as $n\to\infty$, which is of order $n^{-3}h_n^{-2d+1}$ in probability. Therefore, $\mathcal{C}_{n21a}=O_p(n^{-1}h_n^{-d})$. We turn to the double sum $\mathcal{C}_{n21b}$. Let $\varphi_n(Z_i,Z_j)=n^{-2}h_n^{-(d+2)}\mathcal{K}((X_j-X_i)/h_n)e_0'M_{nq}^{-1}(X_j)v_{ji}\varrho_{nj}\varepsilon_i\eta_j$, where $Z_i=(X_i,Y_i)$. Then, $E_X[\varphi_n(Z_i,Z_j)|Z_i]=E_X[\varphi_n(Z_i,Z_j)|Z_j]=0$, and $\mathrm{var}_X[\sum\sum_{i\neq j}\varphi_n(Z_i,Z_j)]\leq 4\sum\sum_{i\neq j}E_X[\varphi_n^2(Z_i,Z_j)]$, by the Cauchy-Schwarz inequality. We now show that $E_X[\varphi_n^2(Z_i,Z_j)]=O(n^{-4}h_n^{-(d+1)})$, which implies that $\mathrm{var}_X[\sum\sum_{i\neq j}\varphi_n(Z_i,Z_j)]=O_p(n^{-2}h_n^{-(d+1)})$. Note that $\sigma_\varepsilon^2(X_i)\sigma_\eta^2(X_j)$ is bounded, while the matrices $M_{nr}(\cdot)$ and $M_{nq}(\cdot)$ are strictly positive definite uniformly in their arguments with probability tending to one. Furthermore, $E[\mathcal{K}^2((X_j-X_i)/h_n)K^2((\lambda_r-r_j)/h_n]=O(h_n^{d+1})$ by the same arguments used above. Likewise,

$$
\begin{aligned}
&E\left[\mathcal{K}^2\left(\frac{X_j-X_i}{h_n}\right)K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r_i}{h_n}\right)\right] \\
&= \int\mathcal{K}^2\left(\frac{X_j-X_i}{h_n}\right)K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r_i}{h_n}\right)f_X(X_i)f_X(X_j)dX_idX_j \\
&= h_n^d\int\mathcal{K}^2(u)\,K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r(X_j+h_nu)}{h_n}\right)f_X(X_j+h_nu)f_X(X_j)dudX_j \\
&= h_n^d\int\mathcal{K}^2(u)\,K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r_j}{h_n}+\nabla r_ju+h_nu'\nabla^2r(X_j^*(u))u\right)f_X(X_j+h_nu)f_X(X_j)dudX_j \\
&\simeq h_n^d\int\mathcal{K}^2(u)\,K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r_j}{h_n}+\nabla r(X_j)u\right)f_X^2(X_j)dudX_j.
\end{aligned}
$$

Here, $\nabla r(\cdot)$ and $\nabla^2 r(\cdot)$ are $1 \times d$ and $d \times d$ matrices containing the first and second order partials of the function r, $\nabla r_j = \nabla r(X_j)$, while $X_j^*(u)$ are intermediate values. The last two lines follow from a mean value expansion and the Lipschitz continuity of the kernel, i.e., for any positive function φ with integrable second moments,

$$\begin{aligned}\int \left| K\left(\frac{\lambda_r - r(X_j + h_n u)}{h_n}\right) - K\left(\frac{\lambda_r - r_j}{h_n} + \nabla r(X_j) u\right)\right| \varphi(u) du &\leq h_n K_{lip} \cdot \int \left| u' \nabla^2 r(X_j^*(u)) u \right| \varphi(u) du \\ &\leq h_n K_{lip} \overline{\lambda}_r \cdot \int u' u \varphi(u) du \\ &= O(h_n),\end{aligned}$$

where K_{lip} is the Lipschitz constant for the kernel and $\overline{\lambda}_r = \sup_x \max\{|\lambda_{r\max}(\nabla^2 r(x))|, |\lambda_{r\min}(\nabla^2 r(x))|\}$. Finally, $\int \mathcal{K}^2(u) K\left(\frac{\lambda_r - r(X)}{h_n}\right) K\left(\frac{\lambda_r - r(X)}{h_n} + \nabla r(X) u\right) f_X^2(X) du dX = O(h_n)$ by the law of iterated expectation and change of variables. In conclusion, $\mathcal{C}_{n21} = O_p(n^{-1} h_n^{-d}) + O_p(n^{-3/2} h_n^{-(2d+1)/2}) + O_p(n^{-1} h_n^{-(d+1)/2}) = o_p\left(n^{-1/2} h_n^{-d/2}\right)$.

■

Proof of Lemma 3.4. Substituting the leading terms of M_{nq}^{-1} and $M_{nr}^{-1}(s)$, we have $\mathcal{C}_{n22} \simeq n^{-1} h_n^{-1} \sum_{j=1}^n B_{q0}(X_j) \eta_j \varrho_n^0(r_j)$, which is $O_p(h_n^2 n^{-1/2} h_n^{-1/2})$.

■

Proof of Lemma 3.5. The arguments are very simiar to Lemma 3.1. Let $\theta_{nj} = \int_{r(x)}^{\lambda_r} q^{-2}(s) e_0' M_{nr}^{-1}(s) \times (\partial B_{nr}(s)/\partial r_j) ds$ and $V_{ni} = n^{-1} h_n^{-d} \sum_{j=1}^n \mathcal{K}\left(\frac{X_j - X_i}{h_n}\right) e_0' M_{nq}^{-1}(X_j) v_{ji} \theta_{nj}$. We can now write $\mathcal{C}_{n31} = \sum_{i=1}^n \varepsilon_i V_{ni}$, where the weights V_{ni} only depend on $X_1, \ldots, X_n$. Since $E(\varepsilon_i | X_1, \ldots, X_n) = 0$, $\mathcal{C}_{n31}$ has conditional mean zero and conditional variance $\sum_{i=1}^n V_{ni}^2 \sigma_\varepsilon^2(X_i)$. We substitute the leading terms of M_{nr}^{-1} and M_{nq}^{-1} and do a partial integration to replace θ_{nj} by

$$\theta_{nj}^0 = \frac{1}{n h_n} \left[\frac{K\left(\frac{s - r_j}{h_n}\right) \Delta_j(s)}{f_r(s) q^2(s)} \right]_{r(x)}^{\lambda_r} + \frac{1}{n h_n} \int_{r(x)}^{\lambda_r} K\left(\frac{s - r_j}{h_n}\right) \Delta_j'(s) \left[\frac{1}{f_r(s) q^2(s)} - \left(\frac{1}{f_r(s) q^2(s)}\right)' \right] ds$$

and V_{ni} by $V_{ni}^0 = n^{-1} h_n^{-d} \sum_{j=1}^n \mathcal{K}\left(\frac{X_j - X_i}{h_n}\right) f^{-1}(X_j) \theta_{nj}^0$. The magnitude of V_{ni}^0 is the same as the magnitude of quantities like $\overline{V}_{ni}^0 = n^{-2} h_n^{-(d+1)} \sum_{j=1}^n |\mathcal{K}\left(\frac{X_j - X_i}{h_n}\right)| |K\left(\frac{\lambda_r - r_j}{h_n}\right) \Delta_j(\lambda_r)|$. This is $O_p(h_n^2/n)$. In conclusion, $\mathcal{C}_{n31} = o_p(\delta_n)$.

■

PROOF OF LEMMA 3.6. Replacing M_{nq}^{-1} and M_{nr}^{-1} by their probability limits and subsituting $\Delta_j(s) \simeq q''(s)((r_j - s)/h_n)^2/2$ and $\Delta_j'(s) \simeq q'''(s)((r_j - s)/h_n)^2/2$, we get

$$\begin{aligned}
\mathcal{C}_{n32} &\simeq \frac{1}{n}\sum_{j=1}^{n}\frac{B_{q0}(X_j)}{f_X(X_j)}\int_{r(x)}^{\lambda_r}\frac{K_j'(s)\,\Delta_j(s)+K_j(s)\Delta_j'(s)}{q^2(s)f_r(s)}ds \\
&\simeq \int_{r(x)}^{\lambda_r}\frac{ds}{q^2(s)f_r(s)}\times\frac{1}{n}\sum_{j=1}^{n}E\left[\frac{B_{q0}(X_j)}{f_X(X_j)}|r(X_j)\right]\left[K'\left(\frac{s-r_j}{h_n}\right)\Delta_j(s)+K\left(\frac{s-r_j}{h_n}\right)\Delta_j'(s)\right] \\
&= o_p(h_n^2),
\end{aligned}$$

where the last line follows from a weak law of large numbers and a change of variable argument.

■

PROOF OF LEMMA 3.7. Replacing M_{nr}^{-1} and M_{nq}^{-1} by their probability limits we have

$$\mathcal{C}_{n4} \simeq -\frac{1}{n}\sum_{j=1}^{n}(\widehat{r}_j - r_j)q'(r_j)\times\int_{r(x)}^{\lambda_r}\frac{1}{f_r(s)q^2(s)}\frac{1}{h_n}K\left(\frac{s-r_j}{h_n}\right)ds.$$

This term has bias

$$-h_n^2\int_{r(x)}^{\lambda_r}E\left(B_{q0}(X)\,|r(X)=s\right)\frac{q'(s)}{q(s)^2}ds$$

and variance of order n^{-1}. ■

PROOF OF LEMMA 3.8. Replacing M_{nr}^{-1} and M_{nq}^{-1} by their probability limits we have

$$\mathcal{C}_{n5} \simeq \frac{1}{nh_n^2}\sum_{j=1}^{n}\frac{U_{nq0}(X_j)+B_{nq0}(X_j)}{f_X(X_j)}\int_{r(x)}^{\lambda_r}\frac{K_i'(s)}{f_r^2(s)}\left[U_{nr}(s)+B_{nr}(s)\right]ds.$$

This term is quadratic in the estimation errors and is $o_p(\delta_n)$.

■

PROOF OF LEMMA 4. We must show that $R_{n1} - R_{n6}$ are small. We use the following uniform convergence results

$$\sup_s|\widehat{q}(s)-q(s)| = O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right)+O_p\left(h_n^2\right) \quad (26)$$

$$\sup_x|\widehat{r}(x)-r(x)| = O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right)+O_p\left(h_n^2\right). \quad (27)$$

The result (27) is derived in Masry (1996b); it implies the same rate of convergence for $\widehat{\lambda}_r - \lambda_r$. The two results (26) and (27) imply that the quadratic terms in (19) and (20) are all of smaller order.

■

Proof of Theorem 6. We have

$$\begin{aligned}\widehat{m}(x)-m(x) &= \widehat{\lambda}_R-\lambda_R-\int_{\widehat{R}(x)}^{\widehat{\lambda}_R}\frac{\widehat{U}(s)-s\widehat{U}'(s)}{\widehat{U}(s)-s^2}ds+\int_{R(x)}^{\lambda_R}\frac{U(s)-sU'(s)}{U(s)-s^2}ds\\ &\simeq \widehat{\lambda}_R-\lambda_R-\{\int_{\widehat{R}(x)}^{\widehat{\lambda}_R}-\int_{R(x)}^{\lambda_R}\}\{\frac{U(s)-sU'(s)}{U(s)-s^2}\}ds\\ &\quad+\int_{R(x)}^{\lambda_R}\{\frac{U(s)-sU'(s)}{U(s)-s^2}-\frac{\widehat{U}(s)-s\widehat{U}'(s)}{\widehat{U}(s)-s^2}\}ds.\end{aligned}$$

The omitted terms will be quadratic in the estimation errors $\widehat{\lambda}_R-\lambda_R, \widehat{R}(x)-R(x), \widehat{U}'(s)-U'(s)$, and $\widehat{U}(s)-U(s)$ and so can be shown to be of smaller order. By a Taylor expansion, we get

$$\begin{aligned}\widehat{m}(x)-m(x) &\simeq -(\widehat{\lambda}_R-\lambda_R)(\frac{\lambda_R^2-\lambda_R U'(\lambda_R)}{U(\lambda_R)-\lambda_R^2})+\frac{U(R(x))-R(x)U'(R(x))}{U(R(x))-R(x)^2}(\widehat{R}(x)-R(x))\\ &\quad+\int_{R(x)}^{\lambda_R}\frac{(U(s)-s^2)s(\widehat{U}'(s)-U'(s))+(U(s)-sU'(s))(\widehat{U}(s)-U(s))}{(U(s)-s^2)^2}ds,\end{aligned}$$

where again the omitted terms are quadratic. We now show that

$$\frac{\lambda^2-\lambda_R U'(\lambda)}{U(\lambda)-\lambda^2}=0.$$

We already have that $R(\lambda)=\lambda$, $\mathfrak{F}(\lambda)=\lambda$, $F(\lambda)=1$, and $f(\lambda)=0$. Assume $\mathfrak{F}_2(\lambda)$ is finite. From the proof of Theorem 3, we have

$$U'[R(m)]=\frac{F(m)\mathfrak{F}(m)-f(m)\mathfrak{F}_2(m)}{F(m)^2-f(m)\mathfrak{F}(m)}.$$

Evaluating this expression at $m=\lambda$ gives

$$U'[R(\lambda)]=U'(\lambda)=\frac{1\cdot\lambda-0\cdot\mathfrak{F}_2(\lambda)}{1^2-(0\cdot\ 1)}=\lambda.$$

We also expect that the stochastic part of $\int_{R(x)}^{\lambda_R}w_1(s)(\widehat{U}(s)-U(s))ds$, where $w_2(s)=(U(s)-sU'(s))/(U(s)-s^2)^2$, is of smaller order by the same arguments we used for the censored regression estimator. Furthermore, we can show that the stochastic part of the term $\int_{R(x)}^{\lambda_R}w_2(s)(\widehat{U}'(s)-U'(s))ds$, where $w_2(s)=s/(U(s)-s^2)$, is of smaller order also. Both these random sequences will contributed to the bias of the estimator however at the magnitude of h_n^2. In conclusion,

$$\widehat{m}(x)-m(x)\simeq\frac{U(R(x))-R(x)U'(R(x))}{U(R(x))-R(x)^2}(\widehat{R}(x)-R(x))+\text{bias terms of order } h_n^2,$$

which has the stated limiting variance.

■

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., (1995), "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560–596.

[5] 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.

[6] Bickel, P.J., (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* 70, 428-434.

[7] Breiman, L., Y. Tsur, and A. Zemel (1993): "On a simple estimation procedure for censored regression models with known error distributions," *Annals of Statistics* 21, 1711-1720.

[8] Buchinsky, M. and J. Hahn, "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica,* 66, 653-671.

[9] Buckley, J. and I. James, (1979), "Linear Regression With Censored Data," *Biometrika*, 66, 429–436.

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

[11] Chaudhuri, P., K. Doksum, and A. Samarov, (1997), "On Average Derivative Quantile Regression,". *Annals of Statistics* **25**, 715-744.

[12] Collomb, G. and W. Härdle (1986), "Strong Uniform Convergence Rates in Robust Nonparametric Time Series Analysis and Prediction: Kernel Regression Estimation From Dependent Observations," *Stochastic Processes and Their Applications*, 23, 77–89.

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

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

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

[16] Fan, J., and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications* Chapman and Hall.

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

[18] Härdle, W., and O.B. Linton (1994): "Applied nonparametric methods," *The Handbook of Econometrics*, vol. IV, eds. D.F. McFadden and R.F. Engle III. North Holland.

[19] Härdle, W. and T. M. Stoker (1989), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986–995.

[20] Hausman, J. A. and W. K. Newey (1995), "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63, 1445–1476.

[21] 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.

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

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

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

[25] Horowitz, J. L., (1998), "Nonparametric estimation of a generalized additive model with an unknown link function," Iowa City Manuscript.

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

[27] James, I. R., and Smith, P. J. (1984): "Consistency results for linear regression with censored data," *The Annals of Statistics* 12, 590-600.

[28] van Keilegom, I., and M.G. Akritas (1999): "Transfer of tail information in censored regression models," Forthcoming in the *Annals of Statistics*.

[29] Koul, H., V. Suslara, and J. Van Ryzin (1981), "Regression Analysis With Randomly Right Censored Data," *Annals of Statistics*, 42, 1276–1288.

[30] Levy, A. (1999), "A Simple Conssitent Non-Parametric Estimator for the Regression Function in a Truncated Sample," Manuscript, North Carolina State University.

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

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

[33] Lewbel, A. (1998a), "Semiparametric Latent Variable Model Estimation With Endogenous or Mismeasured Regressors," *Econometrica*, 66, 105–121.

[34] Lewbel, A. (1998b), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," Unpublished Manuscript.

[35] McDonald, J. and R. Moffitt (1980), "The Uses of Tobit Analysis," *Review of Economics*, 62, 318–321.

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

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

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

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

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

[41] Newey, W. K. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

[42] Powell, J. L., J. H. Stock, and T. M. Stoker (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica* 57, 1403–1430.

[43] Powell, J. L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

[44] Powell, J. L. (1986a), "Symmetrically Trimmed Least Squares Estimation For Tobit Models," *Econometrica*, 54, 1435–1460.

[45] Powell, J. L. (1986b), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

[46] Ritov, Y. (1990): "Estimation in a linear regresion model with censored data," *Annals of Statistics* 18, 303-328.

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

[48] Robinson, P. M. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

[49] Rothenberg, T., and C. T. Leenders (1964): "Efficient Estimation of Simultaneous Equation Systems" *Econometrica*

[50] Rosett, R. and F. Nelson (1975), "Estimation of the Two-Limit Probit Regression Model," *Econometrica*, 43, 141–146.

[51] Schmee, J., and Hahn, Gerald J. (1979): "A simple method for regression analysis with censored data" *Technometrics* 21, 417-432.

[52] Silverman, B. W. (1978) "Weak and Strong Uniform Consistency of the Kernel Estimate of a Density Function and its Derivatives," *Annals of Statistics*, 6, 177–184.

[53] Stoker, Thomas M. (1991), "Equivalence of Direct, Indirect and Slope Estimators of Average Derivatives," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, W. A. Barnett, J. Powell, and G. Tauchen, Eds., Cambridge University Press.

[54] Stone, C.J. (1980), "Optimal rates of convergence for nonparametric estimators," *Annals of Statistics,* 8, 1348-1360.

[55] Stone, C.J. (1982). "Optimal global rates of convergence for nonparametric regression," *Annals of Statistics, 8,* 1040-1053.

[56] Tsiatis, A. A. (1990): "Estimating regression parameters using linear rank tests for censored data," *The Annals of Statistics* 18, 354-372.