NONPARAMETRIC CENSORED AND TRUNCATED REGRESSION*

Arthur Lewbel† Boston College

Oliver Linton‡ LSE and Yale University

January 5, 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 $\partial E(y^k I(y > 0)|x)/\partial m(x) = kE[y^{k-1} I(y > 0)|x]$, 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 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.

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), and Buchinsky and Hahn (1998). Unlike the present paper, most of these models either assume $m(x) = \beta' x$ or some other parametric form, or they provide estimates of average derivatives only up to an unknown scale. 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_i^* 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_i is independent of X_i, 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_i equals the latent variable censored at zero, so $Y_i = I(Y_i^* \geq 0)Y_i^*$, 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:

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

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*

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

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 $E[Y^\kappa I(Y > 0)|m(X) = e] = \int_{-\infty}^{e} \kappa E[Y^{\kappa-1} I(Y > 0)|m(X) = e]de = \int_{-\infty}^{e} \kappa(\kappa - 1)!\mathfrak{F}_{\kappa-1}(e)de = \kappa!\mathfrak{F}_\kappa(e)$. ■

For the special case of $m(x) = \beta' x$ and $\kappa = 1$, the fact that equation (1) holds has long been known. 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:

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

where x_k is the k'th element of x.

Assumption A1. *Assume* $Y^* = m(X) - e$ *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* $\lambda = \sup(\Omega_e)$.

Theorem 2 *Let Assumption A.1 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* $F[m(x)] \neq 0$,

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

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, $r_k(x) = \partial \mathfrak{F}[m(x)]/\partial x_k = F[m(x)]m_k(x) = s(x)m_k(x)$.

■

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 $\widehat{\lambda}_r = \max_{i=1,...,n} \widehat{r}(X_i)$. Note that $m(x) = \lambda_* - \int_{r(x)}^{\lambda_*} \frac{1}{q(r)} dr$ for any $\lambda_* \geq \lambda$. It is a standard result that $\widehat{\lambda}_r \longrightarrow_p \lambda_r$, where $\lambda_r = \max_{i=1,...,n} r(X_i)$. Furthermore, under our assumptions, $\lambda_r \geq \lambda$. 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

2.1 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 $j \leq d$, $m(x) = \beta_1 x_1 + ... + \beta_j x_j + \widetilde{m}(x_{j+1}, \ldots, x_k)$, 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.

2.2 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 $\widehat{m}(X_i)$. In addition, if $m(x)$ can take on any value in the support of e, then Lemma 1 in Lewbel (1997) can 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 $\delta_{(i)}$ is an indicator for censoring, i.e., $\delta_{(i)} = 0$ when observation $Y_{(i)}$ is censored, and $\delta_{(i)} = 1$ otherwise. Then let

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

2.3 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 δ_i is an indicator for censoring, i.e., $\delta_i = 0$ 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 explicitly 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 a constant 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 $q < 1 - \rho(x)$.

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, though we show later that our estimator is more efficient in a simple normal example. However, the main advantage of our estimator over quantiles is that ours does not require knowledge about the degree of censoring at each point x 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:

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

where x_k is the k'th element of x.

Assumption A1*. *Assume $Y^* = m(X) - e$ and $Y = Y^*|Y^* > 0$. 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 $\lambda = \sup_{e \in \Omega_e} |e|$.*

Theorem 3 *Let Assumption A.1* 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 $x \in \Omega$,*

$$m(x) = \lambda - \int_{R(x)}^{\lambda} \frac{U(R) - RU'(R)}{U(R) - R^2} dR, \tag{4}$$

and for each continuously distributed element X_k of X,

$$m_k(x) = \frac{R(x)T_k(x) - T(x)R_k(x)}{R(x)^2 - T(x)}. \tag{5}$$

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, $F(\lambda) = \lambda$, 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 $\widehat{R}_k(x) = \partial\widehat{R}(x)/\partial x_k$ and $\widehat{T}_k(x) = \partial\widehat{T}(x)/\partial x_k$ 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.

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.

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

$$\sum_{i=1}^{n}\left[Y_i - \sum_{0\leq|k|\leq p} b_{\mathbf{k}}(x)(X_i - x)^{\mathbf{k}}\right]^2 \mathcal{K}((X_i - x)/h_n), \tag{6}$$

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 (6) 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 $\widehat{g}_k(x) = (D^{e_k}g)\hat{}(x)$ 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 (6). We then let

$$\widehat{m}(x) = \widehat{\lambda}_r - \int_{\widehat{r}(x)}^{\widehat{\lambda}_r} \frac{1}{\widehat{q}(s)} ds, \tag{7}$$

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 $\widehat{\lambda}_r = \max_{1\leq i\leq n} \widehat{r}(X_i)$. The integral can be evaluated numerically using quadrature. We later show that, under suitable regularity conditions, the limiting distribution of $\widehat{m}(x)$ takes the simple form

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

with bandwidth h_n and kernel $\mathcal{K}$.

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

$$\hat{m}_k(x) = \frac{\widehat{r}_k(x)}{\widehat{s}(x)}, \quad k = 1, \ldots, d. \tag{8}$$

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, $\widehat{\delta}_{wk}$ is root n consistent and asymptotically normal.

4.5 The Truncated Regression Function

Let $\widehat{R}(x)$ be the nonparametric regression of Y_i on X_i constructed as in (6), 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 $\widehat{R}(X_i)$, again using only observations having $Y_i > 0$. Then

$$\widehat{m}(x) = \widehat{\lambda}_R - \int_{\widehat{R}(x)}^{\widehat{\lambda}_R} \frac{\widehat{U}(s) - s\widehat{U}'(s)}{\widehat{U}(s) - s^2} ds, \tag{9}$$

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

5 Asymptotic Properties

5.1 Assumptions

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

$$N_\ell = \begin{pmatrix} \ell + d - 1 \\ d - 1 \end{pmatrix}$$

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

$$i = \phi_{|j|}^{-1}(j) + \sum_{k=0}^{|j|-1} N_k. \tag{10}$$

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

$$M = \begin{bmatrix} M_{0,0} & M_{0,1} & \cdots & M_{0,p} \\ M_{1,0} & M_{1,1} & \cdots & M_{1,p} \\ \vdots & & & \vdots \\ M_{p,0} & M_{p,1} & \cdots & M_{p,p} \end{bmatrix}, \; \Gamma = \begin{bmatrix} \Gamma_{0,0} & \Gamma_{0,1} & \cdots & \Gamma_{0,p} \\ \Gamma_{1,0} & \Gamma_{1,1} & \cdots & \Gamma_{1,p} \\ \vdots & & & \vdots \\ \Gamma_{p,0} & \Gamma_{p,1} & \cdots & \Gamma_{p,p} \end{bmatrix}, \quad B = \begin{bmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{bmatrix}, \tag{11}$$

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 $0 \le |j| \le 2p+1$ define the function

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

We make the following assumptions on the kernel $\mathcal{K}$.

Assumption A2

(a) The kernel $\mathcal{K}$ is symmetric about zero, bounded, and has compact support ($\mathcal{K}(u) = 0$ for $||u|| > A_0$).

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

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

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 $u = x$.

Remark. By dominated convergence, Assumption A3(b) implies that for each $L > 0$, the functions $E[Y1(|Y| < L)|X = u]$, $E[Y^2 1(|Y| < L)|X = u]$, are continuous at the point x. Hence for each $L > 0$, $\tilde{\sigma}_L^2(u) = \text{var}[Y \cdot 1(|Y| > L)|X = u]$ is continuous at the point x provided $m(\cdot)$ and $\sigma(\cdot)$ are continuous at the point x. This is needed in the proof of Theorem 2 where a truncation argument is employed and the continuity of $\tilde{\sigma}_L^2(u)$ at $u = x$ is required.

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

(c) The Lebesgue density f of X and the regression function s satisfy

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

on some compact subset $\mathcal{X}$ of $\mathbb{R}^d$.

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 $\widehat{s}(x)$. Define $\sigma_r^2(x) = \mathrm{var}(Y|X = x)$ and $\sigma_s^2(x) = \mathrm{var}[1(Y > 0)|X = x]$.

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

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

at continuity points x of $\{\sigma_r^2, \sigma_s^2, f, s\}$ whenever $f(x), s(x) > 0$. Here, $(M^{-1}\Gamma M^{-1})_{1,1}$ and $(M^{-1} B \mathcal{D}_{p+1}(x; r))_1$ are the corresponding [as in (11)] 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 (12) below is $o(1)$. Then, we have with probability one

$$\sup_{x \in \mathcal{X}} |\widehat{Dm}(x) - (Dm)(x)| = O \left\{ \left(\frac{\ln n}{nh_n^{d+2}} \right)^{1/2} \right\} + O(h_n^p). \tag{12}$$

The proof of this theorem involves a standard linearization argument.

REMARKS A.

1. The optimal bandwidth for estimating the j^{th} 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

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

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.

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 and B hold except that $r(x)$ has three continuous derivatives, and that $\limsup_{n\to\infty} nh_n^{d+4} < \infty$. Then*

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

where $b_m(x)$ is a bounded continuous function.

Note that the bias term goes to zero provided $nh_n^{d+4} \to 0$.

5.4 Distribution of Censored Regression Function Averages Derivatives

5.5 Distribution of The Truncated Regression Function Estimator

6 Numerical Results

7 Extensions and Conclusions

We have a provided estimators for the nonparametric censored and truncated regression models with fixed censoring. The estimators are based on the conditional means $E(Y^\kappa|X = x)$ for low integers k.. Higher moments of Y could also be employed. For example, for any integer $\kappa \geq 2$,

$$m_k(x) = \frac{\partial E(Y^\kappa|X = x)/\partial x_k}{\kappa E[I(Y^{\kappa-1})|X = x]}, \quad k = 1, \ldots, d. \tag{13}$$

The proof works in exactly the same way as Theorem 2. These higher moment based estimates could either be combined with estimators based on Theorem 2 to improved efficiency, or compared to those estimators as a test of the (nonparametric) model specification.

We will end by listing some two other extensions of the estimators.

7.1 Heteroscedastic errors

Assume $F(e|x) = F[e|w(x)]$ and $E[e|w(x)] = 0$ for some known, vector valued function w. Assume $\mathrm{supp}[e|w(x)] = \mathrm{supp}(e) \subseteq \mathrm{supp}[m(x)|w(x)]$. 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.

7.2 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 then apply local linear regression to the observations $\{\widehat{Y}_i^{BJ}, X_i\}$. The integration in (15) can be done numerically. The revised estimate of m is 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, 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 $\widehat{m}$.

A Appendix

We first give some facts about 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 Theorem 4. 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 $\widehat{g}_k(x) - g_k(x) = e_k' M_n^{-1}(x) U_n(x) + e_k' M_n^{-1}(x) B_n(x)$, where $e_k = (0, 0, \ldots, 0, 1, 0, \ldots, 0)'$ is the $d + 1$ vector with the one in the $k + 1$

position, while 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

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

where $\epsilon_i=Y_i-g(X_i)$ is the mean zero error term, while the bias term is

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

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

$$\begin{aligned}\Delta_i(x) &= \frac{h_n^2}{2}\sum_{j=1}^d\sum_{l=1}^d\frac{\partial^2 g}{\partial x_j\partial x_l}(x)\left(\frac{X_{ji}-x_j}{h_n}\right)\left(\frac{X_{li}-x_l}{h_n}\right)\\ &\quad+\frac{h_n^3}{6}\sum_{j=1}^d\sum_{l=1}^d\sum_{m=1}^d\frac{\partial^3 g}{\partial x_j\partial x_l\partial x_m}(x)\left(\frac{X_{ji}-x_j}{h_n}\right)\left(\frac{X_{li}-x_l}{h_n}\right)\left(\frac{X_{mi}-x_m}{h_n}\right)+o(h_n^3).\end{aligned}$$

Here, the remainder is of the stated order on the set $\{X_i:\|X_i-x\|<h_n\}$. Furthermore, by the law of large numbers and symmetry, we have

$$B_{n0}(x) = \frac{h_n^2}{2}\mu_2(K)\sum_{j=1}^d g_{jj}(x)\{1+o_p(1)\} \tag{16}$$

$$B_{nk}(x) = \frac{h_n^3}{6}\sum_{j=1}^d\sum_{l=1}^d\sum_{m=1}^d\int\mathcal{K}(u)u_ku_ju_lu_m du\left\{3g_{jl}(x)f_m(x)+g_{jlm}(x)f(x)\right\}\{1+o_p(1)\}. \tag{17}$$

Denote the right hand side 'probability limits' by $h_n^2B_0(x)$ and $h_n^3B_j(x)$, $j=1,\ldots,k$.

We have

$$M_n^{-1}(x) - f^{-1}(x)M^{-1} = -M_n^{-1}(x)[M_n(x) - f(x)M]f^{-1}(x)M^{-1} = O_p(h_n) + O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right)$$

uniformly in x.

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 $r_j = r(X_j)$. 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, these matrices are both diagonal. Similarly for U_{nr}, U_{nq}, B_{nr}, and B_{nq} etc. Then define $\varepsilon_i = Y_i - r_i$ 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, write $u_i = g_u(X_i) + \eta_i$, where $E(\eta_i|X_i) = 0$ by construction. Let $\sigma_\eta(X_i) = E(\eta_i^2|X_i)$, $\sigma_{\varepsilon\eta}(X_i) = E(\varepsilon_i\eta_i|X_i)$, $\sigma_\varepsilon^2(X_i) = E(\varepsilon_i^2|X_i)$, and $\sigma_u^2(X_i) = E(u_i^2|X_i)$. Finally, 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.

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 a nonparametric regression, while the terms (19) and (20) are both quadratic in such errors, and can thus be expected to be of smaller order. We shall first address the terms in (18). Since $q(\lambda_r) = 1$, the first term in (18) is zero. The second term is just a constant times the estimation error of $\widehat{r}(x)$. To analyze the third term we make another Taylor series expansion

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

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

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

We first examine the terms $\mathcal{A}_n$, $\mathcal{B}_n$, and $\mathcal{C}_n$, and then the remainder term $\mathcal{R}_n$. Let $\delta_n = \max\{1/\sqrt{nh_n^d}, h_n^2\}$.

LEMMA 1. *The term $\mathcal{A}_n$ is just a constant times the error in the nonparametric regression estimate of r; this term is $O_p(\delta_n)$.*

We next consider the terms $\mathcal{B}_n$ and $\mathcal{C}_n$. For this we need the following decomposition 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 just functions of $X_1, \ldots, X_n$.

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

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

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

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

where

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

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

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 $\partial \widehat{q}(s; r_1, \ldots, r_n)/\partial r_j$, and 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 $\partial U_{nr}(s)/\partial r_i$ with $g_u(X_i)$ substituting for u_i, while $\partial U_{nr}^\eta(s)/\partial r_i$ is $\partial U_{nr}(s)/\partial r_i$ with η_i substituting for u_i. We 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,

$$\mathcal{C}_{n4} = \sum_{j=1}^n (\widehat{r}_j - r_j) \times \int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s) \frac{\partial M_{nr}(s)}{\partial r_j} M_{nr}^{-1}(s) \left[U_{nr}(s) + B_{nr}(s)\right] ds.$$

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)

$$\mathcal{C}_{n12} = h_n^2 \left[\frac{E\left(B_{q0}(X) g_u(X)\,|r(X) = \lambda_r\right)}{q^2(\lambda_r)} - \frac{E\left(B_{q0}(X) g_u(X)\,|r(X) = r(x)\right)}{q^2(r(x))} \right] (1 + o_p(1));$$

(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) $\mathcal{C}_{n4} = o_p(\delta_n)$;

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

A.2 Proofs of Lemmas

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

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

Now note that there is some finite positive constant c such that with probability tending to one

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

where the last line follows by a change of variables $s \longmapsto t = (s - r_i)/h_n$ and dominated convergence. We use the fact that $1/q^2 f_r(s)$ is bounded. The first inequality is because $M_{nr}(s)$ converges uniformly to a finite positive definite matrix. Therefore, $\mathcal{B}_{n1} = O_p(n^{-1/2})$.

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

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

This is justified by dominated convergence and the uniform convergence. ■

Proof of Lemma 3.1. Let

$$\begin{aligned}\varrho_{ni} &= \frac{1}{h_n}\int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s)\begin{bmatrix} K_i'(s) \\ L_i'(s) + K_i(s)\end{bmatrix}\frac{1}{q^2(s)} ds \\ Z_{ni} &= \frac{1}{n^2 h_n^{d+1}}\sum_{j=1}^{n} \mathcal{K}\left(\frac{X_j - X_i}{h_n}\right) g_u(X_j) e_0' M_{nq}^{-1}(X_j) v_{ji}\varrho_{nj},\end{aligned}$$

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 $\mathrm{var}_X(\mathcal{C}_{n11}) = \sum_{i=1}^{n} \sigma_\varepsilon^2(X_i) Z_{ni}^2$.

We have with probability tending to one

$$\left|e_0' M_{nq}^{-1}(X_j) v_{ji}\right| \leq (d+1)^{1/2}\lambda_{\min}(M)^{-1}\left(\min_{1\leq j\leq n} f_X(X_j)\right)^{-1}$$

on the set where $\mathcal{K}((X_j - X_i)/h_n \neq 0$. This is because $M_{nq}(x)$ converges to $f_X(x)M$ uniformly in x, while for any vectors a, b, and nonsingular 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 $\lambda_{\max}(A^{-1}) = \lambda_{\min}^{-1}(A)$. Also write

$$\begin{aligned}\varrho_{ni} &= \frac{1}{h_n}\int_{r(x)}^{\lambda_r} \frac{K'(\frac{s-r_i}{h_n})}{f_r(s)q^2(s)}ds - \frac{1}{h_n}\int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s)\left[M_{nr}(s) - f_r(s)M\right]M^{-1}\begin{bmatrix} K_i'(s) \\ L_i'(s) + K_i(s)\end{bmatrix}\frac{ds}{f_r(s)q^2(s)} \\ &\equiv \varrho_n^0(r_i) + \varrho_{ni}^1.\end{aligned}$$

Using integration by parts and change of variables, we can establish that

$$\varrho_n^0(r_i) = \left[\frac{K\left(\frac{\lambda_r - r_i}{h_n}\right)}{f_r(\lambda_r)q^2(\lambda_r)} - \frac{K\left(\frac{r(x)-r_i}{h_n}\right)}{f_r(r(x))q^2(r(x))}\right] - \int_{r(x)}^{\lambda_r} K\left(\frac{s-r_i}{h_n}\right)\left(\frac{1}{f_r(s)q^2(s)}\right)' ds. \quad (23)$$

The last term is of smaller order, $O_p(h_n)$ for each i. We will also replace ϱ_{nj}^1 by an upper bound that only depends on r_j and n, thus for some constant c

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

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 $M_{nr}^{-1}(s)\left[M_{nr}(s) - f_r(s)M\right]M^{-1}$ are all bounded by some constant times h_n with probability tending to one. Combining these relations and using the triangle inequality, we have on a set whose probability tends to one,

$$\sum_{i=1}^n \sigma_\varepsilon^2(X_i)Z_{ni}^2 \leq c\sum_{i=1}^n \sigma_\varepsilon^2(X_i)\left[(Z_{ni}^0)^2 + (Z_{ni}^1)^2\right], \quad (24)$$

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 $|\varrho_n^0(r_j)|$.

Note that 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 $E[n(Z_{ni}^j)^2] = E^2[\sqrt{n}Z_{ni}^j] + \mathrm{var}[\sqrt{n}Z_{ni}^j]$. We first show that $EZ_{ni}^j = O(n^{-1})$ for each i. 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 two similar terms involving $K((r(x) - r_i)/h_n)$ and the integral term. We have

$$E\left[\left|\mathcal{K}\left(\frac{X_j - X_i}{h_n}\right)\right|\left|K\left(\frac{\lambda_r - r_j}{h_n}\right)\right|\right] = \int\left|\mathcal{K}\left(\frac{X_j - X_i}{h_n}\right)\right|\left|K\left(\frac{\lambda_r - r_j}{h_n}\right)\right| f_X(X_i)f_X(X_j)dX_idX_j$$

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

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.] 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 $\mathrm{var}[\sqrt{n} Z_{ni}^0] = E\mathrm{var}_i[\sqrt{n} Z_{ni}^0] + \mathrm{var}[E_i \sqrt{n} Z_{ni}^0] = O(1/n^2 h_n^{(d+1)})$. We now conclude that $\sum_{i=1}^n \sigma_\varepsilon^2(X_i)(Z_{ni}^0)^2 = O_p(n^{-1} h_n^{-(d+1)/2})$. Similar arguments can be applied to Z_{ni}^1. 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 $M_{nr}^{-1}(s)$ and using the representation (23) we have

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

by the law of large numbers.

■

PROOF OF LEMMA 3.3. Write

$$
\gamma_{ni} = \int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s) \begin{bmatrix} K_i'(s) \\ L_i'(s) + K_i(s) \end{bmatrix} \frac{ds}{q^2(s)} \quad ; \quad \gamma_n^0(r_i) = \int_{r(x)}^{\lambda_r} e_0' f_r^{-1}(s) M^{-1} \begin{bmatrix} K_i'(s) \\ L_i'(s) + K_i(s) \end{bmatrix} \frac{ds}{q^2(s)}.
$$

Here, $\gamma_n^0(r_i)$ is the probability limit of γ_{ni}. Dividing into $j = i$ and $j \neq i$ terms, we get $\mathcal{C}_{n21} = \mathcal{C}_{n21a} + \mathcal{C}_{n21b}$, where $\mathcal{C}_{n21a} = n^{-2} h_n^{-(d+2)} \mathcal{K}(0) \sum_{j=1}^n \varepsilon_j \eta_j e_0' M_{nq}^{-1}(X_j) e_0 \gamma_{nj}$ and $\mathcal{C}_{n21b} = n^{-2} h_n^{-(d+2)} \sum\sum_{i \neq j} \mathcal{K}((X_j -$

$X_i)/h_n)\varepsilon_i\eta_j \times e_0' M_{nq}^{-1}(X_j)v_{ji}\gamma_{nj}$. Taking expectations conditional on $X_1,\ldots,X_n$, we find that $E_X(\mathcal{C}_{n21a}) = n^{-2}h_n^{-(d+2)}\mathcal{K}(0)\sum_{j=1}^n \sigma_{\varepsilon\eta}(X_j)\times e_0' M_{nq}^{-1}(X_j)e_0\gamma_{nj}$. This term is bounded by some constant times $n^{-2}h_n^{-(d+2)}\sum_{j=1}^n \gamma_n^0(r_j)$, 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 of even smaller order. Therefore, $\mathcal{C}_{n21a} = O_p(n^{-1}h_n^{-d})$. We turn to the double sum $\mathcal{C}_{n21b}$. Let $\varphi_n(Z_i,Z_j) = n^{-2}h_n^{-(d+2)}\mathcal{K}((X_j-X_i)/h_n)e_0' M_{nq}^{-1}(X_j)v_{ji}\gamma_{nj}\varepsilon_i\eta_j$, 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

$$
\begin{aligned}
\mathrm{var}_X[\sum_{j=1}^n\sum_{\substack{i=1\\ i\neq j}}^n \varphi_n(Z_i,Z_j)] &= \sum_{j=1}^n\sum_{\substack{i=1\\ i\neq j}}^n E_X[\varphi_n^2(Z_i,Z_j)] + 3E_X[\varphi_n(Z_i,Z_j)\varphi_n(Z_j,Z_i)] \\
&\leq \sum_{j=1}^n\sum_{\substack{i=1\\ i\neq j}}^n E_X[\varphi_n^2(Z_i,Z_j)] + 3E_X^{1/2}[\varphi_n^2(Z_i,Z_j)]E_X^{1/2}[\varphi_n^2(Z_i,Z_j)] \\
&= 4\sum_{j=1}^n\sum_{\substack{i=1\\ i\neq j}}^n E_X[\varphi_n^2(Z_i,Z_j)],
\end{aligned}
$$

where the inequality used Cauchy-Schwarz. 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 a bounded sequence, while $M_{nr}(\cdot)$ and $M_{nq}(\cdot)$ are strictly positive definite with probability tending to one uniformly in their arguments. 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,

$$
\begin{aligned}
& E\left[\mathcal{K}^2\left(\frac{X_j-X_i}{h_n}\right)K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r_i}{h_n}\right)\right] \\
= & \int \mathcal{K}^2\left(\frac{X_j-X_i}{h_n}\right)K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r_i}{h_n}\right)f_X(X_i)f_X(X_j)dX_idX_j \\
= & h_n^d\int \mathcal{K}^2(u)\,K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r(X_j+h_nu)}{h_n}\right)f_X(X_j+h_nu)f_X(X_j)dudX_j \\
= & h_n^d\int \mathcal{K}^2(u)\,K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r_j}{h_n}+\nabla r_ju+h_nu'\nabla^2r(X_j^*(u))u\right)f_X(X_j+h_nu)f_X(X_j)dudX_j \\
= & h_n^d\int \mathcal{K}^2(u)\,K\left(\frac{\lambda_r-r_j}{h_n}\right)K\left(\frac{\lambda_r-r_j}{h_n}+\nabla r(X_j)u\right)f_X^2(X_j)dudX_j\{1+o(1)\}.
\end{aligned}
$$

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,

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

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 $\mathcal{C}_{n22} = n^{-1} h_n^{-1} \sum_{j=1}^n B_{q0}(X_j) \eta_j \varrho_n^0(r_j) \times (1 + o_p(1))$, which is $O_p(h_n^2 n^{-1/2} h_n^{-1/2})$.

■

PROOF OF LEMMA 3.5. Substituting the leading terms of $M_{nr}^{-1}(s)$, we have

$$\int_{r(x)}^{\lambda_r} e_0' M_{nr}^{-1}(s) \frac{\partial B_{nr}(s)}{\partial r_j} \frac{1}{q^2(s)} ds = \int_{r(x)}^{\lambda_r} \frac{\frac{1}{nh_n^2} K_j'(s) \Delta_j(s) + \frac{1}{nh_n^2} K_j(s) \Delta_j'(s)}{q^2(s) f_r(s)} ds \times (1 + o_p(1))$$
$$= \frac{1}{nh_n} \int_{\frac{r(x) - r_j}{h_n}}^{\frac{\lambda_r - r_j}{h_n}} \frac{K'(t) \Delta_j(r_j) + K(t) \Delta_j'(r_j)}{q^2(r_j) f_r(r_j)} dt \times (1 + o_p(1)),$$

by dominated convergence. Likewise substituting the leading terms of M_{nq}^{-1}, we have

$$\mathcal{C}_{n31} = \sum_{j=1}^n f_X^{-1}(X_j) \sum_{i=1}^n \mathcal{K}\left(\frac{X_j - X_i}{h_n}\right) \varepsilon_i \frac{1}{nh_n} \int_{\frac{r(x) - r_j}{h_n}}^{\frac{\lambda_r - r_j}{h_n}} \frac{K'(t) \Delta_j(r_j) + K(t) \Delta_j'(r_j)}{q^2(r_j) f_r(r_j)} dt \times (1 + o_p(1))$$
$$= \frac{1}{n} \sum_{i=1}^n \varepsilon_i \frac{1}{nh_n^d} \sum_{j=1}^n f_X^{-1}(X_j) \mathcal{K}\left(\frac{X_j - X_i}{h_n}\right) \frac{1}{h_n} \int_{\frac{r(x) - r_j}{h_n}}^{\frac{\lambda_r - r_j}{h_n}} \frac{K'(t) \Delta_j(r_j) + K(t) \Delta_j'(r_j)}{q^2(r_j) f_r(r_j)} dt \times (1 + o_p(1)).$$

Therefore, $\mathcal{C}_{n31} = O_p(h_n^2 n^{-1/2} h^{-1/2})$. ■

Proof of Lemma 3.6. Replacing M_{nq}^{-1} and M_{nr}^{-1} by their probability limits, we have

$$\begin{aligned}\mathcal{C}_{n32} &= \frac{1}{nh_n}\sum_{j=1}^{n}\frac{B_{nq0}(X_j)}{f_X(X_j)q^2(r_j)f_r(r_j)}\int_{\frac{r(x)-r_j}{h_n}}^{\frac{\lambda_r-r_j}{h_n}}[K'(t)\,\Delta_j(r_j)+K(t)\Delta_j'(r_j)]dt\times(1+o_p(1))\\ &= o_p(h_n^2),\end{aligned}$$

where the last line follows from a weak law of large numbers and the magnitude of $\Delta_j(r_j)$ and $\Delta_j'(r_j)$.

■

Proof of Lemma 3.7. Replacing $M_{nr}^{-1}(s)$ by its probability limit we have

$$\begin{aligned}|\mathcal{C}_{n4}| &\le \left(\max_{1\le j\le n}|\widehat{r}_j-r_j|\right)\left|\sum_{j=1}^{n}\int_{r(x)}^{\lambda_r}e_0'M_{nr}^{-1}(s)\frac{\partial M_{nr}(s)}{\partial r_j}M_{nr}^{-1}(s)\left[U_{nr}(s)+B_{nr}(s)\right]ds\right|\\ &= \left(\max_{1\le j\le n}|\widehat{r}_j-r_j|\right)\left|\frac{1}{nh_n^2}\sum_{j=1}^{n}\int_{r(x)}^{\lambda_r}\frac{1}{f_r^2(s)}\left[\;K_i'(s)\quad\frac{L_i'(s)+K_i(s)}{\mu_2(K)}\;\right]\left[U_{nr}(s)+B_{nr}(s)\right]ds\right|\\ &\quad+\text{higher order terms.}\end{aligned}$$

A standard argument involving interchanging of order of summation and change of variables etc provides a bound of $O_p(n^{-1/2}h_n^{-1})$ on the second term on the right hand side, so that

$$|\mathcal{C}_{n4}|=\left(O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right)+O_p(h_n^2)\right)\left(O_p\left(\sqrt{\frac{1}{nh_n^2}}\right)+O_p(h_n^2)\right).$$

■

Proof of Lemma 4. We must show that $R_{n1}-R_{n6}$ are small. We use the following uniform convergence results

$$\sup_{\underline{r}\le s\le\overline{r}}|\widehat{q}(s)-q(s)| = O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right)+O_p\left(h_n^2\right) \tag{25}$$

$$\sup_{x}|\widehat{r}(x)-r(x)| = O_p\left(\sqrt{\frac{\log n}{nh_n^d}}\right)+O_p\left(h_n^2\right), \tag{26}$$

where $\underline{r},\overline{r}$ are the lower and upper bounds of the random variable $r(X)$. The result (26) is standard; it implies the same rate of convergence for $\widehat{\lambda}_r-\lambda_r$. The proof of (25) improves on Ahn (1995). The two results (25) and (26) imply that the quadratic terms in (19) and (20) are all of smaller order.

■

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

[6] 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.

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

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

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

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

[11] 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.

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

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

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

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

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

[17] 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.

[18] 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.

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

[20] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[49] 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.

[50] 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.

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

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

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