Nonparametric Identification and Estimation of a Censored Location-Scale Regression Model

Songnian CHEN, Gordon B. DAHL, and Shakeeb KHAN

In this article we consider identification and estimation of a censored nonparametric location scale-model. We first show that in the case where the location function is strictly less than the (fixed) censoring point for all values in the support of the explanatory variables, the location function is not identified anywhere. In contrast, when the location function is greater or equal to the censoring point with positive probability, the location function is identified on the entire support, including the region where the location function is below the censoring point. In the latter case we propose a simple estimation procedure based on combining conditional quantile estimators for various higher quantiles. The new estimator is shown to converge at the optimal nonparametric rate with a limiting normal distribution. A small-scale simulation study indicates that the proposed estimation procedure performs well in finite samples. We also present an empirical illustration on unemployment insurance duration using administrative-level data from New Jersey.

KEY WORDS: Censored regression; Nonparametric quantile regression; Unemployment insurance.

1. INTRODUCTION

The nonparametric location-scale model is usually of the form

$$y_i = \mu(\mathbf{x}_i) + \sigma_0(\mathbf{x}_i)\epsilon_i, \tag{1}$$

where $\mathbf{x}_i$ is an observed d-dimensional random vector and ϵ_i is an unobserved random variable, distributed independently of $\mathbf{x}_i$ and assumed to be centered around 0 in some sense. The functions $\mu(\cdot)$ and $\sigma_0(\cdot)$ are unknown. This location-scale model has received a great deal of attention in the statistics and econometrics literature (see, e.g., Fan and Gijbels 1996, chap. 3; Ruppert and Wand 1994), and existing nonparametric methods, such as kernel, local polynomial, and series estimators, can be used to estimate $\mu(\cdot)$ from a random sample of observations of the vector $(y_i, \mathbf{x}_i')'$.

In this article we consider extending the nonparametric location-scale model to accommodate censored data. Semiparametric (fixed) censored regression models, where $\mu(\mathbf{x}_i)$ is known up to a finite-dimensional parameter, have been studied extensively in the econometrics literature (see Powell 1994 for a survey). The advantage of our nonparametric approach here is that economic theory rarely provides any guidance on functional forms in relationships between variables.

Censoring occurs in many types of economic data, because of either nonnegativity constraints or top coding. To allow for censoring, we work within the latent dependent-variable framework, as is typically done for parametric and semiparametric models. We thus consider a model of the form

$$y_i^* = \mu(\mathbf{x}_i) + \sigma_0(\mathbf{x}_i)\epsilon_i, \tag{2}$$

$$y_i = \max(y_i^*, 0), \tag{3}$$

where y_i^* is the unobserved latent dependent variable and y_i is the observed dependent variable, which is equal to the latent variable y_i^* when it exceeds the fixed censoring point, which we assume, without loss of generality, to be 0. If the latent variable does not exceed the censoring point, then the observed dependent variable is equal to the censoring value. We note that in models with fixed censoring, there is usually no need to include a censoring indicator variable, because the latent variable y_i^* is assumed to be continuously distributed.

We consider identification and estimation of $\mu(\mathbf{x}_i)$ after imposing the location restriction that the median of $\epsilon_i = 0$. We emphasize that our results allow for identification of $\mu(\mathbf{x}_i)$ on the entire support of $\mathbf{x}_i$. This is in contrast to identifying and estimating $\mu(\mathbf{x}_i)$ only in the region where it exceeds the censoring point, which could be easily done by extending Powell's (1984) CLAD estimator for the semiparametric censored regression model to a nonparametric setting.

Our work is motivated by the fact that there are often situations where the econometrician is interested in estimating the location function in the region where it is less than the censoring point. One situation is when the dataset is heavily censored. In this case, $\mu(\mathbf{x}_i)$ will be less than the censoring point for a large portion of the support of $\mathbf{x}_i$, requiring estimation at these points sufficient to draw meaningful inference regarding its shape.

Another situation would be estimating relationships in the presence of some sort of constraint. Of interest from, say, a policy perspective would be to estimate how an economic agent would behave if the constraint were lifted. For example, a labor economist would be interested in estimating how long the unemployed would stay on unemployment insurance if the maximum time allowed were increased.

Our approach is based on a structural relationship between the conditional median and upper quantiles that holds for observations where $\mu(\mathbf{x}_i) \geq 0$. This relationship can be used to motivate an estimator for $\mu(\mathbf{x}_i)$ in the region where it is negative. Our results are thus based on the condition

$$P_{\mathbf{X}}\big(\mathbf{x}_i : \mu(\mathbf{x}_i) \geq 0\big) > 0, \tag{4}$$

where $P_{\mathbf{X}}(\cdot)$ denotes the probability measure of the random variable $\mathbf{x}_i$.

Songnian Chen is Professor of Economics, Department of Economics, Hong Kong University of Science and Technology, Hong Kong (E-mail: *snchen@ust.hk*). Gordon B. Dahl is Associate Professor of Economics (E-mail: *dahl@troi.cc.rochester.edu*) and Shakeeb Khan is Associate Professor of Economics (E-mail: *skhan@troi.cc.rochester.edu*), Department of Economics, University of Rochester, Rochester, NY 14627. An earlier version of this article was presented at the 2000 MEG at the University of Chicago, the 2001 CEME at the University of Rochester, the 2001 Summer Meetings of the Econometric Society at the University of Maryland, and the 2001 CESG at the University of Waterloo. The authors thank D. Card, B. E. Honoré, A. Lewbel, O. B. Linton, L. Lochner, J. L. Powell, and seminar participants at Boston College, Brown University, Duke University, Harvard/MIT, Syracuse, and the University of Rochester for their helpful comments. They also thank D. Card and P. Levine for generously providing the dataset. Chen gratefully acknowledges support from the RGC grant HKUST6070/01H, and Khan gratefully acknowledges support from the National Science Foundation through grant SES-0213621.

Journal of the American Statistical Association
March 2005, Vol. 100, No. 469, Theory and Methods
DOI 10.1198/016214504000000836

Variations of censored nonparametric models have been studied elsewhere in the literature. Lewbel and Linton (2002) estimated a nonparametric censored regression model with a fixed censoring point based on a mean restriction on the disturbance term. Their approach does not require the heteroscedasticity to have the multiplicative structure imposed here, but generally requires it to satisfy some sort of exclusion restriction. Van Keilegom and Akritas (1999) estimated a nonparametric conditional distribution function in a censored model without imposing a location restriction, and consequently estimate a general conditional location functional. An estimator of $\mu(\mathbf{x})$ can be naturally recovered from their estimator by using the identification conditions established in this article. In contrast, the estimator proposed here makes explicit use of this identification condition to directly identify and estimate the parameter of interest $\mu(\mathbf{x})$.

The article is organized as follows. The next section explains the key identification condition, and motivates a way to estimate the function $\mu(\cdot)$ at each point in the support of $\mathbf{x}_i$. Section 3 introduces the new estimation procedure and establishes the asymptotic properties of this estimator when the identification condition is satisfied. Section 4 explores the finite-sample properties of the estimator through the results of a simulation study. Section 5 presents an empirical illustration on unemployment insurance duration. Section 6 concludes by summarizing results and discussing extensions for future research. An Appendix presents proofs of the theorems.

2. IDENTIFICATION OF THE LOCATION FUNCTION

In this section we consider conditions necessary for identifying $\mu(\cdot)$ on $\mathcal{X}$, the support of x_i. Our identification results are based on the following assumptions:

- I1. The disturbance term ϵ_i is distributed independently of $\mathbf{x}_i$ and has a density function with respect to Lebesgue measure that is positive on $\mathbb{R}$.
- I2. ϵ_i has median 0.
- I3. $\mathcal{X}$ is a subset of $\mathbb{R}^d$, and the components of $\mathbf{x}_i$ may be discretely or continuously distributed. Without loss of generality, we assume that the vector $\mathbf{x}_i$ can be partitioned as $\mathbf{x}_i = (\mathbf{x}_i^{(ds)}, \mathbf{x}_i^{(c)})$, where $\mathbf{x}_i^{(ds)}$ is discretely distributed and $\mathbf{x}_i^{(c)}$ is continuously distributed.
- I4. The scale function $\sigma_0(\cdot)$ is continuous in $\mathbf{x}_i^{(c)}$ for all possible values of $\mathbf{x}_i^{(ds)}$, strictly positive and bounded on every bounded subset of $\mathcal{X}$.
- I5. The location function $\mu(\cdot)$ is continuous in $\mathbf{x}_i^{(c)}$ for all possible values of $\mathbf{x}_i^{(ds)}$, and $|\mu(\cdot)| < \infty$ on every bounded subset of $\mathcal{X}$.

The main result of this article establishes the sufficiency of (4) for identification of $\mu(\cdot)$ on every point in $\mathcal{X}$. The proof of the theorem suggests a natural estimator of $\mu(\cdot)$, so it is included in the main text.

Theorem 1 (Sufficiency). Suppose assumptions I1–I5 hold, and condition (4) holds. Then $\mu(\cdot)$ is identified for all $\mathbf{x} \in \mathcal{X}$.

Proof. We show identification sequentially. We first show identification for all points where $\mu(\cdot)$ is nonnegative. We then show how identification of μ in this range of the support of $\mathbf{x}_i$ can be used to identify μ where it is negative. To show identification in the nonnegative region, we let $\mathbf{x}_0$ be any point that satisfies $\mu(\mathbf{x}_0) \geq 0$. Suppose first that $\mu(\mathbf{x}_0) = 0$. We show that $\tilde{\mu}(\mathbf{x}_0) < 0$ or $\tilde{\mu}(\mathbf{x}_0) > 0$ leads to a contradiction. If $\tilde{\mu}(\mathbf{x}_0) = -\delta < 0$, then let $\tilde{\sigma}(\mathbf{x}_0)$ be a positive, finite number. Let c_α denote the αth quantile of ϵ_i, and let $q_\alpha(\cdot)$ denote the α conditional quantile of y_i as a function of $\mathbf{x}_i$. We note by assumption I1 that c_α, when viewed as a function of α, is continuous on $[0, 1]$ and has bounded derivative on any compact subset of $(0, 1)$. Thus if we let $\tilde{\epsilon}_i$ denote an alternative error term, then, by assumption I2, it must follow that $\tilde{c}_{.5} = 0$ and $0 < \tilde{c}_\alpha < \delta/\tilde{\sigma}(\mathbf{x}_0)$ for $\alpha \in (.5, .5 + \varepsilon)$, where $\delta = -\tilde{\mu}(\mathbf{x}_0)$ and ε is an arbitrarily small positive constant. Noting that $c_\alpha > 0$ for $\alpha \in (.5, .5 + \varepsilon)$, we have for $\alpha \in (.5, .5 + \varepsilon)$, $q_\alpha(\mathbf{x}_0) = \max(\mu(\mathbf{x}_0) + c_\alpha\sigma_0(\mathbf{x}_0), 0) = \max(c_\alpha\sigma_0(\mathbf{x}_0), 0) > 0$. Alternatively, we have

$$\tilde{q}_\alpha(\mathbf{x}_0) = \max\bigl(\tilde{\mu}(\mathbf{x}_0) + \tilde{\sigma}(\mathbf{x}_0)\tilde{c}_\alpha, 0\bigr) \tag{5}$$

$$< \max(-\delta + \delta, 0) = 0. \tag{6}$$

Thus we have found quantiles where $q_\alpha(\mathbf{x}_0) \neq \tilde{q}_\alpha(\mathbf{x}_0)$, which shows that $\mu(\mathbf{x}_0) = 0$ is distinguishable from negative alternatives. A similar argument can be used to show that it is distinguishable from positive alternatives, establishing its identification. It is even simpler to show that points $\mathbf{x}$ where $\mu(\mathbf{x}) > 0$ are identified. If $\mu(\mathbf{x}_0) > 0$ and $\tilde{\mu}(\mathbf{x}_0) \neq \mu(\mathbf{x}_0)$, then $q_{.5}(\mathbf{x}_0) = \mu(\mathbf{x}_0)$ and $\tilde{q}_{.5}(\mathbf{x}_0) = \max(\tilde{\mu}(\mathbf{x}_0), 0) \neq \mu(\mathbf{x}_0)$.

We next show how to identify $\mu(\mathbf{x})$ when $\mu(\mathbf{x}) < 0$ given that we have identified $\mu(\mathbf{x}_0)$ for $\mu(\mathbf{x}_0) \geq 0$. We first note that $\mu(\mathbf{x})$ and $\sigma_0(\mathbf{x})$ are finite by assumptions I5 and I4, and by assumption I1 there exists quantiles $\alpha_1 < \alpha_2 < 1$ such that

$$q_{\alpha_1}(\mathbf{x}) = \mu(\mathbf{x}) + c_{\alpha_1}\sigma_0(\mathbf{x}) > 0 \tag{7}$$

and

$$q_{\alpha_2}(\mathbf{x}) = \mu(\mathbf{x}) + c_{\alpha_2}\sigma_0(\mathbf{x}) > 0. \tag{8}$$

Note that because $q_{\alpha_1}(\mathbf{x})$, $q_{\alpha_2}(\mathbf{x})$ are observable, we can identify the values of $\mu(\mathbf{x})$ and $\sigma_0(\mathbf{x})$ from the values of c_{α_1} and c_{α_2} by solving the foregoing system of equations. These values are unknown, but we use the identification of $\mu(\mathbf{x}_0)$ to identify their values.

Before illustrating how to do this, we note that the distribution of ϵ_i is identified only up to scale, because $\sigma_0(\mathbf{x}_i)$ is unknown. For ease of exposition, we impose the scale normalization $c_{\alpha_1} \equiv 1$, so we need solve only for c_{α_2}.

We combine the following values of the conditional quantile function evaluated at the three distinct quantiles .5, α_1, and α_2, at the regressor value $\mathbf{x}_0$:

$$q_{.5}(\mathbf{x}_0) = \mu(\mathbf{x}_0), \tag{9}$$

$$q_{\alpha_1}(\mathbf{x}_0) = \mu(\mathbf{x}_0) + c_{\alpha_1}\sigma_0(\mathbf{x}_0), \tag{10}$$

and

$$q_{\alpha_2}(\mathbf{x}_0) = \mu(\mathbf{x}_0) + c_{\alpha_2}\sigma_0(\mathbf{x}_0). \tag{11}$$

This enables us to identify c_{α_2} as

$$c_{\alpha_2} = \frac{q_{\alpha_2}(\mathbf{x}_0) - q_{.5}(\mathbf{x}_0)}{q_{\alpha_1}(\mathbf{x}_0) - q_{.5}(\mathbf{x}_0)}, \tag{12}$$

which immediately translates into identification of $\mu(\mathbf{x})$ and $\sigma_0(\mathbf{x})$ from the relationships

$$q_{\alpha_\ell}(\mathbf{x}) = \mu(\mathbf{x}) + c_{\alpha_\ell}\sigma_0(\mathbf{x}), \qquad \ell = 1, 2, \tag{13}$$

by solving the two-equation system for the two unknowns. This completes the proof of the theorem.

Remark 1. Although the foregoing theorem establishes the sufficiency of (4), it can also be shown that this condition is necessary for identification. That is, identification of $\mu(\cdot)$ is impossible anywhere on the support of $\mathbf{x}_i$ if $\mu(\cdot)$ is negative everywhere on $\mathcal{X}$.

Remark 2. The (unbounded) support condition on ϵ_i in assumption I1 is not always necessary. In the proof it was used only to ensure that the quantile function exceeded the censoring point for a large enough quantile. Therefore, given the assumptions on $\mu(\cdot)$ and $\sigma_0(\cdot)$ (I4 and I5), ϵ_i can have a bounded support if the support of $\mathbf{x}_i$ is bounded.

Remark 3. Identification of $\mu(\cdot)$ where it is negative involves identification of the quantiles of the homoscedastic component of the disturbance term. Thus an additional consequence of condition (4) being satisfied is that the quantiles of ϵ_i are identified for all $\alpha \geq \alpha_0 \equiv \inf\{\alpha : \sup_{\mathbf{x}\in\mathcal{X}} q_\alpha(\mathbf{x}) > 0\}$. [We note that $\alpha_0 \leq .5$ when condition (4) holds.] This result can be used to estimate and construct hypothesis tests regarding the distribution of ϵ_i, and we consider this when we estimate $\mu(\cdot)$ and $\sigma_0(\cdot)$. We also note that if the econometrician were to impose a distributional form on ϵ_i, then the (known) values of c_{α_1} and c_{α_2} could be used to identify and estimate the location and scale functions, without requiring condition (4).

We conclude this section by noting that our identification result involved only the identification of one additional disturbance quantile. Identification of further quantiles would overidentify the parameters of interest and will be incorporated in the estimation procedure to improve efficiency.

3. ESTIMATION PROCEDURE AND ASYMPTOTIC PROPERTIES

3.1 Estimation Procedure

In this section we consider estimation of the functions $\mu(\cdot)$ and $\sigma_0(\cdot)$. Our procedure is based on our identification results in the previous section, but modified to incorporate more information in the model. Identification is shown through the use of two quantiles corresponding to quantile functions exceeding the censoring point. The estimation procedure uses several quantiles to improve the efficiency of estimating the functions, as was alluded to at the end of the previous section.

Our procedure involves nonparametric quantile regression at different quantiles and different points in the support of the regressors. Our asymptotic arguments are based on the local polynomial estimator for conditional quantile functions introduced by Chaudhuri (1991a,b). For expositional ease, we describe this nonparametric estimator only for a polynomial of degree 0 (see Chaudhuri 1991a,b; Chaudhuri, Doksum, and Samarov 1997, Chen and Khan 2000, 2001; Khan 2001; Khan and Powell 2001 for the additional notation involved for polynomials of arbitrary degree).

First, recall that we assumed that the regressor vector $\mathbf{x}_i$ can be partitioned as $(\mathbf{x}_i^{(ds)}, \mathbf{x}_i^{(c)})$, where the d_{ds}-dimensional vector $\mathbf{x}_i^{(ds)}$ is discretely distributed and the d_c-dimensional vector $\mathbf{x}_i^{(c)}$ is continuously distributed. We let $C_n(\mathbf{x}_i)$ denote the cell of observation $\mathbf{x}_i$ and let h_n denote the sequence of bandwidths that govern the size of the cell. For some observation $\mathbf{x}_j$, $j \neq i$, we let $\mathbf{x}_j \in C_n(\mathbf{x}_i)$ denote that $\mathbf{x}_j^{(ds)} = \mathbf{x}_i^{(ds)}$, and $\mathbf{x}_j^{(c)}$ lies in the d_c-dimensional cube centered at $\mathbf{x}_i^{(c)}$ with side length h_n.

Let $I[\cdot]$ be an indicator function, taking the value 1 if its argument is true and 0 otherwise. Our estimator of the conditional αth quantile function at a point $\mathbf{x}_i$ for any $\alpha \in (0, 1)$ involves αth quantile regression (see Koenker and Bassett 1978) on observations that lie in the defined cells of $\mathbf{x}_i$. Specifically, let $\hat{\theta}$ minimize

$$\sum_{j=1}^{n} I[\mathbf{x}_j \in C_n(\mathbf{x}_i)]\rho_\alpha(y_j - \theta), \tag{14}$$

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

Our estimation procedure is based on a random sample of n observations of the vector $(y_i, \mathbf{x}_i')'$ and involves applying the local polynomial estimator at three stages. As we explain, each of the stages is analogous to a step in the identification proof in the previous section. Throughout our description of the three stages, "$\hat{\cdot}$" denotes estimated values.

Stage 1. Local Constant Estimation of the Conditional Median Function. In the first stage, we estimate the conditional median at each point in the sample, using a polynomial of degree 0. We let h_{1n} denote the bandwidth sequence used in this stage. Following the terminology of Fan (1992), we refer to this as a local constant estimator, and denote the estimated values by $\hat{q}_{.5}(\mathbf{x}_i)$. These estimated values enable us to determine the observation(s) whose median exceeds the censoring value, analogous to $\mathbf{x}_0$ in the proof of Theorem 1.

Stage 2. Weighted Average Estimation of the Disturbance Quantiles. The second stage is the estimation analog of (12). We generalize the results from the previous section in two ways, by using several quantiles (as opposed to two), and by using all of the observations whose median exceeds the censoring point. These extensions serve to improve the precision of the estimator.

Specifically, for $\alpha_1 < \alpha_2 < \cdots < \alpha_{\mathcal{N}}$, we estimate the unknown disturbance quantiles c_{α_ℓ}, $\ell = 1, 2, \ldots, \mathcal{N}$. To impose a scale normalization that is invariant to quantiles on the grid, we select a quantile, $\alpha_s > .5$, which need not be on the grid, and set $c_{\alpha_s} \equiv 1$ and estimate c_{α_ℓ} up to scale by a weighted average of local polynomial estimators. The estimator is based on (12). In this stage we use a polynomial of degree k, and denote the second-stage bandwidth sequence by h_{2n}. We use the superscript (p) to distinguish the estimator of the median function in this stage from that in the first stage. Letting $\hat{c}_{\alpha_\ell}$ denote the estimators of the unknown constants c_{α_ℓ}, we define them as

$$\hat{c}_{\alpha_\ell} = \frac{1}{n}\sum_{i=1}^{n} \tau(\mathbf{x}_i) w\big(\hat{q}_{.5}(\mathbf{x}_i)\big) \times \frac{(\hat{q}_{\alpha_\ell}(\mathbf{x}_i) - \hat{q}_{.5}^{(p)}(\mathbf{x}_i))}{(\hat{q}_{\alpha_s}(\mathbf{x}_i) - \hat{q}_{.5}^{(p)}(\mathbf{x}_i))} \Big/ \frac{1}{n}\sum_{i=1}^{n} \tau(\mathbf{x}_i) w\big(\hat{q}_{.5}(\mathbf{x}_i)\big), \tag{15}$$

where $\tau(x_i)$ is a trimming function whose support, denoted by $\mathcal{X}_\tau$, is a compact set that lies strictly in the interior of $\mathcal{X}$. The trimming function serves to eliminate "boundary effects" that arise in nonparametric estimation. The function $w(\cdot)$ "selects" the observations whose median is positive. We detail its properties in the regularity conditions in the next section.

Stage 3. Local Polynomial Estimation at the Point of Interest. The third stage is an estimation analog of (13) generalized to allow for more than two quantiles. We can extend (13) to

$$q_{\alpha_\ell}(\mathbf{x}) = \mu(\mathbf{x}) + c_{\alpha_\ell}\sigma_0(\mathbf{x}), \tag{16}$$

which now holds true for any α_ℓ such that $q_{\alpha_\ell}(\mathbf{x}) > 0$. Letting $\mathbf{x}$ denote the point at which the functions $\mu(\cdot)$ and $\sigma_0(\cdot)$ are to be estimated, we combine the local polynomial estimator (with polynomial order k and bandwidth sequence h_{3n}) of the conditional quantile function at $\mathbf{x}$ using quantiles α_ℓ, along with the estimator of the unknown disturbance quantiles, $\hat{c}_{\alpha_\ell}$ to yield the estimator of $\mu(\mathbf{x}), \sigma_0(\mathbf{x})$. Let $\theta_0(\mathbf{x})$ denote the vector $(\mu(\mathbf{x}), \sigma_0(\mathbf{x}))$, and let $\hat{\theta}(\mathbf{x})$ denote its estimator, which we define as the least squares estimator based on the $\mathcal{N}$ "observations," which treat $\hat{q}_{\alpha_\ell}(\mathbf{x})$ as "dependent variables," and $\hat{c}_{\alpha_\ell}$ as "independent variables." Specifically, letting $\hat{\mathbf{c}}_{\alpha_\ell}$ denote the vector $(1, \hat{c}_{\alpha_\ell})$ and $\hat{\hat{d}}_{\alpha_\ell}$ denote the indicator $I[\hat{q}_{\alpha_\ell}(\mathbf{x}) \geq \varepsilon]$, where $\varepsilon > 0$ is a small constant, we can define the estimator as

$$\hat{\theta}(\mathbf{x}) = \left(\sum_{\ell=1}^{\mathcal{N}} \hat{\hat{d}}_{\alpha_\ell}\hat{\mathbf{c}}_{\alpha_\ell}\hat{\mathbf{c}}'_{\alpha_\ell}\right)^{-1}\left(\sum_{\ell=1}^{\mathcal{N}} \hat{\hat{d}}_{\alpha_\ell}\hat{\mathbf{c}}_{\alpha_\ell}\hat{q}_{\alpha_\ell}(\mathbf{x})\right). \tag{17}$$

Remark 4. We note that the asymptotic results in this article can be obtained using any nonparametric procedure for estimating conditional quantiles of the observed dependent variable in the various stages. Examples of such procedures include those proposed by Beran (1981), Dabrowska (1987, 1992), and Fan and Gijbels (1994, 1996). These procedures have the advantage of accommodating models with random censoring, which often occurs in biostatistics. However, under their assumptions, models with fixed censoring are ruled out because they assume that the censoring variable is continuously distributed. Nonetheless, the identification results in this article carry over to randomly censored models, and implementation would require replacing the local polynomial estimator used here with one of these estimators.

Remark 5. We note here that a different order polynomial is used in the first stage than in the other two stages. The reason for this is that even though the functions $\mu(\cdot)$ and $\sigma_0(\cdot)$ are assumed to be k-times differentiable, the quantile functions in general will not be smooth at the censoring point. Thus a local polynomial estimator may not work well when the quantile function is in a neighborhood of the censoring point. However, once points in the sample that are greater than the censoring point are "selected" in the first stage, the quantile function at these points are sufficiently smooth for the local polynomial estimators to be used in the second and third stages.

Remark 6. We note that the second stage of the procedure estimated the disturbance quantiles c_{α_ℓ}. Interestingly, because these estimators involve averaging quantile estimates over different value of $\mathbf{x}_i$, these estimators can converge at the parametric (root-n) rate, as can be proven following the arguments in the Appendix.

3.2 Asymptotic Properties

In this section we establish the asymptotic properties of our estimation procedure. Our results are based on the following assumptions:

Assumption A1 (Identification). The weighting function is positive with positive probability

$$P_X\big(\tau(\mathbf{x}_i)w\big(q_{.5}(\mathbf{x}_i)\big) > 0\big) > 0.$$

Assumption A2 (Random sampling). The sequence of $(d+1)$-dimensional vectors $(y_i, \mathbf{x}_i)$ are independent and identically distributed.

Assumption A3 (Weighting function properties). The weighting function, $w(\cdot): \mathbb{R} \to \mathbb{R}^+$ has the following properties:

A3.1. $w(\cdot) \in [0, 1]$ and is continuously differentiable with bounded derivative.

A3.2. $w \equiv 0$ if its argument is less than η, a small positive constant.

Assumption A4 (Regressor distribution). We let $f_{\mathbf{X}^{(c)}|\mathbf{X}^{(ds)}}(\cdot|\mathbf{x}^{(ds)})$ denote the conditional density function of $\mathbf{x}_i^{(c)}$ given $\mathbf{x}_i^{(ds)} = \mathbf{x}^{(ds)}$, and assume that it is positive and finite on $\mathcal{X}_\tau$.

We let $f_{\mathbf{X}^{(ds)}}(\cdot)$ denote the mass function of $\mathbf{x}_i^{(ds)}$, and assume a finite number of mass points on $\mathcal{X}_\tau$.

Hereafter, we let $f_{\mathbf{X}}(\cdot)$ denote $f_{X^{(c)}|\mathbf{X}^{(ds)}}(\cdot|\cdot)f_{\mathbf{X}^{(ds)}}(\cdot)$.

Assumption A5 (Disturbance density). The disturbance terms ϵ_i are assumed to have a continuous distribution with density function that is bounded, positive, and continuous on $\mathbb{R}$.

Assumption A6 (Orders of smoothness). We assume the following smoothness conditions on the regressor density, trimming, location, and scale functions:

A6.1. $f_{\mathbf{X}}(\cdot)$ and $\tau(\cdot)$ are continuous on $\mathcal{X}_\tau$.

A6.2. $\mu(\cdot)$ and $\sigma_0(\cdot)$ are differentiable in $\mathbf{x}_i^{(c)}$ of order $p = k+1$, with pth order derivatives being continuous on $\mathcal{X}_\tau$.

Assumption A7 (Bandwidth conditions). The bandwidths used in each of the three stages are assumed to satisfy the following conditions:

A7.1. h_{1n} satisfies $\frac{\log n}{nh_{1n}^{d_c}} \to 0$, $h_{1n} \to 0$.

A7.2. h_{2n} satisfies $\frac{\log n}{nh_{2n}^{d_c}} \to 0$, $n^{\frac{p}{2p+d_c}}h_{2n}^p \to 0$.

A7.3. h_{3n} is of the form $h_{3n} = \kappa_0 n^{\frac{-1}{2p+d_c}}$, where κ_0 is a positive constant.

Remark 7. The weighting function $w(\cdot)$ in Assumption A3 serves as a smooth approximation to an indicator function, selecting those observations for which the estimated value of the conditional median function is positive. For technical reasons, we require that the weighting function assign only positive weight to estimated conditional median values that are bounded away from 0.

Remark 8. The bandwidth sequences h_{1n}, h_{2n}, and h_{3n} in Assumption A7 are required to satisfy different conditions. The conditions on h_{1n} in Assumption A7.1 ensure consistency of the first-stage estimator. The conditions on h_{2n} in Assumption A7.2, reflect "undersmoothing," implying that the bias of the nonparametric estimators used in this stage converges to 0 at a faster rate than the standard deviation. In contrast, Assumption A7.3 imposes the optimal rate for h_{3n}, so that the estimator of $\theta_0(\cdot)$ will converge at the optimal nonparametric rate.

We now characterize the limiting distribution for the proposed estimator of $\theta_0(\mathbf{x})$, where $\mathbf{x}$ is assumed to lie in the interior of the support of $\mathbf{x}_i$. Before stating the theorem, we explain the notation that we adopt to characterize higher-order derivatives of quantile functions that arise in the limiting distribution.

For a d_c-dimensional vector of integers $\mathbf{A}$, let $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_{d_c}$ denote its individual components, whose sum we denote by $s(\mathbf{A})$. For a regressor vector value $\mathbf{x}$, denote the individual components of its continuously distributed subvector by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{d_c}$. For any quantile function of the regressors, $q_{\alpha_\ell}(\mathbf{x})$, we let $\nabla_{\mathbf{A}} q_{\alpha_\ell}(\mathbf{x})$ denote the $s(\mathbf{A})$th order derivative of $q_{\alpha_\ell}(\mathbf{x})$,

$$q_{\alpha_\ell}(\mathbf{x}) = \frac{\partial^{s(\mathbf{A})}}{\partial x_1^{\mathbf{A}_1}\,\partial x_2^{\mathbf{A}_2}\cdots\partial x_{d_c}^{\mathbf{A}_{d_c}}} q_{\alpha_\ell}(\mathbf{x}). \tag{18}$$

Also, we define the constant $\mathcal{K}_{\mathbf{A}}$ as

$$\mathcal{K}_{\mathbf{A}} = \int_{[-\frac{1}{2},\frac{1}{2}]^{d_c}} u_1^{\mathbf{A}_1}\cdots u_{d_c}^{\mathbf{A}_{d_c}}\, du_1\cdots du_{d_c}, \tag{19}$$

where $[-\frac{1}{2},\frac{1}{2}]^{d_c}$ denotes the Cartesian product of the d_c intervals $[-\frac{1}{2},\frac{1}{2}]$.

With this notation, we can state the main theorem, which establishes that the proposed estimator converges at the optimal nonparametric rate and has a limiting noncentered normal distribution. The proof is deferred to the Appendix.

Theorem 2. If Assumptions A1–A7 hold, then

$$n^{\frac{p}{2p+d_c}}\big(\hat\theta(\mathbf{x}) - \theta_0(\mathbf{x})\big) \Rightarrow \mathrm{N}(\mathbf{C}^{-1}\mathbf{B}, \mathbf{C}^{-1}\mathbf{V}\mathbf{C}^{-1}), \tag{20}$$

where

$$\mathbf{C} = \sum_{\ell=1}^{\mathcal{N}} \tilde d_{\alpha_\ell}\mathbf{c}_{\alpha_\ell}\mathbf{c}'_{\alpha_\ell} \tag{21}$$

and

$$\begin{aligned}\mathbf{V} = {}& \sum_{\ell=1}^{\mathcal{N}} \tilde d_{\alpha_\ell}\mathbf{c}_{\alpha_\ell}\mathbf{c}'_{\alpha_\ell}\kappa_0^{-d_c} f_{Y|\mathbf{X}}\big(q_{\alpha_\ell}(\mathbf{x})|\mathbf{x}\big)^{-2}\alpha_\ell(1-\alpha_\ell)\\ &+ \sum_{\alpha_{\ell_1}<\alpha_{\ell_2}} \big(2\tilde d_{\alpha_{\ell_1}}\tilde d_{\alpha_{\ell_2}}\mathbf{c}_{\alpha_{\ell_1}}\mathbf{c}'_{\alpha_{\ell_2}}\kappa_0^{-d_c} f_{Y|\mathbf{X}}\big(q_{\alpha_{\ell_1}}(\mathbf{x})|\mathbf{x}\big)^{-1}\\ &\qquad\times f_{Y|\mathbf{X}}\big(q_{\alpha_{\ell_2}}(\mathbf{x})|\mathbf{x}\big)^{-1}\alpha_{\ell_1}(1-\alpha_{\ell_2})\big),\end{aligned} \tag{22}$$

with $\tilde d_{\alpha_\ell}$ denoting the indicator based on true quantile function values, $f_{Y|\mathbf{X}}(\cdot)$ denoting the conditional density function of y_i, and $\sum_{\alpha_{\ell_1}<\alpha_{\ell_2}}$ denoting the sum of pairs of quantiles on the grid satisfying $\alpha_{\ell_1}<\alpha_{\ell_2}$.

The limiting bias of the proposed estimator of $\theta_0(\mathbf{x})$ is a weighted sum of limiting biases associated with each quantile,

$$\mathbf{B} = \sum_{\ell=1}^{\mathcal{N}} \tilde d_{\alpha_\ell}\mathbf{c}_{\alpha_\ell} B_{\alpha_\ell}, \tag{23}$$

where the limiting bias of each local polynomial quantile function estimator is of the form

$$B_{\alpha_\ell} = \frac{\kappa_0^p}{p!}\sum_{\mathbf{A}:\,s(\mathbf{A})=p} \mathcal{K}_{\mathbf{A}}\nabla_{\mathbf{A}} q_{\alpha_\ell}(\mathbf{x}). \tag{24}$$

For conducting inference, one approach would be to consistently estimate the components of the variance matrix with consistent estimators of its components $\mathbf{C}$ and $\mathbf{V}$. A simple estimator of $\mathbf{C}$ replaces unknown values $\mathbf{c}_{\alpha_\ell}$ and $\tilde d_{\alpha_\ell}$ with their estimators,

$$\hat C = \sum_{\ell=1}^{\mathcal{N}} \hat{\tilde d}_{\alpha_\ell}\hat{\mathbf{c}}_{\alpha_\ell}\hat{\mathbf{c}}'_{\alpha_\ell}. \tag{25}$$

Estimating $\mathbf{V}$ is more difficult due to the presence of the conditional density functions $f_{Y|\mathbf{X}}(\cdot|\cdot)$. To estimate this, we propose a Nadaraya–Watson estimator,

$$\hat f_{Y|\mathbf{X}}\big(\hat q_{\alpha_\ell}(\mathbf{x})|\mathbf{x}\big) = \frac{\frac{1}{n\hbar_n^{d+1}}\sum_{i=1}^n k_1(\frac{\mathbf{x}_i-\mathbf{x}}{\hbar_n})k_2(\frac{y_i-\hat q_{\alpha_\ell}(\mathbf{x})}{\hbar_n})}{\frac{1}{n\hbar_n^{d}}\sum_{i=1}^n k_1(\frac{\mathbf{x}_i-\mathbf{x}}{\hbar_n})},$$

where $k_1(\cdot)$ and $k_2(\cdot)$ are kernel functions that are continuously differentiable with compact supports and $\hbar_n$ is a bandwidth sequence. Our estimator of $\mathbf{V}$ is

$$\begin{aligned}\hat{\mathbf{V}} = {}& \sum_{\ell=1}^{\mathcal{N}} \hat{\tilde d}_{\alpha_\ell}\hat{\mathbf{c}}_{\alpha_\ell}\hat{\mathbf{c}}'_{\alpha_\ell}\hat f_{Y|\mathbf{X}}\big(\hat q_{\alpha_\ell}(\mathbf{x})|\mathbf{x}\big)^{-2}\alpha_\ell(1-\alpha_\ell)\\ &+ \sum_{\alpha_{\ell_1}<\alpha_{\ell_2}} \big(2\hat{\tilde d}_{\alpha_{\ell_1}}\hat{\tilde d}_{\alpha_{\ell_2}}\hat{\mathbf{c}}_{\alpha_{\ell_1}}\hat{\mathbf{c}}'_{\alpha_{\ell_2}}\hat f_{Y|\mathbf{X}}\big(\hat q_{\alpha_{\ell_1}}(\mathbf{x})|\mathbf{x}\big)^{-1}\\ &\qquad\times \hat f_{Y|\mathbf{X}}\big(\hat q_{\alpha_{\ell_2}}(\mathbf{x})|\mathbf{x}\big)^{-1}\alpha_{\ell_1}\big(1-\alpha_{\ell_2}\big)\big).\end{aligned}$$

The following theorem establishes the consistency of the proposed estimator. Its proof is deferred to the Appendix.

Theorem 3. If Assumptions A1–A7 hold, and $\hbar_n$ satisfies $\hbar_n \to 0$, $n\hbar_n^{d_c+1}\to\infty$ and $n^{p/(2p+d_c)}\hbar_n\to\infty$, then

$$\hat{\mathbf{C}}^{-1}\hat{\mathbf{V}}\hat{\mathbf{C}}^{-1} \xrightarrow{p} \mathbf{C}^{-1}\mathbf{V}\mathbf{C}^{-1}.$$

We conclude this section with a brief discussion on the structure of the asymptotic variance of the proposed estimator, and how efficiency may be improved. The ordinary least squares structure of the third stage of our estimator results in a "sandwich" form of the limiting variance matrix. Efficiency can be improved by a generalized least squares approach, which would weight the data in accordance with the structure of the variance matrix. Feasibly implementing this procedure would require estimates of the conditional density of the dependent variable at the selected quantiles. The choice of such weights would make our procedure loosely analogous to that proposed by Newey and Powell (1990).

Finally, we note that Theorem 2 provides only the limiting distribution for a single point, $\mathbf{x}$. However, this will suffice for deriving the joint limiting distribution for several distinct points. This is because the joint distribution of a local polynomial estimator at various points exhibits asymptotic independence across points, because the rate at which the bandwidth converges to 0 ensures that distinct observations are used for the different points, analogous to results found in theorem 2.2.3 of Bierens (1987).

4. MONTE CARLO RESULTS

In this section we explore the finite-sample properties of the proposed estimator by way of a small-scale simulation study. We simulated from designs of the form

$$y_i = \max\bigl(\mu(x_i) + \sigma_0(x_i)\epsilon_i, 0\bigr),$$

where x_i is a random variable distributed uniformly between -1 and 1, ϵ_i is distributed standard normal, and the scale function $\sigma_0(x_i)$ is set to $e^{.15x_i}$. We considered four different functional forms for $\mu(x_i)$ in our study:

1. $\mu(x) = x$
2. $\mu(x) = x^2 - .35$
3. $\mu(x) = x^3$
4. $\mu(x) = e^x - 1.05$.

Here we chose the constants .35 and 1.05 in the second and fourth designs so that the censoring level was 50%, as it was for the other two designs.

To implement the estimator, we used a grid of 25 quantiles from .5 to .98. For the quantile estimators, we fitted a local constant used in the first stage, using a bandwidth of $n^{-1/5}$, and used a local linear estimator in the second and third stages, using a bandwidth of the form $\kappa_0 n^{-1/5}$. We selected the constant κ_0 using the "rule of thumb" approach detailed Fan and Gijbels (1996, p. 202).

The results, given in Figure 1, are based on sample sizes of $n = 100$ and $n = 400$, with 401 replications. The function $\mu(\cdot)$ was estimated at 100 equispaced points, and the figure plots the average value of the estimated function, denoted by $m_1(x)$. Also reported is an analogous plot for results obtained from implementing the Van Keilegom and Arkritis (1999) estimator, referred to hereafter as the VK estimator, and on the figure as $m_2(x)$. Because the VK estimator imposed no location restriction on ϵ_i, we impose a median 0 condition and extend the identification results in this article to modify the VK estimator accordingly to identify $\mu(\cdot)$. The VK estimator involves implementing the conditional Kaplan–Meier estimator of Beran (1981), and to do so we used a Gaussian kernel function and a bandwidth sequence of $s_1 n^{-1/5}$, where we determined the constant s_1 using the rule-of-thumb approach for density estimation introduced by Silverman (1986). We chose the rate of $n^{-1/5}$ to coincide with the rate used by Van Keilegom, Arkritis, and Veraverbeke (2001), who conducted a thorough simulation study of the performance of the VK estimator. Plots of mean values for each of the estimators are presented alongside the true function, denoted in the figure by $\mu(x)$. Also reported (in parentheses) is the average mean squared error (AMSE) for each estimator.

As indicated by the figure, for $n = 100$, our estimator performs very well at points where $\mu(x) \geq 0$, and is further away from the truth the further $\mu(x)$, in its negative range, is from 0. We note that the VK estimator generally exhibits a larger bias than our estimator, and this is especially true for the quadratic function. Both estimators perform better for $n = 400$, where they are on average closer to the true function value on its entire support. We note that although neither estimator dominates

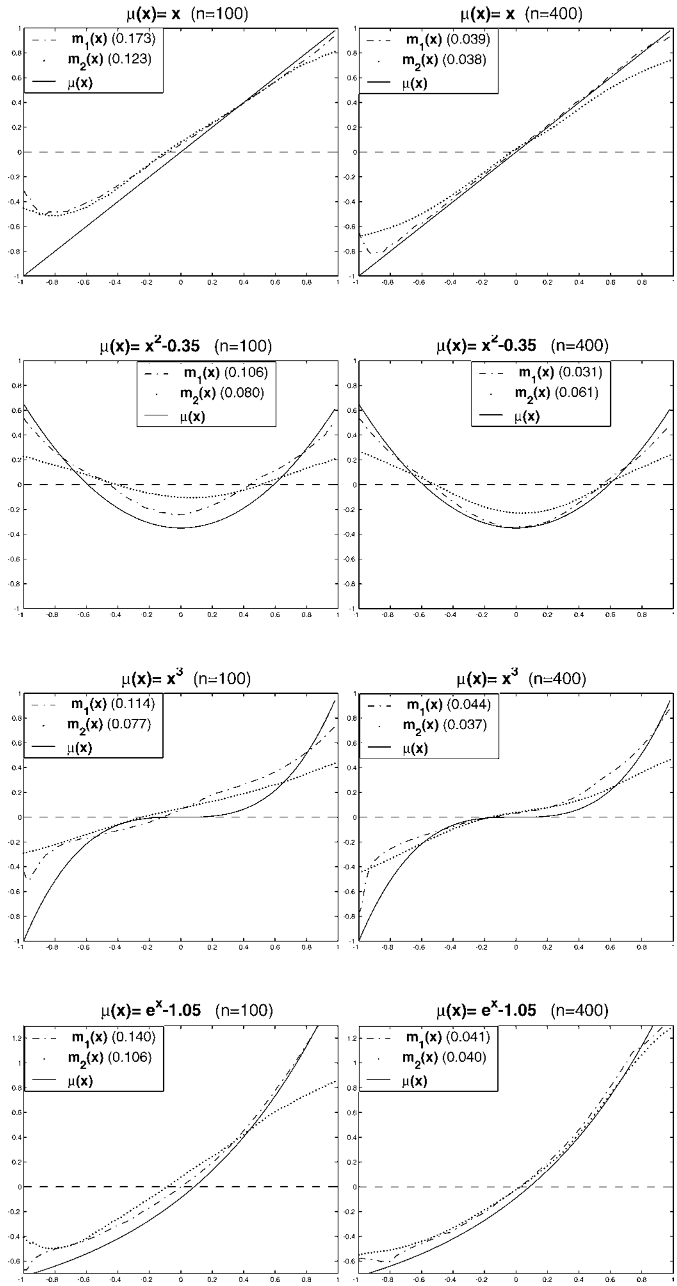

Figure 1. Simulation Results.

in terms of AMSE at $n = 400$, the VK estimator has a smaller AMSE for $n = 100$. This may be due to the fact that the local linear estimator estimates an additional parameter.

Although our results are very encouraging in general, we would expect a worse finite-sample performance when more regressors are present, because the rate of convergence would be slower.

5. ILLUSTRATION

The estimation procedure developed in this article applies to a variety of statistical problems with censoring. To illustrate how the method might be implemented in practice, we consider the example of unemployment insurance (UI) receipt. Previous research has focused on the responses of individuals who collect UI benefits for less than the time limit and generally based predictions on strong parametric or functional form assumptions (Card and Levine 2000; Ham and Rea 1987; Katz and Meyer 1990; Meyer 1990; Moffit and Nicholson 1982; Moffit 1985; Woodbury and Murray 1997).

We model the log of the number of weeks that an individual would like to collect UI benefits (i.e., log failure time) in the general framework described in Section 1. The censoring point is fixed at the time limit of 26 weeks; the latent number of weeks a claimant would like to collect benefits is observed only when a claimant collects benefits for less than 26 weeks. Our approach allows estimation of the location function $\mu(\mathbf{x})$ without parametric and structural assumptions. We caution, however, that the estimators developed in this article do not necessarily predict what will happen if maximum durations increase, because both censored and uncensored individuals may exhibit a behavioral response to a longer time limit. A conservative interpretation of our results is that our estimates represent lower bounds on the effect of increasing the UI time limit. This interpretation makes sense, because time limit increases should not shorten spell lengths.

The data are taken from individual-level administrative records for New Jersey's UI program in 1996 and 1997. (Details on New Jersey's UI system are given in Card and Levine 2000.) For the purpose of illustration, we restrict our sample to claimants who are male, white, are not in a union, are between age 20 and 65, have 6–18 years of completed education, and are eligible for 26 weeks of UI receipt. This results in a dataset containing 56,938 individuals, 38% of whom are censored because they exhaust their 26-week benefit eligibility. After restricting the sample, the two remaining characteristics that we hypothesize might influence median spell length are age and education. Following Chamberlain (1994), we use age and education, both measured in years, to partition the sample into 598 age–education cells. This represents the complete interaction of age (46 different years) with education (13 different years). We then discard cells with fewer than 30 observations, leaving 282 cells and reducing the sample size to 53,373 observations.

Estimation follows the stages described in Section 3. Because the data are discrete, local polynomials of degree 0 (i.e., constant terms) are used in all stages. In the first stage, for each cell with less than 50% censoring, cell medians are calculated. In the second stage, estimates of the unknown disturbance quantiles are estimated up to scale. We first obtain cell-level estimates for quantiles below the median. We use a grid of 25 quantiles evenly spaced between .02 and .50. The scale normalization that we chose was to set $c_{.3} \equiv 1$ (i.e., $\alpha_s = .3$); other choices do not alter the general findings. Estimates of the c_{α_ℓ}'s are then formed by taking the weighted average defined in (15). The third stage regresses the estimated quantiles $\hat{q}_{\alpha_\ell}(\mathbf{x})$ on the estimated constants $\hat{c}_{\alpha_\ell}$ and a vector of ones, as described in (16). Standard errors for $\hat{\mu}(\mathbf{x})$ are calculated using the results from Theorems 2 and 3. To estimate the conditional density functions $f_{Y|\mathbf{X}}(q_{\alpha_\ell}(\mathbf{x})|\mathbf{x})$ that appear in (19), we use a normal kernel and Silverman's (1986) normal bandwidth reference rule.

To illustrate how median spell length varies with age, consider the group of claimants with exactly 16 years of education (i.e., roughly a college degree). This subset of the data comprises 13,070 individuals. Before discussing the location function estimates, we note that the estimated $\sigma_0(\mathbf{x})$'s for this group exhibit a U-shaped pattern in age. This suggests the presence of conditional heteroscedasticity, a possibility allowed for by our estimator. Figure 2 displays the estimated location function (measured in log weeks) by age, for individuals with exactly 16 years of education. The horizontal line in the figure marks the censoring value. Pointwise 95% confidence intervals based on asymptotic normality are included in the graph as well. The figure reveals that the estimated median spell length generally increases with age. The impact of age does not appear to be linear, however, with the estimated median rising faster with age for older claimants. The change in slope with age occurs near the point when the median is estimated to be past the censoring point, illustrating the importance of a flexible estimation approach.

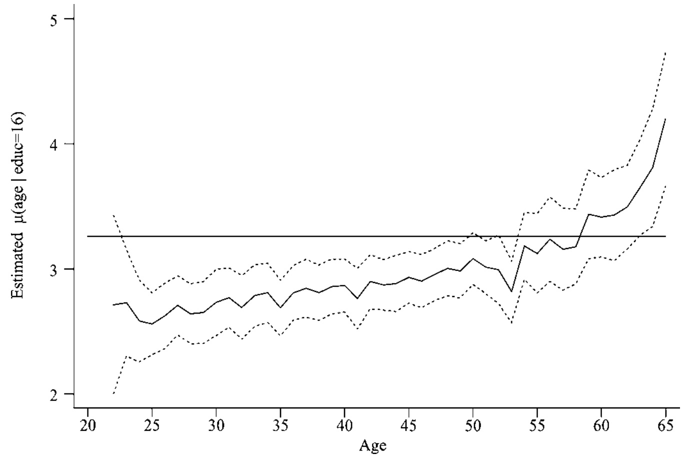

Figure 2. Estimated Median UI Spell Length by Age, for Claimants With 16 Years of Education. Spell length is measured in log weeks. The horizontal line indicates the censoring value, and the dashed lines indicate pointwise 95% confidence intervals.

6. CONCLUSIONS

In this article we have established conditions for nonparametric identification of the location and scale functions in a censored regression model. An estimation procedure was proposed and was shown to have desirable asymptotic properties. The procedure is simple to implement, because it is based on various quantiles of the conditional distribution of the dependent variable, and can be computed by linear programming methods. A Monte Carlo study indicates that the estimator performs well in finite samples. In an empirical illustration using UI spell data, we estimate the effects of extending benefits beyond the current 26-week maximum in New Jersey.

The results in this article suggest areas for future research. The estimator introduced here suggest testing parametric forms of the regression function against nonparametric alternatives in the censored regression model, as has been done in standard regression models (Bierens and Ploberger 1997; Horowitz and Spokoiny 2001). Also, following Remark 4, another important extension would be to allow for randomly censored datasets, as given by Buckley and James (1979), Koul, Suslara, and Van Ryzin (1981), Ying, Jung, and Wei (1995), and Honoré, Khan, and Powell (2002) for semiparametric models and by Van Keilegom and Akritas (1999) for nonparametric models.

APPENDIX: PROOFS

A.1 Proof of Theorem 2

In this section we prove the limiting distribution results stated in Theorem 2. Throughout this section, we adopt new notation. Here we let $\tau_i, \sigma_i, w_i, \hat{w}_i, \hat{q}_{0i}, \hat{q}_{\ell i}, \hat{q}_{si}, q_{0i}, q_{\ell i}, q_{si}, \hat{q}_0, \hat{q}_\ell, \hat{q}_s, q_0, q_\ell, q_s, C_{ni}, C_n$, and N_{ni} denote $\tau(\mathbf{x}_i)$, $\sigma_0(\mathbf{x}_i)$, $w(q_{.5}(\mathbf{x}_i))$, $w(\hat{q}_{.5}(\mathbf{x}_i))$, $\hat{q}_{.5}(\mathbf{x}_i)$, $\hat{q}_{\alpha_\ell}(\mathbf{x}_i)$, $\hat{q}_{\alpha_s}(\mathbf{x}_i)$, $q_{.5}(\mathbf{x}_i)$, $q_{\alpha_\ell}(\mathbf{x}_i)$, $q_{\alpha_s}(\mathbf{x}_i)$, $\hat{q}_{.5}(\mathbf{x})$, $\hat{q}_{\alpha_\ell}(\mathbf{x})$, $\hat{q}_{\alpha_s}(\mathbf{x})$, $q_{.5}(\mathbf{x}), q_{\alpha_\ell}(\mathbf{x}), q_{\alpha_s}(\mathbf{x}), C_n(\mathbf{x}_i), C_n(\mathbf{x})$, and $\sum_{j\neq i} I[\mathbf{x}_j \in C_n(\mathbf{x}_i)]$. Noting that the conditional median function is estimated in both the first and second stages, we let $\hat{q}_{0i}^{(p)}$ denote the second-stage local polynomial estimator, to distinguish it from the first-stage local constant estimator. Also, we let $\hat{\mu}$ and μ denote $\hat{\mu}(\mathbf{x})$ and $\mu(\mathbf{x})$. For a matrix $\mathbf{A}$, with elements $\{a_{ij}\}$, we let $\|\mathbf{A}\|$ denote $(\sum_{i,j} a_{ij}^2)^{1/2}$.

We note that because we aim to prove that the estimator converges at the optimal nonparametric rate of $O_p(n^{-p/(2p+d_c)})$, we use the term "asymptotically negligible" when referring to remainder terms which are $o_p(n^{-p/(2p+d_c)})$. Our proof relies rely heavily on three previously established properties of the nonparametric conditional quantile estimator used. The first property is a uniform rate of convergence of the local constant estimator used in the first stage. The rate is uniform over regressor values for which the conditional median function is bounded away from the censoring point. We denote this set of regressor values by $\mathcal{X}_\eta \equiv \{\mathbf{x}_i \in \mathcal{X}_\tau : q_{0i} \geq \eta\}$.

Lemma A.1 (From Chaudhuri et al. 1997, lemma 4.3a). Under Assumptions A2, A4–A6, and A7.1,

$$\sup_{\mathbf{x}_i\in\mathcal{X}_\eta} |\hat{q}_{0i} - q_{0i}| = o_p(1). \tag{A.1}$$

The second previously established property is an exponential bound for the local constant and local polynomial estimators for regressor values in a neighborhood of the censoring point.

Lemma A.2 (From lemma 2 in Chen and Khan 2000). Let $\mathcal{X}_{\eta/2}^c$ denote the set $\{\mathbf{x}_i \in \mathcal{X}_\tau, q_{0i} \leq \eta/2\}$, and let A_n denote the event $\{\hat{q}_{0i} \geq \eta \text{ for all } \mathbf{x}_i \in \mathcal{X}_{\eta/2}^c\}$, then under Assumptions A2, A4–A6, and A7.1, there exists constants C_1 and C_2 such that $P(A_n) \leq C_1 e^{-C_2 n h_{1n}^{d_c}}$.

The third property of the conditional quantile estimator is the local Bahadur representation developed by Chaudhuri (1991a) and Chaudhuri et al. (1997).

Lemma A.3 (From lemmas 4.1 and 4.2 in Chaudhuri et al. 1997). Let $q_\alpha^*(\mathbf{x}_i, \mathbf{x})$ denote the kth-order Taylor polynomial approximation of $q_\alpha(\mathbf{x}_i)$ for $\mathbf{x}_i$ close to $\mathbf{x}$. Under Assumptions A2, A4–A6, and A7.2, A7.3, for all $\alpha \geq .5$, $\mathbf{x}: q_{.5}(\mathbf{x}) \geq \eta$, we have the following linear representation for the local polynomial estimator used in the second and third stages:

$$\begin{aligned}\hat{q}_\alpha(\mathbf{x}) - q_\alpha(\mathbf{x}) &= \frac{1}{nh_{(2,3)n}^{d_c} f_{Y,\mathbf{X}}(q_\alpha(\mathbf{x}), \mathbf{x})} \\ &\quad \times \sum_{i=1}^n \big\{\big(I[y_i \leq q_\alpha^*(\mathbf{x}_i, \mathbf{x})] - \alpha\big) I[\mathbf{x}_i \in C_n(\mathbf{x})]\big\} + R_n(\mathbf{x}),\end{aligned} \tag{A.2}$$

where $h_{(2,3)n}$ denotes the bandwidth used either in the second or third stages and the remainder term satisfies.

$$\sup_{\mathbf{x}\in\mathcal{X}_\eta} R_n(\mathbf{x}) = o_p\big(n^{-p/(2p+d_c)}\big). \tag{A.3}$$

The main step in the proof is to show that the difference between the constants c_{α_ℓ}, $\ell = 1, 2, \ldots, \mathcal{N}$, and their estimators $\hat{c}_{\alpha_\ell}$, $\ell = 1, 2, \ldots, \mathcal{N}$, are asymptotically negligible. We let β_i denote $\frac{q_{\ell i} - q_{0i}}{q_{si} - q_{0i}}$ and let $\hat{\beta}_i$ denote its estimated value, obtained by replacing quantile functions with their local polynomial estimators.

We adopt the convention $0/0 = 0$, and define

$$c_{\alpha_\ell}^\dagger = \frac{\sum_{i=1}^n \tau_i \hat{w}_i c_{\alpha_\ell}}{\sum_{i=1}^n \tau_i \hat{w}_i}, \tag{A.4}$$

and we note that it can be easily be shown that

$$P\big(c_{\alpha_\ell}^\dagger \neq c_{\alpha_\ell}\big) \to 0 \tag{A.5}$$

by Assumptions A1 and A3 and Lemma A.1. Thus it will suffice to show that $\hat{c}_{\alpha_\ell} - c_{\alpha_\ell}^\dagger$ is asymptotically negligible. This difference is of the form

$$\hat{c}_{\alpha_\ell} - c_{\alpha_\ell}^\dagger = \frac{\frac{1}{n}\sum_{i=1}^n \tau_i \hat{w}_i (\hat{\beta}_i - c_{\alpha_\ell})}{\frac{1}{n}\sum_{i=1}^n \tau_i \hat{w}_i}. \tag{A.6}$$

The following lemma shows that the denominator of the above expression converges in probability to a positive constant.

Lemma A.4. Under Assumptions A1, A3–A6, and A7.1,

$$\frac{1}{n}\sum_{i=1}^n \tau_i \hat{w}_i \xrightarrow{p} E[\tau_i w_i]. \tag{A.7}$$

Proof. A mean value expansion of $\hat{w}_i$ around w_i yields

$$\frac{1}{n}\sum_{i=1}^n \tau_i \hat{w}_i = \frac{1}{n}\sum_{i=1}^n \tau_i w_i + \frac{1}{n}\sum_{i=1}^n \tau_i w_i'^*(\hat{q}_{0i} - q_{0i}), \tag{A.8}$$

where $\tau_i w_i'^*$ denotes the derivative of the weighting function evaluated at an intermediate value. We can decompose the summation involving this intermediate value as:

$$\frac{1}{n}\sum_{i=1}^n \tau_i w_i'^*(\hat{q}_{0i} - q_{0i}) I[q_{0i} \geq \eta/2] \tag{A.9}$$

$$+ \frac{1}{n}\sum_{i=1}^n \tau_i w_i'^*(\hat{q}_{0i} - q_{0i}) I[q_{0i} < \eta/2]. \tag{A.10}$$

It follows by the bound on the derivative of the weighting function and Lemmas A.1 and A.2 that each of these terms is $o_p(1)$. The LLN implies that $\frac{1}{n}\sum_{i=1}^n \tau_i w_i \xrightarrow{p} E[\tau_i w_i]$.

Thus it will suffice to show the numerator term in (A.6) is $o_p(n^{-p/(2p+d_c)})$. To do so, we take a mean value expansion of $\hat{w}_i$ around w_i, yielding the terms

$$\frac{1}{n}\sum_{i=1}^n \tau_i w_i (\hat{\beta}_i - c_{\alpha_\ell}) + \frac{1}{n}\sum_{i=1}^n \tau_i w_i^{*'}(\hat{q}_{0i} - q_{0i})(\hat{\beta}_i - c_{\alpha_\ell}), \tag{A.11}$$

where $\tau_i w_i^{*'}$ again denotes the derivative of the weighting function evaluated at an intermediate value. The following lemma establishes the asymptotic negligibility of the first part of (A.11).

Lemma A.5. Under Assumptions A3–A6 and A7.2,

$$\frac{1}{n}\sum_{i=1}^n \tau_i w_i \big(\hat{\beta}_i - c_{\alpha_\ell}\big) = o_p\big(n^{-p/(2p+d_c)}\big). \tag{A.12}$$

Proof. Note that $\tau_i w_i c_{\alpha_\ell} = \tau_i w_i \beta_i$. We linearize $\hat{\beta}_i - \beta_i$. Here we let $\Delta q_{\ell i}$ and Δq_{si} denote $q_{\ell i} - q_{0i}$ and $q_{si} - q_{0i}$. $\Delta\hat{q}_{\ell i}$ and $\Delta\hat{q}_{si}$ denote corresponding differences of (local polynomial) estimated quantile functions,

$$\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i(\hat\beta_i-\beta_i)=\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i \Delta q_{si}^{-1}(\Delta\hat q_{\ell i}-\Delta q_{1i}) \quad (A.13)$$

$$-\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i \frac{\Delta q_{\ell i}}{(\Delta q_{si})^2}(\Delta\hat q_i-\Delta q_i) \quad (A.14)$$

$$+R_n,$$

where

$$R_n=O_p\left(\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i(\Delta\hat q_{\ell i}-\Delta q_{\ell i})(\Delta\hat q_{si}-\Delta q_{si})\right). \quad (A.15)$$

It follows by lemma 4.1 of Chaudhuri et al. (1997) (after using the Cauchy–Schwartz inequality) that

$$R_n=O_p\left(\left(\sqrt{\frac{\log n}{nh_{2n}^{d_c}}}+h_{2n}^{p}\right)^2\right), \quad (A.16)$$

and thus is asymptotically negligible by Assumption A7.2. The expressions in (A.13) and (A.14) are sample averages of differences of undersmoothed conditional quantile estimators. We thus only show that

$$\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i(\hat q_{\ell i}-q_{\ell i})=o_p\big(n^{-p/(2p+d_c)}\big), \quad (A.17)$$

because similar arguments may used for the other terms. Equation (A.17) follows from the same arguments used in lemma 2 of Chen and Khan (2000). The only difference is in that lemma, the smoothness and bandwidth conditions implied that the bias term was $o_p(n^{-1/2})$, whereas in this case, using Assumptions A6 and A7.2, the bias term is $o_p(n^{-p/(2p+d_c)})$.

The following lemma shows that the second piece in (A.11) is also asymptotically negligible.

Lemma A.6. Under Assumptions A3–A6, A7.1, and A7.2,

$$\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i'^{*}(\hat q_{0i}-q_{0i})\big(\hat\beta_i-c_{\alpha_\ell}\big)=o_p\big(n^{-p/(2p+d_c)}\big). \quad (A.18)$$

Proof. We multiply the left side of the foregoing expression by $I[q_{0i}\ge\eta/2]+I[q_{0i}<\eta/2]$ to separate the terms where the median function is bounded away from 0 from the terms where it is not. Terms where $q_{0i}<\eta/2$ are asymptotically negligible by Lemma A.2, because $\tau_i w_i'^{*}>0\Rightarrow\hat q_{0i}\ge\eta$. For the terms where $q_{0i}\ge\eta/2$, note that $c_{\alpha_\ell}=\beta_i$, and we can apply the uniform rates of convergence of Chaudhuri (1991a,b) and Chaudhuri et al. (1997) after linearizing the difference $\hat\beta_i-\beta_i$ as before. We note that the uniform rates for the local constant estimator and the local polynomial estimator are different, but it will follow by Assumptions A7.1, A7.2, and A6 that their product will be asymptotically negligible. To make this argument precise, we note from the arguments used in lemma 4.1 of Chaudhuri et al. (1997) that the uniform rate for the local constant and local polynomial estimators are

$$O_p\left(\sqrt{\frac{\log n}{nh_{1n}^{d_c}}}+h_{1n}\right)\quad\text{and}\quad O_p\left(\sqrt{\frac{\log n}{nh_{2n}^{d_c}}}+h_{2n}^{p}\right). \quad (A.19)$$

Letting $\|\cdot\|_\infty$ denote $\max_{1\le i\le n}|\cdot|$, we note that $\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i'^{*}\times I[q_{0i}\ge\eta/2](\hat q_{0i}-q_{0i})(\hat\beta_i-\beta_i)$ is of order

$$\|\hat q_{0i}-q_{0i}\|_\infty\|\hat q_{\ell i}-q_{\ell i}\|_\infty+\|\hat q_{0i}-q_{0i}\|_\infty\|\hat q_{si}-q_{si}\|_\infty + \|\hat q_{0i}-q_{0i}\|_\infty\big\|\hat q_{0i}^{(p)}-q_{0i}\big\|_\infty,$$

which, by the stated uniform rates, is

$$O_p\left(\left(\sqrt{\frac{\log n}{nh_{1n}^{d_c}}}+h_{1n}\right)\left(\sqrt{\frac{\log n}{nh_{2n}^{d_c}}}+h_{2n}^{p}\right)\right)$$

which is $o_p(n^{-p/(2p+d_c)})$ by Assumptions A6, A7.1, and A7.2.

Combining all of our results, we can now replace the estimated constants $\hat c_{\alpha_\ell}$, $\ell=1,2,\dots,\mathcal{N}$, with their true values; furthermore, from Lemma A.2, we can replace $\hat{\tilde d}_{\alpha_\ell}$ with $\tilde d_{\alpha_\ell}$ without affecting the limiting distribution. Thus we have

$$\hat\theta(\mathbf{x})=\left(\sum_{\ell=1}^{\mathcal{N}}\tilde d_{\alpha_\ell}\mathbf{c}_{\alpha_\ell}\mathbf{c}_{\alpha_\ell}'\right)^{-1}\left(\sum_{\ell=1}^{\mathcal{N}}\tilde d_{\alpha_\ell}\mathbf{c}_{\alpha_\ell}\hat q_{\alpha_\ell}(\mathbf{x})\right)+o_p\big(n^{-p/(2p+d_c)}\big) \quad (A.20)$$

Noting that for $\tilde d_{\alpha_\ell}=1$ and $q_{\alpha_\ell}(\mathbf{x})=\mu(\mathbf{x})+c_{\alpha_\ell}\sigma_0(\mathbf{x})$, we have

$$\hat\theta(\mathbf{x})-\theta_0(\mathbf{x}) =\left(\sum_{\ell=1}^{\mathcal{N}}\tilde d_{\alpha_\ell}\mathbf{c}_{\alpha_\ell}\mathbf{c}_{\alpha_\ell}'\right)^{-1}\left(\sum_{\ell=1}^{\mathcal{N}}\tilde d_{\alpha_\ell}\mathbf{c}_{\alpha_\ell}\big(\hat q_{\alpha_\ell}(\mathbf{x})-q_{\alpha_\ell}(\mathbf{x})\big)\right) +o_p\big(n^{-p/(2p+d_c)}\big). \quad (A.21)$$

The limiting distribution of the estimator follows from (A.2) and the Lindeberg theorem.

A.2 Proof of Theorem 3

We note the consistency of $\hat C$ follows from the previously established results that $\hat{\tilde d}_{\alpha_\ell}$ and $\hat{\mathbf{c}}_{\alpha_\ell}$ converge in probability to $\tilde d_{\alpha_\ell}$ and $\mathbf{c}_{\alpha_\ell}$. To show consistency of $\hat V$, given previous results, we need only show that

$$\hat f_{Y|\mathbf{X}}\big(\hat q_{\alpha_\ell}(\mathbf{x})|\mathbf{x}\big)\xrightarrow{p} f_{Y|\mathbf{X}}\big(q_{\alpha_\ell}(\mathbf{x})|\mathbf{x}\big). \quad (A.22)$$

A mean value expansion of $k_2(\frac{y_i-\hat q_{\alpha_\ell}(\mathbf{x})}{\hbar_n})$ around $k_2(\frac{y_i-q_{\alpha_\ell}(\mathbf{x})}{\hbar_n})$ yields a remainder term of

$$\hbar_n^{-1}k_2'\left(\frac{y_i-q_{\alpha_\ell}^{*}(\mathbf{x})}{\hbar_n}\right)\big(\hat q_{\alpha_\ell}(\mathbf{x})-q_{\alpha_\ell}(\mathbf{x})\big),$$

where $q_{\alpha_\ell}^{*}(x)$ denotes an intermediate value and $k_2'(\cdot)$ denotes the derivative of $k_2(\cdot)$. This remainder term is $o_p(1)$ because $k_2'(\cdot)$ is bounded and $\frac{(\hat q_{\alpha_\ell}(x)-q_{\alpha_\ell}(x))}{\hbar_n}$ is $o_p(1)$ by Theorem 2 and the conditions assumed on $\hbar_n$. Thus, defining $\tilde f_{Y|\mathbf{X}}(q_{\alpha_\ell}(\mathbf{x})|\mathbf{x})$ as

$$\frac{\frac{1}{n\hbar_n^{d+1}}\sum_{i=1}^{n}k_1(\frac{\mathbf{x}_i-\mathbf{x}}{\hbar_n})k_2(\frac{y_i-q_{\alpha_\ell}(\mathbf{x})}{\hbar_n})}{\frac{1}{n\hbar_n^{d}}\sum_{i=1}^{n}k_1(\frac{\mathbf{x}_i-\mathbf{x}}{\hbar_n})}, \quad (A.23)$$

(A.22) follows from showing that

$$\tilde f_{Y|\mathbf{X}}\big(q_{\alpha_\ell}(\mathbf{x})|\mathbf{x}\big)\xrightarrow{p} f_{Y|\mathbf{X}}\big(q_{\alpha_\ell}(\mathbf{x})|\mathbf{x}\big). \quad (A.24)$$

The foregoing result follows from standard results on nonparametric density estimation (see, e.g. Härdle and Linton 1994; and Newey and McFadden 1994).

[Received August 2002. Revised April 2004.]

REFERENCES

Beran, R. (1981), "Nonparametric Regression With Randomly Censored Survival Data," technical report, University of California, Berkeley.

Bierens, H. J. (1987), "Kernel Estimators of Regression Functions," in *Advances in Econometrics, Fifth World Congress*, Vol. 1., ed. T. F. Bewley, Cambridge, U.K.: Cambridge University Press, pp. 99–144.

Bierens, H. J., and Ploberger, W. (1997), "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

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

Card, D., and Levine, P. B. (2000), "Extended Benefits and the Duration of UI Spells: Evidence From the New Jersey Extended Benefit Program," *Journal of Public Economics*, 78, 107–138.

Chamberlain, G. (1994), "Quantile Regression, Censoring, and the Structure of Wages," in *Advances in Econometrics*, Vol. I, ed. C. A. Sims, Cambridge, U.K.: University Press, pp. 171–210.

Chaudhuri, P. (1991a), "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," *The Annals of Statistics*, 19, 760–777.

——— (1991b), "Global Nonparametric Estimation of Conditional Quantiles and Their Derivatives," *Journal of Multivariate Analysis*, 39, 246–269.

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

Chen, S., and Khan, S. (2000), "Estimation of Censored Regression Models in the Presence of Nonparametric Multiplicative Heteroskedasticity," *Journal of Econometrics*, 98, 283–316.

——— (2001), "Semiparametric Estimation of a Partially Linear Censored Regression Model," *Econometric Theory*, 17, 567–590.

Dabrowska, D. M. (1987), "Non-Parametric Regression With Censored Survival Time Data," *Scandinavian Journal of Statistics*, 14, 181–197.

——— (1992), "Nonparametric Quantile Regression With Censored Data," *Sankhyā*, Ser. A, 54, 252–259.

Fan, J. (1992), "Design-Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

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

——— (1996), *Local Polynomial Modelling and Its Applications*, New York: Chapman & Hall.

Ham, J. C., and Rea, S. A., Jr. (1987), "Unemployment Insurance and Male Unemployment Duration in Canada," *Journal of Labor Economics*, 5, 325–353.

Härdle, W., and Linton, O. (1994), "Applied Nonparametric Methods," in *Handbook of Econometrics*, Vol. 4, eds. R. F. Engle and D. McFadden, Amsterdam: North-Holland, pp. 2297–2339.

Honoré, B. E., Khan, S., and Powell, J. L. (2002), "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 109, 67–105.

Horowitz, J. L., and Spokoiny, V. G. (2001), "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599–631.

Katz, L. F., and Meyer, B. D. (1990), "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment," *Journal of Public Economics*, 105, 973–1002.

Khan, S. (2001), "Two-Stage Rank Estimation of Quantile Index Models," *Journal of Econometrics*, 100, 319–355.

Khan, S., and Powell, J. L. (2001), "Two-Step Quantile Estimation of Semiparametric Censored Regression Models," *Journal of Econometrics*, 103, 73–110.

Koenker, R., and Bassett, G. S., Jr. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koul, H., Suslara, V., and Van Ryzin, J. (1981), "Regression Analysis With Randomly Right Censored Data," *The Annals of Statistics*, 9, 1276–1288.

Lewbel, A., and Linton, O. B. (2002), "Nonparametric Censored Regression," *Econometrica*, 70, 765–779.

Meyer, B. D. (1990), "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58, 757–782.

Moffitt, R. (1985), "Unemployment Insurance and the Distribution of Unemployment Spells," *Journal of Econometrics*, 28, 85–101.

Moffitt, R., and Nicholson, W. (1982), "The Effect of Unemployment Insurance on Unemployment: The Case of Federal Supplemental Compensation," *Review of Economics and Statistics*, 64, 1–11.

Newey, W. K., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. 4, eds. R. F. Engle and D. L. McFadden, Amsterdam: Elsevier, pp. 2111–2245.

Newey, W. K., and Powell, J. L. (1990), "Efficient Estimation of Linear and Type 1 Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317.

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

——— (1994), "Estimation of Semiparametric Models," in *Handbook of Econometrics*, Vol. 4, eds. R. F. Engle and D. McFadden, Amsterdam: North-Holland, pp. 2444–2521.

Ruppert, D., and Wand, M. P. (1994), "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346–1370.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

Van Keilegom, I., and Akritas, M. G. (1999), "Transfer of Tail Information in Censored Regression Models," *The Annals of Statistics*, 27, 1745–1784.

Van Keilegom, I., Akritis, M. G., and Veraverbeke, N. (2001), "Estimation of the Conditional Distribution in Regression With Censored Data: A Comparitive Study," *Computational Statistics & Data Analysis*, 35, 487–500.

Woodbury, S. A., and Murray, A. R. (1997), "The Duration of Benefits," in *Unemployment Insurance in the United States: Analysis of Policy Issues*, eds. C. I. O'Leary and S. A. Wandner, Kalamazoo, MI: W. E. Upjohn Institute for Employment Research, pp. 211–283.

Ying, Z., Jung, S. H., and Wei, L. J. (1995), "Survival Analysis With Median Regression Models," *Journal of the American Statistical Association*, 90, 178–184.