CENTER FOR LABOR ECONOMICS
UNIVERSITY OF CALIFORNIA, BERKELEY
WORKING PAPER NO. 54

Nonparametric Identification and Estimation of a Censored Regression Model with an Application to Unemployment Insurance Receipt*

Songnian Chen
Hong Kong University of Science and Technology

Gordon B. Dahl
University of Rochester and UC Berkeley

Shakeeb Khan
University of Rochester

September 2002

Abstract

In this paper 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, then the location function is not identified anywhere. In contrast, if the location function is greater or equal to the censoring point with positive probability, then 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 three distinct 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 application on unemployment insurance duration using administrative level data from New Jersey. The survival curve for benefit receipt based on our new estimator closely matches the Kaplan-Meier estimate in the non-censored region and is relatively flat past the censoring point. We find that incorrect distributional assumptions can significantly bias the results for estimates past the censoring point.

*We are grateful to 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. An early version of this paper 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. We also thank D. Card and P. Levine for generously providing us with the dataset.

1 Introduction

The nonparametric location-scale model is usually of the form:

$$y_i = \mu(x_i) + \sigma(x_i)\epsilon_i \tag{1.1}$$

where x_i is an observed $d-$dimensional random vector and ϵ_i is an unobserved random variable, distributed independently of x_i, and assumed to be centered around zero in some sense. The functions $\mu(\cdot)$ and $\sigma(\cdot)$ are unknown. This location-scale model has received a great deal of attention in the statistics and econometrics literature (see for example Fan and Gijbels(1996), Chapter 3, and 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 $(y_i, x_i')'$.

In this paper, we consider extending the nonparametric location-scale model to accommodate censored data. Semiparametric (fixed) censored regression models, where $\mu(x_i)$ is known up to a finite-dimensional parameter, has 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, either because of non-negativity 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:

$$y_i^* = \mu(x_i) + \sigma(x_i)\epsilon_i \tag{1.2}$$

$$y_i = \max(y_i^*, 0) \tag{1.3}$$

where y_i^* is a latent dependent variable, which is only observed if it exceeds the fixed censoring point, which we assume without loss of generality is 0.

We consider identification and estimation of $\mu(x_i)$ after imposing the location restriction that the median of $\epsilon_i = 0$. We emphasize that our results allow for identification of $\mu(x_i)$ on the entire support of x_i. This is in contrast to identifying and estimating $\mu(x_i)$ only in the region where it exceeds the censoring point, which could be easily done by extending Powell's(1984) CLAD estimator 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 data set is heavily censored. In this case, $\mu(x_i)$ will be less than the censoring point for a large portion of the support of x_i, requiring estimation at these points necessary 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 which holds for observations where $\mu(x_i) \geq 0$. This relationship can be used to motivate an estimator for $\mu(x_i)$ in the region where it is negative. Our results are thus based on the condition

$$P_X(x_i : \mu(x_i) \geq 0) > 0 \tag{1.4}$$

where $P_X(\cdot)$ denotes the probability measure of the random variable x_i.

Variations of censored nonparametric models have been studied elsewhere in the literature. Lewbel and Linton(2002) estimate a nonparametric censored regression model with a fixed censoring point that is based on a mean restriction on the disturbance term. As conditional mean restrictions are generally not sufficient for identification in censored regression models (see Powell(1994)), their approach either only attains identification up to an additive term, or requires much stronger conditions on the tail behavior of the random variables ϵ_i and x_i than assumed here. Van Keilegom and Akritas(1999) estimate a nonparametric regression model with random censoring under a mean restriction, and face similar difficulties when the censoring variable is a fixed point.

The paper 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 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 considers an extension of the estimation procedure to estimate the distribution of the disturbance term. Section 5 explores the finite sample properties of the estimator through the results of a simulation study. Section 6 presents an empirical application to unemployment insurance

(UI) duration, in which we estimate the survivor function in the region beyond the censoring point. Section 7 concludes by summarizing results and discussing extensions for future research. An appendix contains 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 x_i, and has a density function with respect to Lebesgue measure, that is positive and bounded on $\mathbf{R}$.

I2 ϵ_i has median 0.

I3 The scale function $\sigma(\cdot)$ is continuous, strictly positive and bounded on $\mathcal{X}$.

I4 The location function $\mu(\cdot)$ is continuous on $\mathcal{X}$.

Remark 2.1 *The median restriction in Assumption I2 is different from the usual 0 mean assumption imposed in the location scale model. Censoring introduces a non-linearity which makes identification of $\mu(\cdot)$ impossible without further assumptions on ϵ_i. Conditional mean restrictions were imposed in Lewbel and Linton(2002) and Van Keilegom and Akritas(1999). The latter only estimated the location function up to an additive constant, which prevents using the estimator for prediction/forecasting. The former required strong tail behavior restrictions and support conditions on one of the components of x_i. Such restrictions rule out the classical tobit model with bounded regressors, as well as censored models with only discrete covariates.*

The first result is that the location function is not identified anywhere on $\mathcal{X}$ if $\mu(\cdot) < 0$ everywhere on $\mathcal{X}$. Its proof is left to the appendix.

Theorem 2.1 *(Necessity) Suppose Assumptions I1-I4 hold, and that* $\max_{x\in\mathcal{X}} \mu(x) < 0$. *Then there exists a function $\tilde{\mu}(\cdot) \neq \mu(\cdot)$ and a random variable $\tilde{\epsilon}_i$, where Assumptions I1-I4 still hold with $\tilde{\mu}(\cdot), \tilde{\epsilon}_i$ replacing $\mu(\cdot), \epsilon_i$ respectively, such that if we define:*

$$\tilde{y}_i = \max(\tilde{\mu}(x_i) + \sigma(x_i)\tilde{\epsilon}_i, 0)$$

then

$$\mathcal{L}(y_i|x_i) = \mathcal{L}(\tilde{y}_i|x_i) \quad \forall x_i \in \mathcal{X}$$

where $\mathcal{L}(y_i|x_i)$ denotes the conditional distribution of y_i given x_i.

Remark 2.2 *This result is in contrast to the result in Chen and Khan(2000), who studied a semiparametric model with $\mu(x_i) = x_i'\beta_0$, and considered identification of the finite dimensional parameter β_0. They show identification of β_0 was possible without the conditional median ever exceeding the censoring point. The above theorem shows their result cannot be extended to the nonparametric setting.*

Our next result establishes the sufficiency of (1.4) 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 2.2 *(Sufficiency) Suppose Assumptions I1-I4 hold, and condition (1.4) holds. Then $\mu(\cdot)$ is identified for all $x \in \mathcal{X}$.*

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 x_i can be used to identify μ where it is negative. To show identification in the nonnegative region, we let x_0 be any point which satisfies $\mu(x_0) \geq 0$. Suppose first that $\mu(x_0) = 0$. We will show that $\tilde{\mu}(x_0) < 0$ or $\tilde{\mu}(x_0) > 0$ leads to a contradiction. If $\tilde{\mu}(x_0) = -\delta < 0$, let $\tilde{\sigma}(x_0)$ be a positive, finite number. 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, by Assumption I2 it must follow that $\tilde{c}_{0.5} = 0$, and $0 < \tilde{c}_\alpha < \delta/\tilde{\sigma}(x_0)$ for $\alpha \in (0.5, 0.5 + \varepsilon)$ where recall $\delta = -\tilde{\mu}(x_0)$ and ε is an arbitrarily small positive constant. Noting that $c_\alpha > 0$ for $\alpha \in (0.5, 0.5 + \varepsilon)$, we have for $\alpha \in (0.5, 0.5 + \varepsilon)$, $q_\alpha(x_0) = \max(\mu(x_0) + c_\alpha\sigma(x_0), 0) = \max(c_\alpha\sigma(x_0), 0) > 0$. Alternatively we have:

$$\begin{aligned} \tilde{q}_\alpha(x_0) &= \max(\tilde{\mu}(x_0) + \tilde{\sigma}(x_0)\tilde{c}_\alpha, 0) && (2.1) \\ &< \max(-\delta + \delta, 0) = 0 && (2.2) \end{aligned}$$

Thus we have found quantiles where $q_\alpha(x_0) \neq \tilde{q}_\alpha(x_0)$, which shows that $\mu(x_0) = 0$ 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 x where $\mu(x) > 0$ are identified. If $\mu(x_0) > 0$, and $\tilde{\mu}(x_0) \neq \mu(x_0)$, then $q_{0.5}(x_0) = \mu(x_0)$ and $\tilde{q}_{0.5}(x_0) = \max(\tilde{\mu}(x_0), 0) \neq \mu(x_0)$.

We next show how to identify $\mu(x)$ when $\mu(x) < 0$ given that we have identified $\mu(x_0)$ for $\mu(x_0) \geq 0$. We first note that since $\mu(x)$ and $\sigma(x)$ are finite by Assumptions I4 and I3 respectively, there exists quantiles $\alpha_1 < \alpha_2 < 1$ such that:

$$q_{\alpha_1}(x) = \mu(x) + c_{\alpha_1}\sigma(x) > 0 \tag{2.3}$$

$$q_{\alpha_2}(x) = \mu(x) + c_{\alpha_2}\sigma(x) > 0 \tag{2.4}$$

Thus we have the relationships:

$$\Delta q(x) = \Delta c\sigma(x) \tag{2.5}$$

$$\bar{q}(x) = \mu(x) + \bar{c}\sigma(x) \tag{2.6}$$

where $\Delta q(x) = q_{\alpha_2}(x) - q_{\alpha_1}(x)$, $\bar{q}(x) = (q_{\alpha_2}(x) + q_{\alpha_1}(x))/2$, $\Delta c = c_{\alpha_2} - c_{\alpha_1}$, $\bar{c} = (c_{\alpha_2} + c_{\alpha_1})/2$. Combining the two previous relationships, if we could identify the fraction $\frac{\bar{c}}{\Delta c}$, then we could identify $\mu(x)$ as:

$$\mu(x) = \bar{q}(x) - \frac{\bar{c}}{\Delta c}\Delta q(x) \tag{2.7}$$

We use identification of $\mu(x_0) \geq 0$ to identify $\frac{\bar{c}}{\Delta c}$ in the following manner. We combine the following values of the conditional quantile function evaluated at the three distinct quantiles $0.5, \alpha_1, \alpha_2$.

$$q_{0.5}(x_0) = \mu(x_0) \tag{2.8}$$

$$q_{\alpha_1}(x_0) = \mu(x_0) + c_{\alpha_1}\sigma(x_0) \tag{2.9}$$

$$q_{\alpha_2}(x_0) = \mu(x_0) + c_{\alpha_2}\sigma(x_0) \tag{2.10}$$

This enables us to identify $\frac{\bar{c}}{\Delta c}$ as

$$\frac{\bar{c}}{\Delta c} = \frac{\bar{q}(x_0) - q_{0.5}(x_0)}{\Delta q(x_0)} \tag{2.11}$$

which immediately translates into identification of $\mu(x)$ from the relationship:

$$\mu(x) = \bar{q}(x) - \frac{\bar{q}(x_0) - q_{0.5}(x_0)}{\Delta q(x_0)} \Delta q(x) \tag{2.12}$$

This completes the proof of the theorem. ■

Remark 2.3 *Identification on all points first involves identification of a point where $\mu(x) \geq 0$. As is apparent from the proof, identification is much simpler for points where $\mu(x) > 0$, and we note that the argument for identification of a point where $\mu(x) = 0$ would be difficult to translate into an estimator. In the next section, where we propose an estimator for $\mu(\cdot)$ based on our identification results, we therefore assume $P_X(x_i : \mu(x_i) > 0) > 0$.*

Remark 2.4 *Identification of $\mu(\cdot)$ where it is negative involves identification of the quantiles of the homoskedastic component of the disturbance term. Thus an additional consequence of condition (1.4) being satisfied is that the quantiles of ϵ_i are identified for all $\alpha \geq \alpha_0 \equiv \inf\{\alpha : \sup_{x \in \mathcal{X}} q_\alpha(x) > 0\}$. This result can be used to estimate and construct hypothesis tests regarding the distribution of ϵ_i, as is considered in Section 5. We also note that if the econometrician were to impose a distributional form on ϵ_i, the (known) values of $c_{\alpha_1}, c_{\alpha_2}$ could be used in (2.7) to identify and estimate the location function, without requiring condition (1.4).*

3 Estimation Procedure and Asymptotic Properties

3.1 Estimation Procedure

In this section we consider estimation of the function $\mu(\cdot)$. Our procedure will be based on our identification results in the previous section, and 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 in Chaudhuri(1991a,b). For expositional ease, we only describe this nonparametric estimator for a polynomial of degree 0, and refer readers to Chaudhuri(1991a,b), Chaudhuri et al.(1997), Chen and Khan(2000,2001), and Khan(2001) for the additional notation involved for polynomials of arbitrary degree.

First, we assume the regressor vector x_i can be partitioned as $(x_i^{(ds)}, x_i^{(c)})$, where the $d_{ds}-$dimensional vector $x_i^{(ds)}$ is discretely distributed, and the d_c-dimensional vector $x_i^{(c)}$ is continuously distributed.

We let $C_n(x_i)$ denote the cell of observation x_i and let h_n denote the sequence of bandwidths which govern the size of the cell. For some observation $x_j,\ j \neq i$, we let $x_j \in C_n(x_i)$ denote that $x_j^{(ds)} = x_i^{(ds)}$ and $x_j^{(c)}$ lies in the d_c-dimensional cube centered at $x_i^{(c)}$ with side length $2h_n$.

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

$$\sum_{j=1}^{n} I[x_j \in C_n(x_i)]\rho_\alpha\left(y_j - \theta\right) \tag{3.1}$$

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

Our estimation procedure will be based on a random sample of n observations of the vector $(y_i, x_i')'$ and involves applying the local polynomial estimator at three stages. Throughout our description, $\hat{\cdot}$ will denote estimated values.

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 will 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 $\hat{q}_{0.5}(x_i)$. Recalling that our identification result is based on observations for which the median function is positive, we assigns weights to these estimated values using a weighting function, denoted by $w(\cdot)$. Essentially, $w(\cdot)$ assigns 0 weight to observations in the sample for which the estimated value of the median function is 0, and assigns positive weight for estimated values which are positive.

2. **Weighted Average Estimation of the Disturbance Quantiles** In the second stage, the unknown quantiles $c_{\alpha_1}, c_{\alpha_2}$ are estimated (up to the scalar constant Δc) by a weighted average of local polynomial estimators of the quantile functions for the higher quantiles α_1, α_2. The estimator of these constants is based on (2.11). In this stage, we use a polynomial of degree k, and denote the second stage bandwidth sequence by h_{2n}.

We let $\hat{c}_1, \hat{c}_2$ denote the estimators of the unknown constants $\frac{c_{\alpha_1}}{\Delta c}, \frac{c_{\alpha_2}}{\Delta c}$, and define them as:

$$\hat{c}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} \tau(x_i) w(\hat{q}_{0.5}(x_i)) \cdot \frac{(\hat{q}_{\alpha_1}(x_i) - \hat{q}_{0.5}^{(p)}(x_i))}{(\hat{q}_{\alpha_2}(x_i) - \hat{q}_{\alpha_1}(x_i))}}{\frac{1}{n}\sum_{i=1}^{n} \tau(x_i) w(\hat{q}_{0.5}(x_i))} \tag{3.2}$$

$$\hat{c}_2 = \frac{\frac{1}{n}\sum_{i=1}^{n} \tau(x_i) w(\hat{q}_{0.5}(x_i)) \cdot \frac{(\hat{q}_{\alpha_2}(x_i) - \hat{q}_{0.5}^{(p)}(x_i))}{(\hat{q}_{\alpha_2}(x_i) - \hat{q}_{\alpha_1}(x_i))}}{\frac{1}{n}\sum_{i=1}^{n} \tau(x_i) w(\hat{q}_{0.5}(x_i))} \tag{3.3}$$

where $\tau(x_i)$ is a trimming function, whose support, denoted by $\mathcal{X}_\tau$, is a compact set which lies strictly in the interior of $\mathcal{X}$. The trimming function serves to eliminate "boundary effects" that arise in nonparametric estimation. We use the superscript (p) to distinguish the estimator of the median function in this stage from that in the first stage.

3. **Local Polynomial Estimation at the Point of Interest** The third stage is based on (2.12). Letting x denote the point at which the function $\mu(\cdot)$ is to be estimated at, we combine the local polynomial estimator, with polynomial order k and bandwidth sequence h_{3n}, of the conditional quantile function at x using quantiles α_1, α_2, with the estimator of the unknown disturbance quantiles, to yield the estimator of $\mu(x)$:

$$\hat{\mu}(x) = \hat{c}_2 \hat{q}_{\alpha_1}(x) - \hat{c}_1 \hat{q}_{\alpha_2}(x) \tag{3.4}$$

Remark 3.1 *We note here that a different order polynomial is used in first stage than in the other two stages. The reason for this is that even though the functions $\mu(\cdot), \sigma(\cdot)$ are assumed to be $k-$times differentiable, the quantile functions will not in general be smooth at the censoring point. Thus a local polynomial estimator may not be consistent when the quantile function is in a neighborhood of the censoring point. However, once points in the sample which 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 3.2 *The three stage estimation procedure described is based on the identification results in the previous section. It is not efficient as it is only based on the information in the two quantiles α_1, α_2. However, efficiency can be gained by combining various quantiles in a GMM framework, as was done in the linear quantile regression model in Buchinsky(1995). Also, implementation requires a rule for selecting the quantile pair. We discuss this matter in the simulation study later in the paper.*

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 ID (Identification) The weighting function is positive with positive probability:

$$P_X\left(\tau(x_i)w(q_{0.5}(x_i)) > 0\right) > 0$$

and the α_1 quantile at the point of interest is positive:

$$q_{\alpha_1}(x) > 0$$

Assumption RS (Random sampling) The sequence of $d+1$ dimensional vectors (y_i, x_i) are independent and identically distributed.

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

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

WF.2 $w \equiv 0$ if its argument is less than η, an arbitrarily small positive constant.

Assumption RD (Regressor Distribution) We let $f_{X^{(c)}|X^{(ds)}}(\cdot|x^{(ds)})$ denote the conditional density function of $x_i^{(c)}$ given $x_i^{(ds)} = x^{(ds)}$, and assume it is bounded away from 0 and infinity on $\mathcal{X}_\tau$.

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

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

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

Assumption OS (Orders of Smoothness). For some $\varrho \in (0, 1]$, and any real valued function $\mathcal{F}$ of x_i, we adopt the notation $\mathcal{F} \in \mathcal{C}^\varrho(\mathcal{X}_\tau)$ to mean there exists a positive constant $K < \infty$ such that:

$$|\mathcal{F}(x_1) - \mathcal{F}(x_2)| \leq K\|x_1 - x_2\|^\varrho$$

for all $x_1, x_2 \in \mathcal{X}_\tau$. With this notation, we assume the following smoothness conditions

OS.1 $f_X(\cdot), \quad \tau(\cdot) \in \mathcal{C}^{\varrho}(\mathcal{X}_\tau)$

OS.2 $\mu(\cdot)$ and $\sigma(\cdot)$ are differentiable in $x_i^{(c)}$ of order k, with k^{th} order derivatives $\in \mathcal{C}^{\varrho}(\mathcal{X}_\tau)$. We let $p = k + \varrho$ denote the order of smoothness of this function.

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

BC.1 h_{1n} satisfies $\frac{\log n}{nh_{1n}^{d_c}} \to 0$, $n^{\frac{p}{2p+d_c}} h_{1n}^2 \to 0$.

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

BC.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 3.3 *The weighting function $w(\cdot)$ In Assumption WF 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 only assign positive weight to estimated conditional median values which are bounded away from 0.*

Remark 3.4 *The bandwidth sequences h_{1n},h_{2n}, h_{3n} in Assumption BC are required to satisfy different conditions. The conditions on h_{1n} and h_{2n} in Assumptions BC.1, BC.2, reflect "undersmoothing", implying that the bias of the nonparametric estimators used in the first two stages converges to 0 at a faster rate than the standard deviation. In contrast, Assumption BC.3 imposes the optimal rate for h_{3n}, so that the estimator of $\mu(\cdot)$ will converge at the optimal nonparametric rate.*

We now characterize the limiting distribution for the proposed estimator of $\mu(x)$, where x is assumed to lie in the interior of the support of x_i. The following theorem establishes that the proposed estimator converges at the optimal nonparametric rate, and has a limiting non-centered normal distribution. The proof is left to the appendix.

Theorem 3.1 *If Assumptions ID,RS,WF,RD,ED,OS,BC hold, then*

$$n^{\frac{p}{2p+d_c}}(\hat{\mu}(x) - \mu(x)) \Rightarrow N(B, V) \tag{3.5}$$

where

$$V = \frac{c_{\alpha_2}^2}{(\Delta c)^2 f_{Y|X}(q_{\alpha_1}(x)|x)^2}\alpha_1(1-\alpha_1) + \frac{c_{\alpha_1}^2}{(\Delta c)^2 f_{Y|X}(q_{\alpha_2}(x)|x)^2}\alpha_2(1-\alpha_2) - \frac{2c_{\alpha_2}c_{\alpha_1}}{(\Delta c)^2 f_{Y|X}(q_{\alpha_1}(x)|x) f_{Y|X}(q_{\alpha_2}(x)|x)}\alpha_1(1-\alpha_2) \tag{3.6}$$

with $f_{Y|X}(\cdot)$ denoting the conditional density function of y_i. The form of the limiting bias requires introducing new notation. For any quantile α, we let

$$q^*_{\alpha n}\left(x^{(c)} + \mathbf{t}h_{3n}, x^{(c)}, x^{(ds)}\right)$$

denote the k^{th} order Taylor polynomial approximation of

$$q_\alpha\left(x^{(c)} + \mathbf{t}h_{3n}, x^{(ds)}\right)$$

where here $\mathbf{t}$ is a d_c-dimensional vector of constants, and h_{3n} is as defined in Assumption BC.3. We define

$$B_\alpha = \lim_{n\to\infty}\sqrt{nh_{3n}^{d_c}}\cdot\int_{\left[\frac{1}{2},\frac{1}{2}\right]^{d_c}}\left(q_\alpha\left(x^{(c)} + \mathbf{t}h_{3n}, x^{(ds)}\right) - q^*_{\alpha n}\left(x^{(c)} + \mathbf{t}h_{3n}, x^{(c)}, x^{(ds)}\right)\right)d\mathbf{t}$$

The limting bias of the proposed estimator is of the form

$$B = \frac{c_{\alpha_2}}{\Delta c}B_{\alpha_1} - \frac{c_{\alpha_1}}{\Delta c}B_{\alpha_2} \tag{3.7}$$

4 Estimating the Distribution of ϵ_i

As mentioned in Section 2, the distribution of the random variable ϵ_i is identified for all quantiles exceeding $\alpha_0 \equiv \inf\{\alpha : \sup_{x\in\mathcal{X}} q_\alpha(x) > 0\}$. In this section we consider estimation of these quantiles, and the asymptotic properties of the estimator. Estimating the distribution of ϵ_i is of interest for two reasons. First, the econometrician may be interested in estimating the entire model, which would require estimators of $\sigma(x_i)$ and the distribution of ϵ_i as well as of $\mu(x_i)$. Second, the estimator can be used to construct tests of various parametric forms of the distribution of ϵ_i, and the results of these tests could then be used to adopt a (local) likelihood approach to estimating the function $\mu(x_i)$.

Before proceeding, we note that the distribution of ϵ_i is only identified up to scale, and we impose the scale normalization that $c_{0.75} - c_{0.25} \equiv 1$. We also assume without loss of generality that $\alpha_0 \leq 0.25$. To estimate c_α for any $\alpha \geq \alpha_0$, we let $\alpha_- = \min(\alpha, 0.5)$ and define our estimator as

$$\hat{c}_\alpha = \frac{\frac{1}{n}\sum_{i=1}^n \tau(x_i) w(\hat{q}_{\alpha_-}(x_i)) \cdot (\hat{q}_\alpha(x_i) - \hat{q}_{0.5}^{(p)}(x_i))}{\frac{1}{n}\sum_{i=1}^n \tau(x_i) w(\hat{q}_{\alpha_-}(x_i)) \cdot (\hat{q}_{0.75}(x_i) - \hat{q}_{0.25}(x_i))} \tag{4.1}$$

The proposed estimator, which involves averaging nonparametric estimators, will converge at the parametric ($\sqrt{n}$) rate and have a limiting normal distribution, as can be rigorously shown using similar arguments found in Chen and Khan(1999b).

5 Monte Carlo Results

In this section the finite sample properties of the proposed estimator are explored by way of a small scale simulation study. We simulated from designs of the form:

$$y_i = \max(\mu(x_i) + \sigma(x_i)\epsilon_i, 0)$$

where x_i was a random variable distributed uniformly between -1 and 1, ϵ_i was distributed standard normal, and the scale function $\sigma(x_i)$ was set to $e^{0.15x_i}$. We considered four different functional forms for $\mu(x_i)$ in our study:

1. $\mu(x) = x$
2. $\mu(x) = x^2 - \mathcal{C}_1$
3. $\mu(x) = 0.5 \cdot x^3$
4. $\mu(x) = e^x - \mathcal{C}_2$

where the constants $\mathcal{C}_1, \mathcal{C}_2$ were chosen so that the censoring level was 50%, as it was for the other two designs.

We adopted the following data-driven method to select the quantile pair. For a given point x, we note that the estimator requires that $q_{\alpha_1}(x), q_{\alpha_2}(x)$ both be strictly positive for identification, requiring that the quantiles be sufficiently close to 1. On the other hand, efficiency concerns would suggest that the quantiles not be at the extreme, as the quantile regression estimator becomes imprecise. We thus let the probability of being censored, or the "propensity score" (see Rosenbaum and Rudin(1983)) govern the choice of quantiles for estimating the function $\mu(\cdot)$ at the point x. Letting d_i denote an indicator function which takes the value 1 if an observation is uncensored, we note that

$$1 - E[d_i|x_i = x] = F_\epsilon\left(\frac{-\mu(x)}{\sigma(x)}\right)$$

where $F_\epsilon(\cdot)$ denotes the c.d.f. of ϵ_i. Letting $\alpha^* = F_\epsilon\left(\frac{-\mu(x)}{\sigma(x)}\right)$, we note that

$$\begin{aligned} q_{\alpha^*}(x) &= \max(\mu(x) + c_{\alpha^*}\sigma(x), 0) \\ &= \max(\mu(x) + \frac{-\mu(x)}{\sigma(x)}\sigma(x), 0) \\ &= 0 \end{aligned}$$

Thus if one knew the propensity score value, identification would require that α^* be a lower bound for the choice of quantile pair. The propensity score can be easily estimated using kernel methods, suggesting an estimator of α^*:

$$\hat{\alpha}^* = 1 - \frac{\frac{1}{n}\sum_{i=1}^n d_i K_\hbar(x_i^{(c)} - x^{(c)}) I[x_i^{(d)} = x^{(d)}]}{\frac{1}{n}\sum_{i=1}^n K_\hbar(x_i^{(c)} - x^{(c)}) I[x_i^{(d)} = x^{(d)}]}$$

where $K_\hbar(\cdot) = \hbar^{-d_c} K(\frac{\cdot}{\hbar})$ where $\hbar$ is a bandwidth sequence, and $K(\cdot)$ is a kernel function.

Our proposed choice of quantile pair takes into account this lower bound as well as the efficiency loss of estimating quantiles at the extreme. We set:

$$\alpha_1 = \frac{2\hat{\alpha}^* + 1}{3} \qquad \alpha_2 = \frac{2 + \hat{\alpha}^*}{3}$$

which divides the interval $[\hat{\alpha}^*, 1]$ into three equal spaces. In implementing this procedure in the Monte Carlo study, the propensity scores were estimated using a normal kernel function and a bandwidth of $n^{-1/5}$.

For the quantile estimators, a local constant was fit in the first stage, using a bandwidth of $n^{-1/5}$, and a local linear estimator was used in the second and third stages, using a bandwidth of the form $\kappa n^{-1/5}$. The constant κ was selected using the "rule of thumb" approach detailed on page 202 in Fan and Gijbels(1996).

The results 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 figures plot the average value of the estimated function, denoted by $m(x)$, alongside the true function. Also reported (in parentheses) are the average mean squared errors (AMSE) for the estimator.

As indicated by the figures, the results are pretty much as expected. For $n = 100$, 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. The estimator performs much better for $n = 400$, where it is close to true function value on its entire support. For both sample sizes, the estimator performs better in terms of AMSE for $\mu(x) = x^2 - \mathcal{C}_1$ and $\mu(x) = e^x - \mathcal{C}_2$. This is because the location function is negative with smaller probability than for the other two designs. While our results are very encouraging in general, we would expect a worse finite sample performance when more regressors are present, as the rate of convergence would be slower.

6 Application to Unemployment Insurance

As discussed in the introduction, the estimation approach developed in this paper applies to a variety of economic problems with censoring. Since the conditional distribution of the latent dependent variable can be estimated using our approach, it is particularly well-suited for estimating failure time models.

One survival function of particular interest to economists is the duration of unemployment insurance (UI) spells. In the United States, many claimants exhaust their UI benefits, resulting in a significant fraction of censored observations. We apply the estimator developed in this paper to a large dataset of individual-level administrative records for New Jersey's UI program in the late 1990s. The maximum number of weeks an individual can collect UI benefits varies from state to state and is based on an individual's work history, but the limit is typically 26 weeks. For the New Jersey data analyzed in this paper, 43 percent of claimants exhaust their benefits at 26 weeks. One important question is how long these censored claims would remain active if UI benefits were made available for a longer time period. The answer would be particularly useful in evaluating the costs and benefits of extending the maximum duration of UI benefits.

6.1 Modeling Unemployment Insurance Receipt

Previous research on the determinants of UI spell length has focused on the effect of benefit levels (Anderson and Meyer(1997), Ham and Rea(1987), Hunt(1995), McCall(1995), and Meyer(1992)) and the maximum allowed benefit duration (Card and Levine(2000), Katz and Meyer(1990), Meyer(1990), Moffit and Nicholson(1982), Moffit(1985), Woodbury and Murray(1997)). This literature generally concludes that both higher benefit levels and higher maximum durations significantly increase the length of an individual's spell of unemployment insurance receipt. The empirical findings mainly rely on variation in benefit amounts and maximum benefit lengths over time, across states, or between individuals for identification.

While such studies have helped researchers partly understand the determinants of UI spell length, one drawback is the endogeneity of the identifying variation. (Two exceptions are Card and Levine (2000) and Meyer (1992) which use quasi-experimental policy changes.) Time variation is not likely to be exogeneous since benefits are usually extended when labor market conditions are poor (See Blaustein, et al(1993) and Blank and Card (1991)). Individual states also sometimes extend the maximum duration of benefits in response to a slack labor market. These changes are endogeneous since they occur precisely when spell lengths and benefit exhaustions would otherwise be predicted to increase. Some individuals are eligible for reduced benefits or less than 26 weeks due to a limited work history. These claimants may respond very differently to legislated changes in the UI system compared to those eligible for the full 26 weeks.

Prior research has concentrated more heavily on the responses of individuals with non-censored spells. Any predictions about spell durations past the censoring point are made using parametric assumptions about the distribution of failure times or assumptions about the functional form for how covariates affect failure times past the censoring point. In contrast, our approach allows estimation of the survival function beyond the censoring point with a minimum of parametric and structural assumptions.

We model log-failure time (i.e., when an individual stops collecting UI) in the framework described in Section 1. The model of interest is

$$\log t_i^* = \mu(x_i) + \sigma(x_i)\epsilon_i \tag{6.1}$$

$$t_i = \min(t_i^*, 26). \tag{6.2}$$

The variable t_i is the observed number of weeks a claimant collects unemployment insurance benefits, the exogenous vector x_i contains claimant characteristics, and ϵ_i is a homoskedastic and median zero error term. The censoring point is fixed at 26 weeks; the latent number of weeks a claimant would like to collect benefits, t_i^*, is observed only when a claimant collects benefits for less than this censoring point.

Our approach can estimate the location function $\mu(x_i)$ as well as the conditional distribution of failure times beyond the censoring point. We caution the estimators do not necessarily predict what will happen if maximum durations increase, since both censored and non-censored individuals may exhibit a behavioral response to a longer time limit. A conservative interpretation of our results, and the one favored in this paper, is that our estimates represent lower bounds on the effect of increasing the UI time limit. This interpretation is valid as long as time limit increases do not shorten individual spell lengths. Conceptually, the effect of extending the maximum duration can be separated into the sum of claimants (potentially censored) desired number of weeks under the current time limit plus any behavioral shift associated with the incentive effects of a longer program. The lower bound estimates obtained from our approach should be closer to the overall effects for short extensions, since the behavioral shift should be relatively small for marginal changes.

A common approach to estimating failure time models is the accelerated failure time (AFT) model. Our approach embeds a semiparametric AFT model. Although convenient, there is little theoretical justification for an AFT model based on a particular distribution and faulty distributional assumptions can significantly bias the results. For example, Moffitt(1985) reports that estimates of UI spell length for Tobit-type models are very sensitive to the assumed distribution. Any inconsistencies generated from incorrectly specifying the distribution are likely to be further compounded when predicting survival times beyond the censoring point.

Researchers have relaxed distributional assumptions using a variety of semiparametric approaches for censored duration models (see for example Cox(1972), Powell(1984), Horowitz(1996)). While useful, these semiparametric approaches cannot identify the shape of the survival function beyond the point of censoring without further assumptions. The method proposed in this paper does not specify the distribution of the homoskedastic component of the error term or the form of the location function, but still permits estimation of the conditional distribution of failure times beyond the censoring point.

6.2 Data

We estimate our model using individual-level administrative records from New Jersey's UI system. Our dataset is drawn from the 322,907 individuals who received their first payment between June 1, 1996 and October 25, 1997. Of these claimants, 63 percent were eligible for the legislated maximum of 26 weeks of UI benefits. We restrict our sample to claimants between the ages of 18 and 65, with complete demographic information, with no more than one week of partial UI benefits, and who were eligible for 26 weeks of UI. These restrictions result in a useable dataset containing 192,162 observations (see Card and Levine (2000) for further details of this administrative-level dataset).

New Jersey's UI system is administered at the state level, with benefits being financed by a tax on both workers and firms. These taxes are subject to maximum and minimum rates, and are partially experienced rated. Individuals can collect unemployment benefits if they have a sufficiently long work history and they continue to actively seek work. Benefits are paid weekly and are based on an individual's previous earnings with a maximum benefit of \$362 in 1996. During the period of our data, New Jersey experienced a strong labor market; for the individuals in our dataset, the median unemployment rate measured at the county level was 5.5 percent. During the period of our data, the maximum duration of UI benefits remained constant at 26 weeks. Of those individuals eligible for 26 weeks in our dataset, 43 percent exhausted their benefits.

A variety of characteristics influence the length of time an individual remains on unemployment insurance. In our analysis, we include demographic information, characteristics of a claimant's previous job, and a measure of local labor market conditions. The curse of dimensionality quickly makes local polynomial estimators computationally difficult with more than one or two continuous variables in the model. To make estimation feasible, we discretize the most important continuous covariates so the data can be grouped into a set of mutually exclusive and exhaustive cells. Many important characteristics such as race, gender, and union status are already discrete. Other characteristics, such as years of schooling, previous weekly earnings, age, or tenure, can be broken up into a few defining categories. Nonparametric quantile regression applied to this type of cell-grouped data is simple and computationally fast. With such cell-grouped data, our estimator requires only easily computed estimates of the median and other quantiles at the cell level combined with simple algebra.

Estimation proceeds as follows. First, median log failure times are calculated for each cell. In the second stage, estimates of the unknown constants $\frac{c_{\alpha_1}}{\Delta c}$ and $\frac{c_{\alpha_2}}{\Delta c}$ are calculated. This requires first estimating the α_1 and α_2 cell quantiles (two quantiles lower than the median). For each cell where the median is below the censoring point, the expressions $\hat{q}_{\alpha_1} - \hat{q}_{0.5}$ and $\hat{q}_{\alpha_2} - \hat{q}_{0.5}$ are divided by $\hat{q}_{\alpha_2} - \hat{q}_{\alpha_1}$. The sums of these cell calculations, using only cells where the median is below the censoring point, yield the estimates $\hat{c}_1$ and $\hat{c}_2$. The third stage estimates the median separately for each cell by taking a simple algebraic combination of the estimated constants $\hat{c}_1$ and $\hat{c}_2$ (common to all cells) and the estimated quantiles $\hat{q}_{\alpha_1}$ and $\hat{q}_{\alpha_2}$ (specific to each cell), as described in equation (3.4).

Once medians have been estimated for all cells, other quantiles can be estimated for each cell as well since the conditional distribution of ϵ_i can be identified up to scale. The α_j quantile of a cell is estimated as $\hat{q}_{\alpha_j} = \hat{c}_j(\hat{q}_{\alpha_2} - \hat{q}_{\alpha_1}) + \hat{\mu}$ where the estimator $\hat{c}_j$ is the average of the expression $(\hat{q}_{\alpha_j} - \hat{q}_{0.5})/(\hat{q}_{\alpha_2} - \hat{q}_{\alpha_1})$ over cells where both the estimated median and the α_j quantile lie below the censoring point. It should be noted that the estimated quantiles $\hat{q}_{\alpha_1}$ and $\hat{q}_{\alpha_2}$ appearing in these expressions need not be the same as those used to calculate $\hat{\mu}$. We keep them the same largely for convenience; we find that the estimates are not overly sensitive to the choice of these two quantiles.

Table 1 contains summary statistics for the discrete characteristics, including the fraction of claimants who exhaust their UI benefits. A typical UI claimant in New Jersey is male, white, middle-aged, not a union member, and has a high school degree or less. Claimants have varied previous earnings histories and a large fraction of UI claimants have been at their job for less than two years. Exhaustion rates differ widely across characteristics. For example, the exhaustion rate for whites is 40 percent compared to 53 percent for blacks and the exhaustion rate for union members is 34 percent compared to 45 percent for non-union members. Exhaustion rates are also higher for women, older workers, workers with long tenure, and in counties with high unemployment rates.

The complete interaction of all the characteristics in Table 1 yields a possible 864 non-overlapping cells, 862 of which are nonempty. In 598 of the cells, the median time on UI benefits is below the censoring point of 26 weeks. These cells contain approximately 75 percent of all observations. The remaining 25 percent of the data are grouped in 264 cells where more than 50 percent of claimants are observed to exhaust benefits.

6.3 Results

6.3.1 Estimates at the Cell Level

Estimation parallels the stages described in Section 3. In the first two stages, for each cell where the median lies below the censoring point, we calculate cell medians and obtain cell-level estimates of the two unknown lower quantiles of the error term up to scale. We chose the 30^{th} and 40^{th} quantiles for these two quantiles below the median. To construct our estimates, the 30^{th} and 40^{th} quantiles must be less than 26 weeks and the 40^{th} quantile cannot equal the median (to avoid division by zero). The first restriction eliminates approximately seven percent of the data and the second restriction eliminates less than one quarter of one percent of the data. Of course, other quantiles could be used as well. We find that other choices, such as the 17^{th} and 33^{rd} quantiles, do not alter the general findings.

One gauge of the method's applicability to UI claims is how well our estimate of the location function compares to the observed median in cells with less than 50 percent censoring. Figure 2 plots the observed median versus our estimate of the location function for cells where the observed median is less than 26 weeks. For cells with few observations, the estimate $\hat{\mu}$ is highly variable and has a large bootstrapped standard error. To simplify the graph, we plot only those cells with more than 100 observations, or about 75 percent of the data with observed cell medians below the censoring point. The observations are generally clustered around the forty-five degree line, although the estimate $\hat{\mu}$ is somewhat higher in the weeks immediately prior to censoring.

A primary objective of the empirical application is to estimate the location function for cells with more than a 50 percent exhaustion rate. Since it is impractical to present estimates for all 264 of these cells, in Table 2 we provide estimates for the subset of cells with more than 250 observations. Since only 13 percent of claimants belong to a union and union members have low exhaustion rates to begin with, all of the cells appearing in Table 2 contain only non-union claimants. Otherwise, there is a rich variety of characteristics defining cells. The estimated medians vary substantially, ranging from approximately 27 weeks (male, black, young, college, middle earnings, short tenure, high unemployment rate) to 40 weeks (female, black, mid-age, high school, high earnings, long tenure, high unemployment rate). The standard errors appearing in Table 2 are estimated using the bootstrap. The bootstrap estimates are based on 400 replications for each cell, with samples equal in size to the number of observations in a cell and drawn with replacement.

Conditional quantiles for all cells can be estimated up to the largest probability α_j for which sufficiently many cells have α_j^{th} quantiles observed below the censoring point. In our application, we chose 85 percent as the cutoff probability. There are 17 cells with an observed 85^{th} quantile less than 26 weeks, with a total of 2,526 observations in these cells. Using information from these observations, we are able to estimate up to the 85^{th} quantile for all cells in our dataset. For lower quantiles, of course, we are able to use information from more than just these 17 cells. For example, there are 119 cells (18,688 total observations) with an observed 70^{th} quantile less than 26 weeks.

Figure 3 graphs the estimated quantiles of UI receipt for eight cells with a diverse set of characteristics. We point out the varied shapes of the quantile functions suggest the presence of conditional heteroskedasticity. The top four panels show the estimates for four cells with observed medians well below the censoring point. The Kaplan-Meier estimates (with the axes reversed) closely track the estimates obtained using our method and generally lie within the bootstrapped 95 percent confidence intervals. As expected, these confidence intervals fan out as the fraction of active claims falls, i.e., for estimates of the higher quantiles. Notice that in panel 1, the observed 85^{th} quantile occurs at less than 26 weeks. Intuitively, the structure of our assumptions allows cells with more severe censoring to take advantage of information on the error distribution from cells like those found in panel 1 to estimate quantiles beyond the censoring point.

The bottom four panels in Figure 3 depict estimates for cells with more severe censoring. For example, in panel 5 close to 60 percent of the claimants in this cell exhaust benefits. Even with such severe censoring, however, we are able to estimate quantile values greater than 26 weeks. Although we could plot up to the 85^{th} quantile for each of these cells as we did in the top panels, we choose instead to include the estimated quantiles until the point estimate exceeds 52 weeks. We made this choice so the scale of the graphs would more clearly illustrate the comparison to the Kaplan-Meier estimate and what is happening immediately after the 26 week censoring point. As before, the Kaplan-Meier estimates are similar to our estimates.

6.3.2 Aggregate Survival Functions Based on Cell-Level Estimates

While cell-level estimates of medians and other quantiles beyond the censoring point are useful, they do not provide a concise summary of what is happening at a more aggregate level. In this section we describe how to combine cell-level estimates to create aggregate survival functions. The idea behind aggregation is to take a weighted vertical sum of the survival curves of individual cells.

To calculate the aggregate survival function, first notice the failure time distribution for all the data, $F(\cdot)$, can be written as a weighted average of the failure time distributions for each cell. The weights are merely the fraction of observations belonging to each cell and do not depend on the failure time. Let $F_i(\cdot)$ denote the distribution of failure times and w_i denote the weight for cell i. The p^{th} quantile is the number t such that

$$F(t) = \sum_i w_i F_i(t) = p \tag{6.3}$$

For each cell, we have already obtained quantile estimates up to the 85^{th} quantile. These cell quantiles are just the inverse distribution functions, $F_i^{-1}(\cdot)$, evaluated at various probabilities. To calculate the unconditional p^{th} quantile for the entire population, we need to find $p_1, p_2, \ldots, p_N$ such that

$$F_1^{-1}(p_1) = F_2^{-1}(p_2) = \ldots = F_N^{-1}(p_N) \tag{6.4}$$

and

$$\sum_i w_i p_i = p \tag{6.5}$$

where N indicates the total number of cells the data have been divided into.

We use a grid method to compute probabilities associated with a given value for the inverse distribution function, $F_i^{-1}(\cdot)$. For each cell we estimate the quantiles at 1,000 probabilities evenly spaced between 0 and 1. To calculate the aggregate survival function at a specific failure time, for each cell i we first find the probability p_i which corresponds as closely as possible to the quantile equaling this failure time. That is, we take a weighted average of probabilities straddling integer failure times, where the weights on the probabilities depend on the distance of the associated quantiles from the integer failure time. We then take the weighted sum of these cell probabilities, where the weights are estimated by the fraction of all observations belonging to cell i. We evaluate the overall survival function at failure times which take on positive integer values, although it should be noted the survival function could be evaluated at other points as well.

One difficulty arises in the current application since we are limited to quantile estimates below our chosen cutoff, i.e., the 85^{th} quantile. For cells where the 85^{th} quantile occurs relatively early, for example at 23 weeks, what probability should be assigned for failure times greater than 23 weeks? We adopt a very conservative approach. The maximum probability for failure times beyond the 85^{th} quantile would be one, implying all remaining observations fail immediately after the 85^{th} quantile. The minimum probability would not increase at all, implying that no observations fail after the 85^{th} quantile. These two extreme assumptions provide an upper and lower bound on the cell-specific probabilities associated with failure times beyond the 85^{th} quantile in a cell.

We point out that aggregate estimates can be calculated for subsets of the data as well as for the entire dataset. For example, unconditional survival functions can be calculated for different races, men and women, or union and non-union members separately. In this paper we present just one such example, for the subset of data with severe cell-level censoring. Figure 4 displays the aggregate survival functions for claimants in cells with an observed median greater than 26 weeks. For this group of heavily censored claimants, the predicted median spell length is approximately 32 weeks, or 6 weeks longer than the actual censoring point. Hence, this group is of special interest when considering extensions to UI benefits. The figure includes the upper and lower bound estimates of the survival function calculated as described above, as well as the Kaplan-Meier estimate. The standard errors for the Kaplan-Meier estimate are relatively small, so to keep things visually simple confidence intervals for the Kaplan-Meier estimate are excluded from the graph.

The upper and lower bound estimates of the survival curve are very similar, and do not begin to noticeably separate until 40 weeks. The estimates track the Kaplan-Meier estimate in the non-censored region fairly well. In the figure we have also added the estimated survival curve using a Weibull model, a model which has both a proportional hazards and an accelerated failure time (AFT) interpretation. We point out that our approach embeds semiparametric AFT models, so if the Weibull model is correct it should yield a similar estimate. The pointwise confidence intervals appearing in the graph for the Weibull estimate are based on standard errors calculated using the delta method. While the Weibull estimate and our upper and lower bound estimates are similar immediately prior to the censoring point, they have very different shapes to the right of 26 weeks. In particular, the Weibull estimate is noticeably lower for failure times past the censoring point. For claimants belonging to heavily censored cells, our approach predicts between 34.6 (upper bound estimate) and 32.7 (lower bound estimate) percent of claims would still be active at 52 weeks compared to 26.3 percent for the Weibull model. Put another way, the median residual life (i.e., median$[t - 26 \mid t > 26]$) for our model is between 35 and 39 weeks compared to only 24 weeks for the Weibull model–a difference of approximately 50 percent. These inconsistencies illustrate the costs of making incorrect distributional assumptions.

Figure 5 graphs the aggregate survival functions for all claimants in our dataset. To the left of the censoring point, our upper and lower bound estimates are similar to the Kaplan-Meier estimate. Table 3 lists the upper and lower bound estimates of the survival function, along with pointwise standard errors calculated using the bootstrap. As before, the bootstrap estimates are based on 400 replications, with samples equal in size to the total number of observations and drawn with replacement. Immediately prior to the censoring point at 25 weeks, between 43.2 (upper bound) and 41.0 (lower bound) percent of claims are estimated to be active compared to 41.4 percent for the Kaplan-Meier estimate. To the right of 26 weeks, the shape of the survival curve flattens out. Six weeks after the censoring point, between 35.5 (upper bound) and 33.2 (lower bound) percent of claims are estimated to still be ongoing. Starting around 34 weeks, the upper and lower bound point estimates start to diverge as a larger fraction of cell-level estimates exceeds the 85^{th} quantile. Although the bounds become wide in latter weeks, our estimates suggest a significant fraction of claimants would like to continue collecting UI benefits beyond the legally-specified maximum duration.

The information contained in the survival functions can readily be used to derive estimates of the monetary cost of extending UI benefits in New Jersey. For each cell and each week, cell-level costs are calculated by multiplying the average weekly benefit for individuals in a cell by the number of claims estimated to be ongoing in a cell. Average payouts vary across cells since benefit amounts depend on pre-displacement wages and work history. Aggregate cost estimates are then calculated by summing these cell-level cost estimates. As Table 3 documents, the first week of UI claims is estimated to cost around 49 million dollars in benefit payouts. By the censoring point of 26 weeks, this aggregate weekly payout falls to approximately 20 million dollars with an estimated cumulative cost of between 875 and 883 million dollars. Since the survival curve flattens out after the censoring point, predicted costs decline slowly. Increasing the maximum duration by fifty percent to 39 weeks is predicted to cost between an extra 199 and 222 million dollars. We remind the reader that these estimates represent aggregate costs assuming no behavioral shift associated with the incentive effects of a longer program. Since time limit increases should not decrease the length of time an individual receives UI, we interpret our estimates as the minimum cost of extending the maximum allowed duration. Any behavioral response would further add to the costs of extending UI benefits.

7 Conclusions

This paper has established conditions for nonparametric identification of the location function in a censored regression model. An estimation procedure was proposed, and shown to have desirable asymptotic properties. The procedure is simple to implement, as 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 the estimator performs well in finite samples. In an empirical application to unemployment insurance spells we estimate the effects of extending benefits beyond the current 26 week maximum in New Jersey.

The results in this paper suggest both empirical and theoretical areas for future research. The estimator introduced here would suggest testing parametric forms of the regression function against nonparametric alternatives in the censored regression model, as was done in standard regression models (Bierens and Ploberger(1997) and Horowitz and Spokoiny (2001)). Another important extension would be to allow for randomly censored data sets, as in Buckley and James(1979), Koul et al.(1981), Ying et al.(1995), Honoré et al.(2001) for semiparametric models. Furthermore, a more formal data driven approach needs to be developed for selection of the quantiles used in the second and third stages, and the asymptotics of such an approach needs to be derived. An empirical application we are currently pursuing is estimating the demand for retirement savings beyong the current Roth IRA and 401K limits, which are scheduled to increase.

References

[1] Anderson, P. and B.D. Meyer (1997) "Unemployment Insurance Take Up Rates and the After-Tax Value of Benefits", *Quarterly Journal of Economics*, 112, 913-938.

[2] Beran, R. (1981), "Nonparametric Regression with Randomly Censored Survival Data", *Technical Report, U.C.-Berkeley.*

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

[4] Blank, R. M. and D. Card (1991) "Recent Trends in Insured and Uninsured Unemployment: Is There an Explanation?", *Quarterly Journal of Economics*, 106, 1157-1190.

[5] Blaustein, S.J., W.J. Cohen and W. Haber (1993) *Unemployment Insurance in the United States: The First Half Century*, W.E. Upjohn Institute for Employment Research, Kalamazoo, MI.

[6] Buchinsky, M. (1998) "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research." *Journal of Human Resources* 33, 88-126.

[7] Buchinsky, M. and J. Hahn (1998) "An Alternative Estimator for the Censored Quantile Regression Model", *Econometrica*, 66, 653-672.

[8] Buckley, J. and I. James (1979), "Linear Regression with Censored Data," *Biometrika*, 66, 429-436.

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

[10] Chaudhuri, P. (1991a) "Nonparametric Quantile Regression", *Annals of Statistics*, 19, 760-777.

[11] Chaudhuri, P. (1991b) "Global Nonparametric Estimation of Conditional Quantiles and their Derivatives", *Journal of Multivariate Analysis*, 39, 246-269.

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

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

[14] Chen, S. and S. Khan (2001) "Semiparametric Estimation of a Partially Linear Censored Regression Model", *Econometric Theory*, 17, 567-590.

[15] Cox, D.R. (1972) "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society, Series B*, 34, 187-220.

[16] Fan, J. (1992) "Design-adaptive Nonparametric Regression", *Journal of the American Statistical Association*, 87, 998-1004.

[17] Fan, J. and I. Gijbels (1996) *Local Polynomial Modelling and its Applications*, New York: Chapman and Hall.

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

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

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

[21] Honoré, B.E., Khan, S. and J.L. Powell (2001), "Quantile Regression under Random Censoring", *Journal of Econometrics*, forthcoming.

[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. (1996) "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable", *Econometrica*, 64, 103-137

[25] Horowitz, J.L. (1998) *Semiparametric Methods in Econometrics*, New York: Springer.

[26] Horowitz, J.L. (2001) "Nonparametric Estimation of a Generalized Additive Model with an Unknown Link Function", *Econometrica*, 69, 499-513.

[27] 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.

[28] Hunt, J. (1995) "The Effect of Unemployment Compensation on Unemployment Duration in Germany", *Journal of Labor Economics* , 13, 88-120.

[29] Jarque, C. (1987) "An Application of LDV Models to Household Expenditure Analysis in Mexico", *Journal of Econometrics*, 36, 31-54.

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

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

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

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

[34] Koul, H., V. Suslara, and J. Van Ryzin (1981) "Regression Analysis with Randomly Right Censored Data," *Annals of Statistics*, 9, 1276-1288.

[35] Lewbel, A. and O.B. Linton (2001) "Nonparametric Censored Regression", *Econometrica, forthcoming.*

[36] Linton, O.B. and J. P. Nielsen (1995) "Estimating Structured Nonparametric Regression by the Kernel Method", *Biometrika*, 83, 529-540.

[37] Manski, C.F. (1985) "Semiparametric Analysis of Discrete Response: Asymptotic Properties of Maximum Score Estimation", *Journal of Econometrics*, 27, 313-334

[38] Manski, C.F. (1988) "Identification of Binary Response Models", *Journal of the American Statistical Association*, 83, 729-738.

[39] McCall, B.P. (1995) "The Impact of Unemployment Insurance Benefit Levels on Recipiency", *Journal of Business and Economic Statistics,* 12, 189-198.

[40] Melenberg, B. and A. Van Soest (1996) "Parametric and Semiparametric Modelling of Vacation Expenditures", *Journal of Applied Econometrics*, 11, 59-76.

[41] Meyer, B.D. (1990) "Unemployment Insurance and Unemployment Spells", *Econometrica,* 58, 757-782.

[42] Meyer, B.D. (1992) "Quasi-Experimental Evidence on the Effects of Unemployment Insurance from New York State", Unpublished Manuscript.

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

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

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

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

[47] Newey, W.K. (1994) "Kernel Estimation of Partial Means and a General Variance Estimator", *Econometric Theory*, 10, 233-253.

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

[49] Powell, J.L. (1986a) "Censored Regression Quantiles", *Journal of Econometrics*, 32, 143-155.

[50] Powell, J.L. (1994) "Estimation of Semiparametric Models", in Engle, R.F. and D. McFadden (eds.) , *Handbook of Econometrics, Vol. 4*, Amsterdam: North-Holland.

[51] Rosenbaum, P.R. and D.B. Rubin (1983) "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrica*, 70, 41-55

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

[53] Van Keilegom, I. and M.G. Akritas (1999) "Transfer of Tail Information in Censored Regression Models", *Forthcoming, Annals of Statistics.*

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

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

A Appendix

A.1 Proof of Theorem 2.1

Let x_0 satisfy $\mu(x_0) = \max_{x\in\mathcal{X}} \mu(x)$, and let x_1 satisfy $\sigma(x_1) = \max_{\mathcal{X}} \sigma(x)$. Note by the assumption in the theorem, $\mu(x_0) < 0$ and by Assumption I4 $\sigma(x_1)$ is positive and bounded. Let $\delta > 0$ and let $\tilde{\mu}(x) = \mu(x) - \frac{\delta\sigma(x)}{\sigma(x_0)}$. We define the distribution of $\tilde{\epsilon}_i$ through its quantiles $\alpha \in [0,1]$. Let $\alpha^* = \inf\{\alpha : \mu(x_0) + c_\alpha \sigma(x_1) > 0\}$. Note by Assumptions I2-I5, $\alpha^* \in (0.5, 1)$. Letting $\tilde{c}_\alpha$ denote the quantiles of $\tilde{\epsilon}_i$, we define $\tilde{c}_\alpha$ to take the following values:

$$\begin{aligned}
\tilde{c}_\alpha &= c_\alpha \quad 0 \leq \alpha \leq 1/2 && \text{(A.1)}\\
&= c_\alpha + \frac{\delta}{\sigma(x_0)} \quad \alpha^* \leq \alpha \leq 1 && \text{(A.2)}\\
&= \frac{\alpha - 0.5}{\alpha^* - 0.5} \cdot \tilde{c}_{\alpha^*} \quad \alpha \in (0.5, \alpha^*) && \text{(A.3)}
\end{aligned}$$

Let x be an arbitrary point in $\mathcal{X}$. We will show that $\mathcal{L}(y_i|x_i = x) = \mathcal{L}(\tilde{y}_i|x_i = x)$ by showing all the conditional quantiles are the same. Let $q_\alpha(x)$ and $\tilde{q}_\alpha(x)$ denote the respective quantiles. Let $\alpha^\dagger = \inf\{\alpha : \mu(x) + c_\alpha\sigma(x) > 0\}$. Note that $\alpha^\dagger \geq \alpha^* > 0.5$. Note that for $\alpha \in [0, \alpha^\dagger)$, we have $q_\alpha(x_0) = 0$. We also have:

$$
\begin{aligned}
\tilde{q}_\alpha(x) &= \max\left(\tilde{\mu}(x) + \tilde{c}_\alpha\sigma(x), 0\right) && \text{(A.4)}\\
&= \max\left(\mu(x) - \frac{\delta\sigma(x)}{\sigma(x_0)} + \tilde{c}_\alpha\sigma(x), 0\right) && \text{(A.5)}\\
&\leq \max\left(\mu(x) - \frac{\delta\sigma(x)}{\sigma(x_0)} + \tilde{c}_{\alpha^\dagger}\sigma(x), 0\right) && \text{(A.6)}\\
&= \max\left(\mu(x) - \frac{\delta\sigma(x)}{\sigma(x_0)} + \frac{\delta\sigma(x)}{\sigma(x_0)} + c_{\alpha^\dagger}\sigma(x), 0\right) && \text{(A.7)}\\
&= 0 && \text{(A.8)}
\end{aligned}
$$

where the inequality follows from $\tilde{c}_\alpha\sigma(x) < \tilde{c}_{\alpha^\dagger}\sigma(x)$, as $\alpha < \alpha^*$ and $\sigma(x) > 0$. The last equality follows from $\mu(x) + c_\alpha\sigma(x) \leq 0$ when $\alpha < \alpha^\dagger$.

For $\alpha \geq \alpha^\dagger$, we have $q_\alpha(x) = \max(\mu(x) + c_\alpha\sigma(x), 0) = \mu(x) + c_\alpha\sigma(x)$. We also have:

$$
\begin{aligned}
\tilde{q}_\alpha(x) &= \max\left(\tilde{\mu}(x) + \tilde{c}_\alpha\sigma(x), 0\right) && \text{(A.9)}\\
&= \max\left(\mu(x) - \frac{\delta\sigma(x)}{\sigma(x_0)} + \frac{\delta\sigma(x)}{\sigma(x_0)} + c_\alpha\sigma(x), 0\right) && \text{(A.10)}\\
&= \mu(x) + c_\alpha\sigma(x) && \text{(A.11)}
\end{aligned}
$$

Thus we have shown all the quantiles match up, making $\mu(\cdot)$ indistinguishable from $\tilde{\mu}(\cdot)$, establishing the desired result. ■

A.2 Proof of Theorem 3.1

In this section, we prove the limiting distribution results stated in the theorem. Throughout this section, we adopt new notation. Here we let $\tau_i, \sigma_i, w_i, \hat{w}_i, \hat{q}_{0i}, \hat{q}_{1i}, \hat{q}_{2i}, q_{0i}, q_{1i}, q_{2i}, \hat{q}_0, \hat{q}_1, \hat{q}_2, q_0, q_1, q_2, C_{ni}, C_n, N_{ni}$ denote $\tau(x_i), \sigma(x_i), w(q_{0.5}(x_i)), w(\hat{q}_{0.5}(x_i)), \hat{q}_{0.5}(x_i), \hat{q}_{\alpha_1}(x_i), \hat{q}_{\alpha_2}(x_i), q_{0.5}(x_i), q_{\alpha_1}(x_i), q_{\alpha_2}(x_i), \hat{q}_{0.5}(x), \hat{q}_{\alpha_1}(x), \hat{q}_{\alpha_2}(x)$, $q_{0.5}(x), q_{\alpha_1}(x), q_{\alpha_2}(x), C_n(x_i), C_n(x), \sum_{j\neq i} I[x_j \in C_n(x_i)]$, respectively. 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 $\hat{\mu}, \mu$ denote $\hat{\mu}(x), \mu(x)$ respectively. For a matrix A, with elements $\{a_{ij}\}$, we let $\|A\|$ denote $\left(\sum_{i,j} a_{ij}^2\right)^{1/2}$.

We note that since we aim to prove that the estimator converges at the optimal nonparametric rate of $O_p\left(n^{-\frac{p}{2p+d_c}}\right)$, we will use the term "asymptotically negligible" when referring to remainder terms which are $o_p\left(n^{-\frac{p}{2p+d_c}}\right)$. Our proof will rely heavily on three previously established properties of the nonparametric conditional quantile estimator used. The first 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 as $\mathcal{X}_\eta \equiv \{x_i \in \mathcal{X}_\tau : q_{0i} \geq \eta\}$.

Lemma A.1 *(From Chaudhuri et al. Lemma 4.3a) Under Assumptions RS, RD, ES, OS, BC.1,*

$$\sup_{x_i \in \mathcal{X}_\eta} |\hat{q}_{0i} - q_{0i}| = o_p(1)$$

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 $\mathcal{X}^c_{\eta/2}$ denote the set*

$$\{x_i \in \mathcal{X}_\tau, q_{0i} \leq \eta/2\}$$

and let A_n denote the event:

$$\{\hat{q}_{0i} \geq \eta \text{ for all } x_i \in \mathcal{X}^c_{\eta/2}\}$$

then under Assumptions RS,RD,ED,OS,BC.1, there exists constants C_1, 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 in Chaudhuri(1991a) and Chaudhuri et al.(1997).

Lemma A.3 *(From Lemmas 4.1 and 4.2 in Chaudhuri et al.(1997)) Let $q^*_\alpha(x_i, x)$ denote the k^{th} order Taylor polynomial approximation of $q_\alpha(x_i)$ for x_i close to x. Under assumptions RS, RD, ED, OS, and BC, for all $\alpha \geq 0.5$, $x : q_{0.5}(x) \geq \eta$, we have the following linear representation for the local polynomial estimator used in the second and third stages:*

$$\hat{q}_\alpha(x) - q_\alpha(x) = \frac{1}{n h_{(2,3)n}^{d_c} f_{Y,X}(q_\alpha(x), x)} \sum_{i=1}^{n} \left(I[y_i \leq q^*_\alpha(x_i, x)] - \alpha\right) I[x_i \in C_n(x)] + R_n(x) \tag{A.12}$$

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

$$\sup_{x \in \mathcal{X}_\eta} R_n(x) = o_p\left(n^{\frac{-p}{2p+d_c}}\right)$$

The main step in the proof is to show that the difference between the constants $\frac{c_{\alpha_1}}{\Delta c}, \frac{c_{\alpha_2}}{\Delta c}$ and their estimators $\hat{c}_1, \hat{c}_2$, are asymptotically negligible. We only show this result for the first quantile, as the same arguments can be used for the second quantile. We let c_1 denote the constant $\frac{c_{\alpha_1}}{\Delta c}$ and let β_i denote $\frac{q_{1i} - q_{0i}}{q_{2i} - q_{1i}}$ and $\hat{\beta}_i$ its estimated value, obtained by replacing quantile functions with their local polynomial estimators.

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

$$c_1^\dagger = \frac{\sum_{i=1}^n \tau_i \hat{w}_i c_1}{\sum_{i=1}^n \tau_i \hat{w}_i}$$

and we note that it can be easily be shown that

$$P(c_1^\dagger \neq c_1) \to 0$$

by Assumption ID, WF and Lemma A.1. Thus it will suffice to show that $\hat{c}_1 - c_1^\dagger$ is asymptotically negligible. This difference is of the form:

$$\hat{c}_1 - c_1^\dagger = \frac{\frac{1}{n}\sum_{i=1}^n \tau_i \hat{w}_i(\hat{\beta}_i - c_1)}{\frac{1}{n}\sum_{i=1}^n \tau_i \hat{w}_i} \tag{A.13}$$

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

Lemma A.4 *Under Assumptions WF,ID,ED,OS,RD,BC.1,*

$$\frac{1}{n}\sum_{i=1}^n \tau_i \hat{w}_i \xrightarrow{p} E[\tau_i w_i] \tag{A.14}$$

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

$$\frac{1}{n}\sum_{i=1}^n \tau_i w_i + \frac{1}{n}\sum_{i=1}^n \tau_i w_i'^*(\hat{q}_{0i} - q_{0i})$$

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:

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

It follows by the bound on the derivative of the weighting function and Lemmas A.1, 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.13) is $o_p\left(n^{\frac{-p}{2p+d_c}}\right)$. To do so, we take a mean value expansion of $\hat{w}_i$ around w_i, yielding the terms:

$$\frac{1}{n}\sum_{i=1}^n \tau_i w_i(\hat{\beta}_i - c_1) \tag{A.15}$$

$$+ \quad \frac{1}{n}\sum_{i=1}^n \tau_i w_i^{*'}(\hat{q}_{0i} - q_{0i})(\hat{\beta}_i - c_1) \tag{A.16}$$

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 (A.15).

Lemma A.5 *Under Assumptions WF, RD, ED, OS, BC1,*

$$\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i(\hat{\beta}_i - c_1) = o_p\left(n^{\frac{-p}{2p+d_c}}\right) \tag{A.17}$$

Proof : Note that $\tau_i w_i c_1 = \tau_i w_i \beta_i$. We linearize $\hat{\beta}_i - \beta_i$ to yield:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i(\hat{\beta}_i - \beta_i) &= \frac{1}{n}\sum_{i=1}^{n}\tau_i w_i \Delta q_i^{-1}\left(\hat{\bar{q}}_i - \hat{q}_{0i}^{(p)} - \bar{q}_i - q_{0i}\right) &\text{(A.18)}\\
&- \frac{1}{n}\sum_{i=1}^{n}\tau_i w_i \frac{\bar{q}_i - q_{0i}}{(\Delta q_i)^2}\left(\Delta\hat{q}_i - \Delta q_i\right) &\text{(A.19)}\\
&+ R_n &\text{(A.20)}
\end{aligned}$$

where

$$R_n = O_p\left(\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i\left(|\hat{q}_{2i} - q_{2i}|^2 + |\hat{q}_{1i} - q_{1i}|^2 + |\hat{q}_{0i}^{(p)} - q_{0i}|^2\right)\right)$$

It follows by Lemma 4.1 in Chaudhuri et al.(1997) that

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

and is thus asymptotically negligible by Assumption BC.2. The expressions in (A.18) and (A.19), are sample averages of undersmoothed conditional quantile estimators. We will thus only show that

$$\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i(\hat{q}_{1i} - q_{1i}) = o_p\left(n^{\frac{-p}{2p+d_c}}\right) \tag{A.21}$$

as similar arguments may used for the other terms. (A.21) follows from the same arguments used in Lemma 2 of Chen and Khan(2000). The only difference is in that paper, the smoothness and bandwidth conditions implied that the bias term was $o_p(n^{-1/2})$, whereas in this case, using Assumptions OS, BC.2, the bias term is $o_p\left(n^{\frac{-p}{2p+d_c}}\right)$. ∎

The following lemma shows that (A.16) is also asymptotically negligible.

Lemma A.6 *Under Assumptions WF,RD,ED,OS,BC.1,BC.2,*

$$\frac{1}{n}\sum_{i=1}^{n}\tau_i w_i'^{*}(\hat{q}_{0i} - q_{0i})(\hat{\beta}_i - c_1) = o_p\left(n^{\frac{-p}{2p+d_c}}\right) \tag{A.22}$$

Proof: We multiply the left hand side of the above expression by $I[q_{0i} \geq \eta/2] + I[q_{0i} < \eta/2]$, 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, since $\tau_i w_i'^* > 0 \Rightarrow \hat{q}_{0i} \geq \eta$. For the terms where $q_{0i} \geq \eta/2$, note that $c_1 = \beta_i$, and we can apply the uniform rates of convergence in 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 BC.1,BC.2, and OS, 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 estimator is

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

and for the local polynomial estimator it is

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

Letting $\|\cdot\|_\infty$ denote $\max_{1\leq i\leq n}|\cdot|$, we note that

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

is of order:

$$\|\hat{q}_{0i} - q_{0i}\|_\infty \|\hat{q}_{1i} - q_{1i}\|_\infty + \|\hat{q}_{0i} - q_{0i}\|_\infty \|\hat{q}_{2i} - q_{2i}\|_\infty + \|\hat{q}_{0i} - q_{0i}\|_\infty \|\hat{q}_{0i}^{(p)} - q_{2i}\|_\infty$$

which by the states uniform rates is

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

which is $o_p\left(n^{\frac{-p}{2p+d_c}}\right)$ by Assumptions OS,BC.1,BC.2. ■

Combining all our results, we can now replace the estimated constants with their true values:

$$\hat{\mu}(x) - \mu(x) = \frac{c_{\alpha_2}}{\Delta c}(\hat{q}_1 - q_1) - \frac{c_{\alpha_1}}{\Delta c}(\hat{q}_2 - q_2) + o_p\left(n^{\frac{-p}{2p+d_c}}\right) \tag{A.23}$$

The limiting distribution of the estimator follows from (A.12).

Table 1. Characteristics of Unemployment Insurance Recipients in New Jersey.

Characteristic	Percent	Percent Exhausting Benefits
Gender		
male	56.7	40.1
female	43.3	47.1
Race		
white (not hispanic)	62.7	39.8
black (not hispanic)	18.0	53.4
hispanic and other	19.3	44.4
Age		
age ≤ 35	37.8	40.0
35 < age ≤ 50	40.3	43.0
50 < age ≤ 65	21.9	48.5
Education		
high school or less	60.3	43.9
some college or more	39.7	41.8
Weekly Earnings at Previous Job		
earnings ≤ $375	33.2	42.1
$375 < earnings ≤ $625	33.5	47.6
earnings > $625	33.3	39.5
Tenure at Previous Job		
less than 2 years	43.3	39.6
greater than 2 years	56.7	45.8
Union Status at Previous Job		
union member	12.7	33.5
not a union member	87.3	44.5
County Unemployment Rate		
less than 5.5%	49.6	40.9
greater than or equal to 5.5%	50.4	45.3
All Observations	100.0	43.1

Notes: Data consists of 192,162 individuals from administrative records of the New Jersey Department of Labor. Sample is restricted to claimants age 18 to 65 who were eligible for 26 weeks of benefits and received their first payment between June 1, 1996 and October 25, 1997. Sample also excludes claimants with missing information on age, earnings, or UI claim characteristics.

Table 2. Estimated Medians for Cells with More Than a Fifty Percent Exhaustion Rate and More Than 250 Observations.

Gender	Race	Age	Education	Earnings	Tenure	Unemp. Rate	Estimated Median	Std. Error	Obs. in Cell
male	white	young	H.S.	middle	long	low	33.50	1.95	992
male	white	young	H.S.	middle	long	high	32.33	1.92	731
male	white	young	H.S.	middle	short	high	32.72	2.26	477
male	white	young	H.S.	high	long	low	28.74	2.38	304
male	white	mid-age	H.S.	middle	long	low	29.76	1.59	1500
male	white	mid-age	H.S.	mid-age	long	high	35.82	2.49	1135
male	white	mid-age	H.S.	high	long	low	34.31	2.49	819
male	white	mid-age	college	mid-age	long	high	31.25	3.02	712
male	white	older	H.S.	mid-age	short	high	33.50	4.98	258
male	white	older	college	mid-age	long	low	38.47	3.36	607
male	white	older	college	mid-age	long	high	39.70	3.84	353
male	white	older	college	high	long	low	33.50	2.85	857
male	black	young	H.S.	low	long	low	34.52	5.56	253
male	black	young	H.S.	low	short	high	34.31	1.91	1091
male	black	young	college	mid-age	long	high	35.14	2.24	429
male	black	young	college	mid-age	short	high	26.95	2.96	387
male	black	mid-age	H.S.	low	short	high	34.31	2.65	524
male	black	mid-age	college	low	short	high	32.72	4.01	251
male	hispanic	young	H.S.	mid-age	long	high	37.28	3.59	262
male	hispanic	young	college	mid-age	long	high	30.57	2.37	253
female	white	older	H.S.	mid-age	long	low	32.72	2.70	699
female	white	older	H.S.	high	long	low	28.94	1.97	1345
female	white	older	H.S.	high	long	high	32.72	3.63	981
female	white	older	college	mid-age	long	low	31.46	3.35	276
female	white	older	college	mid-age	short	low	35.14	4.95	277
female	white	older	college	high	long	low	31.97	1.79	2114
female	white	older	college	high	long	high	29.76	2.28	1085
female	black	young	H.S.	low	long	high	33.50	2.81	649
female	black	young	H.S.	mid-age	short	high	28.54	1.64	415
female	black	young	college	mid-age	short	high	26.95	4.40	321
female	black	mid-age	H.S.	low	long	high	33.22	2.34	488
female	black	mid-age	H.S.	low	short	low	31.97	4.23	301
female	black	mid-age	H.S.	mid-age	long	low	31.50	2.23	288
female	black	mid-age	H.S.	mid-age	short	high	31.50	2.24	344
female	black	mid-age	H.S.	high	long	high	39.70	4.13	297
female	black	mid-age	college	low	short	high	33.22	3.06	252
female	black	mid-age	college	high	long	high	33.50	4.67	339
female	black	older	H.S.	mid-age	long	high	36.53	4.58	337

Notes: All cells contain non-union claimants. Standard errors are calculated using the bootstrap, based on 400 replications with samples equal in size to the number of observations in a cell and drawn with replacement.

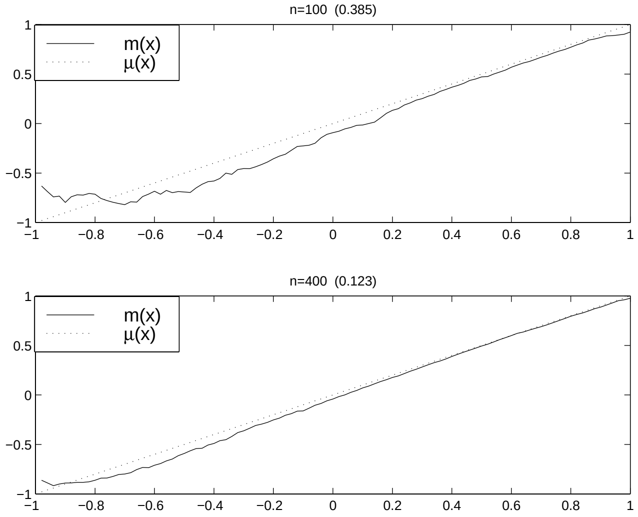
Table 3. Aggregate Survival Function and Total Cost Estimates for All Claimants.

Prior to Censoring Point					Censoring Point and After				
	Survival Function		Total Cost			Survival Function		Total Cost	
Week	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Week	Upper Bound	Lower Bound	Upper Bound	Lower Bound
0	1.0000	0.9592	48.64	46.55	26	0.4212	0.3990	20.53	19.41
	(0.0000)	(0.0013)	(0.70)	(0.68)		(0.0040)	(0.0067)	(0.38)	(0.47)
1	0.9724	0.9592	47.22	46.55	27	0.4109	0.3887	19.98	18.80
	(0.0015)	(0.0013)	(0.69)	(0.68)		(0.0044)	(0.0053)	(0.38)	(0.42)
2	0.9423	0.9395	45.71	45.55	28	0.3967	0.3745	19.43	18.22
	(0.0014)	(0.0014)	(0.67)	(0.66)		(0.0053)	(0.0053)	(0.41)	(0.42)
3	0.9174	0.9164	44.45	44.39	29	0.3851	0.3627	18.85	17.63
	(0.0014)	(0.0015)	(0.65)	(0.65)		(0.0062)	(0.0059)	(0.44)	(0.42)
4	0.8931	0.8925	43.23	43.19	30	0.3771	0.3538	18.29	17.04
	(0.0014)	(0.0015)	(0.64)	(0.64)		(0.0066)	(0.0072)	(0.46)	(0.48)
5	0.8679	0.8673	41.98	41.95	31	0.3650	0.3417	17.76	16.49
	(0.0013)	(0.0014)	(0.62)	(0.62)		(0.0070)	(0.0079)	(0.47)	(0.51)
6	0.8413	0.8410	40.68	40.67	32	0.3554	0.3321	17.27	15.92
	(0.0014)	(0.0014)	(0.61)	(0.61)		(0.0072)	(0.0106)	(0.48)	(0.63)
7	0.8139	0.8138	39.35	39.34	33	0.3455	0.3215	16.81	15.28
	(0.0013)	(0.0013)	(0.59)	(0.59)		(0.0069)	(0.0146)	(0.46)	(0.82)
8	0.7857	0.7857	37.99	37.99	34	0.3378	0.3050	16.42	14.64
	(0.0013)	(0.0013)	(0.57)	(0.57)		(0.0067)	(0.0173)	(0.45)	(0.95)
9	0.7570	0.7569	36.62	36.62	35	0.3313	0.2908	16.07	14.01
	(0.0014)	(0.0014)	(0.55)	(0.55)		(0.0066)	(0.0181)	(0.44)	(0.99)
10	0.7283	0.7282	35.25	35.25	36	0.3253	0.2772	15.75	13.38
	(0.0014)	(0.0014)	(0.53)	(0.53)		(0.0068)	(0.0172)	(0.44)	(0.94)
11	0.6999	0.6999	33.91	33.91	37	0.3180	0.2598	15.45	12.86
	(0.0015)	(0.0015)	(0.52)	(0.52)		(0.0071)	(0.0157)	(0.44)	(0.87)
12	0.6730	0.6728	32.62	32.61	38	0.3103	0.2521	15.15	12.38
	(0.0016)	(0.0016)	(0.50)	(0.50)		(0.0075)	(0.0140)	(0.45)	(0.79)
13	0.6468	0.6466	31.38	31.36	39	0.3042	0.2459	14.84	11.96
	(0.0017)	(0.0017)	(0.48)	(0.48)		(0.0077)	(0.0128)	(0.46)	(0.72)
14	0.6234	0.6231	30.26	30.22	40	0.2985	0.2394	14.53	11.57
	(0.0017)	(0.0018)	(0.46)	(0.47)		(0.0079)	(0.0120)	(0.46)	(0.68)
15	0.6008	0.6002	29.19	29.13	41	0.2904	0.2307	14.23	11.20
	(0.0018)	(0.0018)	(0.45)	(0.45)		(0.0081)	(0.0110)	(0.47)	(0.63)
16	0.5795	0.5787	28.18	28.09	42	0.2845	0.2248	13.95	10.86
	(0.0020)	(0.0022)	(0.43)	(0.43)		(0.0082)	(0.0106)	(0.47)	(0.60)
17	0.5598	0.5575	27.25	27.10	43	0.2806	0.2199	13.68	10.52
	(0.0019)	(0.0027)	(0.42)	(0.42)		(0.0081)	(0.0114)	(0.47)	(0.63)
18	0.5411	0.5378	26.36	26.18	44	0.2764	0.2155	13.44	10.17
	(0.0018)	(0.0025)	(0.41)	(0.41)		(0.0080)	(0.0127)	(0.46)	(0.66)
19	0.5223	0.5190	25.49	25.30	45	0.2721	0.2112	13.21	9.84
	(0.0019)	(0.0023)	(0.40)	(0.39)		(0.0081)	(0.0140)	(0.46)	(0.72)
20	0.5044	0.5010	24.61	24.42	46	0.2662	0.2021	13.00	9.49
	(0.0024)	(0.0024)	(0.40)	(0.39)		(0.0082)	(0.0150)	(0.46)	(0.77)
21	0.4875	0.4827	23.76	23.55	47	0.2597	0.1850	12.79	9.14
	(0.0031)	(0.0033)	(0.40)	(0.40)		(0.0084)	(0.0156)	(0.47)	(0.79)
22	0.4723	0.4675	22.98	22.69	48	0.2545	0.1789	12.58	8.77
	(0.0036)	(0.0054)	(0.40)	(0.45)		(0.0086)	(0.0157)	(0.48)	(0.80)
23	0.4580	0.4532	22.30	21.82	49	0.2492	0.1745	12.37	8.41
	(0.0036)	(0.0086)	(0.39)	(0.54)		(0.0089)	(0.0152)	(0.49)	(0.78)
24	0.4451	0.4271	21.70	20.94	50	0.2457	0.1614	12.16	8.07
	(0.0038)	(0.0097)	(0.39)	(0.58)		(0.0090)	(0.0145)	(0.49)	(0.74)
25	0.4317	0.4096	21.11	20.11	51	0.2418	0.1580	11.97	7.71
	(0.0040)	(0.0083)	(0.38)	(0.52)		(0.0091)	(0.0130)	(0.49)	(0.67)
					52	0.2385	0.1408	11.79	7.34
						(0.0091)	(0.0119)	(0.50)	(0.62)

Notes: Total cost is measured in millions of dollars. Standard errors are calculated using the bootstrap, based on 400 replications with samples equal in size to the number of observations and drawn with replacement.

Figure 1

$\mu(x) = x$



$\mu(x) = x^2 - \mathcal{C}_1$

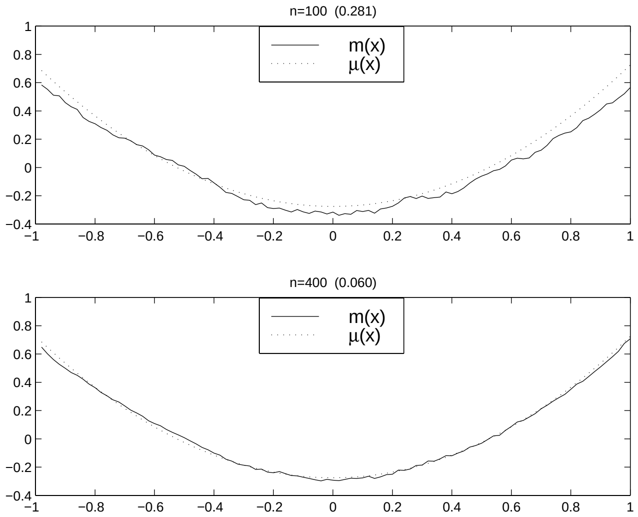
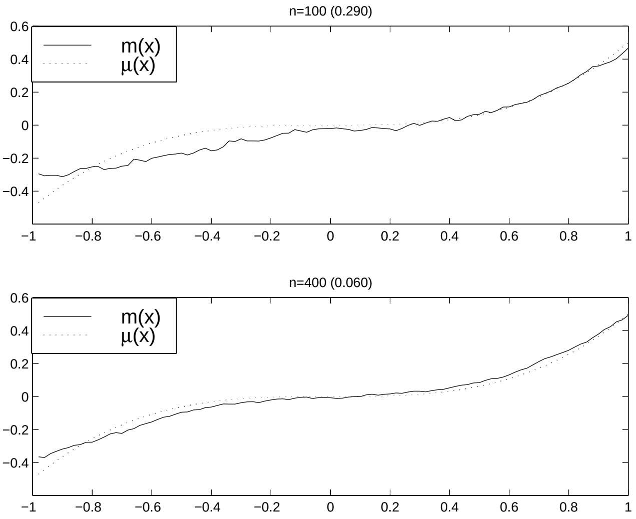


Figure 1

$\mu(x) = 0.5 \cdot x^3$



$\mu(x) = e^x - \mathcal{C}_2$

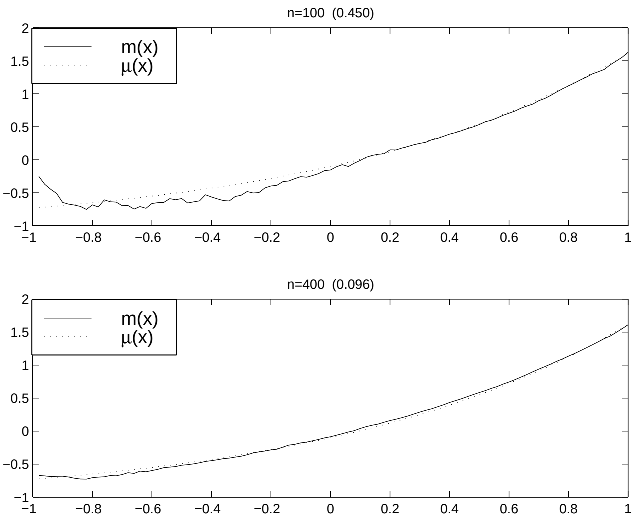
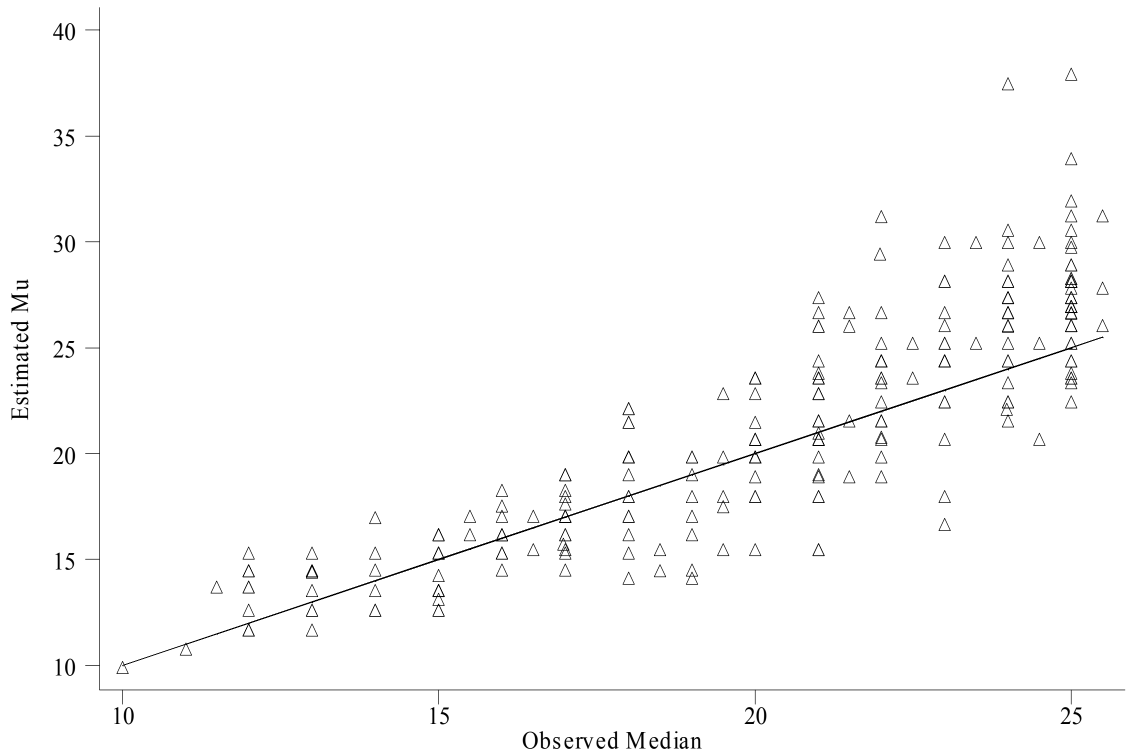
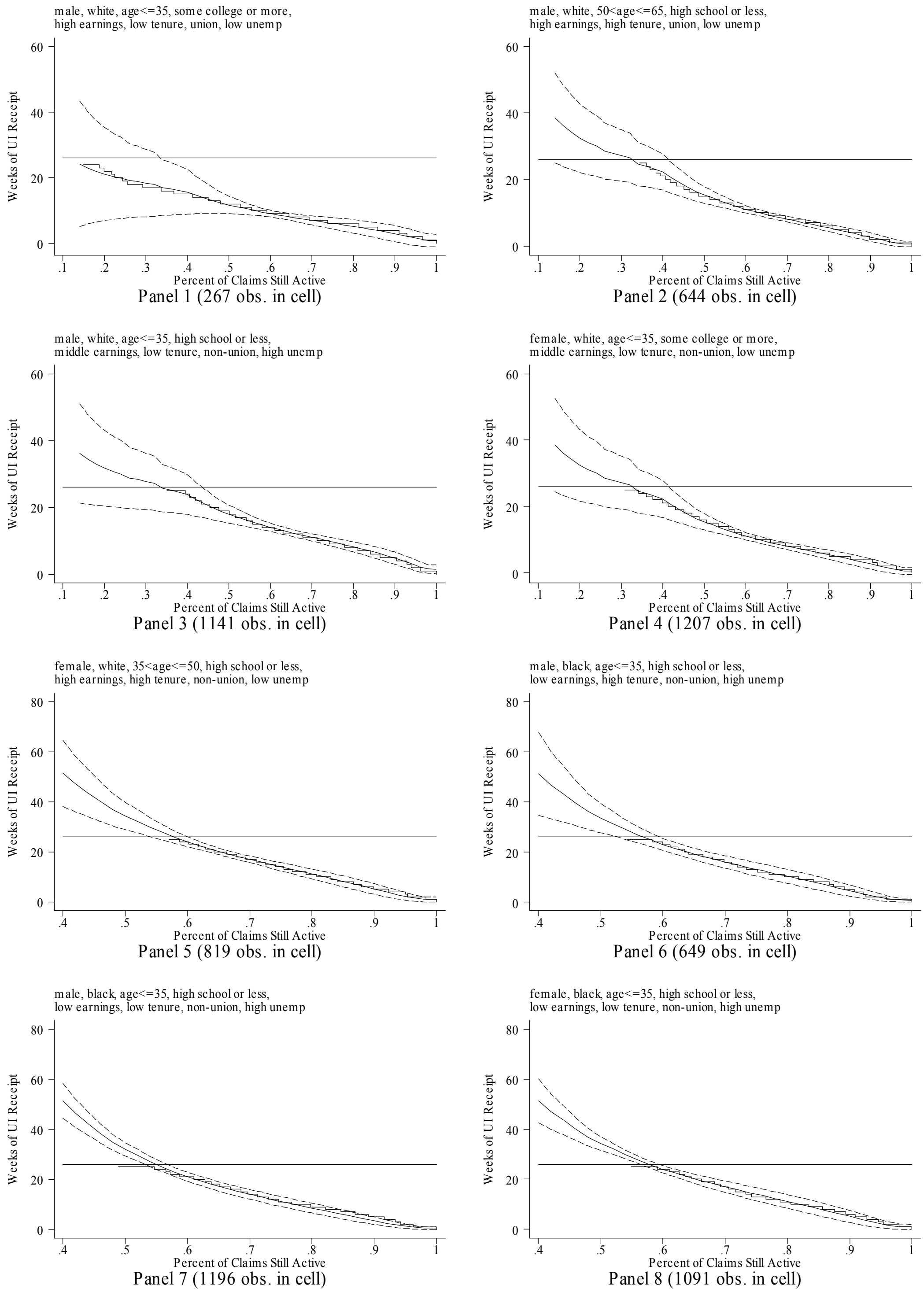


Figure 2. Comparison of the Estimated Location Function $\hat{\mu}$ to the Observed Median for Cells with Less than Fifty Percent Censoring.



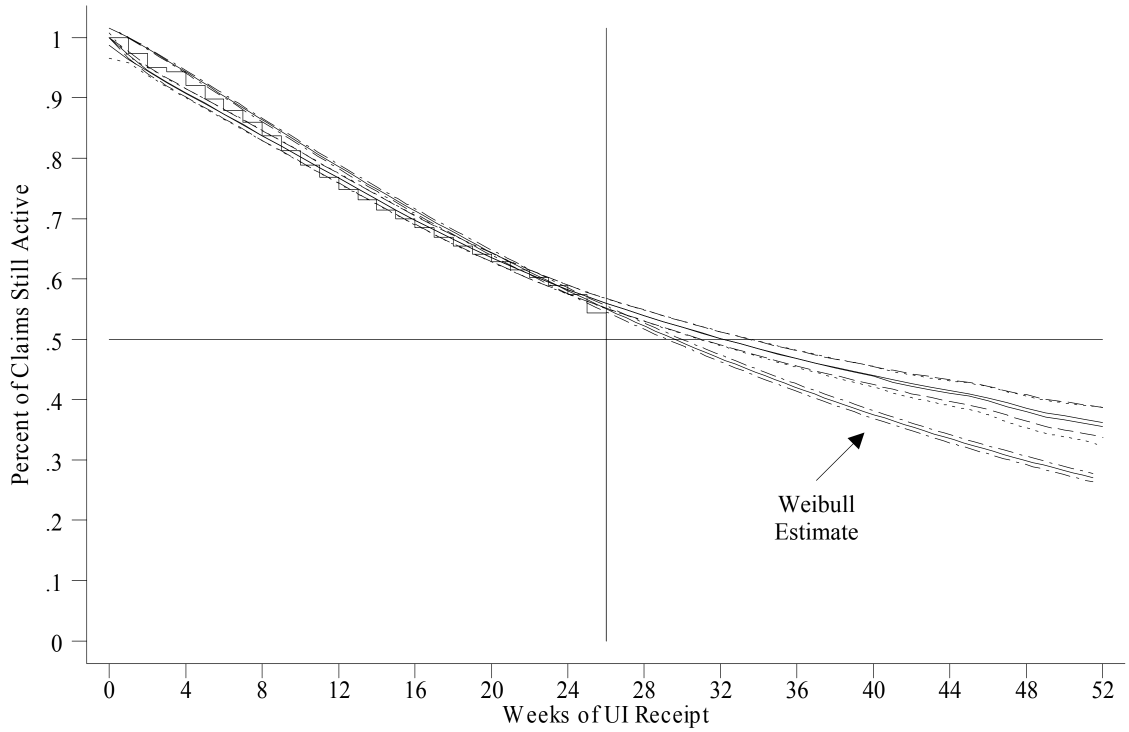
Notes: Solid line indicates the 45 degree line. Sample restricted to cells with at least 100 observations.

Figure 3. Estimated Quantiles of Unemployment Insurance Receipt for Selected Cells, with a Comparison to the Kaplan-Meier Estimates.



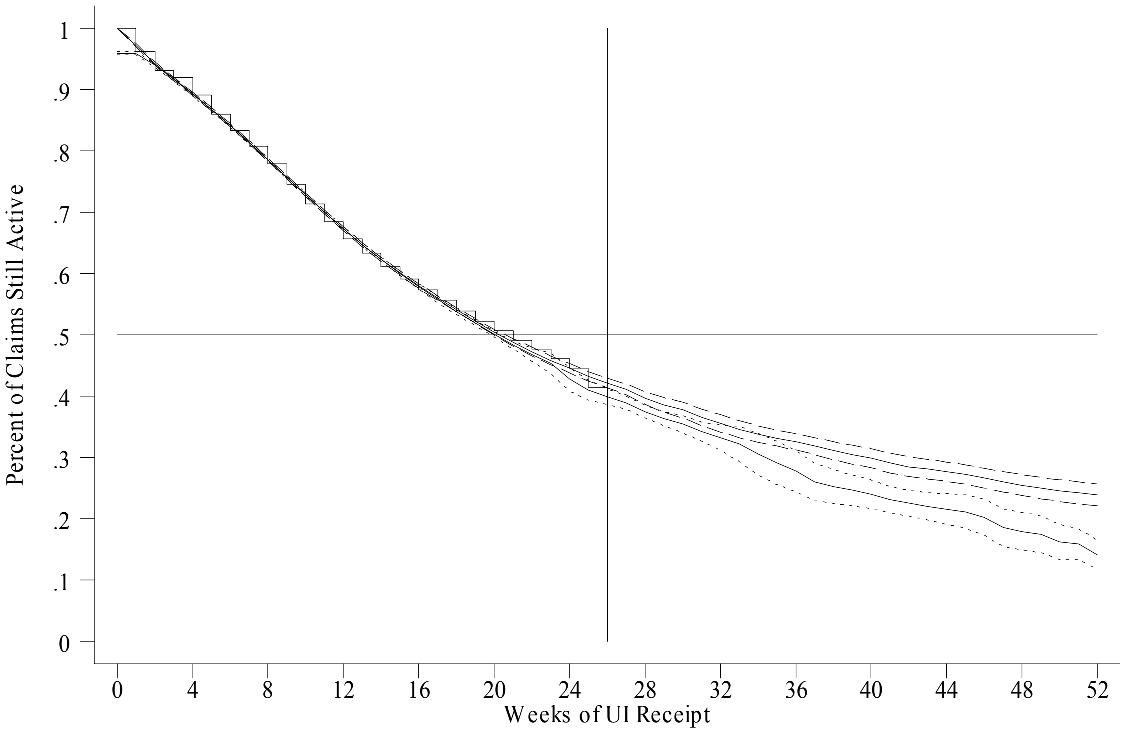
Notes: Smooth lines indicate estimated quantiles and broken lines indicate pointwise 95 percent confidence intervals. The step function is the Kaplan-Meier estimate (with the axes reversed). A line is drawn at 26 weeks to indicate the censoring point. Panels 1-4 plot the estimated quantiles up to the point where 15 percent of claims are predicted to be active. Panels 5-8 plot the estimated quantiles until the point estimate exceeds 52 weeks.

Figure 4. Upper and Lower Bound Estimates of the Aggregate Survival Function for Claimants in Cells with More than Fifty Percent Censoring, with a Comparison to the Kaplan-Meier and Weibull Accelerated Failure Time Estimates.



Notes: Two solid lines indicate the upper and lower bound estimates of the aggregate survival function. The Weibull estimate is labeled on the graph. Broken lines indicate pointwise 95 percent confidence intervals for the estimates. The step function is the Kaplan-Meier estimate. A vertical line is drawn at 26 weeks to indicate the censoring point.

Figure 5. Upper and Lower Bound Estimates of the Aggregate Survival Function for All Claimants, with a Comparison to the Kaplan-Meier Estimate.



Notes: Two solid lines indicate the upper and lower bound estimates of the aggregate survival function. Broken lines indicate pointwise 95 percent confidence intervals for the estimates. The step function is the Kaplan-Meier estimate. A vertical line is drawn at 26 weeks to indicate the censoring point.