

Heteroskedastic Transformation Models with Covariate Dependent Censoring

Shakeeb Khan, Youngki Shin, Elie Tamer

August 2007

Abstract

In this paper we propose an inferential procedure for transformation models with conditional heteroskedasticity in the error terms. The proposed method is robust to covariate dependent censoring of arbitrary form. We provide sufficient conditions for point identification. We then propose a consistent estimator and show that it is asymptotically \sqrt{n} normal. We conduct a simulation study that reveals adequate finite sample performance. We also use the estimator in an empirical illustration, where we estimate the effect of UI benefits.

JEL Classification: C13, C14, C41

Keywords: Rank estimation, transformation model, covariate dependent censoring, conditional heteroskedasticity.

1 Introduction

The monotone transformation model is usually expressed as:

$$T(y_i) = x_i' \beta_0 + \varepsilon_i \tag{1}$$

where y_i is a scalar dependent variable, x_i is a d -dimensional vector of covariates, ε_i is an unobservable error term whose distribution is unknown. The function $T(\cdot)$ is unspecified, but assumed

to be monotone. The object of interest is the d -dimensional parameter vector β_0 . We assume that we have a sample of iid observations.

In this paper we consider the monotone transformation model with a *general form of random censoring* and *conditional heteroskedasticity*. We motivate this problem by illustrating a duration analysis. It can be shown that popular duration models such as an accelerated failure time model, a proportional hazard model, and a mixed proportional hazard model, are specific examples of the transformation model (see ?). Thus, applying the transformation model will reduce model specification error that might be caused by a specific hazard assumption. In duration data, however, censoring arises naturally because of a data collection procedure, and it may easily depend on covariates. For instance, a censoring variable of unemployment duration data may possibly depend on individual characteristics such as age, occupation, education etc. Also, conditional heteroskedasticity is quite common in such an economic data set. Moreover, the presence of heteroskedasticity in nonlinear models can lead to inconsistency. Hence, the estimator proposed here fills a gap in the literature on censored models in that it is both robust to random censoring and heteroskedasticity.

There is a large volume of the literature on estimating the transformation model, but no estimator allows for both a general form of random censoring and conditional heteroskedasticity simultaneously. The maximum rank correlation (MRC) estimator by ? and the Monotone Rank Estimator (MRE) by ? can only be applied to special forms of conditional heteroskedasticity, and do not allow for a censoring variable to be dependent on covariates in an arbitrary way. ? proposed the Quantile Rank Estimator (QRE) for a heteroskedastic transformation model, but it still does not allow for random censoring. Recently, ? proposed the Partial Rank Estimator (PRE) that allows for random censoring, but it assumes that the error term ε_i be distributed independently of the covariates x_i .

The two-stage partial rank estimation procedure can accommodate data with random censoring and conditional heteroskedasticity. The new estimator shares the same advantages of usual rank estimators, so it does not require a specific error distribution, and satisfies \sqrt{n} -consistency. In

addition, it is consistent for random covariate dependent censoring and heteroskedastic error terms under mild regularity conditions. The key condition is the median independence that is also adopted by ? for the maximum score estimator. The new estimator exploits the monotone property of a conditional median function that is implied by the median independence. Without any difficulty, it can be extended to any other quantile that is independent of covariates.

The rest of the paper is organized as follows. In section 2, we introduce the model and estimation method. The asymptotic properties are also proposed in the section. The finite sample properties are investigated by means of Monte Carlo simulations in section 3. Section 4 applies our estimator to unemployment duration data and compare our estimator with existing ones. Section 5 concludes and suggests future research areas. Technical proofs are presented in the appendix unless it is helpful to understand the argument.

2 Estimation Procedure and Asymptotic Properties

We modify the model (1) slightly to introduce censoring problem.¹ The right censored transformation model can be expressed as:

$$T(v_i) = \min(x_i'\beta_0 + \varepsilon_i, c_i) \quad (2)$$

$$d_i = 1(x_i'\beta_0 + \varepsilon_i \leq c_i) \quad (3)$$

where $1(\cdot)$ is an indicator function, c_i is a random censoring variable that may depend on x_i in an arbitrary way, d_i is a binary random variable that indicates if the data is censored or not, and v_i is a new dependent variable. So, the variable v_i is $T^{-1}(x_i'\beta_0 + \varepsilon_i)$ for uncensored observations, which is y_i in the original transformation model. For censored observations, v_i is $T^{-1}(c_i)$. Observations are composed of $(p + 2)$ -dimensional vectors (v_i, d_i, x_i') , and satisfy the *i.i.d.* assumption.

Random censoring models has been studied widely both in econometrics and statistics, espe-

¹We illustrate here identification for the univariate censoring case. Similar arguments can be used to attain point identification results for the double censoring case. See ?.

cially related to covariate dependent censoring (see ? and references therein). Although the PRE proposed by ? is robust to covariate dependent censoring in the censored transformation model, it requires that the error term ε_i be distributed independently of the covariate x_i . This assumption may be overly restrictive in the sense that it rules out any form of conditional heteroskedasticity. In this paper we relax the independence assumption by assuming only a quantile of ε_i , say the median, is independent of the covariates. To permit random covariate dependent censoring, we now make the assumption that the random variables c_i and ε_i are statistically independent given x_i .

First, we establish identification result and propose a Quantile Partial Rank Estimator (QPRE). Then, we show that it satisfies \sqrt{n} -consistency and asymptotic normality. The new estimator exploits monotone property of the conditional quantile function. Based on monotone quantile functions, we can construct a rank estimator analogously to ? and ?. The key condition for point identification of the MRC is

$$P(f_{ij} \geq 0 | x_i, x_j) \geq P(f_{ji} \geq 0 | x_i, x_j) \Leftrightarrow x'_i \beta_0 \geq x'_j \beta_0. \quad (4)$$

To motivate the new estimator, we first define the random variables:

$$y_{1i} = d_i \cdot v_i + (1 - d_i) \cdot (+\infty) \quad (5)$$

$$y_{0i} = v_i. \quad (6)$$

Then, we can derive the condition (4) by defining

$$f_{ij} = \text{med}(T(y_{1i})) - \text{med}(T(y_{0j})). \quad (7)$$

Let $m_1(x_i)$ and $m_0(x_i)$ denote conditional median functions of $T(y_{1i})$ and $T(y_{0i})$ respectively:

$$m_1(x_i) = \text{med}(T(y_{1i}) | x_i) \quad (8)$$

$$m_0(x_i) = \text{med}(T(y_{0i}) | x_i). \quad (9)$$

Then, we can rewrite the conditional probability $P(f_{ij} \geq 0 | x_i, x_j)$ as:

$$\Pr(f_{ij} \geq 0 | x_i, x_j) = \Pr(m_1(x_i) \geq m_0(x_j)), \quad (10)$$

and the following inequality always holds:

$$m_0(x_i) \leq x_i' \beta_0 \leq m_1(x_i). \quad (11)$$

Therefore, we can see the following rank correlation:

$$x_i \beta_0 \geq x_j \beta_0 \Rightarrow \Pr(m_1(x_i) \geq m_0(x_j)) = 1.$$

We next summarize the related regularity conditions and propose the detailed identification result in the Theorem 2.1.

I1 Let S_X denote the support of x_i , and let \mathcal{X} denote the set

$$\mathcal{X} = \{x \in S_X : \Pr(x_i' \beta_0 \leq c | x_i = x) = 1\}. \quad (12)$$

Then, \mathcal{X} has positive measure.

I2. The random variable ε_i is distributed independently of the random variable c_i conditional on

x_i , and satisfies the median independence:

$$\text{med}(\varepsilon_i|x_i) = 0. \tag{13}$$

- I3. The support of x_i is not contained in a proper linear subspace of \mathbf{R}^d and the d th component of x_i has an everywhere positive Lebesgue density conditional on the remaining components.

Condition I1 requires that there exists some x with positive measure, such that the index value, $x\beta_0$, is less than the censoring variable for all support of c . The condition can be seen as an extension of the Assumption R.1 in ? to a random censoring case. This condition always holds if the censoring variable is truncated and Condition I3 is true. Condition I2 requires the conditional independence between the error term and the censoring variable. Also it assumes that median of the error term is distributed independently of the covariate x_i . Since this condition permits any relationship between ε_i and x_i for other quantiles, conditional heteroskedasticity is allowed in the model. Condition I3 is the standard regularity condition for point identification in semiparametric literature. This condition is also adopted by the MRC and the maximum score estimator.

Point identification is characterized by the following theorem, whose proof is left to the appendix:

Theorem 2.1 *Suppose that the Assumptions I1–I3 hold. Let y_i denote $T^{-1}(x'_i\beta_0 + \varepsilon)$. Then we have that*

$$m_0(x_i) = m_1(x_i) = \text{med}(T(y_i)|x_i) = x_i\beta_0 \tag{14}$$

if and only if $x \in \mathcal{X}$.

The above result, along with the invariance of medians, suggests an (infeasible) rank estimator based on the conditional medians of y_{0i} and y_{1i} . Thus, we would estimate β_0 by maximizing the following objective function

$$Q_n(\beta) = \frac{1}{n(n-1)} \sum_{i \neq j} 1[m_1(x_i) \geq m_0(x_j)]1[x'_i\beta \geq x'_j\beta]. \tag{15}$$

The function $Q_n(\beta)$ exploits Spearman's measure for rank correlation between median functions and indices.

To construct a feasible estimation procedure, we replace the unknown median functions in the above estimator with their nonparametric estimators. For these first stage estimators, we adopt the local polynomial approach introduced in Deaton (2002). For a detailed description of the estimator, see Deaton (2002). Here, we simply let $\hat{m}_0^{\delta_n, p}(x_i)$, $\hat{m}_1^{\delta_n, p}(x_i)$ denote the local polynomial estimators where the superscripts denote the bandwidth sequence (δ_n) , and order of polynomial (p) used. Conditions on δ_n and p are stated in the theorem below characterizing the limiting distribution of our estimator of β_0 . To avoid the technical difficulty of dealing with a smoothing parameter inside an indicator function, we define our heteroskedasticity robust estimator of β_0 , denoted here as $\hat{\beta}_{ht}$ as follows:

$$\hat{\beta}_{ht} = \arg \max_{\beta \in \mathcal{B}} \frac{1}{n(n-1)} \sum_{i \neq j} K_{h_n}(\hat{m}_1^{\delta_n, p}(x_i) - \hat{m}_0^{\delta_n, p}(x_j)) 1[x'_i \beta \geq x'_j \beta] \quad (16)$$

where $K_{h_n}(\cdot) \equiv K(\cdot/h_n)$, with $K(\cdot)$ denoting a smooth approximating function to an indicator function (i.e. a cumulative distribution function), and h_n denotes a sequence of positive constants, converging to 0, such that in the limit we have an indicator function. This smoothing technique was introduced in the seminal work of Deaton (2002).

We next state the limiting distribution theory for $\hat{\beta}_{ht}$. Our limiting distribution theory for this estimator is based on the following assumptions:

Assumptions on the Median Functions

Q1. For any value $x^{(d)}$ in the support of $x_i^{(d)}$, $m_j(\cdot)$ $j = 0, 1$ is k times differentiable in $x_i^{(c)}$.

Letting $\nabla_k m_j(x^{(c)}, x^{(d)})$ denote the vector of k^{th} order derivatives of $m_j(\cdot)$ in $x_i^{(c)}$, we assume the following Lipschitz condition:

$$\|\nabla_k m_j(x_1^{(c)}, x^{(d)}) - \nabla_k m_j(x_2^{(c)}, x^{(d)})\| \leq \mathcal{K} \|x_1^{(c)} - x_2^{(c)}\|^\gamma$$

for all values $x_1^{(c)}, x_2^{(c)}$ in the support of $x_i^{(c)}$, where $\|\cdot\|$ denotes the Euclidean norm, $\gamma \in (0, 1]$, and \mathcal{K} is some positive constant. In the theorems to follow, we will let $p = k + \gamma$ denote the

order of smoothness of the quantile function.

Assumptions on the Trimming Function

- T.** The trimming function $\tau : \mathfrak{R}^d \mapsto \mathfrak{R}^+$ is continuous, bounded, and bounded away from zero on its support, denoted by \mathcal{X}_τ , a compact subset of \mathfrak{R}^d .

Assumptions on the Regressors

- B1.** The sequence of $d + 2$ dimensional vectors (v_i, d_i, x_i) are independent and identically distributed.
- B2.** The regressor vector x_i has support which is a subset of \mathfrak{R}^d .

We order the components of x_i so it can be written as $x_i = (x_i^{(d)}, x_i^{(c)})'$. Let d_c denote $\dim(x_i^{(c)})$. Assume that $1 \leq d_c \leq d$ and that the support $x_i^{(c)}$ is a convex subset of \mathfrak{R}^{d_c} and has nonempty interior. Assume that the support of $x_i^{(d)}$ is a finite number of points lying in \mathfrak{R}^{d-d_c} . We will let $f_X(x)$ denote the product of the conditional (Lebesgue) density of $x_i^{(c)}$ given $x_i^{(d)}$ (denoted by $f_{X^{(c)}|X^{(d)}=x^{(d)}}(x^{(c)})$) and the marginal probability mass function of $X^{(d)}$ (denoted by $f_{X^{(d)}}(x^{(d)})$).

- B3.** $f_{X^{(c)}|X^{(d)}}(x^{(c)})$ is continuous and bounded on the support of $x_i^{(c)}$.
- B4.** Assume that $\mathcal{X}_t = \mathcal{X}_{t(d-1)} \times \mathcal{X}_{td}$ where $\mathcal{X}_{t(d-1)}$ and \mathcal{X}_{td} are compact subsets with non-empty interiors of the supports of the first $d - 1$ components, and the d^{th} component of x_i , respectively. For each $x \in \mathcal{X}_t$, denote its first $d - 1$ components by $x_{(d-1)}$. \mathcal{X}_t will be assumed to have the following properties:

B4.1. \mathcal{X}_t is not contained in any proper linear subspace of \mathfrak{R}^d .

B4.2. $f_X(x) \geq \epsilon_0 > 0 \quad \forall x \in \mathcal{X}_t$, for some constant ϵ_0 .

Assumptions on the Median Residual Terms

D1. Let $u_{1i} = y_{1i} - m_1(x_i)$; in a neighborhood of 0, u_{1i} has a conditional (Lebesgue) density, denoted by $f_{u_1|X_i=x}(\cdot)$ which is continuous, and bounded away from 0 and infinity for all values of $x \in \mathcal{X}_t$. As a function of x , $f_{u_1|X_i=x}$ is Lipschitz continuous for all values of u_{1i} in a neighborhood of 0. Define u_{0i} analogously and assume it has analogous properties.

Furthermore, we require conditions on the smoothness of the median functions. Let

$$\begin{aligned} \tau_{q1}(x, \theta) &= \int 1[x \in \mathcal{X}]I[u \in \mathcal{X}]\tau(x)1[m_1(x) \geq m_0(u)]I[x'\beta(\theta) > u'\beta(\theta)]dF_X(u) \\ &+ \int 1[x \in \mathcal{X}]I[u \in \mathcal{X}]\tau_q(u)1[m_1(u) \geq m_0(x)]I[u'\beta(\theta) > x'\beta(\theta)]dF_X(u) \end{aligned}$$

and let

$$\tau_{q2}(x, \theta) = \int 1[x \in \mathcal{X}]I[u \in \mathcal{X}]I[x'\beta(\theta) > u'\beta(\theta)]dF_X(u)$$

let \mathcal{N} be a neighborhood of the $d-1$ dimensional vector θ_0 . Then we impose the following additional assumptions:

E1. For each x in the support of x_i , $\tau_{q1}(x, \cdot)$ is differentiable of order 2, with Lipschitz continuous second derivative on \mathcal{N} .

E2. $E[\nabla_2 \tau_{q1}(\cdot, \theta_0)]$ is negative definite

E3. For each x in the support of x_i , $\tau_{q2}(x, \cdot)$ is continuously differentiable on \mathcal{N} .

E4. $E[\|\nabla_1 \tau_{q2}(\cdot, \theta_0)\|^2] < \infty$

Finally, we impose conditions on the second stage smoothed indicator function and bandwidth:

SI1. The function $K(\cdot)$ is positive, strictly increasing, twice differentiable with bounded first and second derivatives, and satisfies the following:

SI1.1. $\lim_{x \rightarrow +\infty} K(x) = 1$, $\lim_{x \rightarrow -\infty} K(x) = 0$

$$\mathbf{SI1.2.} \int_{-\infty}^{\infty} K'(x)dx = 1$$

SI2. $h_n > 0$ and $h_n \rightarrow 0$.

The following theorem establishes that these additional assumptions, along with a stronger smoothness condition on the quantile function and further restrictions on the bandwidth sequence, are sufficient for root- n consistency and asymptotic normality of the proposed estimator:

Theorem 2.2 *Assume that $p > 3d_c/2$, and that in the first stage, k is set to $\text{int}(p)$ and the bandwidth sequences satisfy $\sqrt{n}\delta_n^p \rightarrow 0$, $\log n \sqrt{n^{-1}\delta_n^{-3d_c}} \rightarrow 0$ and*

$$\sqrt{nh_n^{-2}}(\delta_n^{2p} + \log n \cdot n^{-1}\delta_n^{-d_c}) \rightarrow 0$$

. Define

$$\begin{aligned} \delta(y_{1i}, y_{0i}, x_i) &= \tau(x_i) f_{u_{1i}|x_i}^{-1}(0) f'_{m_0}(m_1(x_i)) (1[y_{1i} \leq m_1(x_i)] - 0.5) \nabla_1 \tau_{q2}(x_i, \theta_0) \\ &+ \tau(x_i) f_{u_{0i}|x_i}^{-1}(0) f'_{m_1}(m_0(x_i)) (1[y_{0i} \leq m_0(x_i)] - 0.5) \nabla_1 \tau_{q2}(x_i, \theta_0) \end{aligned}$$

where $f'_{m_1}(\cdot), f'_{m_0}(\cdot)$ denote derivatives of density functions of the median functions; then under Assumptions A, B, Q, T, E, SI

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V_q^{-1} \Delta_q V_q^{-1}) \quad (17)$$

where $\Delta_q = E[\delta_q(v_i, x_i) \delta_q(v_i, x_i)']$ and $V_q = \frac{1}{2} E[\nabla_2 \tau_{q1}(x_i, \theta_0)]$.

The variance matrix can be consistently estimated by a numerical derivative form (see ? and ?) or a kernel method. We may also adopt resampling methods in case of a small sample.

3 Monte Carlo Simulation

In this section we investigate small sample properties of the proposed estimator by conducting a simple Monte Carlo simulation study. The base design is a censored transformation model with

two covariates as below:

$$\begin{aligned} T(v_i) &= x_{1i} + x_{2i}\beta_0 + \varepsilon_i \\ d_i &= 1(x_{1i} + x_{2i}\beta_0 + \varepsilon_i \leq c_i) \end{aligned}$$

where x_{1i} and x_{2i} are distributed as a chi-squared with one degree of freedom, and standard normal. The coefficient of x_{1i} is normalized as one. We consider the following two simulation models:

1. $T(v) = v$; ε_i follows the standard normal divided by 2; c_i follows a log-normal distribution.
2. $T(v) = \log v$; $\varepsilon_i = \exp(-0.5 \cdot (x_{1i}^2)) \eta_i$, where η_i follows the standard normal; $c_i = 0.5 \cdot (x_{1i}^2 + x_{2i}^2)$.

The first model has homoskedastic error terms and covariate independent censoring variables. In the second mode, we have both conditional heteroskedastic errors and covariate dependent censoring.

Tables 1 and 2 show the simulation results. For each design, we conducted 401 replications with sample size of 100, 200 and 400. In the first stage of QPRE, we estimated conditional median functions by applying the nonparametric method in ?. We fitted the model as a constant function and used the optimal bandwidth $c_1 n^{-6.5}$ for a constant c_1 . In the second stage, we estimated β_0 by maximizing the objective function of the QPRE in (16). The cumulative density function of the standard normal distribution was used for the kernel function. For the bandwidth of Kernel function, we again used the optimal bandwidth $c_2 (n(n-1))^{-1/5}$. Constants c_1 and c_2 was chosen by the rule of the thumb explained in ?. A grid search method was applied for an optimization algorithm. We evaluated the objective function on 501 equispaced points on the interval $[-5, 5]$. For comparison, the performance of the PRE and the MRC estimator was also examined. Again, the same grid search method was used to determine estimation values.

Overall, the QPRE estimator shows good finite sample performance in both simulation experiments. In the first model, the QPRE performs better than the PRE or the MRC when the sample

size is 100, but the RMSEs of the latter decrease faster. The MRC has the smallest RMSE when the sample size is 400. On the other hand, the QPRE shows better performance than the other two estimators at all sample sizes in the second model as we expected. Table 2 shows that in the log-linear regression model with heteroskedasticity the RMSEs of the MRC and the PRE decrease very slowly while the QPRE shows \sqrt{n} -convergence rate.

Consequently, the results from our simulation study indicate that the QPRE introduced in this paper performs well for heteroskedastic transformation models with covariate dependent random censoring. Thus, it can be applied to empirical settings with flexible restriction as we will see in the next section.

4 Empirical Illustration: UI Benefits and Unemployment Duration

In this section we analyze the effects of UI benefits on unemployment duration by using the censored transformation model. We use the data set from ? submitted to the U.S. Department of Labor. They constructed this data set by individual-level surveys, but the initial sample was generated from administrative records. So, key variables such as unemployment duration and UI benefits do not suffer from severe measurement error as in other household surveys. From the data set, we choose seven interesting variables whose descriptive statistics are summarized in Table 3.

We consider the following censored transformation model:

$$\begin{aligned} T(y_i) &= \min(x_i'\beta + \varepsilon_i, c_i) \\ d_i &= 1(x_i'\beta + \varepsilon_i \leq c_i) \end{aligned}$$

where y_i is the length of unemployment and x_i is covariates whose columns correspond to *age*, *race*, *highschool dropout*, *recall*, and *log weekly UI benefit level* respectively. We normalize the coefficient of age to be one. The variable c_i is a random censoring that may depend on regressors in an

arbitrary way, and the error term ε_i allows conditional heteroskedasticity. Finally, the variable d_i is an indicator of right censoring. Covariate dependent censoring and conditional heteroskedasticity is quite plausible for this model. If there were idiosyncratic shocks in the market, then it would be highly probable that workers with a similar characteristic might lose their jobs at the same time. For instance, we can think of an industry-specific shock that might induce high unemployment rate in a certain group of workers. Considering quite heterogeneous individuals in the data set, the error term is easily to be heteroskedastic.

Table 4 summarizes estimation results. For comparison purpose, we also provide estimates using Cox’s proportional hazard (PH), PRE and MRC. The PH model was estimated by the standard maximum likelihood method, and we changed its scale to compare them with results of rank estimators. The simulated annealing method was adopted for point estimates of rank estimators, and 95% confidence intervals denoted in square brackets were constructed by using the bootstrapping method. Besides the *high school dropout*, the QPRE gives the largest coefficient values in absolute term. The Scaled PH based on MLE gives the smallest interval estimate as expected. The confidence interval of QPRE is a little larger than the other two rank estimators. This might be the result of the nonparametric estimation procedure that the QPRE involves in the first stage. All estimators give the same sign for coefficients and they are significant with 95% confidence level. Thus, we may conclude that covariate dependent censoring and conditional heteroskedasticity are not so relevant for this specific data set. Finally, the QPRE shows that the additional UI benefits decrease the length of unemployment duration in this data set, which was also found in ? and ?. The result in this paper indicates that any form of conditional heteroskedasticity is not the reason inducing minus sign of the coefficient.

5 Conclusion

In this paper we consider a censored transformation model with conditional heteroskedasticity. To estimate the model, we propose the quantile partial rank estimation procedure. It is shown that

the proposed estimator satisfies \sqrt{n} -consistency and asymptotic normality. We also investigate its finite sample properties through a Monte Carlo simulation study and show that it works appropriately in small samples. For empirical illustration, we estimate the effect of UI benefits on unemployment duration.

We conclude this paper by suggesting areas of future research. First, it would be useful to formally construct estimator for the transformation function $T(\cdot)$. It could be achieved by modifying the rank estimation procedure in ?. Second, we can think of an extended model with functional coefficients, i.e. linear coefficients are nonparametric functions of additional covariates. These nonparametric functional coefficients can be estimated by localizing the QPRE similarly to ?.

Appendix

5.1 Proof of Theorem 2.1

First we show the necessity part. Suppose that $x_i \in \mathcal{X}$. Then, we have

$$\begin{aligned}\Pr(T(y_{1i}) - x'_i\beta_0 \leq 0|x_i) &= \Pr(T(y_{1i}) - x'_i\beta_0 \leq 0, d_i = 1|x_i) + \Pr(T(y_{1i}) - x'_i\beta_0 \leq 0, d_i = 0|x_i) \\ &= \Pr(\varepsilon_i \leq 0, \varepsilon_i \leq c_i - x'_i\beta_0|x_i) \\ &= \Pr(\varepsilon_i \leq 0|x_i) \\ &= \frac{1}{2}\end{aligned}$$

where the third equality follows from the hypothesis that $x_i \in \mathcal{X}$, i.e. $\Pr(c_i - x'_i\beta_0 \geq 0|x_i) = 1$.

Now we turn our attention to $m_0(x)$ keeping the hypothesis.

$$\begin{aligned}\Pr(T(y_{0i}) - x'_i\beta_0 \leq 0|x_i) &= \Pr(T(y_{0i}) - x'_i\beta_0 \leq 0, d_i = 1|x_i) + \Pr(T(y_{0i}) - x'_i\beta_0 \leq 0, d_i = 0|x_i) \\ &= \Pr(\varepsilon_i \leq 0, \varepsilon_i \leq c_i - x'_i\beta_0|x_i) + \Pr(x'_i\beta_0 \geq c_i, \varepsilon_i \geq c_i - x'_i\beta_0|x_i) \\ &= \Pr(\varepsilon_i \leq 0|x_i) \\ &= \frac{1}{2}\end{aligned}$$

where the third equality again follows from the hypothesis. Therefore, we can conclude that

$$x_i \in \mathcal{X} \Rightarrow m_1(x_i) = m_0(x_i) = x'_i\beta_0 = \text{med}(T(y_i)|x_i).$$

Next we look at the sufficiency part. Suppose that $m_1(x_i) = m_0(x_i) = x'_i\beta_0 = \text{med}(T(y_i)|x_i)$.

Then, we have

$$\begin{aligned}
\Pr(\varepsilon_i \geq 0|x_i) &= \Pr(T(y_{0i}) - x'_i\beta_0 \geq 0|x_i) \\
&= \Pr(T(y_{0i}) - x'_i\beta_0 \geq 0, d_i = 1|x_i) + \Pr(T(y_{0i}) - x'_i\beta_0 \geq 0, d_i = 0|x_i) \\
&= \Pr(0 \leq \varepsilon_i \leq c_i - x'_i\beta_0|x_i) + \Pr(0 \leq c_i - x'_i\beta_0 \leq \varepsilon_i|x_i) \\
&= \Pr(\varepsilon_i \geq 0, c_i - x'_i\beta_0 \geq 0|x_i) \\
&= \Pr(\varepsilon_i \geq 0|x) \cdot \Pr(x'_i\beta_0 \leq c_i|x_i)
\end{aligned}$$

where the last equality follows from the hypothesis that $\varepsilon_i \perp c_i|x_i$ which is the maintained assumption. From the equation $\Pr(\varepsilon_i \geq 0|x_i) = \Pr(\varepsilon_i \geq 0|x) \cdot \Pr(x'_i\beta_0 \leq c_i|x_i)$, we can conclude that $\Pr(x'_i\beta_0 \leq c_i|x_i) = 1$. \blacksquare

5.2 Proof of Theorem 2.2

The asymptotic properties follow from arguments that are very similar to those used in ?, so we only provide a sketch of the steps involved. First we expand the kernel function of the estimated median functions around the kernel of the true median functions in (16), yielding the sum of the three components

$$\Gamma_n(\beta) \equiv \frac{1}{n(n-1)} \sum_{i \neq j} K_{h_n}(m_{1i} - m_{0j}) 1[x'_i\beta \geq x'_j\beta] \quad (18)$$

$$H_n(\beta) \equiv \frac{1}{n(n-1)} \sum_{i \neq j} K'_{h_n}(m_{1i} - m_{0j}) h_n^{-1}((\hat{m}_{1i} - m_{1i}) - (\hat{m}_{0j} - m_{0j})) 1[x'_i\beta \geq x'_j\beta] \quad (19)$$

$$R_n(\beta) \equiv \frac{1}{n(n-1)} \sum_{i \neq j} K''_{h_n}(m_{1i}^* - m_{0j}^*) h_n^{-2}(\hat{m}_{1i} - m_{1i} - \hat{m}_{0j} + m_{0j})^2 1[x'_i\beta \geq x'_j\beta] \quad (20)$$

where we have adopted the shorthand notation \hat{m}_{1i}, m_{1i} denotes $\hat{m}_1^{\delta_n, p}(x_i), m_1(x_i)$ respectively, and $*$ denotes intermediate values.

First we deal with (19). It follows by uniform rates of convergence for median function estimators over compact sets, (see, e.g. ?) where these rates depend on p, δ_n , Assumptions SI1, SI2, and the

rates imposed on δ_n, h_n stated in the theorem $R_n(\beta)$ is $o_p(1/n)$ uniformly over β within an $O_p(1/\sqrt{n})$ neighborhood of β_0 .

Turning attention to $H_n(\beta)$, with the properties of $K(\cdot)$ in Assumption SI1, we apply the arguments in Lemma A.4 in ? that uniformly over β within $o_p(1)$ neighborhoods of β_0 , we have

$$H_n(\beta) = (\beta - \beta_0)' \frac{1}{n} \sum_{i=1}^n \delta(y_{1i}, y_{0i}, x_i) + o_p(1/n) \quad (21)$$

Finally, with regard to $\Gamma_n(\beta)$, we have by the properties of $K(\cdot), h_n$ in Assumption SI1, SI2, using identical arguments as in Lemma A.3 in ?, that uniformly over β within $o_p(1)$ neighborhoods of β_0 , we have

$$\Gamma_n(\beta) = \frac{1}{2} (\beta - \beta_0)' V_q (\beta - \beta_0) + o_p(1/n) \quad (22)$$

Combining these three results, the limiting distribution of the estimator follows by applying Lemma A.2 in ?. ■

Table 1: Simulations Results: Linear Model with Homoskedasticity and CI censoring

	Mean Bias	Median Bias	RMSE	MAD
<i>100 obs.</i>				
QPRE	-0.1555	-0.1600	0.2220	0.1864
PRE	0.0495	0.0200	0.3280	0.2416
MRC	0.0185	0.0000	0.2505	0.1935
<hr/>				
<i>200 obs.</i>				
QPRE	-0.14951	-0.1600	0.1855	0.1597
PRE	-0.0043	-0.0200	0.1954	0.1526
MRC	-0.0097	-0.0200	0.1538	0.1217
<hr/>				
<i>400 obs.</i>				
QPRE	-0.1304	-0.1400	0.1560	0.1350
PRE	0.0086	0.0000	0.1378	0.1065
MRC	0.0047	0.0000	0.1077	0.0865

Table 2: Simulations Results: Log-linear Model with Heteroskedasticity and CD censoring

	Mean Bias	Median Bias	RMSE	MAD
<i>100 obs.</i>				
QPRE	-0.1104	-0.1200	0.2375	0.1876
PRE	0.4263	0.3800	0.4814	0.4271
MRC	0.4107	0.4000	0.4435	0.4107
<hr/>				
<i>200 obs.</i>				
QPRE	-0.0999	-0.1000	0.1735	0.1406
PRE	0.3980	0.3800	0.4241	0.3980
MRC	0.3867	0.3800	0.4007	0.3867
<hr/>				
<i>400 obs.</i>				
QPRE	-0.0659	-0.0600	0.1248	0.0999
PRE	0.3817	0.3800	0.3933	0.3817
MRC	0.3803	0.3800	0.3878	0.3803

Table 3: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
Unemployment duration	48.90	45.05	0.14	185.71
Censor (censored=0)	0.79	0.41	0	1
Age	39.75	11.90	17	83
Race (white=1)	0.66	0.47	0	1
High school dropout	0.17	0.38	0	1
Recall	0.44	0.50	0	1
Log UI benefit level	5.27	0.45	2.71	6.02

Data Source: ?, n=2736.

Table 4: Estimation Results of Unemployment Duration Data

	PH	Scaled PH	MRC	PRE	QPRE
<i>Age</i>	0.016	1	1	1	1
	[0.012, 0.019]	[0.77, 1.23]	-	-	-
Race (White=1)	-0.174	-11.03	-13.32	-12.51	-16.61
	[-0.267, -0.081]	[-16.96, -5.11]	[-18.02, 5.00]	[-18.67, -5.30]	[-25.93, -8.86]
High school dropout	0.289	18.32	11.69	17.88	18.12
	[0.163, 0.415]	[10.32, 26.32]	[2.13, 23.72]	[2.96, 25.34]	[2.44, 34.18]
Recall	-0.271	-17.16	-31.51	-33.86	-34.91
	[-0.356, -0.185]	[-22.58, -11.73]	[-36.42, -19.00]	[-36.02, -18.71]	[-46.52, -21.40]
Log weekly UI benefit level	-0.299	-18.94	-21.05	-19.48	-23.82
	[-0.398, -0.200]	[-25.23, -12.66]	[-30.03, -14.44]	[-27.96, -14.19]	[-39.45, -12.08]