Partial Rank Estimation of Transformation Models with General forms of Censoring*

Shakeeb Khan[a] and **Elie Tamer**[b†]

[a] *Department of Economics, University of Rochester, Rochester NY 14627, USA*

[b] *Department of Economics, Northwestern University, 2001 Sheridan Rd, Evanston IL 60208, USA.*

Submitted Version: February 2004

Current Version: March 2005

Abstract

In this paper we propose estimators for the regression coefficients in censored duration models which are distribution free, impose no parametric specification of the baseline hazard function, and can accommodate general forms of censoring. The estimators are shown to have desirable asymptotic properties and Monte Carlo simulations demonstrate good finite sample performance. Among the data features the new estimators can accommodate are covariate dependent censoring, double censoring, heteroskedasticity, and fixed (individual or group specific) effects.

JEL Classification: C13; C31

Keywords: duration analysis, rank correlation, median regression, panel data.

1 Introduction

This paper considers estimation of regression coefficients in censored duration models. Duration models have seen widespread use in empirical work in various areas of economics. This is because many time-to-event variables are of interest to researchers conducting empirical studies in labor economics, development economics, public finance and finance. For example, the time-to-event of interest may be the length of an unemployment spell, the time between purchases of a particular good, time intervals between child births, and insurance claim durations, to name a few. (Van den Berg(2001) surveys the many applications of duration models.)

*We thank the co-editor Takeshi Amemiya and three referees for extremely useful suggestions (the report of one very diligent referee was particularly helpful in the revision). We also thank J. Abrevaya, S. Chen, H. Hong, B. Honoré, J. Powell as well as seminar participants at various institutions and the European and North American Summer Meetings of the Econometric Society for many helpful comments. Both authors also gratefully acknowledge the National Science Foundation for financial support.

†Corresponding Author: Department of Economics, Northwestern University, 2001 Sheridan Rd, Evanston IL 60208; e-mail: tamer@northwestern.edu.

Since the seminal work in Cox(1972,1975), the most widely used models in duration analysis are the proportional hazards model, and its extension, the mixed proportional hazards model, introduced in Lancaster(1979). These models can be represented as monotonic transformation models, where an unknown, monotonic transformation of the dependent variable is a linear function of observed covariates plus an unobserved error term, subject to restrictions that maintain the (mixed) proportional hazards assumption. Relaxing these restrictions gives the Generalized Accelerated Failure Time (GAFT) model introduced in Ridder(1990).

The GAFT model in its most basic form is usually expressed as

$$T(y_i) = x_i'\beta_0 + \epsilon_i \qquad i = 1, 2, ...n \tag{1.1}$$

where $(y_i, x_i')'$ is a $(k+1)$ dimensional observed random vector, with y_i denoting the dependent variable, usually a time to event, and x_i denoting a vector of observed covariates. The random variable ϵ_i is unobserved and independent of x_i with an unknown distribution. The function $T(\cdot)$ is assumed to be strictly monotonic, but otherwise unspecified. The k-dimensional vector β_0 is unknown, and is often the object of interest to be estimated from a random sample of n observations.

Duration data is often subject to right censoring for a variety of reasons that are usually a consequence of the empirical researcher's observation or data collection plan. For example, unemployment spell-length may be censored because the agent is lost from the sample, or to control data collection costs, unemployed agents are only followed for short period of time. If they are still unemployed at the end of this period, their spell length is censored. This paper considers right censoring as the base case and shows how our approach can be extended to cover double censoring (and naturally left censoring).

When the data is subject to censoring the variable y_i is no longer always observed. Instead one observes the pair (v_i, d_i) where v_i is a scalar random variable, and d_i is a binary random variable. We express the *right censored* transformation model as[1]:

$$T(v_i) = \min(x_i'\beta_0 + \epsilon_i, c_i) \tag{1.2}$$
$$d_i = I[x_i'\beta_0 + \epsilon_i \leq c_i] \tag{1.3}$$

where $I[\cdot]$ denotes the indicator function, and c_i denotes the random censoring variable. So, here $v_i = y_i$ for uncensored observations, and $v_i = c_i$ otherwise. We note the censoring variable need not always be observed, as would occur in a *competing risks* type setting (see, e.g. Heckman and Honoré(1990)).

The primary aim of this paper is to provide an estimator of β_0 in the above model with few restrictions on c_i. Specifically, we wish to allow for the presence of *covariate dependent* censoring,

[1]We can also express the censored model in the latent dependent variable framework. Letting $y_i^* = T^{-1}(x_i'\beta_0 + \epsilon_i)$ and $\tilde{c}_i = T^{-1}(c_i)$, one observes the covariates and the pair (y_i, d_i) where $y_i = \min(y_i^*, \tilde{c}_i)$ and d_i can now be expressed as $I[y_i^* \leq \tilde{c}_i]$.

i.e., in the case where c_i can be arbitrarily correlated with x_i. This would be in line with the form of censoring allowed for in the Partial Maximum Likelihood Estimator (PMLE) introduced in Cox(1972,1975), and several other estimators (to be mentioned below) in the duration literature. Outside the proportional hazards framework, covariate dependent censoring also arises in the biostatistics literature on competing risks, and survival analysis (even for randomized clinical trials - see Chen, Jin and Ying(2002)).

We motivate the construction of such an estimator by two ways- first by illustrating the relevance of the censored transformation model in empirical settings, and second, by showing that the problem of (distribution free) estimation of β_0 in a censored transformation model has not been completely solved.

Turning first to relevance of the model in empirical work we note the censored transformation model has become increasingly popular in the applied econometrics literature. This is because economic theory rarely provides guidelines on how to specify functional form relationships among variables while (1.1) can accommodate many functional relationships used in practice such as linear, log-linear, or the parametric transformation in Box-Cox models, without suffering from the dimensionality problems encountered when adopting a fully nonparametric approach.

Next, we explain why the problem of estimating β_0 has not been completely solved despite the extensive literature (both in econometrics and in biostatistics) and much recent research progress. We note that with $T(\cdot)$ known there exists many distribution free estimators for β_0- examples include Buckley James(1979), Koul, Susarla and Van Rysin(1981), Tsiatis(1990), Ying, Jung and Wei(1995), Yang(1999), Honoré, Khan and Powell(2002) and Portnoy(2003), some of which allow for the distribution of the censoring variable to depend on the covariates. Bijwaard (2001) imposes parametric restrictions on $T(\cdot)$. For $T(\cdot)$ unknown except for a strict monotonicity assumption, we can divide the existing literature into two groups. One group allows for covariate dependent censoring but require a known distribution of ϵ_i, and can be inconsistent if this distribution is misspecified. See for example Cox(1975)'s partial maximum likelihood estimator (PMLE), Cuzick(1988), and more recently Chen, Jin and Ying (2002). Cheng, Wei and Ying(1995), Fine, Ying and Wei(1998), Cai, Wei and Wilcox(2000) are more restrictive in the sense that in addition to parametrically specifying the distribution of ϵ_i, they do not allow the censoring variable to depend on x_i. The other group, of which examples include the important single-index estimators in Han(1987) and Cavanagh and Sherman(1998), do not impose distributional assumptions on ϵ_i, but for consistency requires that the censoring variable be independent of the covariates. However, as mentioned above, this assumption is often too restrictive. Attempting to remedy this problem using conditional Kaplan-Meier methods would require smoothing parameters, trimming procedures, and tail behavior restrictions. In summary, the literature lacks an estimator for β_0 that is distribution free and permits covariate dependent censoring[2]. The estimator we provide in this

[2]If the censoring distribution depends on the regressors through the index $x_i'\beta_0$, we note some single index estimators may be applied, though we consider this too restrictive of a condition. We also note that general covariate censoring is permitted in Gorgens and Horowitz(1998) in their estimator of the link function $T(\cdot)$. They do not provide an estimator of β_0 assuming it is known, and only suggest estimation by an existing single index estimator.

paper is a partial rank estimator that allows for general forms of censoring, including left, right, and double sided censoring, where the censoring can possibly depend on the regressors. This leads to new identification strategies that are emphasized in the text (and the proofs).

The rest of the paper is organized as follows. In the next section, we introduce an estimator for β_0 in transformation models with covariate dependent censoring that aims to fill the gap in the literature, and provide conditions for parameter identification. We then show that the estimator is consistent and derive its asymptotic distribution. Sections 3-5 extend our estimator to cases of doubly censored, heteroskedastic, and fixed effect panel data. Section 6 explores the finite sample properties of the new estimators by means of a small scale simulation study. Section 7 concludes by summarizing results and discussing areas for future research. The proofs of the main results are collected in an appendix.

2 Estimation Procedure

Our estimator for β_0 in the censored GAFT model is motivated by existing rank estimators for β_0 in uncensored transformation models, specifically Han (1987)'s maximum rank correlation (MRC) estimator[3]. This estimator, like other estimators in the single index literature (e.g. Powell, Stock and Stoker (1989), Ichimura(1993), Cavanagh and Sherman(1997)), is inconsistent when the censoring variable depends on the covariates in an arbitrary way.

Before introducing our distribution free estimator that accommodates covariate dependent censoring, we define the vector $\mathbf{y}_i = (v_i, d_i)'$. To construct a rank regression estimator analogous to Han (1987)'s, we wish to construct a function:

$$f_{ij} \equiv f(\mathbf{y}_i, \mathbf{y}_j)$$

which satisfies the property

$$E[I[f_{ij} \geq 0]|x_i, x_j] \geq E[I[f_{ji} \geq 0]|x_i, x_j] \text{ iff } x_i'\beta_0 \geq x_j'\beta_0 \tag{2.1}$$

For the uncensored transformation model, Han(1987) sets $f_{ij} = y_i - y_j$. For the problem at hand with covariate dependent censoring, we propose an alternative form for f_{ij} that satisfies (2.1). First define the random variables

$$\begin{aligned} y_{0i} &= v_i \\ y_{1i} &= d_i v_i + (1 - d_i) \cdot (+\infty) \end{aligned} \tag{2.2}$$

where by definition we have

$$y_{0i} \leq y_i^* \leq y_{1i} \tag{2.3}$$

As mentioned, this will not yield consistent estimates if the censoring depends arbitrarily on x_i.

[3] A similar rank estimator was introduced in Cavanagh and Sherman(1998). Their Monotone Rank Estimator (MRE) is computationally simpler than the MRC, but also does not allow for covariate dependent censoring.

where $y_i^* \equiv T^{-1}(x_i'\beta_0 + \epsilon_i)$. We can then define f_{ij} and consequently $I[f_{ij} \geq 0]$ as

$$\begin{aligned} f_{ij} &= y_{1i} - y_{0j} && (2.4) \\ I[f_{ij} \geq 0] &= I[y_{1i} - y_{0j} \geq 0] = (1 - d_i) + d_i I[v_i \geq v_j] && (2.5) \end{aligned}$$

Our choice of f_{ij} is motivated by the following inequalities which follow from (2.3)

$$T(y_{0i}) \leq x_i'\beta_0 + \epsilon_i \leq T(y_{1i})$$

Heuristically, by monotonicity of $T(.)$, we should have

$$x_i'\beta \geq x_j'\beta \;\; \Rightarrow \;\; P(y_{1i} \geq y_{0j}) \;\; \geq \frac{1}{2} \tag{2.6}$$

We first show that (2.1) holds for the censored transformation model. Our result is based on the following assumptions.

I1 Letting S_X denote the support of x_i, and let $\mathcal{X}_{uc}$ denote the set

$$\mathcal{X}_{uc} = \{x \in S_X : P(d_i = 1 | x_i = x) > 0\}$$

Then $\mathcal{X}_{uc}$ has positive measure.

I2 The random variable ϵ_i is distributed independently of the random vector $(c_i, x_i')'$.

I3 The first component of x_i has everywhere positive Lebesgue density, conditional on the other components.

Condition **I1** requires that the probability of censoring is not equal to one for all x. The independence assumption **I2** requires that ϵ be independent of both x and c. This assumption is a natural starting point for examining the identification of β in this class of models. The independence between ϵ and x is similar to independence assumption in the MRC estimator. In section 4, we relax this assumption where we allow for (conditional) heteroskedasticity. This will come at the expense of stronger (sufficient) point identification condition[4]. Finally, the last assumption **I3** provides sufficient condition for point identification. This support condition is a widely used identification condition in semiparametric econometric models such as the MRC and the maximum score estimator in Manski(1975,1985). The main result of this section is stated in the next lemma whose proof is in the appendix.

Lemma 2.1 *Under Assumptions I1-I3, (2.1) holds.*

[4]On the other hand, it is not generally possible to relax the conditional independence of c and ϵ without additional assumption like exclusion restrictions (or functional form assumptions). This independence assumption is natural and has been widely adotpted in the duration literature.

It is this result which motivates our estimator. Before describing it in detail, we note that the object of interest β_0 is only identified up to scale as the function $T(\cdot)$ is unknown. Following convention, we set the first component of the vector β_0 to 1, express $\beta_0 = (1, \theta_0')'$ and consider estimation of θ_0. Adopting standard notation, for any $\theta \in \Theta$, we let β denote $(1, \theta')'$.

Our censoring robust rank estimator, which we refer to hereafter as the *partial rank estimator* (PRE), is of the form:

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta\in\Theta} \frac{1}{n(n-1)} \sum_{i\neq j} I[f_{ij} \geq 0] I[x_i'\beta \geq x_j'\beta] & (2.7) \\ &= \arg\max_{\theta\in\Theta} \frac{1}{n(n-1)} \sum_{i\neq j} (d_i I[v_i \geq v_j] + (1-d_i)) I[x_i'\beta \geq x_j'\beta] & (2.8) \end{aligned}$$

where Θ denotes the parameter space.

Remark 2.1 *We note the following particular features of the estimation procedure:*

- *The above estimator is numerically equivalent to maximizing the objective function:*

$$\frac{1}{n(n-1)} \sum_{i\neq j} I[y_{0i} \geq y_{1j}] I[x_i'\beta \geq x_j'\beta] = \frac{1}{n(n-1)} \sum_{i\neq j} d_i I[v_i \leq v_j] I[x_i'\beta \leq x_j'\beta] \quad (2.9)$$

  Expressed this way, a loose analogy can be drawn to the partial maximum likelihood estimator (PMLE) introduced in Cox(1972,1975). In the PMLE only uncensored observations enter the likelihood function, and for a given such observation, all the observations in its risk set (i.e. observations whose spell length is known not to be less that i's spell length) are used in evaluating the likelihood. Analogously, for the PRE only uncensored observations enter the rank correlation function, and for a given uncensored observation, all the observations in its risk set are used to determine its rank.

- *The PRE is numerically equivalent to the MRC in the absence of censoring, as the censoring indicators are always 1. PRE is also numerically equivalent to the MRC, and hence is consistent in the case of fixed censoring, (e.g. $c_i \equiv 0$), which arises often in economic models. The estimator can therefore be applied without change to fixed and randomly censored data. This is in contrast to procedures dividing by Kaplan-Meier estimators of the censoring variable's survivor function (see, e.g. Koul, Susarla and Van Rysin(1980)), which cannot in general be applied in the fixed censoring case.*

- *Though the PRE was defined in terms of the right censored model, it can be easily modified for the left censored model. The form of this modification is illustrated in the next section where we explore the doubly censored model.*

We first establish consistency of the PRE. For this we impose the additional conditions

I4 The vector $\mathbf{z}_i = (d_i, v_i, x_i),\ i = 1, 2, ..n$ are i.i.d.

I5 Θ is a compact subset of $\mathbf{R}^{k-1}$.

I6 S_X is not contained in any proper linear subspace of R^k.

The following theorem, whose proof is left to the appendix, establishes the consistency of the PRE.

Theorem 2.1 *Under Assumptions* **I1**-**I6**,

$$\hat{\theta} \xrightarrow{p} \theta_0$$

We now establish the limiting distribution theory of the PRE. The arguments are analogous to those used in Sherman(1993) for establishing the asymptotic distribution of the MRC. Our results are based on a set of similar assumptions and we deliberately choose notation to match his as closely as possible.

Recalling that $\mathbf{z}_i$ denotes the vector $(d_i, v_i, x_i')'$, we define

$$\begin{aligned}\tau(\mathbf{z_i}, \theta) &= E[(d_i I[v_i \geq v] + (1 - d_i))I[x'\beta \geq x_i'\beta]] \\ &+ E[(dI[v \geq v_i] + (1 - d))]I[x_i'\beta \geq x'\beta]]\end{aligned}$$

Finally, we let $\mathcal{N}$ denote a neighborhood of θ_0.

A1 θ_0 lies in the interior of Θ, a compact subset of R^{k-1}.

A2 For each $\mathbf{z}$, the function $\tau(\mathbf{z}, \cdot)$ is twice differentiable in a neighborhood of θ_0. Furthermore, the vector of second derivatives of $\tau(\mathbf{z}, \cdot)$ satisfies the following Lipschitz condition:

$$\|\nabla_2 \tau(\mathbf{z}, \theta) - \nabla_2 \tau(\mathbf{z}, \theta_0)\| \leq M(\mathbf{z})\|\theta - \theta_0\|$$

where ∇_2 denotes the second derivative operator and $M(\cdot)$ denotes an integrable function of $\mathbf{z}$.

A3 $E[\|\nabla_1 \tau(\mathbf{z}_i, \theta_0)\|^2]$ and $E[\|\nabla_2 \tau(\mathbf{z}_i, \theta_0)\|]$ are finite, where ∇_1 denotes the first derivative operator.

A4 $E[\nabla_2 \tau(\mathbf{z}_i, \theta_0)]$ is non-singular.

We now state the main theorem, characterizing the asymptotic distribution of the PRE; its proof is left to the appendix.

Theorem 2.2 *Under Assumptions I1-I5, A1-A4,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V^{-1}\Delta V^{-1}) \tag{2.10}$$

where $V = E[\nabla_2 \tau(\mathbf{z}_i, \theta_0)]/2$ *and* $\Delta = E[\nabla_1 \tau(\mathbf{z}_i, \theta_0)\nabla_1 \tau(\mathbf{z}_i, \theta_0)']$.

We conclude this section with a brief discussion on conducting inference with the PRE. The asymptotic variance matrix can be estimated in a similar fashion to the estimator in Sherman(1993). As is the case with that estimator, the selection of smoothing parameters will be required.

Unfortunately, it has not been formally established that the bootstrap is asymptotically valid in this setting, or else inference could be conducted without the selection of smoothing parameters. However, other sampling procedures are possible. For one, a recent "wild" sampling procedure introduced in Jin, Ying, and Wei (2001) appears likely to be applicable (after appropriate modifications) to the problem at hand.

Separately, the PRE can be used to construct model specification tests by comparing its value to those of existing estimators. For example, the PRE may be compared to the MRC to test for the presence of covariate dependent censoring. We can compare the PRE to the relative coefficients obtained from Cox's partial likelihood estimator (PMLE) or those found using the estimator in Ying, Jung and Wei (2002) to test for the presence of unobserved heterogeneity, or more generally, to test for particular distributions of ϵ_i. Also, we can compare the PRE to relative coefficients obtained from the Tsiatsis(1990) and/or Ying, Jung and Wei (1995) estimators, to test for particular functional forms of the transformation.

3 Extension I: Doubly Censored Data

Many data sets in both biostatistics and economics are subject to double (i.e. left and right) random censoring. Examples are when the dependent variable is the age of the individual at which a particular event(e.g. cancerous tumor, change in employment status) occurs, and individuals are regularly and frequently surveyed or tested for an interval of time. If the occurrence of the event is detected on the first survey/test, the dependent variable (age) is left censored, as the recorded value is greater than the actual (latent) value. If no such events have occurred by the last survey/test, the dependent variable is right censored, as the recorded value is exceeded by the actual value.

In the monotonic transformation framework, the doubly censored regression model can be expressed as follows. (1.1) still holds, but the econometrician does not always observe the dependent variable $y_i \equiv T^{-1}(x_i'\beta_0 + \epsilon_i)$. Instead one observes the doubly censored sample, which we can express as the pair (v_i, d_i) where

$$d_i = I[c_{1i} < x_i'\beta_0 + \epsilon_i \leq c_{2i}] + 2 \cdot I[x_i'\beta_0 + \epsilon_i \leq c_{1i}] + 3 \cdot I[c_{2i} < x_i'\beta_0 + \epsilon_i]$$

$$v_i = I[d_i = 1] \cdot (x_i'\beta_0 + \epsilon_i) + I[d_i = 2]c_{1i} + I[d_i = 3]c_{2i}$$

where $I[\cdot]$ denotes the usual indicator function, c_{1i}, c_{2i} denote left and right censoring variables, whose distributions may depend on the covariates x_i and who satisfy $P(c_{1i} < c_{2i}) = 1$.

For a regression model with double censoring, estimators have been proposed by Zhang and Li(1996), Ren and Gu(1997) to name a few. Both of these require a linear regression specification and the censoring variables to be independent of the covariates. With $T(\cdot)$ unknown, one can again perform MRC using v_i as the dependent variable if x_i is independent of (c_{1i}, c_{2i}). However in the doubly censored case the efficiency loss can be very severe due to ignoring the value of d_i.

To estimate β_0 in the general model with $T(\cdot)$ and the distribution of ϵ_i unknown, as well as covariate dependent censoring, we first define y_{1i}, y_{0i} as

$$\begin{aligned} y_{1i} &= I[d_i < 3]v_i + I[d_i = 3] \cdot +\infty && (3.1) \\ y_{0i} &= I[d_i \neq 2]v_i + I[d_i = 2] \cdot -\infty && (3.2) \end{aligned}$$

and accordingly we may define $f_{ij}, I[f_{ij} \geq 0]$ as:

$$\begin{aligned} f_{ij} &= y_{1i} - y_{0j} \\ I[f_{ij} \geq 0] &= I[d_i = 3] + I[d_j = 2] - (I[d_i = 3] * I[d_j = 2]) \\ &+ (I[d_i = 1] + I[d_i = 2]) * (I[d_j = 1] + I[d_j = 3])I[v_i \geq v_j] \end{aligned}$$

Letting d_{1i}, d_{2i}, d_{3i} denote $I[d_i = 1], I[d_i = 2], I[d_i = 3]$, respectively, we can express the PRE for doubly censored data as:

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} I[y_{1i} \geq y_{0j}] I[x_i'\beta \geq x_j'\beta] && (3.3) \\ &= \arg\max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} ((d_{1i} + d_{2i}) \cdot (d_{1j} + d_{3j}) I[v_i \geq v_j] \\ &+ (d_{3i} + d_{2j} - d_{3i}d_{2j})) I[x_i'\beta \geq x_j'\beta] && (3.4) \end{aligned}$$

which, as before, is numerically equivalent to

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} I[y_{0i} \geq y_{1j}] I[x_i'\beta \geq x_j'\beta] && (3.5) \\ &= \arg\max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} (1 - I[d_i = 2])(1 - I[d_j = 3]) I[v_i \geq v_j] I[x_i'\beta \geq x_j'\beta] && (3.6) \end{aligned}$$

We first establish the analogous identification result for the PRE in the doubly censored case. The identification result is again based on Assumptions I1-I3, where it is now understood that the event $d_i = 1$ is defined for doubly censored data. The proof is left to the appendix.

Lemma 3.1 *Under assumptions analogous to I1-I3, if either the covariates are independent of the regressors or we have the condition:*

$$P(c_{1i} > c_{2j} | x_i, x_j) = P(c_{1j} > c_{2i} | x_i, x_j) = 0 \quad x_i, x_j \text{ a.s.} \tag{3.7}$$

then we have

$$P(y_{0i} \geq y_{1j} | x_i, x_j) \geq P(y_{0j} \geq y_{1i} | x_i, x_j) \quad \text{iff } x_i'\beta_0 \geq x_j'\beta_0 \tag{3.8}$$

In the above lemma, we impose the additional condition (3.7) which is a sufficient condition for point identification. Consistency of the estimator follows by including assumptions analogous to **I4-I6**. In the appendix, we also provide the asymptotic distribution of (3.6) above.

4 Extension II: Heteroskedastic Data

One of the assumptions that we used above the independence between the disturbance term ϵ_i and the covariates x_i. This assumption may be overly restrictive; for example, it rules out any form of conditional heteroskedasticity which is important in some data sets. In this section we relax the independence assumption by assuming only one of the quantiles of ϵ_i, say the median, is independent of the covariates. Khan(2000) proposed a two step rank estimator for a heteroskedastic transformation model, but did not allow for random censoring. In contrast, Honoré, Khan and Powell(2002) and Portnoy(2004) allow for unknown heteroskedasticity and random censoring, but require the transformation function to be known. For point identification in models with random covariate dependent censoring, heteroskedasticity and an unknown transformation function, we assume that the random variables c_i, ϵ_i are statistically independent[5]given x_i.

Next, we provide identification conditions for the univariate censoring case. Similar arguments can be used to attain point identification results for the double censoring case. The results in the next lemma, whose proof is in the appendix, provide sufficient conditions for regular point identification.

Lemma 4.1 *Define the set $\mathcal{X}$ such that*

$$\mathcal{X} = \{x : \Pr(c - x\beta \geq 0|x) = 1\}$$

Assume further that $\Pr_x(\mathcal{X}) > 0$. Moreover, the random variable c is such that $\epsilon \perp c|x$. Finally, define the random variables $y_{0i} = v_i$ and $y_{1i} = d_i v_i + (1 - d_i) \cdot +\infty$. Then we have that

$$Med(T(y_0)|x) = Med(T(y)|x) = Med(T(y_1)|x) = x\beta$$

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

The above identification result[6], along with the invariance of medians, suggests an (infeasible) rank estimator based on the conditional medians of y_{0i} and y_{1i}. Letting $m_0(x_i), m_1(x_i)$ denote these conditional median functions, we would estimate β_0 by maximizing the function

$$Q(\beta) = \frac{1}{n(n-1)} \sum_{i \neq j} I[m_1(x_i) \geq m_0(x_j)] I[x_i'\beta \geq x_j'\beta] \tag{4.1}$$

[5]Without this assumption, the identified features of the model is a set of parameters (as opposed to a unique parameter.).

[6]The lemma above provides sufficient conditions for *regular* point identification or β_0, i.e., conditions that allow for consistent estimation of β_0 at the parametric rate. Notice that, the sufficient condition rules out the case where x and c are jointly normal. One can easily show that this is a case of *identification at infinity* where the parameter β_0 is point identified but can only be estimated at rates slower than $\sqrt{n}$.

To construct a feasible estimation procedure, we replace the unknown median functions in the above estimator with their nonparametric estimators. To construct these estimators, we adopt the local polynomial approach introduced in Chaudhuri(1991). For a detailed description of the estimator, see Chaudhuri(1991). 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:

$$\hat{\beta}_{ht} = \arg\max_{\beta\in\mathcal{B}} \frac{1}{n(n-1)} \sum_{i\neq j} K_{h_n}(\hat{m}_1^{\delta_n,p}(x_i) - \hat{m}_0^{\delta_n,p}(x_j))I[x_i'\beta \geq x_j'\beta] \tag{4.2}$$

where $K_{h_n}(\cdot) \equiv K(\cdot/h_n)/h_n$, 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 Horowitz(1992). In the appendix, we provide the asymptotic distribution of our estimator in (4.2) and state the required regularity conditions necessary for its well behavior.

5 Extension III: Panel Data

As is the case with duration data[7], panel data has received an increasing amount of attention in the econometric literature - see Arellano and Honoré(2001) for a recent survey. In the duration context, a panel data set usually refers to a cross section for which we observe multiple time-to-events, or spells. This may refer to multiple spells by the same individual, or spells for different individuals in a group or family.

Empirical examples in the first case include include unemployment spells (Heckman and Borjas(1980)), time intervals between child births (Newman and McCullogh (1984)) and car insurance claim durations (Abbring, Chiappori and Pinquet(2003)). Examples in the second case would include survival times of children in a family in a developing country (Sastry (1997), Ridder and Tunah(1999)), lifetimes of machines grouped by a firm, or unemployment spells grouped by a family or region (e.g. Fitzgerald(1992)).

In this section we consider estimation of a right censored duration model with fixed effects. As in the previous sections, we allow for general forms of censoring. Of particular interest in the panel data setting is to permit the distribution of the censoring variable to be spell-specific and individual/group specific.

The vast existing literature does not address this type of problem. Honoré at al.(2002) allows for random censoring, but requires a linear transformation, and the censoring variables to be distributed independently of the covariates with the same distribution across spells. Exten-

[7]We thank Bo Honoré for suggesting this extension to us.

sions of the linear specification can be found in Abrevaya(1999,2000), which allow for a generalized transformation function, but rule out fixed and/or general random censoring. Other work in the panel duration literature parametrically specifies the distribution of the error terms. Examples include Chamberlain(1985), Honoré(1993), Ridder and Tunah(1999), Lancaster (2000), Horowitz and Lee(2003) and Lee(2003). Some of these also rule out censoring distributions that vary across spells and/or are independent of covariates[8].

In the context of multiple spell data, we wish to allow for distribution of the censoring variable to vary across spells, for one of two reasons: for one, the censoring distribution may depend on time-varying covariates. Also, even if the censoring distribution does not depend on the covariates, and is purely a result of the observation plan, the observation plans may vary across spells.

To be precise, we will focus on the following model:

$$\begin{aligned} T_i(v_{it}) &= \min(\alpha_i + x_{it}'\beta_0 + \epsilon_{it}, c_{it}) \\ d_{it} &= I[\alpha_i + x_{it}'\beta_0 + \epsilon_{it} \le c_{it}] \quad i = 1, 2, ...n \quad t = 1, 2, ...\tau \end{aligned}$$

where here the subscript i denotes an economic agent in a cross section of n observations. In the duration model framework studied in this section, the subscript t does not denote the time period, but one of τ spells. $T_i(\cdot)$ is an unknown, strictly monotonic function that varies across individuals, x_{it} is a $k-$dimensional vector of covariates, c_{it} is a censoring variable which is permitted to be random, and whose distribution is permitted to depend on x_{it}. The disturbance term ϵ_{it} is unobserved, and will be assumed to satisfy conditions which will be discussed shortly. The individual specific effect α_i is unobserved, and following the standard fixed effects approach, is permitted to depend on the covariates x_{it} in an arbitrary way. Finally, the $k-$dimensional vector β_0 is the parameter of interest which we wish to estimate.

Following convention in the fixed-effects literature, we regard n to be large and τ small, as many time-to-event panel data sets encountered in practice are characterized by a large cross section but few spells. Without loss of generality, we set $\tau = 2$, as this facilitates description of the new estimation procedure, and allow $n \to \infty$. Consequently, we wish to estimate β_0 from a random sample of pairs of the $(4 + 2k) \times 1$ vector

$$(d_{i1}, d_{i2}, v_{i1}, v_{i2}, x_{i1}', x_{i2}')'$$

As in the previous sections, we let θ_0 denote the remaining components of β_0 after imposing the same scale normalization, and propose an estimator for θ_0. The estimator, denoted here as $\hat{\theta}_p$,

[8] Honoré(1993), Horowitz and Lee (2003) and Lee(2003) do allow for the censoring variable's distribution to depend on the error term, which is not considered here. As mentioned in these papers, dependent censoring can easily occur in multiple data. Therefore the independence assumption considered here is better suited for analyzing data with group specific effects.

and referred to hereafter as the *censored duration panel* (CDP) estimator, is of the form[9]:

$$\hat{\theta}_p = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} d_{i1} I[v_{i1} < v_{i2}] I[x'_{i1}\beta(\theta) < x'_{i2}\beta(\theta)] \tag{5.1}$$

where Θ denotes the parameter space and for each $\theta \in \Theta$, $\beta(\theta) \equiv (1, \theta')'$.

We establish the consistency of the estimator; this result is based on conditions which are analogous to those imposed in previous sections. To simplify notation, we will let Δx_i denote $x_{i2} - x_{i1}$.

P1 Letting S_{X_t} denote the support of x_{it}, and let $\mathcal{X}_{uct}$ denote the set

$$\mathcal{X}_{uct} = \{x \in S_X : P(d_{it} = 1 | x_{it} = x) > 0\}$$

Then $\mathcal{X}_{uct}$ has positive measure.

P2 The random variables $\epsilon_{i1}, \epsilon_{i2}$ are identically distributed conditional on the vector $(c_{i1}, x'_{i1}, c_{i2}, x'_{i2})$, with common support equal to the real line.

P3 The first component of Δx_i has everywhere positive Lebesgue density, conditional on the other components.

P4 The vector (d_{it}, v_{it}, x_{it}), $i = 1, 2, ..., n$ are i.i.d.

P5 Θ is a compact subset of $\mathbf{R}^{k-1}$.

P6 S_X, the support of Δx_i, is not contained in any proper linear subspace of R^k.

We can now state the theorem establishing consistency in the panel data setting. The proof is left to the appendix.

Theorem 5.1 *Under Assumptions P1-P6, the CDP estimator is consistent:*

$$\hat{\theta}_p \xrightarrow{p} \theta_0 \tag{5.2}$$

Inference on parameters requires limiting distribution theory. We note since the CDP estimator has the same form of a maximum score estimator, the rate of convergence will be slower than the parametric rate, with a non-Gaussian limiting distribution (Kim and Pollard(1990)), making inference difficult to conduct. We also note that one can adopt a smoothed maximum score approach

[9]This estimator can be related to, but is distinct from, existing panel data estimators. The indicator functions comparing values of the observed dependent variables and index values across time originates in Manski(1987), and was also used in Abrevaya(1999). Comparing these values across time intuitively in a "maximum score" (Manski(1975,1985)) type setting leads to consistent estimation of θ_0 for binary choice (Manski(1987)) models and transformation models with fixed censoring (Abrevaya(1999)). However this estimation approach by itself will not consistently estimate θ_0 in the presence of random covariate dependent and/or spell specific censoring, as considered here.

(Horowitz(1992)) and attain a faster rate by imposing stronger smoothness conditions. Since this is now a standard exercise and has been applied in many settings, we omit the details here.

Finally, we conjecture that the slow rate of convergence is a consequence of the generality of the model and not a deficiency of the proposed estimator. This is because there is no common element in our model which permits averaging over individuals in the cross section. Such an averaging would permit attaining the parametric (root-n) rate of convergence for a regression coefficient estimator.

6 Monte Carlo Results

In this section we explore the finite sample properties of the new estimators introduced in this paper by reporting results obtained from a small scale simulation study. Our base design involves two regressors and an additive error term which we express in the absence of censoring as:

$$T(y_i) = \alpha_0 + x_{1i}\beta_0 + x_{2i} + \epsilon_i$$

where x_{1i}, x_{2i} are distributed as a chi-squared with one degree of freedom, and standard normal, respectively; α_0, β_0 were set to 1 and -1 respectively. We considered 2 functional forms for $T(\cdot)$ and the error distribution as follows:

1. $T^{-1}(\nu) = \nu$; $\epsilon_i \sim$ mixture of two normals, centered around -1 and 2 respectively.
2. $T^{-1}(\nu) = exp(\nu)$; $\epsilon_i \sim$ chi squared, 1 degree of freedom.

We simulated four types of censoring:

1. Covariate dependent right censoring: For the exponential design, the censoring variable was distributed as $2.05 \cdot \exp(x_{1i}^2 \cdot x_{2i} + x_{2i})$ and for the linear design it was distributed as $-x_{1i}^2 - x_{2i}$.
2. Covariate independent right censoring: Here for both functional forms the censoring variable was distributed as a chi-squared random variable, with one degree of freedom.
3. Double covariate independent censoring (linear transformation only) : the left censoring variable was distributed as the right censoring variable - 2 times a chi-squared with one degree of freedom - 2, and the right censoring variable was distributed as in 2.
4. Double covariate dependent censoring (linear transformation only): The right censoring variable was the same as in 1. and the relationship between the two censoring variables was the same as in 3.

In tables I through VI we report results for 4 estimators: 1)PRE 2) the MRC 3) the monotone rank estimator (MRE) introduced in Cavanagh and Sherman(1998) 4) the PMLE in Cox(1972,1975). For each estimator and each design the summary statistics mean bias, median bias, root mean squared error (RMSE) and median absolute deviation (MAD) are reported for 100,200, and 400 observations, with 401 replications. As there is only one parameter to compute, each rank estimator was evaluated by means of a grid search of 500 evenly spaced points over the interval [-5,5]. For the PMLE, the intercept and both slope coefficients were evaluated, and the tables report the ratio of the two slope coefficients. Computation of these three values was performed using QNewton in GAUSS, with 10 starting values, which included the true values, least squares estimates, and randomly generated values.

In general, the simulation results are in accordance with the theory. For covariate independent right censoring, all rank estimator perform well in the linear design, and the PRE has the smallest RMSE at all sample sizes. The PMLE performs well, even though the error distribution is misspecified, though its bias values and RMSE do not decline with the sample size. In the exponential design, the only estimators that perform adequately are the PRE and the PMLE, with the PMLE again suffering from RMSE values not shrinking with the sample size. The MRC and MRE only perform adequately at a sample size of 400, which is surprising since they are both theoretically consistent for this design.

For covariate dependent right censoring, the results clearly establish the benefits of the PRE over MRC and MRE. It performs quite well with bias and RMSE values shrinking at the parametric rate. In complete contrast, the MRC and MRE perform very poorly for both functional forms, with RMSE values in most cases not reducing, and sometimes even increasing with the sample size. The PMLE's inconsistency (due to the error distribution misspecification) is also apparent, tough not as pronounced in the linear design at 400 observations. Its RMSE values are much larger than the PRE's for the smaller sample sizes. For the exponential design, the PMLE performs very poorly at all sample sizes.

For double covariate independent censoring, all rank estimators have RMSE's shrinking at the parametric rate, but the efficiency gains of the PRE are very apparent for both functional form error distribution pairs. This is due to the fact that the PRE uses more information on the censoring structure than the other two estimators. The PMLE performs poorly at all sample sizes.

For covariate dependent double censoring, the results are similar to the one sided covariate dependent censoring case, i.e., only the PRE exhibits root-n consistency and the others are clearly inconsistent.

In summary, the results from our simulation indicate that the PRE estimators introduced in this paper perform adequately well in finite samples, so it can be applied in empirical settings, which we turn to in the following section. The results also show how sensitive the other estimators are to model misspecification.

7 Conclusions and further extensions

In this paper, we introduced new estimators for duration models with general forms of covariate dependent censoring. The new estimators have the attractive properties of being distribution free, require no smoothing parameters, and are robust to censoring that depends on the regressors. The estimator is shown to converge at the parametric rate with asymptotically normal distribution.

Extensions were provided for doubly censored, heteroskedastic, and panel data. A simulation study indicated the estimator(s) performed well in finite samples, and also illustrated how erroneous existing estimators can be if the censoring variable depends on covariates or the error distribution is misspecified.

The work in this paper suggest areas for future research. We provide two such examples. For one, it would be useful to construct an estimator for the function $T(.)$ based on y_{0i}, y_{1i} that modifies the rank estimator of $T(\cdot)$ in Chen(2002) to allow for covariate dependent censoring. Finally, another important area for future work would be to formally confirm the conjecture that the proposed panel data estimator attains the fastest rate of convergence possible under the assumptions of the model.

References

[1] Abbring, J.H., P.A. Chiappori, and J. Pinquet (2003), “Moral Hazard and Dynamic Insurance Data”, *Journal of the European Economic Association*, forthcoming.

[2] Abrevaya, J. (1999), ”Leapfrog Estimation of a Fixed-effects Model with Unknown Transformation of the Dependent Variable.” *Journal of Econometrics*, 93, 203-228.

[3] Abrevaya, J. (2000), ”Rank estimation of a generalized fixed-effects regression model”,*Journal of Econometrics*, 95, 1-23.

[4] Bijwaard, G.E. (2001), “Rank Estimation of Duration Models”, *Ph D. Disseration, Tinbergen Institute.*

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

[6] Cai, T., Wei, L.J. and M. Wilcox(2000), “Semiparametric Regression Analysis for Clustered Failure Time Data.” *Biometrika*, 87, 867-878.

[7] Cavanagh, C. and R.P. Sherman (1998), “Rank Estimators for Monotonic Index Models”, *Journal of Econometrics*, 84, 351-381

[8] Chamberlain, G. (1985), “Heterogeneity, Omitted Variable Bias, and Duration Dependence”, in in Heckman, J.J. and B. Singer, eds., *Longitudinal Analysis of Labor Market Data*, Cambridge: Cambridge University Press.

[9] Chen, S. (2002), “Rank Estimation of Transformation Models”, *Econometrica*, 70, 1683-1696.

[10] Chen, J., Jin, Z., and Z. Ying (2002), “Semiparametric Analyisis of Transformation Models with Censored Data”, *Biometrika*, 89, 659-668.

[11] Cheng, S.C., Wei, L.J., and Z. Ying (1995), “Analysis of Transformation Models with Censored Data”, *Biometrika*, 82, 835-845.

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

[13] Cox D.R. (1975), “Partial Likelihood” *Biometrika*, 62, 269-276.

[14] Cuzick, J. (1988), “Rank Regression”, *Annals of Statistics*, 16, 1369-1389.

[15] Fine, J.P., Ying, Z. and L.J. Wei (1998), “On the Linear Transformation Model with Censored Data”, *Biometrika*, 85, 980-986.

[16] Flinn, C. and J. Heckman (1983), “ Are Unemployement and out of the Labor Force Behaviorally Distinct Labor Force States?”, *Journal of Labor Economics*, 1, 28-42.

[17] Gorgens, Tue and J. Horowitz (1998), “Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable”, *Journal of Econometrics*, Volume 90, Issue 2, Pages 155-191.

[18] Han, A. (1987) “Non Parametric Analysis of a Generalized Regression Model”, *Journal of Econometrics*, 35, 303-316

[19] Heckman, J.J. and G.J. Borjas (1980) “Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers for a Continuous Time Model of Heterogeneity and State Dependence”, *Economica*, 47, 247-283.

[20] Honoré, B.E.(1993) “Identification Results for Duration Models with Multiple Spells”, *Review of Economic Studies*, 60, 241-246.

[21] Honoré, B.E., Khan, S. and J.L. Powell (2002) “Quantile Regression under Random Censoring”, *Journal of Econometrics*, 109. 67-105.

[22] Horowitz, J.L. and S. Lee (2003), "Semiparametric Estimation of a Panel Data Proportional Hazards Model with Fixed Effects", *Journal of Econometrics*, forthcoming.

[23] Jin Z., Ying Z., and Wei L.J. (2002) "A Simple Resampling Method by Perturbing the Minimand", *Biometrika*, 88, 381-390.

[24] Kalbfleisch, J.D. and R.L. Prentice (1980), *The Statistical Analysis of Failure Time Data.* New York: Wiley.

[25] Kaplan, E.L. and P. Meier (1958), "Nonparametric Estimation from Incomplete Data," *Journal of the American Statistical Association*, 53, 457-481.

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

[27] Lancaster, T. (2000), "The Incidental Parameter Problem since 1948", *Journal of Econometrics*, 95, 291-413.

[28] Lee, S. (2003), "Estimating Panel Data Duration Models with Censored Data", *manuscript, University College London.*

[29] Manski, C.F. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, 3, 205-228

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

[31] Newey, W.K. and D. McFadden (1994) "Estimation and Hypothesis Testing in Large Samples", in Engle, R.F. and D. McFadden (eds.) , *Handbook of Econometrics, Vol. 4*, Amsterdam: North-Holland.

[32] Newman, J.L. and C.E. McCullogh (1984), "A hazard Rate Approach to the Timing of Births", *Econometrica*, 52, 939-961.

[33] Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57, 1027-1057.

[34] Portnoy, S. (2003), "Censored Regression Quantiles", *Journal of the American Statistical Association*, 98, 1001-1012.

[35] Ridder, G. (1990) "The Non-parametric Identification of Generalized Accelerated Failure-time Models", *Review of Economic Studies*, 57, 167-182

[36] Ridder, G. and I. Tunah(1999) "Stratified Partial Likeliood Estimation", *Journal of Econometrics*, 92, 193-232.

[37] Sherman, R.P. (1993), "The Limiting Distribution of the Maximum Rank Correlation Estimator", *Econometrica*, 61, 123-137

[38] Sherman, R.P. (1994a), "U-Processes in the Analysis of a Generalized Semiparametric Regression Estimator", *Econometric Theory*, 10, 372-395

[39] Sherman, R.P. (1994b), "Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators ", *Annals of Statistics*, 22, 439-459

[40] Tsiatis, A.A. (1990), "Estimating Regression Parameters Using Linear Rank Tests for Censored Data," *Annals of Statistics*, 18, 354-372.

[41] Van den Berg, G.J. (2001), "Duration Models: Specification, Identification and Multiple Durations", in Heckman, J.J. and E. Leamer, eds., *Handbook of Econometrics, Vol. 5*, Amsterdam: North-Holland.

[42] Wei, L.J., Z. Ying, and D.Y. Lin (1990), "Linear Regression Analysis of Censored Survival Data Based on Rank Tests." *Biometrika*, 19, 417-442.

[43] Wang, J.-G. (1987), "A Note on the Uniform Consistency of the Kaplan-Meier Estimator" *Annals of Statistics*, 15, 1313-1316

[44] Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Survival and Hazard Functions", *Journal of the American Statistical Association*, 94, 137-145

[45] 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 lemma 2.1

We need to show that

$$\Pr[y_{1i} \geq y_{0j}|x_i, x_j] \geq \Pr[y_{1j} \geq y_{0i}|x_i, x_j] \text{ iff } \quad x_i'\beta_0 \geq x_j'\beta_0 \tag{A.1}$$

which is equivalent to showing that

$$P(y_{0i} \geq y_{1j}|x_i, x_j) \geq P(y_{0j} \geq y_{1i}|x_i, x_j) \text{ iff } \ x_i'\beta_0 \geq x_j'\beta_0 \tag{A.2}$$

For notational convenience, we let z_i, z_j denote $x_i'\beta_0, x_j'\beta_0$ respectively. We first evaluate

$$P(y_{0i} \geq y_{1j}) \tag{A.3}$$

where we condition on x_i, x_j. This probability can be decomposed into the mutually exclusive cases $c_i > c_j$ and $c_i \leq c_j$. We first focus on the case where $c_i > c_j$, and evaluate the probability conditional on the censoring values c_i, c_j.

Note the probability of (A.3) is zero whenever $d_j = 0$, so we can decompose (A.3) as

$$P(y_{0i} \geq y_{1j}) = P(y_{0i} \geq y_{1j}, d_i = 1, d_j = 1) + P(y_{0i} \geq y_{1j}, d_i = 0, d_j = 1) \tag{A.4}$$

We derive an expression for the first term, which we write here as:

$$P(\epsilon_i \geq \epsilon_j - \Delta z, \epsilon_i \leq c_i - z_i, \epsilon_j \leq c_j - z_j)$$

where here, $\Delta z \equiv z_i - z_j$. Recall that we are assuming for now that $c_i > c_j$, so by the independence assumption, we express the above probability as:

$$\int_{-\infty}^{c_j - z_j} \int_{\epsilon_j - \Delta z}^{c_i - z_i} dF(\epsilon_i) dF(\epsilon_j) \tag{A.5}$$

where $F(\cdot)$ denotes the c.d.f. of ϵ_i and ϵ_j. So (A.5) is

$$F(c_i - z_i)F(c_j - z_j) - \int_{-\infty}^{c_j - z_j} F(\epsilon_j - \Delta z) dF(\epsilon_j) \tag{A.6}$$

Now, turning attention to the second term in (A.4), we express it as:

$$P(\epsilon_j \leq c_j - z_j, \epsilon_i \geq c_i - z_i, \epsilon_j \leq c_i - z_j) = P(\epsilon_i \geq c_i - z_i, \epsilon_j \leq c_j - z_j) \tag{A.7}$$

where the equality follows from $c_i > c_j$. This is equal to

$$(1 - F(c_i - z_i))(F(c_j - z_j)) \tag{A.8}$$

Thus we have that conditioning on x_i, x_j, and $c_i > c_j$, (A.4) can be expressed as:

$$\int_{-\infty}^{c_j - z_j} S(\epsilon_j - \Delta z) dF(\epsilon_j) \tag{A.9}$$

where here $S(\cdot) = 1 - F(\cdot)$.

We next evaluate $P(y_{0j} \geq y_{1i})$, again conditioning on x_i, x_j, $c_i > c_j$. A similar decomposition yields:

$$P(y_{0j} \geq y_{1i}, d_i = 1, d_j = 1) + P(y_{0j} \geq y_{1i}, d_i = 1, d_j = 0) \tag{A.10}$$

The first term is:

$$P(\epsilon_i \le \epsilon_j - \Delta z, \epsilon_i \le c_i - z_i, \epsilon_j \le c_j - z_j) \tag{A.11}$$

which we can decompose into the sum of

$$P(\epsilon_i \le \epsilon_j - \Delta z, \epsilon_i \le c_j - z_i, \epsilon_j \le c_j - z_j) + P(\epsilon_i \le \epsilon_j - \Delta z, c_j - z_i \le \epsilon_i \le c_i - z_i, \epsilon_j \le c_j - z_j) \tag{A.12}$$

Note the second term is 0 (since $\epsilon_i \le \epsilon_j - \Delta z$ and $\epsilon_j \le c_j - z_j$ contradicts the middle event), and the first term is:

$$\int_{-\infty}^{c_j - z_i} \int_{\epsilon_i + \Delta z}^{c_j - z_j} dF(\epsilon_j) dF(\epsilon_i) \tag{A.13}$$

which is equal to

$$F(c_j - z_j)F(c_j - z_i) - \int_{-\infty}^{c_j - z_i} F(\epsilon_i + \Delta z) dF(\epsilon_i)$$

The second term in (A.10) is

$$P(c_j \ge \epsilon_i + z_i, \epsilon_j \ge c_j - z_j, \epsilon_i \le c_i - z_i) = P(\epsilon_j \ge c_j - z_j, \epsilon_i \le c_j - z_i) \tag{A.14}$$

where the equality follows from $c_i > c_j$. This can be expressed as:

$$(1 - F(c_j - z_j))F(c_j - z_i) \tag{A.15}$$

Therefore, (A.10) is $\int_{-\infty}^{c_j - z_i} S(\epsilon_i + \Delta z) dF(\epsilon_i)$, and the difference between (A.4) and (A.10) is

$$\int_{-\infty}^{c_j - z_j} S(\epsilon - \Delta z) dF(\epsilon) - \int_{-\infty}^{c_j - z_i} S(\epsilon + \Delta z) dF(\epsilon) \tag{A.16}$$

We note that the above difference has the same sign as $z_i - z_j$, as the integrand is strictly larger for the first term if and only if $z_i > z_j$, as is the range of integration. This establishes the identification result for the case where $c_i > c_j$. For $c_i \le c_j$, using analogous arguments, we find that

$$P(y_{0i} \ge y_{1j}) = \int_{-\infty}^{c_i - z_j} S(\epsilon_i - \Delta z) dF(\epsilon_i) \tag{A.17}$$

and

$$P(y_{0j} \ge y_{1i}) = \int_{-\infty}^{c_i - z_i} S(\epsilon_i + \Delta z) dF(\epsilon_i) \tag{A.18}$$

and the difference in the two integrals has the same sign as $z_i - z_j$ here as well. Thus after integrating over c_i, c_j we can conclude that

$$P(y_{0i} \ge y_{1j} | x_i, x_j) - P(y_{0i} \ge y_{1j} | x_i, x_j)$$

has the same sign as $x_i'\beta_0 - x_j'\beta_0$. ■

A.2 Proof of Theorem 2.1

To show consistency it suffices to show 4 conditions (see e.g. Newey and McFadden(1994), Theorem 2.1.): compactness, uniform convergence, continuity, identification.

We first turn attention to identification, whose result will be shown to follow Lemma 2.1. The sample objective function (2.7) can be written as

$$\frac{1}{n(n-1)}\sum_{i\neq j} I[y_{1i}\geq y_{0j}]I[x_i'\beta\geq x_j'\beta] \quad = \quad \frac{1}{n(n-1)}\sum_{i<j}\left(I[y_{1i}\geq y_{0j}]I[x_i'\beta\geq x_j'\beta]+I[y_{1j}\geq y_{0i}]I[x_j'\beta\geq x_i'\beta]\right)$$

Then, let $Q(\beta)$ denote the limiting objective function:

$$Q(\beta)=E_X[P(y_{1i}\geq y_{0j}|x_i,x_j)I[x_i'\beta\geq x_j'\beta]+P(y_{1j}\geq y_{0i}|x_i,x_j)I[x_j'\beta\geq x_i'\beta]]$$

where $E_X[\cdot]$ denotes the expectation over x_i, x_j. We need to show that this is uniquely maximized at β_0. We have (suppressing the conditioning on x_i and x_j):

$$\begin{aligned} Q(\beta_0)-Q(\beta) &= E_X[(P(y_{1i}\geq y_{0j})-P(y_{1j}\geq y_{0i}))\left(I[x_i'\beta_0\geq x_j'\beta_0]-I[x_i'\beta\geq x_j'\beta]\right)] \qquad (A.19)\\ &= E_X\left[(P(y_{1i}\geq y_{0j})-P(y_{1j}\geq y_{0i}))\left(I[\Delta x'\beta_0\geq 0>\Delta x'\beta]-I[\Delta x'\beta\geq 0>\Delta x'\beta_0]\right)\right] \end{aligned}$$

By the previous lemma, the above expectation is non-negative, and trivially equal to 0 when $\beta=\beta_0$ by Assumption **I3**. We show for $\beta\neq\beta_0$, the above expectation is strictly positive. Note that $\beta\neq\beta_0$ corresponds to $\theta\neq\theta_0$ since θ and θ_0 have the same first component of 1. It follows from Assumption **I6** that with positive probability, $\Delta x^{(-1)'}\theta\neq x^{(-1)'}\theta_0$, where here $\Delta x^{(-1)}$ denotes the difference in the $k-1$ dimensional vector corresponding to the last $k-1$ components of the regressor vector. By Assumption **I3**, we can find a subset of $\mathcal{X}_{uc}\times\mathcal{X}_{uc}$ where $I[x_i'\beta_0>x_j'\beta_0, x_i'\beta<x_j'\beta]]=1$ or $I[x_i'\beta_0<x_j'\beta_0, x_i'\beta>x_j'\beta]]=1$, and this subset has positive probability. But from Lemma 2.1, on this subset $P(y_{1i}\geq y_{0j}|x_i,x_j)>P(y_{1j}\geq y_{0i}|x_i,x_j)$, so (A.19) is strictly positive, establishing that the limiting objective function is uniquely maximized at β_0 and proving identification.

Turning attention to the other three items, we note that compactness holds by Assumption **I5**. Regarding uniform convergence, we need to show

$$\sup_{\theta\in\Theta} Q_n(\beta)\xrightarrow{p} Q(\beta) \qquad (A.20)$$

where $Q_n(\beta)$ is the sample objective function defined in (2.9). (A.20) follows from uniform laws of large numbers for U-statistics with bounded kernel functions satisfying a Euclidean property. This property (with the constant envelope 1) is shown below, so we can apply Corollary 7 in Sherman(1994) to establish (A.20). The continuity condition that $Q(\beta)$ is continuous at $\beta=\beta_0$ follows from the smoothness of the density of $x_i'\beta_0$ which follows from **I3**. This establishes consistency. ■

A.3 Proof of Theorem 2.2

We note that virtually identical arguments as in Sherman(1993) can be used, as the objective functions of the MRC and the PRE are very similar. The only component of the proof there that does not immediately carry over to the problem at hand is establishing the Euclidean property of the class of functions in the objective function. For the problem at hand, we consider the class of functions:

$$\mathcal{F}=\{f(\cdot,\cdot,\theta):\theta\in\Theta\} \qquad (A.21)$$

where for each $(\mathbf{z}_1,\mathbf{z}_2)\in S\times S$, $\theta\in\Theta$, we can define

$$\begin{aligned} f(\mathbf{z}_1,\mathbf{z}_2,\theta) &= I[y_{01}\geq y_{12}]I[x_1'\beta\geq x_2'\beta] \qquad (A.22)\\ &= (d_2 I[v_1\geq v_2])I[x_1'\beta\geq x_2'\beta] \qquad (A.23) \end{aligned}$$

where with our notation, recall β is a function of θ.

Note the class of functions

$$f_1(\mathbf{z}_1, \mathbf{z}_2, \theta) = I[v_1 \geq v_2]I[x_1'\beta \geq x_2'\beta] \tag{A.24}$$

is Euclidean for envelope 1 from identical subgraph set arguments used in Sherman(1993). The class of functions:

$$f_2(\mathbf{z}_1, \mathbf{z}_2, \theta) = d_2 \tag{A.25}$$

is trivially Euclidean for envelope 1 as it does not depend on θ. The Euclidean property of $f = f_1 \cdot f_2$ follows from Lemma 2.14(ii) in Pakes and Pollard(1989).

A.4 Double Censored Section: Consistency and Normality

First, we prove Lemma 3.1.

Proof of Lemma 3.1: Again, we wish to show that conditional on x_i, x_j,

$$P(y_{0i} \geq y_{1j}) \geq P(y_{0j} \geq y_{1i}) \tag{A.26}$$

iff $z_i \geq z_j$. As with the singly censored data we will separately consider cases involving relationships for the i^{th} and j^{th} censoring values. For double censoring, the problem becomes more involved as there are the following 6 cases to consider:

$$c_{1i} > c_{1j}, c_{2i} > c_{2j}, c_{1i} < c_{2j} \tag{A.27}$$
$$c_{1i} > c_{1j}, c_{2i} > c_{2j}, c_{1i} > c_{2j} \tag{A.28}$$
$$c_{1j} > c_{1i}, c_{2j} > c_{2i}, c_{2i} > c_{1j} \tag{A.29}$$
$$c_{1j} > c_{1i}, c_{2j} > c_{2i}, c_{2i} < c_{1j} \tag{A.30}$$
$$c_{2j} > c_{2i}, c_{1j} < c_{1i} \tag{A.31}$$
$$c_{2i} > c_{2j}, c_{1i} < c_{1j} \tag{A.32}$$

We first show (A.26) holds for the first of the mentioned censoring variable relationships in the theorem. As before, we derive an expression for each of the probabilities being compared. For the left hand side probability, we decompose it as follows:

$$\begin{aligned}
P(y_{0i} \geq y_{1j}) &= P(y_{0i} \geq y_{1j}, d_i = 1, d_j = 1) && \text{(A.33)}\\
&+ P(y_{0i} \geq y_{1j}, d_i = 1, d_j = 2) && \text{(A.34)}\\
&+ P(y_{0i} \geq y_{1j}, d_i = 3, d_j = 2) && \text{(A.35)}\\
&+ P(y_{0i} \geq y_{1j}, d_i = 3, d_j = 1) && \text{(A.36)}
\end{aligned}$$

Before deriving expressions for each of the above terms, for ease of exposition, we introduce some new notation. Again we let $F(\cdot)$ denote the c.d.f. of the error term, for $n = 1, 2$ we let F_{nij} denote $F(c_{ni} - z_j)$.

Turning attention to the first of the four above expressions, it can be expanded as:

$$P(\epsilon_j \le \epsilon_i + \Delta z, c_{1i} - z_i \le \epsilon_i \le c_{2i} - z_i, c_{1j} - z_j \le \epsilon_j \le c_{2j} - z_j) \tag{A.37}$$

which, recalling we are assuming the censoring values relationship in (A.27), can be expressed as:

$$P(c_{1i} - z_i \le \epsilon_i \le c_{2j} - z_i), c_{1j} - z_j \le \epsilon_j \le c_{2j} - z_j, \epsilon_j \le \epsilon_i + \Delta z) + \tag{A.38}$$

$$P(c_{2j} - z_i \le \epsilon_i \le c_{2i} - z_i), c_{1j} - z_j \le \epsilon_j \le c_{2j} - z_j, \epsilon_j \le \epsilon_i + \Delta z) \tag{A.39}$$

where $\Delta z \equiv z_i - z_j$. Note in the first probability, the second bound on ϵ_j is binding, and in the second probability, the first bound on ϵ_j is binding. Therefore we can write the sum of these two terms as (again conditioning on censoring variables and regressors) :

$$\int_{c_{1i}-z_i}^{c_{2j}-z_i} F(\epsilon_i + \Delta z) dF(\epsilon_i) + F_{1jj}F_{1ii} + F_{2ii}F_{2jj} - F_{2ii}F_{1jj} - F_{2ji}F_{2jj} \tag{A.40}$$

This is the expression for (A.33). The terms (A.34), (A.35), (A.36) are easier to deal with so we omit the details. They can be expressed as:

$$(F_{2ii} - F_{1ii})F_{1jj}, \ (1 - F_{2ii})F_{1jj}, \ (1 - F_{2ii})(F_{2jj} - F_{1jj}) \tag{A.41}$$

respectively. Collecting all terms we have

$$P(y_{0i} \ge y_{1j}) = \int_{c_{1i}-z_i}^{c_{2j}-z_i} F(\epsilon_i + \Delta z) dF(\epsilon_i) + F_{2jj} - F_{2jj}F_{2ji} \tag{A.42}$$

We will compare this expression to one for $P(y_{0j} \ge y_{1i})$, which we also decompose into the sum of four probabilities involving 4 censoring pair indicators. As before, the case $d_i = 1, d_j = 1$ is the most involved. Here, the probability is

$$P(c_{1i} - z_i \le \epsilon_i \le \epsilon_j - \Delta z, c_{1j} - z_j \le \epsilon_j \le c_{2j} - z_j) \tag{A.43}$$

which we can express as:

$$\int_{c_{1i}-z_j}^{c_{2j}-z_j} F(\epsilon_j - \Delta z) dF(\epsilon_j) - F_{1ii}F_{2jj} + F_{1ii}F_{1ij} \tag{A.44}$$

The cases $(d_j = 1, d_i = 2)$, $(d_j = 3, d_i = 2)$, $(d_j = 3, d_i = 1)$ are easier to deal with. They are, respectively:

$$F_{1ii}(F_{2jj} - F_{1ij}), \ F_{1ii}(1 - F_{2jj}, \ (1 - F_{2jj})(F_{2ji} - F_{1ii}) \tag{A.45}$$

So collecting terms we have

$$P(y_{0j} \ge y_{1i}) = \int_{c_{1i}-z_j}^{c_{2j}-z_j} F(\epsilon_i + \Delta z) dF(\epsilon_i) + F_{2jj} - F_{2jj}F_{2ji} \tag{A.46}$$

Thus we have that $P(y_{0i} \ge y_{1j}) - P(y_{0j} \ge y_{1i})$ is

$$\int_{c_{1i}-z_i}^{c_{2j}-z_i} F(\epsilon_i + \Delta z) dF(\epsilon_i) - \int_{c_{1i}-z_j}^{c_{2j}-z_j} F(\epsilon_i - \Delta z) dF(\epsilon_i) + F_{2jj} - F_{2ji} \tag{A.47}$$

we now show the above expression has the same sign as $z_i - z_j$. First, we assume that $z_i \ge z_j$ and show the above expression is nonnegative. Note that $F_{2jj} - F_{2ji}$ is nonegative, but unlike in the one-sided censoring case, the difference in integrals is not unambiguously nonnegative, as the ranges of integration differ. Thus we work with the decomposition:

$$\int_{c_{1i}-z_j}^{c_{2j}-z_j} F(\epsilon_i - \Delta z)dF(\epsilon_i) = \int_{c_{1i}-z_j}^{c_{2j}-z_i} F(\epsilon_i - \Delta z)dF(\epsilon_i) + \int_{c_{2j}-z_i}^{c_{2j}-z_j} F(\epsilon_i - \Delta z)dF(\epsilon_i) \tag{A.48}$$

Note the first piece on the right hand side of the above decomposition is less than or equal to the the first term in (A.47), which has a larger integrand and range of integration. (We note that here we are assuming that $c_{2j} - z_i \geq c_{1i} - z_j$. If it is not, then this integral (the first piece on the right hand side of the above decomposition) is negative, and trivially less than the the first term in (A.47).

Also, the second piece on the right hand side of the above decomposition is smaller that $F_{2jj} - F_{2ji}$, which follows by exploiting $F(\cdot) \leq 1$. Thus we have shown that (A.47) is nonnegative when $z_i \geq z_j$. The same arguments can be used to show that (A.47) is nonpositive when $z_i \leq z_j$. This establishes identification for the case (A.27).

Similar arguments may be used for the other censoring variable relationships. We omit the details and only state the difference between $P(y_{0i} \geq y_{1j})$ and $P(y_{0j} \geq y_{1i})$ for the five remaining cases:

1. $F_{2jj}(1 - F_{1ii})$
2. $\int_{c_{1j}-z_i}^{c_{2i}-z_i} F(\epsilon_i + \Delta z)dF(\epsilon_i) - \int_{c_{1j}-z_j}^{c_{2i}-z_j} F(\epsilon_i - \Delta z)dF(\epsilon_i) + F_{2ij} - F_{2ii}$
3. $-F_{2ii}(1 - F_{1jj})$
4. $\int_{c_{1i}-z_i}^{c_{2i}-z_i} F(\epsilon_i + \Delta z)dF(\epsilon_i) - \int_{c_{1i}-z_j}^{c_{2i}-z_j} F(\epsilon_i - \Delta z)dF(\epsilon_i) + F_{2ij} - F_{2ii}$
5. $\int_{c_{1j}-z_i}^{c_{2j}-z_i} F(\epsilon_i + \Delta z)dF(\epsilon_i) - \int_{c_{1j}-z_j}^{c_{2j}-z_j} F(\epsilon_i - \Delta z)dF(\epsilon_i) + F_{2jj} - F_{2ji}$

Note we can use the same arguments to show that, with the exception of cases 1 and 3, the above terms have the same sign as $z_i - z_j$. Thus if were not for 1 and 3 the identification proof is complete. Note that if we have $P(c_{1i} \geq c_{2j}|x_i, x_j) = P(c_{1j} \geq c_{2i}|x_i, x_j) = 0$, then cases 1 and 3 cannot occur and we have identification. Alternatively, if the censoring distributions are independent of the covariates, then we have $P(c_{1i} \geq c_{2j}|x_i, x_j) = P(c_{1j} \geq c_{2i}|x_i, x_j) = p_{12}$ for some constant p_{12} not necessarily equal to 0. Then we can integrate the sum of the term in 1 and the term in 3 with respect to the censoring variables $c_{1i}, c_{2i}, c_{1j}, c_{2j}$, which yields a term whose sign is the same sign as $z_i - z_j$, establishing identification here as well. ■

The asymptotic distribution theory of the estimator in 3.1 above is based on Assumptions AD1-AD4 below. We first need to introduce some further notation for the doubly censored case. Now $\mathbf{z}_i$ denotes the vector $(d_{1i}, d_{2i}, d_{3i}, v_i, x_i')'$, we define

$$\begin{aligned} \tau_d(\mathbf{z}, \theta) &= E[((1-d_2)(1-d_{3i})I[v \geq v_i])I[x'\beta \geq x_i'\beta]] \\ &+ E[((1-d_{2i})(1-d_3)I[v_i \geq v]]I[x_i'\beta \geq x'\beta]] \end{aligned}$$

Finally, we let $\mathcal{N}$ denote a neighborhood of θ_0.

AD1 θ_0 lies in the interior of Θ, a compact subset of R^{k-1}.

AD2 For each $\mathbf{z}$, the function $\tau_d(\mathbf{z}, \cdot)$ is twice differentiable in a neighborhood of θ_0. Furthermore, the vector of second derivatives of $\tau_d(\mathbf{z}, \cdot)$ satisfies the following Lipschitz condition:

$$\|\nabla_2 \tau_d(\mathbf{z}, \theta) - \nabla_2 \tau_d(\mathbf{z}, \theta_0)\| \leq M(\mathbf{z})\|\theta - \theta_0\|$$

where ∇_2 denotes the second derivative operator and $M(\cdot)$ denotes an integrable function of z.

AD3 $E[\|\nabla_1 \tau_d(\mathbf{z}_i, \theta_0)\|^2]$ and $E[\|\nabla_2 \tau_d(\mathbf{z}_i, \theta_0)\|]$ are finite.

AD4 $E[\nabla_2 \tau_d(\mathbf{z}_i, \theta_0)]$ is non-singular.

The following theorem characterizes the asymptotic distribution of the estimator. The proof of the theorem is omitted, as it follows from identical steps used in proving Theorem 2.2.

Theorem A.1 *Under Assumptions AD1-AD4,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V_d^{-1} \Delta_d V_d^{-1}) \tag{A.49}$$

where $V_d = E[\nabla_2 \tau_d(\mathbf{z}_i, \theta_0)]/2$ *and* $\Delta_d = E[\nabla_1 \tau_d(\mathbf{z}_i, \theta_0) \nabla_1 \tau_d(\mathbf{z}_i, \theta_0)']$.

A.5 Heteroskedasticity: Consistency and Normality

We first provide the proof for lemma 4.1.

Proof of Lemma 4.1 (only if) Consider the following

$$\begin{aligned}
\Pr(T(y_1) - x\beta \leq 0|x) &= \Pr(T(y_1) \leq x\beta, d = 1|x) + \Pr(T(y_1) \leq x\beta, d = 0|x) \\
&= \Pr(\epsilon \leq 0, \epsilon \leq c - x\beta|x) \\
&= \Pr(\epsilon \leq 0|x)
\end{aligned}$$

where the second equality follows from the definition of y_1 and the third equality follows from the hypothesis that $x \in \mathcal{X}$.

$$\begin{aligned}
\Pr(T(y_0) = x\beta \leq 0|x) &= \Pr(T(y_0) \leq x\beta, d = 1|x) + \Pr(T(y_0) - x\beta \leq 0, d = 0|x) \\
&= \Pr(\epsilon \leq 0, \epsilon \leq c - x\beta|x) + \Pr(c - x\beta \leq 0, \epsilon > c - x\beta|x) \\
&= \Pr(\epsilon \leq 0, \epsilon \leq c - x\beta|x) \\
&= \Pr(\epsilon \leq 0|x)
\end{aligned}$$

where the second equality follows from the definition of y_0, and the third and fourth equalities follow from the fact that $x \in \mathcal{X}$ (i.e., $c - x\beta \geq 0$). This implies that for $x \in \mathcal{X}$,

$$\Pr(T(y_0) = x\beta \leq 0|x) = \Pr(T(y) = x\beta \leq 0|x) = \Pr(\epsilon \leq 0|x) = \Pr(T(y_1) = x\beta \leq 0|x)$$

(if) Now we have

$$\begin{aligned}
\Pr(T(y) - x\beta \leq 0|x) &= \Pr(\epsilon \leq 0|x) = \frac{1}{2} = \Pr(T(y_0) - x\beta \leq 0|x) \\
&= \Pr(T(y_0) - x\beta \leq 0, d = 1|x) + \Pr(T(y_0) - x\beta \leq 0, d = 0|x) \\
&= \Pr(\epsilon \leq 0, \epsilon \leq c - x\beta|x) + \Pr(c - x\beta \leq 0, \epsilon > c - x\beta|x) \\
&= \Pr(\epsilon \leq 0, \epsilon \leq c - x\beta, c - x\beta \geq 0|x) + \Pr(\epsilon \leq 0, \epsilon \leq c - x\beta, c - x\beta < 0|x) \\
&\quad + \Pr(c - x\beta \leq 0, \epsilon > c - x\beta|x) \\
&= \Pr(\epsilon \leq 0|x) \Pr(c - x\beta \geq 0|x) + \Pr(\epsilon \leq c - x\beta, c - x\beta \leq 0|x) \\
&\quad + \Pr(c - x\beta \leq 0, \epsilon > c - x\beta|x) \\
&= \Pr(\epsilon \leq 0|x) \Pr(c - x\beta \geq 0|x) + \Pr(c - x\beta \leq 0|x) \\
&= \frac{1}{2} \Pr(c - x\beta \geq 0|x) + 1 - \Pr(c - x\beta \geq 0|x) \\
&= 1 - \frac{1}{2} \Pr(c - x\beta \geq 0|x)
\end{aligned}$$

which implies that

$$\Pr(c - x\beta \geq 0|x) = 1$$

i.e., $x \in \mathcal{X}$. The third equality above follows from the definition of y_0, and the sixth equality follows from the independence assumption $\epsilon \perp c|x$. ■

We next state the limiting distribution theory for $\hat{\beta}_{ht}$ we defined in (4.2). For notational ease, in this section we assume the covariates are all continuously distributed. Our limiting distribution theory for this estimator is based on the following assumptions:

Assumptions on the Median Functions

Q1 $m_j(\cdot)$ $\quad j = 0, 1$ is $[p]$ times differentiable in x_i with $[\cdot]$ denoting the integer operator. Letting $\nabla_{[p]} m_j(x)$ denote the vector of $[p]^{th}$ order derivatives of $m_j(\cdot)$ in x_i, we assume the following Lipschitz condition:

$$\|\nabla_{[p]} m_j(x_1) - \nabla_{[p]} m_j(x_2)\| \leq \mathcal{K}\|x_1 - x_2\|^{\gamma}$$

for all values x_1, x_2 in the support of x_i, 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 $p = [p] + \gamma$ denote the *order of smoothness* of the quantile function.

Assumptions on the Trimming Function

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

Assumptions on the Regressors

B1 The sequence of $k + 2$ dimensional vectors (d_i, v_i, x_i) are independent and identically distributed.

B2 The regressor vector x_i has support which is a subset of $\mathbf{R}^k$. We will let $f_X(x)$ denote the (Lebesgue) density function of x_i.

B3 $f_X(x)$ is continuous and bounded on the support of x_i.

B4 Assume that $\mathcal{X}_t = \mathcal{X}_{t(k-1)} \times \mathcal{X}_{tk}$ where $\mathcal{X}_{t(k-1)}$ and $\mathcal{X}_{tk}$ are compact subsets with non-empty interiors of the supports of the first $k-1$ components, and the k^{th} component of x_i, respectively. For each $x \in \mathcal{X}_t$, denote its first $k - 1$ components by $x_{(k-1)}$. $\mathcal{X}_t$ will be assumed to have the following properties:

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

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 $f_{u1|X_i=x}(\cdot)$ which is continuous, and bounded away from 0 and infinity for all values of $x \in \mathcal{X}_t$. As a function of x, $f_{u|X_i=x}$ 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

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

and let

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

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 $\lim_{\nu\to+\infty}K(\nu)=1$, $\lim_{\nu\to-\infty}K(\nu)=0$

SI1.2 $\int_{-\infty}^{\infty}K'(\nu)d\nu=1$

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

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 A.2 *Assume that $p>3k/2$, and that in the first stage, and the bandwidth sequences satisfy $\sqrt{n}\delta_n^p\to 0$, $\log n\sqrt{n^{-1}\delta_n^{-3k}}\to 0$ and*

$$\sqrt{n}h_n^{-2}(\delta_n^{2p}+\log n\cdot n^{-1}\delta_n^{-k})\to 0$$

Define

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

where $f'_{m1}(\cdot), f'_{m0}(\cdot)$ denote derivatives of density functions of the median functions; then under Assumptions A,B,Q,T,E,SI

$$\sqrt{n}(\hat\theta-\theta_0)\Rightarrow N(0,V_q^{-1}\Delta_qV_q^{-1}) \tag{A.50}$$

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

Proof: The asymptotic properties follow from arguments that are very similar to those used in Khan(2001), 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 (4.2), yielding the sum of the three components

$$\Gamma_n(\beta) \equiv \frac{1}{n(n-1)} \sum_{i \neq j} K_{h_n}(m_{1i} - m_{0j}) I[x_i'\beta \geq x_j'\beta] \tag{A.51}$$

$$H_n(\beta) \equiv \frac{1}{n(n-1)} \sum_{i \neq j} K'_{h_n}(m_{1i} - m_{0j}) h_n^{-1}((\hat{m}_{1i} - m_{1i}) - (\hat{m}_{0j} - m_{0j}) I[x_i'\beta \geq x_j'\beta] \tag{A.52}$$

$$R_n(\beta) \frac{1}{n(n-1)} \sum_{i \neq j} K''_{h_n}(m^*_{1i} - m^*_{0j}) h_n^{-2}(\hat{m}_{1i} - m_{1i} - \hat{m}_{0j} + m_{0j})^2 I[x_i'\beta \geq x_j'\beta] \tag{A.53}$$

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 (A.52). It follows by uniform rates of convergence for median function estimators over compact sets, (see, e.g. Chaudhuri(1991)) 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 Khan(2000) that uniformly over β within $o_p(1)$ neighborhoods of β_0, we have

$$H_n(\beta) = (\beta - \beta_0)' \frac{1}{n} \sum_{i=1}^{n} \delta(y_{1i}, y_{0i}, x_i) + o_p(1/n) \tag{A.54}$$

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 Khan(2000), that uniformly over β within $o_p(1)$ neighborhoods of β_0, we have

$$\Gamma_n(\beta) = \frac{1}{2}(\beta - \beta_0)' V_q (\beta - \beta_0) + o_p(1/n) \tag{A.55}$$

Combining these three results, the limiting distribution of the estimator follows by applying Lemma A.2 in Khan(2000). ■

A.6 Panel Data: Consistency:

Here, we prove theorem 5.1. We will only establish the key identification condition as before. We will take the same approach as before, though for the problem at hand we need to allow for the serial dependence in both the error terms and the censoring variables.

We first show that, conditioning on the variables x_{i1}, x_{i2},

$$P(d_{i1} = 1, v_{i1} \leq v_{i2}) \leq P(d_{i2} = 1, v_{i2} \leq v_{i1}) \tag{A.56}$$

iff $x_{i1}'\beta_0 \leq x_{i2}'\beta_0$. For notational convenience, we let z_{i1}, z_{i2} denote $x_{i1}'\beta_0 + \alpha_i, x_{i2}'\beta_0 + \alpha_i$ respectively.

We first evaluate

$$P(d_{i1} = 1, v_{i1} \leq v_{i2}) \tag{A.57}$$

where we condition on x_{i1}, x_{i2}. This probability can be decomposed into the mutually exclusive cases $c_{i1} > c_{i2}$, $c_{i1} \le c_{i2}$. We first focus on the first case $c_{i1} > c_{i2}$, and evaluate the probability conditional on the censoring values c_{i1}, c_{i2}.

Note we can decompose (A.57) as the sum of

$$P(v_{i1} \le v_{i2}, d_{i1} = 1, d_{i2} = 1) + P(v_{i1} \le v_{i2}, d_{i1} = 1, d_{i2} = 0) \tag{A.58}$$

We derive an expression for the first term, which we write here as:

$$P(\epsilon_{i2} \ge \epsilon_{i1} + \Delta z_i, \epsilon_{i1} \le c_{i1} - z_{i1}, \epsilon_{i2} \le c_{i2} - z_{i2}) \tag{A.59}$$

where here, $\Delta z_i \equiv z_{i1} - z_{i2}$. Recall that we are assuming for now that $c_{i1} > c_{i2}$, so we express the above probability as:

$$P(\epsilon_{i2} \ge \epsilon_{i1} + \Delta z_i, c_{i2} - z_{i1} \le \epsilon_{i1} \le c_{i1} - z_{i1}, \epsilon_{i2} \le c_{i2} - z_{i2}) + P(\epsilon_{i2} \ge \epsilon_{i1} + \Delta z_i, \epsilon_{i1} \le c_{i2} - z_{i1}, \epsilon_{i2} \le c_{i2} - z_{i2}) \tag{A.60}$$

The first piece above is zero, so

$$(A.59) = P(\epsilon_{i2} \ge \epsilon_{i1} + \Delta z_i, \epsilon_{i1} \le c_{i2} - z_{i1}, \epsilon_{i2} \le c_{i2} - z_{i2}) = \int_{-\infty}^{c_{i2} - z_{i1}} \int_{e_1 + \Delta z_i}^{c_{i2} - z_{i2}} f(e_1, e_2) de_2 de_1$$

Now, turning attention to the second term in (A.58), we express it as:

$$P(\epsilon_{i1} \le c_{i2} - z_{i1}, \epsilon_{i2} \ge c_{i2} - z_{i2}, \epsilon_{i1} \le c_{i1} - z_{i1}) = P(\epsilon_{i1} \le c_{i2} - z_{i1}, \epsilon_{i2} \ge c_{i2} - z_{i2}) \tag{A.61}$$

where the equality follows $c_{i1} > c_{i2}$. This is equal to

$$\int_{c_{i2} - z_{i2}}^{+\infty} \int_{-\infty}^{c_{i2} - z_{i1}} f(e_1, e_2) de_1 de_2 \tag{A.62}$$

Thus we have that conditioning on x_{i1}, x_{i2}, and $c_{i1} > c_{i2}$, (A.58) can be expressed as:

$$\begin{aligned} (A.58) &= \int_{-\infty}^{c_{i2} - z_{i1}} \int_{e_1 + \Delta z_i}^{c_{i2} - z_{i2}} f(e_1, e_2) de_2 de_1 + \int_{c_{i2} - z_{i2}}^{+\infty} \int_{-\infty}^{c_{i2} - z_{i1}} f(e_1, e_2) de_1 de_2 \\ &= \int_{-\infty}^{c_{i2} - z_{i1}} \int_{e_1 + \Delta z_i}^{\infty} f(e_1, e_2) de_2 de_1 = (A.57) \end{aligned} \tag{A.63}$$

where we have used exchangeability of the distribution of ϵ_1 and ϵ_2.

We next evaluate $P(d_{i2} = 1, v_{i2} \ge v_{i1})$, again conditioning on x_{i1}, x_{i2}, $c_{i1} > c_{i2}$. A similar decomposition yields:

$$P(v_{i2} \le v_{i1}, d_{i1} = 1, d_{i2} = 1) + P(v_{i2} \le v_{i1}, d_{i2} = 1, d_{i1} = 0) \tag{A.64}$$

The first term is:

$$P(\epsilon_{i1} \ge \epsilon_{i2} - \Delta z_i, \epsilon_{i1} \le c_{i1} - z_{i1}, \epsilon_{i2} \le c_{i2} - z_{i2}) = \int_{-\infty}^{c_{i2} - z_{i2}} \int_{e_2 - \Delta z_i}^{c_{i1} - z_{i1}} f(e_1, e_2) de_1 de_2$$

The second term in (A.64) is

$$P(\epsilon_{i2} \le c_{i1} - z_{i2}, \epsilon_{i2} \le c_{i2} - z_{i2}, \epsilon_{i1} \ge c_{i1} - z_{i1}) = P(\epsilon_{i2} \le c_{i2} - z_{i2}, \epsilon_{i1} \ge c_{i1} - z_{i1})$$

where the equality uses $c_{i1} > c_{i2}$. This can be expressed as:

$$\int_{-\infty}^{c_{i2}-z_{i2}} \int_{c_{i1}-z_{i1}}^{\infty} f(e_1, e_2) de_1 de_2$$

Hence, we have

$$\begin{aligned} P(d_{i2} = 1, v_{i2} \geq v_{i1}) &= \int_{-\infty}^{c_{i2}-z_{i2}} \int_{e_2-\Delta z_i}^{\infty} f(e_1, e_2) de_1 de_2 && \text{(A.65)} \\ P(d_{i1} = 1, v_{i1} \leq v_{i2}) &= \int_{-\infty}^{c_{i2}-z_{i1}} \int_{e_2+\Delta z_i}^{\infty} f(e_1, e_2) de_1 de_2 && \text{(A.66)} \end{aligned}$$

As we can see, the difference $P(d_{i2} = 1, v_{i2} \geq v_{i1}) - P(d_{i1} = 1, v_{i1} \leq v_{i2})$ has the same sign as $z_{i1} - z_{i2}$ since for the case where $z_{i1} > z_{i2}$ for example, the area of integration for $P(d_{i2} = 1, v_{i2} \geq v_{i1})$ is strictly larger than the area of integration for $P(d_{i1} = 1, v_{i1} \leq v_{i2})$.

This establishes the identification result for the case where $c_{i1} > c_{i2}$. For $c_{i1} \leq c_{i2}$, using analogous arguments, we find that

$$P(d_{i1} = 1, v_{i1} \leq v_{i2}) = \int_{-\infty}^{c_{i1}-z_{i1}} \int_{e_2+\Delta z_i}^{\infty} f(e_1, e_2) de_1 de_2 \tag{A.67}$$

and

$$P(d_{i2} = 1, v_{i2} \leq v_{i1}) = \int_{-\infty}^{c_{i1}-z_{i2}} \int_{e_2-\Delta z_i}^{\infty} f(e_1, e_2) de_1 de_2 \tag{A.68}$$

and the difference in the two integrals has the same sign as $z_{i1} - z_{i2}$ here as well. Thus after integrating over c_{i1}, c_{i2} we can conclude that

$$P(d_{i1} = 1, v_{i1} \leq v_{i2} | x_{i1}, x_{i2}) - P(d_{i2} = 1, v_{i2} \leq v_{i1} | x_{i1}, x_{i2})$$

has the same sign as $x'_{i1}\beta_0 - x'_{i2}\beta_0$. ■

TABLE I

Simulation Results for Rank Regression Estimators

One Sided CI Censoring Linear

	β			
	Mean Bias	Med. Bias	RMSE	MAD
100 obs.				
PRE	-0.0114	0.0000	0.1310	0.0972
MRC	-0.0934	0.0000	0.4477	0.2091
MRE	-0.1459	-0.0400	0.5408	0.2804
PMLE	0.1728	0.1029	0.5142	0.3650
200 obs.				
PRE	0.0039	0.0000	0.0745	0.0516
MRC	-0.0163	0.0000	0.1289	0.0867
MRE	-0.0442	0.0000	0.2199	0.1499
PMLE	0.2582	0.1569	0.5083	0.3653
400 obs.				
PRE	-0.0000	0.0000	0.0497	0.0347
MRC	-0.0056	0.0000	0.0764	0.0545
MRE	-0.0155	0.0000	0.1259	0.0939
PMLE	0.2713	0.1496	0.5241	0.3703

TABLE II

Simulation Results for Rank Regression Estimators

One Sided CI Censoring Exponential

	β			
	Mean Bias	Med. Bias	RMSE	MAD
100 obs.				
PRE	-0.0075	0.0000	0.1887	0.1407
MRC	-0.6863	-0.0800	1.6583	1.0402
MRE	-0.6388	-0.1200	1.7200	1.1142
PMLE	0.0071	0.0429	0.3426	0.2387
200 obs.				
PRE	-0.0071	0.0000	0.1136	0.0803
MRC	-0.6148	-0.0400	1.4565	0.8524
MRE	-0.5057	-0.0800	1.2994	0.7816
PMLE	0.1147	0.0923	0.3561	0.2454
400 obs.				
PRE	-0.0001	0.0000	0.0600	0.0420
MRC	-0.2794	0.0400	0.9462	0.4595
MRE	-0.2724	0.0400	0.9079	0.4960
PMLE	0.1103	0.0990	0.2885	0.2000

TABLE III

Simulation Results for Rank Regression Estimators

One Sided CD Censoring Linear

	β			
	Mean Bias	Med. Bias	RMSE	MAD
100 obs.				
PRE	-0.1738	0.1200	1.7905	1.3221
MRC	-3.9549	-3.9600	3.9550	3.9549
MRE	-3.9441	-3.9600	3.9445	3.9441
PMLE	-0.7720	0.1061	19.0136	3.4298
200 obs.				
PRE	-0.1423	0.0800	1.2551	0.8821
MRC	-3.9599	-3.9600	3.9599	3.9599
MRE	-3.9593	-3.9600	3.9593	3.9593
PMLE	-1.0761	0.0622	16.1674	1.8040
400 obs.				
PRE	-0.1289	0.0400	0.9102	0.6326
MRC	-3.9600	-3.9600	3.9600	3.9600
MRE	-3.9600	-3.9600	3.9600	3.9600
PMLE	-0.2322	-0.0649	1.0144	0.6270

TABLE IV

Simulation Results for Rank Regression Estimators

One Sided CD Censoring Exponential

	β			
	Mean Bias	Med. Bias	RMSE	MAD
100 obs.				
PRE	-0.0577	0.0000	0.2639	0.1293
MRC	0.3860	0.4000	0.3972	0.3860
MRE	0.7575	0.7200	0.8013	0.7575
PMLE	-0.9946	-0.3277	8.9472	1.6586
200 obs.				
PRE	-0.0237	0.0000	0.1073	0.0712
MRC	0.3867	0.4000	0.3930	0.3867
MRE	0.7992	0.7600	0.8323	0.7992
PMLE	-0.8173	-0.2140	12.3673	1.5778
400 obs.				
PRE	-0.0044	0.0000	0.0590	0.0401
MRC	0.3869	0.4000	0.3901	0.3869
MRE	0.8525	0.8400	0.8759	0.8525
PMLE	-0.2599	-0.2933	2.5544	1.1266

TABLE V

Simulation Results for Rank Regression Estimators

Two Sided Sided CI Censoring Linear

	β			
	Mean Bias	Med. Bias	RMSE	MAD
100 obs.				
PRE	-0.1302	0.0000	0.6601	0.4574
MRC	-0.2029	0.0800	0.9163	0.6302
MRE	-0.2146	0.0000	0.9010	0.6004
PMLE	0.3390	0.4022	0.6808	0.4701
200 obs.				
PRE	-0.0584	0.0000	0.4074	0.2970
MRC	-0.1189	-0.0400	0.5729	0.4084
MRE	-0.1152	0.0000	0.5586	0.3801
PMLE	0.3888	0.4338	0.4785	0.4304
400 obs.				
PRE	-0.0238	0.0000	0.2664	0.2008
MRC	-0.0429	0.0000	0.3850	0.2757
MRE	-0.0584	0.0000	0.3762	0.2676
PMLE	0.4322	0.4508	0.4677	0.4368

TABLE VI

Simulation Results for Rank Regression Estimators

Two Sided CD Censoring Linear

	β			
	Mean Bias	Med. Bias	RMSE	MAD
100 obs.				
PRE	-0.3144	-0.0800	1.0762	0.7481
MRC	-3.9424	-3.9600	3.9428	3.9424
MRE	-3.9331	-3.9600	3.9337	3.9331
PMLE	-0.8615	-0.6787	3.9160	1.6850
200 obs.				
PRE	-0.3187	-0.1600	0.7808	0.5240
MRC	-3.9571	-3.9600	3.9571	3.9571
MRE	-3.9575	-3.9600	3.9575	3.9575
PMLE	-0.8166	-0.7717	1.5573	1.2185
400 obs.				
PRE	-0.1937	-0.1200	0.4757	0.3423
MRC	-3.9600	-3.9600	3.9600	3.9600
MRE	-3.9600	-3.9600	3.9600	3.9600
PMLE	-0.6579	-0.8208	1.2974	1.1231