

Partial rank estimation of duration models with general forms of censoring

Shakeeb Khan^a, 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*

Available online 19 April 2006

Abstract

In this paper we propose estimators for the regression coefficients in censored duration models which are distribution free, impose no parametric specification on 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, and fixed (individual or group specific) effects. We also examine the behavior of the estimator in an empirical illustration.

© 2006 Elsevier B.V. All rights reserved.

JEL classification: C13; C31

Keywords: Rank correlation; Duration analysis; Covariate-dependent censoring; Panel data

1. Introduction

This paper considers estimation of regression coefficients in censored transformation models that arise from duration analysis. 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

*Corresponding author. Tel.: +1 8474918218; fax: +1 8474917001.

E-mail address: tamer@northwestern.edu (E. Tamer).

intervals between child births, and insurance claim durations, to name a few. (Van den Berg, 2001 surveys the many applications of duration models.)

Since the seminal work in Cox (1972, 1975), the most widely used model in duration analysis is 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 is expressed as

$$T(y_i) = x_i' \beta_0 + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (1.1)$$

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 an object of interest to be estimated from a random sample of n observations.

Complications arise since duration data is often subject to right censoring for a variety of reasons that are 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. Also, to control data collection costs, unemployed agents are only followed for a short period of time. If they are still unemployed at the end of this period, their spell length is censored. Another example is data on welfare spells for programs such as TANF (temporary assistance to needy families) where the time in between welfare collection and employment will be censored by the time entitlement expires.

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 as¹

$$T(v_i) = \min(x_i' \beta_0 + \varepsilon_i, c_i), \quad (1.2)$$

$$d_i = I[x_i' \beta_0 + \varepsilon_i \leq c_i], \quad (1.3)$$

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 that the censoring variable need not always be observed.

The primary aim of this paper is to provide an estimator, up to scale, 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, 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,

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

covariate-dependent censoring also arises in the biostatistics literature on competing risks, and survival analysis (even for randomized clinical trials—see Chen, Jin and Yin (2002)). We consider right censoring as the base case and show how our approach can be extended to cover double censoring (and naturally left censoring).

We motivate this paper in 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 and James (1979), Koul, Susarla and Van Ryzin (1981), Tsiatis (1990), Ying, Jung and Wei (1995), Yang (1999), Honoré, Khan and Powel (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 Yin (2002). Cheng, Wei and Ying (1995), Fine, Ying and Wei (1998), and 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),² do not impose distributional assumptions on ε_i , but for consistency requires that the censoring variable be independent of the covariates.³ 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.⁴ The estimator we provide in

²See also Hausman and Woutersen (2005) for recent work in this area.

³One notable exception in the single index literature is the rank estimator proposed in Abrevaya (1999a,b). There he considered a truncated transformation model with observed truncation points that could be correlated with regressors. As the truncation points need always be observed, his procedure is not applicable to the random censoring/competing risk models considered in this paper. He also required the error distribution to satisfy a log-concavity condition not typically imposed in the single index literature. This assumption can be quite restrictive, especially in a duration analysis context—for example, it imposes a monotonic hazard rate.

⁴If the censoring distribution depends on the regressors through the index $x_i'\beta_0$, some single index estimators may be applied, though we consider this too restrictive of a condition. Also note that Gorgens and Horowitz (1998) allow for general covariate-dependent censoring in their estimator of the link function $T(\cdot)$. They do not provide an estimator of β_0 , and only suggest estimation by an existing single index estimator. As mentioned above, this will not yield consistent estimates if the censoring depends arbitrarily on x_i .

this paper allows for general forms of censoring, including left, right, and double sided random 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 up to scale. We then show that the estimator is consistent and derive its asymptotic distribution. Sections 3 and 4 extend our estimator to cases of doubly censored and to group effects data (panel). Section 5 explores the finite sample properties of the new estimators by means of a small scale simulation study. Section 6 applies our estimator to the Stanford heart transplant data and compares our estimator to existing ones. 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 transformation model is motivated by existing rank estimators for β_0 in uncensored transformation models, specifically Han (1987)'s maximum rank correlation (MRC) estimator.⁵ This estimator, like other estimators in the single index literature (e.g. Powell, 1994; 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

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

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 *observed* random variables⁶

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

where multiplication of 0 by $+\infty$ is defined to be 0. Note we have that

$$y_{0i} \leq y_i \leq y_{1i}, \quad (2.3)$$

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

$$f_{ij} = y_{1i} - y_{0j}, \quad (2.4)$$

⁵A 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.

⁶The random variable y_1 is defined on the extended real line with the Borel field extended to include the singletons $\{-\infty\}$ and $\{+\infty\}$.

$$I[f_{ij} \geq 0] = I[y_{1i} - y_{0j} \geq 0] = (1 - d_i) + d_i I[v_i \geq v_j]. \tag{2.5}$$

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

$$T(y_{0i}) \leq x'_i \beta_0 + \varepsilon_i \leq T(y_{1i}).$$

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

$$x'_i \beta \geq x'_j \beta \Rightarrow P(y_{1i} \geq y_{0j}) \geq \frac{1}{2}. \tag{2.6}$$

Before describing the estimator in detail, we note that the object of interest β_0 is only identified up to scale as the function $T(\cdot)$ is unknown. We first assume that the first component of β_0 is not equal to zero and we set it to 1 (normalization). Hence, we express $\beta_0 = (1, \theta'_0)'$ and consider estimation of θ_0 . So, for any $\theta \in \Theta$, we let β denote $(1, \theta)'$.

The censoring robust rank estimator, which we refer to hereafter as the *partial rank estimator* (PRE),⁷ is of the form

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

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:

$$\frac{1}{n(n-1)} \sum_{i \neq j} I[y_{0i} > y_{1j}] I[x'_i \beta > x'_j \beta] = \frac{1}{n(n-1)} \sum_{i \neq j} d_i I[v_i < v_j] I[x'_i \beta < x'_j \beta]. \tag{2.8}$$

To see why, start with $I[y_{1i} \geq y_{0j}] I[x'_i \beta \geq x'_j \beta] = (1 - I[y_{0j} > y_{1i}]) * (1 - I[x'_j \beta > x'_i \beta])$. This is equal to $1 + I[y_{0j} > y_{1i}] * I[x'_j \beta > x'_i \beta] + I[y_{0j} > y_{1i}] - I[x'_j \beta > x'_i \beta]$. This implies that $\sum_{i \neq j} I[y_{1i} \geq y_{0j}] I[x'_i \beta \geq x'_j \beta] = \sum_{i \neq j} 1 + I[y_{0j} > y_{1i}] * I[x'_j \beta > x'_i \beta] - I[y_{0j} > y_{1i}] - I[x'_j \beta > x'_i \beta]$ which is equal to $\sum_{i \neq j} I[y_{0j} > y_{1i}] * I[x'_j \beta > x'_i \beta]$ plus terms that do not vary with the choice of β . (Note even the term $I[x'_j \beta > x'_i \beta]$ does not vary with β when summed across pairs $i \neq j$.) Therefore, maximizing (2.8) is numerically equivalent to maximizing (2.7) above.

Expressed as in (2.8), 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 directly, 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 than i 's spell length) are used in evaluating the likelihood. The risk set may contain censored observations. Analogously, for the PRE only uncensored observations enter the rank correlation function directly, and for a given uncensored observation, all the observations in its risk set are used to determine its rank. Like the PMLE, the PRE uses information on both censored and uncensored observations.

⁷Our motivation for using the term partial rank estimator is that it has properties that are analogous to the PMLE in Cox (1975), as we explain below.

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 Ryzin (1981), 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. This involves showing that the objective function above point identifies β_0 . To do that, we list next a set of assumptions that are sufficient for consistency of the above estimator.

- (I1) The random variable ε_i is distributed independently of the random vector $(c_i, x_i)'$.
- (I2) The first component of x_i has a density with respect to Lebesgue measure that is positive everywhere, conditional on the other components.
- (I3) 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 $\Pr(\mathcal{X}_{uc}) > 0$.

- (I4) The vector $\mathbf{z}_i = (d_i, v_i, x_i)$, $i = 1, 2, \dots, n$ are i.i.d.
- (I5) Θ is a compact subset of \mathbf{R}^{k-1} .
- (I6) \mathcal{X}_{uc} is not contained in any proper linear subspace of R^k .

The independence assumption (I1) 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. Independence between c and ε is important for point identification. Allowing for these variables to be correlated, though interesting, requires additional assumptions like exclusion restrictions. On the other hand, the independence between ε and x is similar to independence assumption in the MRC estimator and can be relaxed. It is possible to allow for (conditional) heteroskedasticity⁸ for example. This will come at the expense of stronger (sufficient) point identification conditions.⁹ Assumption (I2) 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 (1985). Generally, the conditions for point identification of β are similar to ones used in many semiparametric models in econometrics: only the direction of β is identified and one needs some continuity in the support of $x_i'\beta$. Typically, the later condition is guaranteed by assuming that one regressor with a nonzero coefficient has wide support. This condition can be weakened to requiring that $x_i'\beta$ be continuous in a region of the support. Condition (I3) requires that the probability of censoring is not equal to one for all x . Conditions (I4) and (I5) are commonly imposed in these types of models. Condition (I6) is stronger than a full rank condition, but is usually required in settings where censoring is

⁸For results in this case, see Khan and Tamer (2005).

⁹On the other hand, it is not generally possible to relax the conditional independence of c and ε without additional assumptions like exclusion restrictions (or functional form assumptions).

present.¹⁰ Both (I3) and (I6) guarantee that the objective function is *uniquely* minimized at the truth. 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

$$\tau(\mathbf{z}_i, \theta) = E[d_i I[v_i < v] I[x_i' \beta < x' \beta]] + E[d_i I[v < v_i] I[x' \beta < x_i' \beta]].$$

Finally, we let \mathcal{N} denote a neighborhood of θ_0 . Next, we list a set of additional assumptions that are needed in the theorem.

- (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)–(I6), and (A1)–(A4),*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0, V^{-1} \Delta V^{-1}), \quad (2.9)$$

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

Remark 2.2. As is the case with other rank based estimators—e.g. Han (1987), Cavanagh and Sherman (1998), the variance matrix can be expressed in terms of “model primitives”, such as the conditional probability function and the density of the index. Let $g_0(\cdot)$ denote the marginal density function of the index $x_i' \beta_0$, let \tilde{x}_i denote regressors excluding the first one (whose coefficient is normalized to 1), and let $\mu_0(x_i' \beta_0)$ denote the conditional expectation of \tilde{x} given $x_i' \beta_0$. Furthermore, define the function:

$$S(d, v, t) = E[d_i I[v < v_i] - d_i I[v_i < v] | x_i' \beta_0 = t] \quad (2.10)$$

¹⁰See for example condition R1 on page 256 of Honoré and Powell (1994).

and we note, by the proof of Theorem 2.1, that $S(\cdot, \cdot, \cdot)$ satisfies the following moment condition:

$$E[S(d_i, v_i, t)] = 0 \quad \forall t \in \mathbf{R}. \quad (2.11)$$

Then we have the following representation of the variance covariance matrix which can be easily shown using arguments similar to ones given in Theorem 4 in Sherman (1993). Specifically, if $S(\cdot, \cdot, \cdot)$ is differentiable with respect to its third argument, $g_0(\cdot)$ is differentiable, and $E[\|x_i\|^2] < \infty$, then

$$\Delta = E_X[(\tilde{x}_i - \mu(x'_i\beta_0))(\tilde{x}_i - \mu(x'_i\beta_0))' S(d_i, v_i, x'_i\beta_0)^2 g_0(x'_i\beta_0)], \quad (2.12)$$

$$V = \frac{1}{2} E_X[(\tilde{x}_i - \mu_0(x'_i\beta_0))(\tilde{x}_i - \mu_0(x'_i\beta_0))' S_3(d_i, v_i, x'_i\beta_0) g_0(x'_i\beta_0)], \quad (2.13)$$

where $S_3(\cdot, \cdot, \cdot)$ denotes the partial derivative of $S(\cdot, \cdot, \cdot)$ with respect to its third argument. The above representation facilitates illustrating which features of the data effect the variance of the estimator. The components are analogous to the components in the variance matrix of an LAD estimator. This is not surprising as the original MRC can be interpreted as an LAD estimator, with the median of the sign of $(y_i - y_j)$, being equal to the sign of $(x'_i\beta_0 - x'_j\beta_0)$, conditional on $y_i \neq y_j$, see, e.g. Powell, 1994. The analogy is most prevalent in the Hessian term, where the function $S_3(\cdot, \cdot, \cdot)$ is directly related to density of the error difference term $(\varepsilon_i - \varepsilon_j)$ at 0. A similar density function appears in the Hessian term in quantile regressions.

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. Our proposed estimator is of the numerical derivative form in Pakes and Pollard (1989) and Sherman (1993).

First define the function of θ and $\mathbf{z} = (d, v, x)$

$$\tau_n(\mathbf{z}, \theta) = \frac{1}{n} \sum_{i=1}^n d_i I[v_i < v] I[x'_i\beta < x'\beta] + d I[v < v_i] I[x'\beta < x'_i\beta], \quad (2.14)$$

whose numerical derivative we will define as

$$p_{n\tilde{k}}(\mathbf{z}, \theta) = \varepsilon_n^{-1} (\tau_n(\mathbf{z}, \theta + \varepsilon_n \mathbf{1}_{\tilde{k}}) - \tau_n(\mathbf{z}, \theta)), \quad (2.15)$$

where ε_n is a sequence of positive numbers converging to 0, and $\mathbf{1}_{\tilde{k}}$ denotes a vector the same dimension as θ whose \tilde{k} th component is 1 and remaining components are 0. From this we can define our estimator of the matrix Δ as $\hat{\Delta}$, a matrix whose (m, q) component we define as

$$\hat{\delta}_{mq} = \frac{1}{n} \sum_{i=1}^n p_{nm}(\mathbf{z}_i, \hat{\theta}) p_{nq}(\mathbf{z}_i, \hat{\theta}). \quad (2.16)$$

To estimate the Hessian term V , define the function

$$p_{n\tilde{k}l}(\mathbf{z}, \theta) = \varepsilon_n^{-2} (\tau_n(\mathbf{z}, \theta + \varepsilon_n (\mathbf{1}_{\tilde{k}} + \mathbf{1}_l)) - \tau_n(\mathbf{z}, \theta + \varepsilon_n \mathbf{1}_{\tilde{k}}) - \tau_n(\mathbf{z}, \theta + \varepsilon_n \mathbf{1}_l) + \tau_n(\mathbf{z}, \theta)), \quad (2.17)$$

from which we can define the estimator of V as \hat{V} , a matrix whose (m, q) element is

$$\hat{v}_{mq} = \frac{1}{n} \sum_{i=1}^n p_{nmq}(\mathbf{z}_i, \hat{\theta})/2. \quad (2.18)$$

The following theorem, whose proof is left to the appendix, establishes the consistency of the estimator:

Theorem 2.3. *If the sequence ε_n satisfies $\varepsilon_n \rightarrow 0$ and $n^{1/4}\varepsilon_n \rightarrow \infty$, then*

$$\hat{V}^{-1} \hat{\Delta} \hat{V}^{-1} \xrightarrow{p} V^{-1} \Delta V^{-1}. \quad (2.19)$$

An alternative strategy to conducting inference is to adopt resampling methods. Unfortunately, it has not been formally established that the bootstrap is asymptotically valid in this setting. However, other sampling procedures are possible. For one, a recent “wild” sampling procedure introduced in Jin, Ying and Wei (2002) appears likely to be applicable (after appropriate modifications)¹¹ to the problem at hand. A further resampling option for conducting inference would be subsampling, and its validity in this setting would follow from Theorem 2.2.1 in Politis, Romano and Wolf (1999).

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 (1995) 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 Tsiatis (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 $y_i \equiv T^{-1}(x_i'\beta_0 + \varepsilon_i)$. Instead one observes the doubly censored sample, which we can express as the pair (v_i, d_i) where

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

¹¹The approach in Jin, Ying and Wei (2002) applies to M-estimators, whereas our rank estimator minimizes a U-process. Consequently, random draws of order n^2 from a mean 1 random variable should be generated, as opposed to order n .

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

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 even more 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

$$y_{1i} = I[d_i < 3]v_i + I[d_i = 3] \cdot +\infty, \quad (3.1)$$

$$y_{0i} = I[d_i \neq 2]v_i + I[d_i = 2] \cdot -\infty \quad (3.2)$$

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

$$f_{ij} = y_{1i} - y_{0j},$$

$$I[f_{ij} \geq 0] = I[d_i = 3] + I[d_j = 2] - (I[d_i = 3] * I[d_j = 2]) \\ + (I[d_i = 1] + I[d_i = 2]) * (I[d_j = 1] + I[d_j = 3])I[v_i \geq v_j].$$

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

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} I[y_{1i} \geq y_{0j}] I[x_i' \beta \geq x_j' \beta] \quad (3.3)$$

$$= \arg \max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} ((d_{1i} + d_{2i}) \cdot (d_{1j} + d_{3j}) I[v_i \geq v_j] \\ + (d_{3i} + d_{2j} - d_{3i}d_{2j})) I[x_i' \beta \geq x_j' \beta], \quad (3.4)$$

which, as before, is numerically equivalent to

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} I[y_{0i} > y_{1j}] I[x_i' \beta > x_j' \beta] \quad (3.5)$$

$$= \arg \max_{\theta \in \Theta} \frac{1}{n(n-1)} \sum_{i \neq j} (1 - I[d_i = 2])(1 - I[d_j = 3]) I[v_i > v_j] I[x_i' \beta > x_j' \beta]. \quad (3.6)$$

We first establish a result that shows that the objective function is uniquely minimized at the truth. This is similar to the one given in Theorem 2.1 (and its proof). The identification result is again based on Assumptions (I1)–(I6), 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) and (I6), if either the covariates are independent of the censoring variables or we have the condition*

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

then we have

$$\mathbf{P}(y_{0i} > y_{1j} | x_i, x_j) > \mathbf{P}(y_{0j} > y_{1i} | x_i, x_j) \quad \text{iff } x'_i \beta_0 > x'_j \beta_0. \quad (3.8)$$

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) and (I5). To establish limiting distribution theory, based on Assumptions (AD1)–(AD4), we first introduce some further notation for the doubly censored case. Now \mathbf{z}_i denotes the vector $(d_{1i}, d_{2i}, d_{3i}, v_i, x'_i)'$, and we define

$$\begin{aligned} \tau_d(\mathbf{z}, \theta) = & \mathbf{E}[(1 - d_2)(1 - d_3)I[v > v_i]I[x'\beta > x'_i\beta]] \\ & + \mathbf{E}[(1 - d_2)(1 - d_3)I[v_i > v]I[x'_i\beta > x'\beta]]. \end{aligned}$$

Finally, we let \mathcal{N} denote a neighborhood of θ_0 . The assumptions used for the distribution theory are:

- (AD1) θ_0 lies in the interior of Θ , a compact subset of \mathcal{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 \mathbf{z} .

- (AD3) $\mathbf{E}[\|\nabla_1 \tau_d(\mathbf{z}_i, \theta_0)\|^2]$ and $\mathbf{E}[\|\nabla_2 \tau_d(\mathbf{z}_i, \theta_0)\|]$ are finite.
- (AD4) $\mathbf{E}[\nabla_2 \tau_d(\mathbf{z}_i, \theta_0)]$ 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 3.1. *Under Assumptions (AD1)–(AD4),*

$$\sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow \mathbf{N}(0, V_d^{-1} \Delta_d V_d^{-1}), \quad (3.9)$$

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

4. Extension II: group specific effects

As is the case with duration data,¹² panel data or models with group specific effects have received an increasing amount of attention in the econometric literature—see Arellano and Honore, 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 unemployment spells (Heckman and Borjas, 1980), time intervals between child births (Newman and McCulloch, 1984) and car insurance claim durations (Abbring, Chiappori and Pinquet (2003)). Examples in the

¹²We thank Bo Honoré for suggesting the panel data extension to us.

second case would include survival times of children in a family in a developing country (Ridder and Tunali, 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/group 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 et al. (2002) allow 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. Extensions of the linear specification can be found in Abrevaya (1999a,b, 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 Tunali (1999), Lancaster (2000), Horowitz and Lee (2004) and Lee (2003). Some of these also rule out censoring distributions that vary across spells and/or are independent of covariates.¹³

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:

$$T_i(v_{it}) = \min(\alpha_i + x'_{it}\beta_0 + \varepsilon_{it}, c_{it}),$$

$$d_{it} = I[\alpha_i + x'_{it}\beta_0 + \varepsilon_{it} \leq c_{it}], \quad i = 1, 2, \dots, n, \quad t = 1, 2, \dots, \tau$$

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 $n \rightarrow \infty$. Consequently, we wish to

¹³Honoré (1993), Horowitz and Lee (2004) 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 spell data. Therefore, the independence assumption considered here is generally better suited for analyzing data with group specific effects. In the individual specific setting, the independence assumption would be suitable if the censoring mechanism is due not to data collection limitations, but a process such as that described for welfare spells (TANF) mentioned earlier in the paper.

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 $\hat{\theta}_p$, and referred to hereafter as the *censored duration panel* (CDP) estimator, is of the form¹⁴

$$\hat{\theta}_p = \arg \max_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n d_{i1} I[v_{i1} < v_{i2}] I[x'_{i1} \beta(\theta) < x'_{i2} \beta(\theta)] + d_{i2} I[v_{i2} < v_{i1}] I[x'_{i2} \beta(\theta) < x'_{i1} \beta(\theta)], \quad (4.1)$$

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

We establish 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) Let S_{X_i} denote the support of x_{it} , and let \mathcal{X}_{uct} denote the set

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

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

(P2) The random variables $\varepsilon_{i1}, \varepsilon_{i2}$ are exchangeable and are independent of the vector $(c_{i1}, x'_{i1}, c_{i2}, x'_{i2})$.

Each has a distribution with support on 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}) , $i = 1, 2, \dots, n$ are i.i.d.

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

(P6) \mathcal{X}_{uct} is not contained in any proper linear subspace of \mathbf{R}^k .

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

Theorem 4.1. *Under Assumptions (P1)–(P6), the CDP estimator is consistent:*

$$\hat{\theta}_p \xrightarrow{p} \theta_0. \quad (4.2)$$

Inference on parameters requires limiting distribution theory. We note that 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.

¹⁴This 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 (1999a,b). Comparing these values across time intuitively in a “maximum score” (Manski, 1985) type setting leads to consistent estimation of θ_0 for binary choice (Manski, 1987) models and transformation models with fixed censoring (Abrevaya, 1999a,b). 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.

5. 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

$$T(y_i) = \alpha_0 + x_{1i}\theta_0 + x_{2i} + \varepsilon_i,$$

where x_{1i}, x_{2i} are distributed as a chi-squared with one degree of freedom, and standard normal, respectively; α_0, θ_0 were set to 1 and -1 , respectively. We considered two functional forms for $T(\cdot)$ and the error distribution as follows:

1. $T^{-1}(v) = v$; $\varepsilon_i \sim$ mixture of two normals, centered around -1 and 2 , respectively.
2. $T^{-1}(v) = \exp(v)$; $\varepsilon_i \sim$ 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 $2.05 \exp(x_{1i}^2 \cdot x_{2i} + x_{2i})$ 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 1 through 6 we report results for four 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

Table 1
Simulation results for rank regression estimators one sided CI censoring linear

	θ			
	Mean bias	Med. bias	RMSE	MAD
<i>100 observations</i>				
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
<i>200 observations</i>				
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
<i>400 observations</i>				
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 2
Simulation results for rank regression estimators one sided CI censoring exponential

	θ			
	Mean bias	Med. bias	RMSE	MAD
<i>100 observations</i>				
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
<i>200 observations</i>				
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
<i>400 observations</i>				
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

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

Table 3
Simulation results for rank regression estimators one sided CD censoring linear

	θ			
	Mean bias	Med. bias	RMSE	MAD
<i>100 observations</i>				
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
<i>200 observations</i>				
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
<i>400 observations</i>				
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 4
Simulation results for rank regression estimators one sided CD censoring exponential

	θ			
	Mean bias	Med. bias	RMSE	MAD
<i>100 observations</i>				
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
<i>200 observations</i>				
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
<i>400 observations</i>				
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

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 PMLEs inconsistency (due to the error distribution misspecification) is also apparent, though not as pronounced in the linear design at 400

Table 5
Simulation results for rank regression estimators two sided CI censoring linear

	θ			
	Mean bias	Med. bias	RMSE	MAD
<i>100 observations</i>				
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
<i>200 observations</i>				
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
<i>400 observations</i>				
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 6
Simulation results for rank regression estimators two sided CD censoring linear

	θ			
	Mean bias	Med. bias	RMSE	MAD
<i>100 observations</i>				
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
<i>200 observations</i>				
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
<i>400 observations</i>				
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

observations. Its RMSE values are much larger than the PREs 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 RMSEs 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.

6. Empirical illustration: Stanford heart transplant data

We consider the Stanford heart transplant data taken from Kalbfleisch and Prentice (1980, Appendix 1, pp. 230–232). Summarized in this data set are the survival times of 103 patients who received heart transplants at the Stanford University Medical Center, as well as an indicator variable which equals one if the patient was dead (uncensored) at the time the data were collected, the age of the patient (in years) at the time of the transplant and other variables. We estimate the following model of the survival times,

$$T(v_i) = \min\{\alpha_0 + \beta_0 x_i + \gamma_0 x_i^2 + \varepsilon_i, c_i\}, \quad (6.1)$$

where the dependent variable v_i is the observed survival time (in days), x_i is age of patient i and x_i^2 is the age squared. For this model, covariate-dependent censoring seems quite plausible. Larger censoring times correspond to earlier transplants; if transplants for younger or older patients were not typically performed in the earlier years, this would induce a dependence between censoring and the covariate age. We standardize the coefficient on age to one and provide estimates using the PRE, MRC, MRE, and a set of other estimators which require a known $T(\cdot)$. Setting $T(\cdot)$ equal to the logarithmic function, base 10 we compute estimates using procedures introduced in Honoré, Khan and Powell (2002), Yang (1999), Ying, Jung and Wei (1995), and Buckley and James (1979). For detailed description of each of these procedures and algorithms for their computation, see Honoré, Khan and Powell (2002). Table 7 summarizes our results. In addition to providing point estimates, we estimate we provide the 95% confidence interval based on taking 200 bootstrap pseudo samples from the data. The PRE, which imposes the fewest restrictions on the model, has the smallest point estimate of the quadratic effect, though its

Table 7
Stanford heart data estimation results

Age ²	Parameter	95% Bootstrap CI
PRE	-.0178	[-.0243, -.0105]
MRC (1987)	-.0140	[-.0370, -.0105]
MRE (1998)	-.0147	[-.0156, -.0105]
Honoré et al. (2002)	-.0134	[-.0162, -.0111]
Yang (1999)	-.0002	[-.0003, -.0000]
Ying et al. (1995)	-.0001	[-.0003, -.0000]
Buckley–James (1979)	-.0018	[-.0121, -.0008]

interval estimate does not give the smallest bounds, as its interval is contained inside that obtained from the MRC. The PRE, MRC, MRE and Honoré, Khan and Powell (2002) all conclude that the quadratic effect is significantly negative, whereas the remaining procedures yield results that are not significantly different from 0. This contrasting result may be due to misspecification of functional form of $T(\cdot)$ since both the Yang (1999) and the Buckley and James (1979) estimators allow for covariate dependent censoring. The similarity in results for the PRE and MRC seem to indicate that covariate censoring is not that relevant an issue for this particular data set.

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, and group specific 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 three such examples. For one, it would be useful to construct an estimator for the function $T(\cdot)$ based on y_{0i}, y_{1i} that modifies the rank estimator of $T(\cdot)$ in Chen (2002) to allow for covariate-dependent censoring. Second, it is possible to extend the analysis above to include heteroskedasticity. This involves imposing a conditional median assumption on the error. Imposing weaker assumption on the conditional error distribution comes at the cost of stronger sufficient point identification conditions (see Khan and Tamer, 2005 for more on this model). Finally, another area for future work would be to determine if the proposed group specific effects estimator attains the fastest rate of convergence possible under the assumptions of the model.

Acknowledgements

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.

Appendix A

A.1. Proof of Theorem 2.1

A crucial step in showing consistency is identification. To do that, we first show that (2.1) holds, i.e., that

$$\mathbf{P}(y_{1i} \geq y_{0j} | x_i, x_j) - \mathbf{P}(y_{1j} \geq y_{0i} | x_i, x_j) \quad (\text{A.1})$$

relates to the sign of $x'_i\beta_0 - x'_j\beta_0$. This will be equivalent to showing how

$$P(y_{0i} > y_{1j} | x_i, x_j) - P(y_{0j} > y_{1i} | x_i, x_j) \quad (\text{A.2})$$

relates to the sign of $x'_i\beta_0 - x'_j\beta_0$. For notational convenience, we let z_i, z_j denote $x'_i\beta_0, x'_j\beta_0$, respectively. We first evaluate

$$P(y_{0i} > y_{1j}), \quad (\text{A.3})$$

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

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

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

$$P(\varepsilon_i > \varepsilon_j - \Delta z, \varepsilon_i \leq c_i - z_i, \varepsilon_j \leq c_j - z_j),$$

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 (II), we express the above probability as

$$\int_{-\infty}^{c_j - z_j} \int_{\varepsilon_j - \Delta z}^{c_i - z_i} dF(\varepsilon_i) dF(\varepsilon_j), \quad (\text{A.5})$$

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

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

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

$$P(\varepsilon_j \leq c_j - z_j, \varepsilon_i > c_i - z_i, \varepsilon_j \leq c_i - z_j) = P(\varepsilon_i > c_i - z_i, \varepsilon_j \leq c_j - z_j), \quad (\text{A.7})$$

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

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

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

$$\int_{-\infty}^{c_j - z_j} S(\varepsilon_j - \Delta z) dF(\varepsilon_j), \quad (\text{A.9})$$

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

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

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

The first term is

$$P(\varepsilon_i < \varepsilon_j - \Delta z, \varepsilon_i \leq c_i - z_i, \varepsilon_j \leq c_j - z_j), \quad (\text{A.11})$$

which we can decompose into the sum of

$$P(\varepsilon_i < \varepsilon_j - \Delta z, \varepsilon_i \leq c_j - z_i, \varepsilon_j \leq c_j - z_j) + P(\varepsilon_i < \varepsilon_j - \Delta z, c_j - z_i < \varepsilon_i \leq c_i - z_i, \varepsilon_j \leq c_j - z_j). \quad (\text{A.12})$$

Note the second term is 0 (since $\varepsilon_i < \varepsilon_j - \Delta z$ and $\varepsilon_j \leq c_j - z_j$ contradicts the middle event), and the first term is

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

which is equal to

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

The second term in (A.10) is

$$P(c_j \geq \varepsilon_i + z_i, \varepsilon_j > c_j - z_j, \varepsilon_i \leq c_i - z_i) = P(\varepsilon_j > c_j - z_j, \varepsilon_i \leq c_j - z_i), \tag{A.14}$$

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

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

We note that the above difference has the same sign as $z_i - z_j$ if $P(\varepsilon_j \leq c_j - z_j | c_j, z_j) > 0$. Note that letting e_* denote the lower support of ε_j , if $e_* > \max(c_j - z_j, c_j - z_i)$, both of the above integrals are zero as is their difference, regardless of the sign of $z_i - z_j$. However, by assumption (I3), we know there exists (c_j, z_j) with positive probability where $P(\varepsilon_j \leq c_j - z_j | c_j, z_j) > 0$. This establishes the identification result for the case where $c_i > c_j$. For $c_i \leq c_j$, using analogous arguments, we find that

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

and

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

and the difference in the two integrals has the same sign as $z_i - z_j$ here as well whenever the lower support point of ε_i is not strictly greater than $c_i - z_i$. Thus, after integrating over c_i, c_j we can conclude that

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

has the same sign as $x'_i \beta_0 - x'_j \beta_0$.

We know use the above to show that the objective function is uniquely maximized at the truth. The sample objective function (2.7) can be written as

$$\begin{aligned} & \frac{1}{n(n-1)} \sum_{i \neq j} I[y_{1i} \geq y_{0j}] I[x'_i \beta \geq x'_j \beta] \\ &= \frac{1}{n(n-1)} \sum_{i < j} (I[y_{1i} \geq y_{0j}] I[x'_i \beta \geq x'_j \beta] + I[y_{1j} \geq y_{0i}] I[x'_j \beta \geq x'_i \beta]). \end{aligned}$$

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

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

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):

$$\begin{aligned} Q(\beta_0) - Q(\beta) &= E_X[(\mathbb{P}(y_{1i} \geq y_{0j}) - \mathbb{P}(y_{1j} \geq y_{0i})) (I[x'_i \beta_0 \geq x'_j \beta_0] - I[x'_i \beta \geq x'_j \beta])] \\ &= E_X[(\mathbb{P}(y_{1i} \geq y_{0j}) - \mathbb{P}(y_{1j} \geq y_{0i})) (I[\Delta x' \beta_0 \geq 0 > \Delta x' \beta] \\ &\quad - I[\Delta x' \beta \geq 0 > \Delta x' \beta_0])]. \end{aligned} \tag{A.19}$$

By (2.1) and its proof above, 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 even when restricting ourselves to $\mathcal{X}_{uc} \times \mathcal{X}_{uc}$, with positive probability, $\Delta x^{(-1)' \theta \neq \Delta x^{(-1)' \theta_0}$, 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), the 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 $I[x'_i \beta_0 < x'_j \beta_0, x'_i \beta > x'_j \beta] = 1$ has positive probability. But from (A.16) to (A.18), on this subset $\mathbb{P}(y_{1i} \geq y_{0j} | x_i, x_j) > \mathbb{P}(y_{1j} \geq y_{0i} | x_i, x_j)$, (A.19) is strictly positive, and hence establishing that the limiting objective function is uniquely maximized at β_0 and proving identification.

What remains for consistency (see e.g. Newey and McFadden, 1994) is to satisfy compactness, uniform convergence, and continuity. We note that compactness holds by Assumption (I5). Regarding uniform convergence, we need to show

$$\sup_{\beta \in \mathcal{B}} Q_n(\beta) \xrightarrow{p} Q(\beta), \tag{A.20}$$

where $Q_n(\beta)$ is the sample objective function defined in (2.8). Eq. (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 (1994b) 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 (I2). This establishes consistency. \square .

A.2. 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

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

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

$$f(\mathbf{z}_1, \mathbf{z}_2, \theta) = I[y_{01} > y_{12}]I[x'_1\beta > x'_2\beta] \tag{A.22}$$

$$= (d_2 I[v_1 > v_2])I[x'_1\beta > x'_2\beta], \tag{A.23}$$

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

Note the class of functions

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

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). \square

A.3. Proof of Theorem 2.3

We notice immediately that $E[\tau_n(\mathbf{z}, \theta)] = \tau(\mathbf{z}, \theta)$. The function inside the summation in the definition of $\tau_n(\cdot, \theta)$ are Euclidean with respect to a constant envelope by the same arguments as in the proof of the previous theorem, so by letting S denote the support of the random vector \mathbf{z}_i , it follows from Corollary 7 in Sherman (1994a,b) that

$$\sup_{S \times \Theta} \|\tau_n(\mathbf{z}, \theta) - \tau(\mathbf{z}, \theta)\| = O_p(n^{-1/2}). \tag{A.26}$$

Therefore, since $\varepsilon_n = O(n^{-1/4})$, to show consistency of $\hat{\Delta}$ it will suffice to derive the plim of $\tilde{\delta}_{mq}$, where

$$\tilde{\delta}_{mq} = \frac{1}{n} \sum_{i=1}^n \tilde{p}_{nm}(\mathbf{z}_i, \hat{\theta}) \tilde{p}_{nq}(\mathbf{z}_i, \hat{\theta}), \tag{A.27}$$

where

$$\tilde{p}_{nq}(\mathbf{z}, \theta) = \varepsilon_n^{-1}(\tau(\mathbf{z}, \theta + \varepsilon_n \mathbf{1}_q) - \tau(\mathbf{z}, \theta)). \tag{A.28}$$

Note by Assumption (A2), $\tilde{p}_{nq}(\cdot, \cdot)$ is twice continuously differentiable in its second argument so we can expand $\tilde{p}_{nq}(\mathbf{z}, \hat{\theta})$ around $\tilde{p}_{nq}(\mathbf{z}, \theta_0)$, where the resulting remainder term is of the form

$$\left(\frac{1}{n} \sum_{i=1}^n (\nabla \tau(\mathbf{z}_i, \theta^* + \varepsilon_n \mathbf{1}_q) - \nabla \tau(\mathbf{z}_i, \theta^*)) \right) (\hat{\theta} - \theta_0) \varepsilon_n^{-1}. \tag{A.29}$$

The above expression is $o_p(1)$ by the root- n consistency of $\hat{\theta}$, Assumption (A2), and Assumption (A3). Thus, it will suffice to evaluate the plim of

$$\frac{1}{n} \sum_{i=1}^n \tilde{p}_{nm}(\mathbf{z}_i, \theta_0) \tilde{p}_{nq}(\mathbf{z}_i, \theta_0). \tag{A.30}$$

Expand

$$(\tau(\mathbf{z}, \theta_0 + \varepsilon_n \mathbf{1}_q) - \tau(\mathbf{z}, \theta_0))$$

and

$$(\tau(\mathbf{z}, \theta_0 + \varepsilon_n \mathbf{1}_m) - \tau(\mathbf{z}, \theta_0))$$

around $\varepsilon_n = 0$, again using Assumption (A2). With Assumption (A3), we can conclude that (A.30) is

$$\frac{1}{n} \sum_{i=1}^n \nabla_1 \tau(\mathbf{z}_i, \theta_0)^{(q)} \nabla_1 \tau(\mathbf{z}_i, \theta_0)^{(m)} + o_p(1), \quad (\text{A.31})$$

where $\nabla_1 \tau(\mathbf{z}_i, \theta_0)^{(p)}$ denotes the p th component of the gradient vector. The above term converges to the (m, q) component of Δ by the law of large numbers. Using similar arguments, it follows that $p_{mq}(\mathbf{z}, \theta)$ converges in probability to the (mq) second derivative component of $\tau(\mathbf{z}, \theta)$ at $\theta = \theta_0$, so \hat{v}_{mq} converges in probability to the expectation of this second derivative evaluated at \mathbf{z}_i and $\theta = \theta_0$, establishing the consistency of \hat{V} . The conclusion of the theorem follows from the continuous mapping theorem and Slutsky's theorem. \square

A.4. Proof of Lemma 3.1

First, we prove Lemma 3.1.

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

$$P(y_{0i} > y_{1j}) - P(y_{0j} > y_{1i}) \quad (\text{A.32})$$

has the same sign as $z_i - z_j$. 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 six cases to consider:

$$c_{1i} > c_{1j}, c_{2i} > c_{2j}, c_{1i} < c_{2j}, \quad (\text{A.33})$$

$$c_{1i} > c_{1j}, c_{2i} > c_{2j}, c_{1i} \geq c_{2j}, \quad (\text{A.34})$$

$$c_{1j} \geq c_{1i}, c_{2j} \geq c_{2i}, c_{2i} > c_{1j}, \quad (\text{A.35})$$

$$c_{1j} \geq c_{1i}, c_{2j} \geq c_{2i}, c_{2i} \leq c_{1j}, \quad (\text{A.36})$$

$$c_{2j} \geq c_{2i}, c_{1j} < c_{1i}, \quad (\text{A.37})$$

$$c_{2i} > c_{2j}, c_{1i} \leq c_{1j}. \quad (\text{A.38})$$

We first show (A.32) 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:

$$P(y_{0i} > y_{1j}) = P(y_{0i} > y_{1j}, d_i = 1, d_j = 1) \tag{A.39}$$

$$+ P(y_{0i} > y_{1j}, d_i = 1, d_j = 2) \tag{A.40}$$

$$+ P(y_{0i} > y_{1j}, d_i = 3, d_j = 2) \tag{A.41}$$

$$+ P(y_{0i} > y_{1j}, d_i = 3, d_j = 1). \tag{A.42}$$

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

$$P(\varepsilon_j < \varepsilon_i + \Delta z, c_{1i} - z_i \leq \varepsilon_i \leq c_{2i} - z_i, c_{1j} - z_j \leq \varepsilon_j \leq c_{2j} - z_j), \tag{A.43}$$

which, recalling we are assuming the censoring values relationship in (A.33), can be expressed as

$$P(c_{1i} - z_i \leq \varepsilon_i \leq c_{2j} - z_i, c_{1j} - z_j \leq \varepsilon_j \leq c_{2j} - z_j, \varepsilon_j < \varepsilon_i + \Delta z) \tag{A.44}$$

$$+ P(c_{2j} - z_i \leq \varepsilon_i \leq c_{2i} - z_i, c_{1j} - z_j \leq \varepsilon_j \leq c_{2j} - z_j, \varepsilon_j < \varepsilon_i + \Delta z) \tag{A.45}$$

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

$$\int_{c_{1i}-z_i}^{c_{2j}-z_i} F(\varepsilon_i + \Delta z) dF(\varepsilon_i) + F_{1jj}F_{1ii} + F_{2ii}F_{2jj} - F_{2ii}F_{1jj} - F_{2ji}F_{2jj}. \tag{A.46}$$

This is the expression for (A.39). The terms (A.40)–(A.42) are easier to deal with so we omit the details. They can be expressed as

$$(F_{2ii} - F_{1ii})F_{1jj}, \quad (1 - F_{2ii})F_{1jj}, \quad (1 - F_{2ii})(F_{2jj} - F_{1jj}), \tag{A.47}$$

respectively. Collecting all terms we have

$$P(y_{0i} > y_{1j}) = \int_{c_{1i}-z_i}^{c_{2j}-z_i} F(\varepsilon_i + \Delta z) dF(\varepsilon_i) + F_{2jj} - F_{2jj}F_{2ji}. \tag{A.48}$$

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

$$P(c_{1i} - z_i \leq \varepsilon_i < \varepsilon_j - \Delta z, c_{1j} - z_j \leq \varepsilon_j \leq c_{2j} - z_j), \tag{A.49}$$

which we can first decompose as

$$P(c_{1i} - z_i \leq \varepsilon_i < \varepsilon_j - \Delta z, c_{1j} - z_j \leq \varepsilon_j \leq c_{1i} - z_j) \tag{A.50}$$

$$+ P(c_{1i} - z_i \leq \varepsilon_i < \varepsilon_j - \Delta z, c_{1i} - z_j \leq \varepsilon_j \leq c_{2j} - z_j). \tag{A.51}$$

Note the above decomposition is valid since we are assuming $c_{1i} < c_{2j}$. Note also (A.50) is 0, because when ε_j takes its largest value of $c_{1i} - z_j$, the lower bound and the upper bound on

ε_i coincide. Therefore, (A.49) equals (A.51) which can be expressed as

$$\int_{c_{1i}-z_j}^{c_{2j}-z_j} F(\varepsilon_j - \Delta z) dF(\varepsilon_j) - F_{1ii}F_{2jj} + F_{1ii}F_{1ij}. \quad (\text{A.52})$$

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,

$$F_{1ii}(F_{2jj} - F_{1ij}), \quad F_{1ii}(1 - F_{2jj}), \quad (1 - F_{2jj})(F_{2ji} - F_{1ii}). \quad (\text{A.53})$$

So collecting terms we have

$$P(y_{0j} > y_{1i}) = \int_{c_{1i}-z_j}^{c_{2j}-z_j} F(\varepsilon_i + \Delta z) dF(\varepsilon_i) + F_{2ji} - F_{2jj}F_{2ji}. \quad (\text{A.54})$$

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

$$\int_{c_{1i}-z_i}^{c_{2j}-z_i} F(\varepsilon_i + \Delta z) dF(\varepsilon_i) - \int_{c_{1i}-z_j}^{c_{2j}-z_j} F(\varepsilon_i - \Delta z) dF(\varepsilon_i) + F_{2ji} - F_{2ji}, \quad (\text{A.55})$$

we now show the above expression has the same sign as $z_i - z_j$. First, we assume that $z_i \geq z_j$ and show the above expression is nonnegative. Note that $F_{2jj} - F_{2ji}$ is nonnegative, 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

$$\int_{c_{1i}-z_j}^{c_{2j}-z_j} F(\varepsilon_i - \Delta z) dF(\varepsilon_i) = \int_{c_{1i}-z_j}^{c_{2j}-z_i} F(\varepsilon_i - \Delta z) dF(\varepsilon_i) + \int_{c_{2j}-z_i}^{c_{2j}-z_j} F(\varepsilon_i - \Delta z) dF(\varepsilon_i). \quad (\text{A.56})$$

Note the first piece on the right-hand side of the above decomposition is less than or equal to the first term in (A.55), 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 first term in (A.55).

Also, the second piece on the right-hand side of the above decomposition is smaller than $F_{2jj} - F_{2ji}$, which follows by exploiting $F(\cdot) \leq 1$. Thus, we have shown that (A.55) is nonnegative when $z_i \geq z_j$. The same arguments can be used to show that (A.55) is nonpositive when $z_i \leq z_j$. Similar arguments as before can be used to extend these weak inequalities to strict inequalities in the regions where there is not 100% censoring. This establishes identification for the case (A.33).

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

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. \square

A.5. Proof of Theorem 4.1

Here, we prove Theorem 4.1. 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} ,

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

has the same sign as $x'_{i1}\beta_0 - x'_{i2}\beta_0$ (showing that $P(d_{i2} = 1, v_{i2} < v_{i1}) - P(d_{i2} = 1, v_{i2} < v_{i1})$ has the same sign as $x'_{i2}\beta_0 - x'_{i1}\beta_0$ follows immediately). For notational convenience, we let z_{i1}, z_{i2} denote $x'_{i1}\beta_0 + \alpha_i, x'_{i2}\beta_0 + \alpha_i$, respectively and first evaluate

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

where we condition on x_{i1} and x_{i2} . This probability can be decomposed into the mutually exclusive cases $c_{i1} > c_{i2}, c_{i1} \leq c_{i2}$. 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.58) as the sum of

$$P(v_{i1} < v_{i2}, d_{i1} = 1, d_{i2} = 1) + P(v_{i1} < v_{i2}, d_{i1} = 1, d_{i2} = 0). \tag{A.59}$$

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

$$P(\varepsilon_{i2} > \varepsilon_{i1} + \Delta z_i, \varepsilon_{i1} \leq c_{i1} - z_{i1}, \varepsilon_{i2} \leq c_{i2} - z_{i2}), \tag{A.60}$$

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

$$P(\varepsilon_{i2} > \varepsilon_{i1} + \Delta z_i, c_{i2} - z_{i1} < \varepsilon_{i1} \leq c_{i1} - z_{i1}, \varepsilon_{i2} \leq c_{i2} - z_{i2}) + P(\varepsilon_{i2} > \varepsilon_{i1} + \Delta z_i, \varepsilon_{i1} \leq c_{i2} - z_{i1}, \varepsilon_{i2} \leq c_{i2} - z_{i2}). \tag{A.61}$$

The first piece above is zero since the first event $\varepsilon_{i2} > \varepsilon_{i1} + \Delta z_i$ combined with the lhs of the second event, $c_{i2} - z_{i1} \leq \varepsilon_{i1}$, imply that $\varepsilon_{i2} > c_{i2} - z_{i2}$ contradicting the third event. So we have

$$(A.60) = P(\varepsilon_{i2} > \varepsilon_{i1} + \Delta z_i, \varepsilon_{i1} \leq c_{i2} - z_{i1}, \varepsilon_{i2} \leq c_{i2} - z_{i2}) = \int_{-\infty}^{c_{i2}-z_{i1}} \int_{\varepsilon_{i1}+\Delta z_i}^{c_{i2}-z_{i2}} dF(\varepsilon_1, \varepsilon_2),$$

where $F(., .)$ denotes the joint CDF of ε_1 and ε_2 . Now, turning attention to the second term in (A.59), we express it as

$$P(\varepsilon_{i1} < c_{i2} - z_{i1}, \varepsilon_{i2} > c_{i2} - z_{i2}, \varepsilon_{i1} \leq c_{i1} - z_{i1}) = P(\varepsilon_{i1} < c_{i2} - z_{i1}, \varepsilon_{i2} \geq c_{i2} - z_{i2}), \tag{A.62}$$

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

$$\int_{c_{i2}-z_{i2}}^{+\infty} \int_{-\infty}^{c_{i2}-z_{i1}} dF(\varepsilon_1, \varepsilon_2). \tag{A.63}$$

Thus, we have that conditioning on x_{i1}, x_{i2} , and $c_{i1} > c_{i2}$, (A.59) can be expressed as

$$\begin{aligned} \text{(A.59)} &= \int_{-\infty}^{c_{i2}-z_{i1}} \int_{e_1+\Delta z_i}^{c_{i2}-z_{i2}} dF(\varepsilon_1, \varepsilon_2) + \int_{c_{i2}-z_{i2}}^{+\infty} \int_{-\infty}^{c_{i2}-z_{i1}} dF(\varepsilon_1, \varepsilon_2) \\ &= \int_{-\infty}^{c_{i2}-z_{i1}} \int_{e_1+\Delta z_i}^{+\infty} dF(\varepsilon_1, \varepsilon_2) = \text{(A.58)}, \end{aligned} \quad \text{(A.64)}$$

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

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

$$P(v_{i2} < v_{i1}, d_{i1} = 1, d_{i2} = 1) + P(v_{i2} < v_{i1}, d_{i2} = 1, d_{i1} = 0). \quad \text{(A.65)}$$

The first term is

$$P(\varepsilon_{i1} > \varepsilon_{i2} - \Delta z_i, \varepsilon_{i1} \leq c_{i1} - z_{i1}, \varepsilon_{i2} \leq c_{i2} - z_{i2}) = \int_{-\infty}^{c_{i2}-z_{i2}} \int_{e_2-\Delta z_i}^{c_{i1}-z_{i1}} dF(\varepsilon_1, \varepsilon_2).$$

The second term in (A.65) is

$$P(\varepsilon_{i2} < c_{i1} - z_{i2}, \varepsilon_{i2} \leq c_{i2} - z_{i2}, \varepsilon_{i1} > c_{i1} - z_{i1}) = P(\varepsilon_{i2} \leq c_{i2} - z_{i2}, \varepsilon_{i1} \geq c_{i1} - z_{i1}),$$

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

$$\int_{-\infty}^{c_{i2}-z_{i2}} \int_{c_{i1}-z_{i1}}^{+\infty} dF(\varepsilon_1, \varepsilon_2).$$

Hence, we have

$$P(d_{i2} = 1, v_{i2} < v_{i1}) = \int_{-\infty}^{c_{i2}-z_{i2}} \int_{e_2-\Delta z_i}^{+\infty} dF(\varepsilon_1, \varepsilon_2), \quad \text{(A.66)}$$

$$P(d_{i1} = 1, v_{i1} < v_{i2}) = \int_{-\infty}^{c_{i2}-z_{i1}} \int_{e_2+\Delta z_i}^{+\infty} dF(\varepsilon_1, \varepsilon_2). \quad \text{(A.67)}$$

As we can see, the difference $P(d_{i2} = 1, v_{i2} < v_{i1}) - P(d_{i1} = 1, v_{i1} < v_{i2})$ 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 $P(d_{i2} = 1, v_{i2} < v_{i1})$ is strictly larger than the area of integration for $P(d_{i1} = 1, v_{i1} < v_{i2})$ (as long as we are not in the 100% censoring region).

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

$$P(d_{i1} = 1, v_{i1} < v_{i2}) = \int_{-\infty}^{c_{i1}-z_{i1}} \int_{e_2+\Delta z_i}^{+\infty} dF(\varepsilon_1, \varepsilon_2), \quad \text{(A.68)}$$

and

$$P(d_{i2} = 1, v_{i2} < v_{i1}) = \int_{-\infty}^{c_{i1}-z_{i2}} \int_{e_2-\Delta z_i}^{+\infty} dF(\varepsilon_1, \varepsilon_2) \quad \text{(A.69)}$$

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 $x'_{i1}\beta_0 - x'_{i2}\beta_0$. \square

References

- Abbring, J.H., Chiappori, P.A., Pinquet, J., 2003. Moral hazard and dynamic insurance data. *Journal of the European Economic Association* 1, 767–820.
- Abrevaya, J., 1999a. Rank estimation of a transformation model with observed truncation. *Econometrics Journal* 2 (2), 292–305.
- Abrevaya, J., 1999b. Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics* 93, 203–228.
- Abrevaya, J., 2000. Rank estimation of a generalized fixed-effects regression model. *Journal of Econometrics* 95, 1–23.
- Arellano, M., Honore, B. 2001. Panel data models: some recent developments. *Handbook of Econometrics* 5, 3229–3296.
- Bijwaard, G.E., 2001. Rank Estimation of Duration Models. Ph.D. Dissertation, Tinbergen Institute.
- Buckley, J., James, I., 1979. Linear regression with censored data. *Biometrika* 66, 429–436.
- Cai, T., Wei, L.J., Wilcox, M., 2000. Semiparametric regression analysis for clustered failure time data. *Biometrika* 87, 867–878.
- Cavanagh, C., Sherman, R.P., 1998. Rank estimators for monotonic index models. *Journal of Econometrics* 84, 351–381.
- Chamberlain, G., 1985. Heterogeneity, omitted variable bias, and duration dependence. In: Heckman, J.J., Singer, B. (Eds.), *Longitudinal Analysis of Labor Market Data*. Cambridge University Press, Cambridge.
- Chen, S., 2002. Rank estimation of transformation models. *Econometrica* 70, 1683–1696.
- Chen, J., Jin, Z., Ying, Z., 2002. Semiparametric analysis of transformation models with censored data. *Biometrika* 89, 659–668.
- Cheng, S.C., Wei, L.J., Ying, Z., 1995. Analysis of transformation models with censored data. *Biometrika* 82, 835–845.
- Cox, D.R., 1972. Regression models and life tables. *Journal of the Royal Statistical Society Series B* 34, 187–220.
- Cox, D.R., 1975. Partial likelihood. *Biometrika* 62, 269–276.
- Cuzick, J., 1988. Rank regression. *Annals of Statistics* 16, 1369–1389.
- Fine, J.P., Ying, Z., Wei, L.J., 1998. On the linear transformation model with censored data. *Biometrika* 85, 980–986.
- Fitzgerald, J., 1992. A hazard model for welfare durations with unobserved location specific effects. Working paper, Bowdoin College.
- Gorgens, T., Horowitz, J., 1998. Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable. *Journal of Econometrics* 90 (2), 155–191.
- Han, A., 1987. Non-parametric analysis of a generalized regression model. *Journal of Econometrics* 35, 303–316.
- Hausman, J., Woutersen, T., 2005. Estimating a semi-parametric duration model without specifying heterogeneity. Working paper.
- Heckman, J.J., Borjas, G.J., 1980. Does unemployment cause future unemployment? definitions, questions and answers for a continuous time model of heterogeneity and state dependence. *Economica* 47, 247–283.
- Honoré, B.E., 1993. Identification results for duration models with multiple spells. *Review of Economic Studies* 60, 241–246.
- Honoré, B.E., Powell, J.L., 1994. Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics* 64 (1–2), 241–278.
- Honoré, B.E., Khan, S., Powell, J.L., 2002. Quantile regression under random censoring. *Journal of Econometrics* 109, 67–105.
- Horowitz, J.L., Lee, S., 2004. Semiparametric estimation of a panel data proportional hazards model with fixed effects. *Journal of Econometrics* 19, 155–190.
- Ichimura, H. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* 58, 71–120.
- Jin, Z., Ying, Z., Wei, L.J., 2002. A simple resampling method by perturbing the minimand. *Biometrika* 88, 381–390.
- Kalbfleisch, J.D., Prentice, R.L., 1980. *The Statistical Analysis of Failure Time Data*. Wiley, New York.
- Khan, S., Tamer, E., 2005. Heteroskedastic transformation models with covariate dependent censoring. Manuscript, Northwestern University.

- Koul, H., Susarla, V., Van Ryzin, J., 1981. Regression analysis with randomly right censored data. *Annals of Statistics* 9, 1276–1288.
- Lancaster, T., 1979. Econometric methods for the duration of unemployment. *Econometrica* 47(4), 939–956.
- Lancaster, T., 2000. The incidental parameter problem since 1948. *Journal of Econometrics* 95, 291–413.
- Lee, S., 2003. Estimating panel data duration models with censored data. Manuscript, University College London.
- Manski, C.F., 1985. Semiparametric analysis of discrete response: asymptotic properties of maximum score estimation. *Journal of Econometrics* 27, 313–334.
- Manski, C., 1987. Semiparametric analysis of random effects linear models from binary panel data. *Econometrica* 55(2), 357–362.
- Newey, W.K., McFadden, D., 1994. Estimation and hypothesis testing in large samples. In: Engle, R.F., McFadden, D. (Eds.), *Handbook of Econometrics*, vol. 4. North-Holland, Amsterdam.
- Newman, J.L., McCulloch, C.E., 1984. A hazard rate approach to the timing of births. *Econometrica* 52, 939–961.
- Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027–1057.
- Portnoy, S., 2003. Censored regression quantiles. *Journal of the American Statistical Association* 98, 1001–1012.
- Powell, J., 1994. Estimation of semiparametric models. In: Engle, R.F., McFadden, D. (Eds.), *Handbook of Econometrics* vol. IV. North-Holland, Amsterdam, pp. 2444–2514.
- Ren, J., Gu, M., 1997. Regression M-estimators with doubly censored data. *The Annals of Statistics* 25(6), 2638–2664.
- Ridder, G., 1990. The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies* 57, 167–182.
- Ridder, G., Tunalı, I., 1999. Stratified partial likelihood estimation. *Journal of Econometrics* 92, 193–232.
- Sherman, R.P., 1993. The limiting distribution of the maximum rank correlation estimator. *Econometrica* 61, 123–137.
- Sherman, R.P., 1994a. U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* 10, 372–395.
- Sherman, R.P., 1994b. Maximal inequalities for degenerate U-processes with applications to optimization estimators. *Annals of Statistics* 22, 439–459.
- Tsiatis, A.A., 1990. Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* 18, 354–372.
- Van den Berg, G.J., 2001. Duration models: specification, identification and multiple durations. In: Heckman, J.J., Leamer, E. (Eds.), *Handbook of Econometrics*, vol. 5. North-Holland, Amsterdam.
- Yang, S., 1999. Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association* 94, 137–145.
- Ying, Z., Jung, S.H., Wei, L.J., 1995. Survival analysis with median regression models. *Journal of the American Statistical Association* 90, 178–184.
- Zhang, C., Li, X. 1996. Linear regression with doubly censored data. *The Annals of Statistics* 24(6), 2720–2743.

Further reading

- Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete data. *Journal of the American Statistical Association* 53, 457–481.
- Politis, D., Romano, J., Wolf, M., 1999. *Subsampling*. Springer Series in Statistics, New York.
- Wei, L.J., Ying, Z., Lin, D.Y., 1990. Linear regression analysis of censored survival data based on rank tests. *Biometrika* 19, 417–442.
- Wang, J.-G., 1987. A Note on the uniform consistency of the Kaplan–Meier estimator. *Annals of Statistics* 15, 1313–1316.