Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 136 (2007) 251–280
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

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.

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).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.03.003

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

$$T(y_i) = x_i'\beta_0 + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1.1}$$

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 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 as[1]

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

$$d_i = I[x_i'\beta_0 + \varepsilon_i \leqslant c_i], \tag{1.3}$$

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 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,

[1] We 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 $I[y_i \leqslant \tilde{c}_i]$.