Statistics and Probability Letters 95 (2014) 150–154

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A simple root-N-consistent semiparametric estimator for discrete duration models

Sadat Reza [a,*], Paul Rilstone [b]

[a] Nanyang Business School, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

[b] Department of Economics, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

ARTICLE INFO

Article history:
Received 5 December 2013
Received in revised form 19 August 2014
Accepted 19 August 2014
Available online 1 September 2014

Keywords:
Discrete duration model
Semiparametric
Kernel estimator
Least squares estimator
Single index model
Empirical process

ABSTRACT

Incorrect specification of the hazard rate in duration analysis can produce inconsistent estimators of the parameters of the model. We propose a new estimator for discrete duration models in which the hazard rate is comprised of an inner index function of the covariates and time variable and an outer link function. The index function is specified up to a finite dimensional parameter, β_0. β_0 is estimated using a least squares objective function. The link function is left unspecified and estimated by the method of kernels. We demonstrate the consistency and asymptotic normality of the estimator of β_0. Simulations show the efficacy of the estimator.

1. Introduction

Discrete time duration models are widely used in statistical research, particularly in epidemiology and econometrics. The benefits of discrete duration models, particularly in terms of ease of implementation and incorporation of time varying covariates, have been widely discussed (Yamaguchi, 1991; Jenkins, 1995). Let T denote the duration or length of a completed spell. A common way to write the probability mass function of T is

$$P(T=t;\beta_0)=(1-F_t(v(t,X_t;\beta_0)))\prod_{s=1}^{t-1}F_s(v(s,X_s;\beta_0)) \tag{1.1}$$

where $1-F_s(v(s,X_s;\beta_0))$ is the hazard rate, v is an index function of elapsed time, s, an observable covariate, X_s, and a $k\times 1$ parameter β_0. The objective is to estimate β_0.

Almost all applications of this model have used various parametric specifications of F_s. Common specifications for F_s (with recent examples) are Logistic (Caliendo et al., 2010), normal or Probit (Di Porto and Revelli, 2013), Clog–log (Rebollo-Sanz, 2012) and Generalized Pareto (Hess, 2009). In each of these cases, maximum likelihood is used to estimate β_0 basing the likelihood on the distribution in (1.1). Misspecification of a component of the model, including F_s, can lead to inconsistent estimates of β_0.

A number of authors such as Ichimura (1993), Klein and Spady (1993, henceforth KS) and Escanciano et al. (2014,[1] henceforth EJL) have considered similar single-index models using kernel estimators of link functions such as F_s. In particular,

* Corresponding author.
E-mail addresses: sreza@ntu.edu.sg, sadatr@gmail.com (S. Reza), pril@yorku.ca (P. Rilstone).

[1] We thank a referee for providing this reference.

http://dx.doi.org/10.1016/j.spl.2014.08.017

EJL allow for nonparametrically estimated generated regressors and data-driven window width selection in their general results. One of EJL's specific examples is the semiparametric estimation of a binary choice model given sample selection and with a nonparametrically estimated response variable in the index function. Our model differs in that the observations here are over time periods as well as individuals, the time period potentially enters into the index here and in each period s there is a "selection" based on $Y_{s-1} = 1$. We allow for non-linear v. Also, there is no presumption here that a preliminary consistent estimate of β_0 is available as in KS or EJL.

To understand our proposed estimation procedure consider that a common way (e.g., Yamaguchi, 1991) to derive the probability function in (1.1) is to view the spell length as arising from a sequence of latent bivariate choices summarized by the indicator:

$$Y_s = 1[v(s, X_s; \beta_0) - \epsilon_s \geq 0]Y_{s-1} \tag{1.2}$$

where $\epsilon_s \perp X_s$, ϵ_s has the distribution function, F_s, and $v(s, X_s; \beta)$ is known up to the true value of β. Similar to standard binary choice models, we have

$$E[Y_s|X_s, Y_{s-1} = 1] = \Pr[v(s, X_s; \beta_0) > \epsilon_s|Y_{s-1} = 1] = F_s(v(s, X_s; \beta_0)). \tag{1.3}$$

The spell length is simply the sum of the Y_s's. The benefit of this formulation is that the hazard rate is thus written as a conditional mean function, which, with some caveats, can be estimated by nonparametric techniques.

We suggest a simple least squares estimator of β_0, estimating F_s by the method of kernels. To do so, define $G_s(v(s, X_s; \beta)) = E[Y_s|v(s, X_s; \beta), Y_{s-1} = 1]$ and note that $G_s(v(s, X_s; \beta_0)) = F_s(v(s, X_s; \beta_0))$. Let subscript i's, $i = 1, 2, \ldots, N$, denote individual observations on the random quantities. If the functional form of G_s was known we could estimate β_0 by minimizing

$$Q_N(\beta; G) = \frac{1}{2N} \sum_{i=1}^{N} \sum_{s=1}^{S} ((Y_{is} - G_{is}(\beta))Y_{is-1}\rho_{is})^2 \tag{1.4}$$

where ρ_{is} is a trimming or weighting function. With G_s unknown, we replace G_s with a nonparametric estimate, $\widehat{G}_s$. Our approach thus also differs from EJL in that our estimator minimizes a least squares objective function rather than maximizing a likelihood function.

The trimming function, $\rho_{is} = \rho(X_{is})$, is introduced to circumvent certain technical difficulties. In particular, we set ρ to zero outside a compact set, inside of which, the density of $v(s, X_{is}; \beta)$ is uniformly bounded away from zero. Alternatively, ρ can be used to weight observations in a potentially more efficient manner, although we do not explore that here. Since the trimming is fixed, the estimators may not be fully efficient. However, our purpose here is to establish a simple consistent estimator. From a practical perspective, the possible loss in efficiency from fixed trimming may be minimal and the numerical implementation is simpler. In addition, we could further generalize the weighting (such as inversely proportional to the conditional variance of G_s) to potentially increase efficiency. However, our objective is to first establish a simple $\sqrt{N}$-consistent estimate without these complications.

In this note we first derive the consistency and asymptotic normality of the estimator. To do so we adapt the empirical process theory used by EJL. We then discuss a simulation which examines some of the practical aspects of the estimator and also demonstrates its efficacy.

2. Asymptotic properties

EJL use empirical process theory to derive their results. We replicate a certain amount of their notation to facilitate communication of our results. Let $\mathcal{V}$ denote the class of functions of the form $v(s, x; \beta)$ for any $\beta \in \mathcal{B}$ and $s \in 1, \ldots, S$. Define the random variables $V = v(s, X; \beta)$ as measurable functions on an underlying probability space denoted $(\Omega, \mathcal{F}, P)$ on which all the random variables used here are defined.

We use a standard kernel estimator of $G_s(\beta)$:

$$\widehat{G}_s(v(s, x; \beta)|V) = \frac{\widehat{g}_{1s}(v(s, x; \beta)|V)}{\widehat{g}_s(v(s, x; \beta)|V)}, \tag{2.1}$$

where

$$\widehat{g}_{ys}(v(s, x; \beta)|V) = \frac{1}{\gamma N} \sum_{j=1}^{N} 1[Y_{js} = y]K\left(\frac{v(s, x; \beta) - v(s, X_{js}; \beta)}{\gamma}\right)Y_{js-1}, \quad y = 0, 1$$

and $\widehat{g}_s = \widehat{g}_{0s} + \widehat{g}_{1s}$ is an estimator of $g_s(v(s, X_s; \beta))$, the density function of $v(s, X_s; \beta)$ given $Y_{s-1} = 1$.

To keep the discussion succinct, we modify the assumptions in EJL to our estimation situation and indicate how to apply their results to the case here. We introduce the following notations: $\bigtriangledown G_s(\beta)$ denotes the gradient of $G_s(\beta)$, $u_{is} = Y_{is} - G_{is}$, $\eta_{is} = u_{is} \bigtriangledown G_{is}Y_{is-1}\rho_{is}$, $B = \sum_{s=1}^{S} E[\eta_{is}\eta_{is}']$, and $A = \sum_{s=1}^{S} E[\bigtriangledown G_{is} \bigtriangledown G_{is}' Y_{is-1}\rho_{is}]$. Note that $E[\eta_{is}] = 0$.