A Note on Binary Choice Duration Models

Deepankar Basu*   Robert de Jong†

February 28, 2007

Abstract

In this paper we demonstrate that standard methods of asymptotic inference will break down in a binary choice duration model in a time series setting. This comes about because the dependent variable has a degenerate limit distribution, which makes the asymptotic variance-covariance matrix singular. This result has implications for discrete choice duration panel data models under large N and T asymptotics.

1 Introduction

This note points out a problem of internal consistency in a binary choice duration model in a time series framework. The problem emerges whenever duration in the current state is included as a regressor in a binary choice model. In such a situation, the dependent variable converges in probability to unity and standard methods of asymptotic inference breaks down. Recently, Frederiksen, Honoré and Hu (2007) have studied a discrete choice duration model with group level heterogeneity in a panel data setting. The problem that we point out will be relevant for any attempt to extend this model to the case where both N and T are large.

To fix ideas consider the following binary choice probit model

$$y_t = I(\beta_0 z_t + \varepsilon_t > 0) \tag{1}$$

where $I(.)$ is the indicator function, $\beta_0 > 0$ is the parameter of interest, the process starts at time $t = 0$ at an arbitrary starting value y_0 and z_t is defined as the number of consecutive

*Department of Economics, Ohio State University, 317 Arps Hall, Columbus, OH 43210, email: basu.15@osu.edu

†Department of Economics, Ohio State University, 429 Arps Hall, 1945 N High Street, Columbus, OH 43202, email: dejong@econ.ohio-state.edu.

ones of the y_t sequence leading up to the current period, i.e., for $t = 1, 2, \ldots, T$

$$z_t = \sum_{j=1}^{t} \prod_{i=1}^{j} y_{t-i}. \tag{2}$$

In (1) and (2), y_t is a binary variable taking values in $\{0, 1\}$, z_t is the duration in the current state (looking back from the current time period) and the error term ε_t is distributed as i.i.d. $N(0, 1)$. For instance, y_t could capture whether a person is unemployed or not ($y_t = 1$ if unemployed, and $y_t = 0$ otherwise) with z_t measuring the current spell of unemployment; we show below that the structure of the model in (1) and (2) implies that eventually every individual will be unemployed. Our focus on the case $\beta_0 > 0$ is motivated by the assumption that long spells of unemployment increase the probability of further unemployment.

The joint density of the observed sample conditional on y_0 is

$$\begin{aligned} &f(y_1, \ldots, y_T | y_0; \beta) \\ &= f(y_T | y_{T-1}, \ldots, y_1, y_0; \beta) f(y_{T-1} | y_{T-2}, \ldots, y_1, y_0; \beta) \ldots f(y_2 | y_1, y_0; \beta) f(y_1 | y_0; \beta). \end{aligned}$$

From (1) and (2), we see that

$$f(y_t | y_{t-1}, \ldots, y_1, y_0; \beta) = [\Phi(\beta z_t)]^{y_t} [1 - \Phi(\beta z_t)]^{1-y_t} \quad (t = 1, 2, \ldots, T).$$

Hence, the conditional log-likelihood function for the sample is

$$\log L(\beta) = \sum_{t=1}^{T} y_t \log \Phi(\beta z_t) + \sum_{t=1}^{T} (1 - y_t) \log[1 - \Phi(\beta z_t)]. \tag{3}$$

The question that we want to pose and answer is the following: can we consistently estimate and conduct asymptotically valid inference on β_0 using standard procedures?

2 Main Result and Discussion

Theorem 1 *Consider the model in (1) and (2) for y_t. Let $F_\varepsilon(\cdot)$ denote the distribution function of ε_t. If (i) ε_t is i.i.d., (ii) $\forall y \in \mathbb{R}, F_\varepsilon(y) > 0$ and (iii) $E[|\varepsilon_t| I(\varepsilon_t < 0)] < \infty$ then $y_t \xrightarrow{p} 1$ as $t \to \infty$.*

Note that since Theorem 1 does not impose a centering assumption on ε_t, the result will remain true if an intercept is added to the specification of Equation (1). The intuition behind this result is the following. Since $P(y_t = 1) = E[y_t] \geq (1/2)$, the y_t sequence will eventually hit unity. But once it hits unity, the probability of y_t attaining unity in the next period increases because z_t increases by one. Thus, in a sense, the y_t sequence gradually drifts away from zero and towards unity, eventually getting stuck at unity.

The main implication of this result is that standard maximum likelihood-based asymptotic inference on β_0 is problematic. To see this, note that the score function equals

$$s_T(\beta) = \sum_{t=1}^{T} s_t(\beta) = \sum_{t=1}^{T} \frac{\phi(\beta z_t) z_t [y_t - \Phi(\beta z_t)]}{\Phi(\beta z_t)[1 - \Phi(\beta z_t)]} \tag{4}$$

and the Hessian is

$$H_T(\beta) = \sum_{t=1}^{T} H_t(\beta) = -\sum_{t=1}^{T} \frac{\{\phi(\beta z_t)\}^2 z_t^2}{\Phi(\beta z_t)[1 - \Phi(\beta z_t)]} + \sum_{t=1}^{T} [y_t - \Phi(\beta z_t)] L(\beta z_t) \tag{5}$$

where $\phi(.)$ denotes the density and $\Phi(.)$ the distribution function of a standard normal random variable, and $L(\cdot)$ is a known function. Following Wooldridge (2002), it is easy to see that $\mathrm{E}[s_t(\beta_0)|z_t] = 0$, and one might be tempted to conclude that

$$\sqrt{T}(\beta_0 - \beta) \xrightarrow{d} N(0, A_0^{-1}) \tag{6}$$

where

$$A_0 = -\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} \mathrm{E}[H_t(\beta_0)]. \tag{7}$$

However, because z_t diverges as $t \to \infty$,

$$E(H_t(\beta_0)|z_t) = \frac{\{\phi(\beta_0 z_t)\}^2 z_t^2}{\Phi(\beta_0 z_t)[1 - \Phi(\beta_0 z_t)]}) \xrightarrow{p} 0 \quad \text{a}s \quad t \to \infty.$$

Because $E(H_t(\beta_0)|z_t)$ is bounded uniformly in z_t, it follows by the dominated convergence theorem that

$$EH_t(\beta_0) \to 0 \quad \text{a}s \quad t \to \infty.$$

Therefore, it follows that $A_0 = 0$. This violates standard assumptions made in the theory of minimization estimators and obviously, the standard result of Equation (6) ceases to be valid.

In conclusion therefore, standard methods for justifying asymptotic inference for minimization estimators will break down in the binary choice duration model when T is large.

Appendix

To analyze the properties of the stochastic dynamical system in (1) and (2) and to prove Theorem 1, let $m \in \mathbb{N}$ be a positive integer and define a random variable τ_m as

$$\tau_m = \inf\{k : y_{k-m} = 1, y_{k-m+1} = 1, \ldots, y_{k-1} = 1, y_k = 1\}. \tag{8}$$

Then, τ_m is the first time period when the y_t sequence gets a realization of m consecutive ones; thus, τ_m takes values in $\{1, 2, \ldots\} \cup \{\infty\}$. Let $F_\varepsilon(\cdot)$ be the distribution function of ε_t. Note that the event $\{\tau_m > Km\}$ means that K times in a row, the y_t sequence did not come up with even a single m-tuple of ones; thus,

$$\begin{aligned}
&P(\tau_m > Km) \\
&\leq (1 - [F_\varepsilon(-\beta_0 m)]^m)^K \to 0 \quad \text{as} \quad K \to \infty
\end{aligned}$$

because the ε_t are i.i.d. and $\forall y \in \mathbb{R}, F_\varepsilon(y) > 0$. Hence, τ_m is a well-defined random variable.

Proof of Theorem 1: Choose any $j \in \mathbb{N}$ such that $j > m$. Define an index set, $A_m = \{j \in \mathbb{N} : P(\tau_m = j) > 0\}$. Then

$$\begin{aligned}
&P(y_t = 0 \quad \text{i.o.}) \\
&= P(y_t = 0 \quad \text{i.o.} \quad \& \quad \cup_{j \in A_m} \{\tau_m = j\}) \\
&= \sum_{j \in A_m} P(y_t = 0 \quad \text{i.o.} \quad \& \quad \tau_m = j) \\
&= \sum_{j \in A_m} P(y_t = 0 \quad \text{i.o.} | \tau_m = j) P(\tau_m = j)
\end{aligned}$$

The first equality follows because:

$$P(\cup_{j \in A_m} \{\tau_m = j\}) = \sum_{j \in A_m} P(\{\tau_m = j\}) = 1. \tag{9}$$

and the second follows because the events $\{\tau_m = j\}$ are disjoint, i.e., $\{\tau_m = j\} \cap \{\tau_m = k\} = \phi$, for $j \neq k$. Note that we can legitimately condition on the event $\{\tau_m = j\}$ in the third equality because τ_m is a well-defined random variable. Now,

$$\begin{aligned}
&\sum_{j \in A_m} P(y_t = 0 \quad \text{i.o.} | \tau_m = j) P(\tau_m = j) \\
&\leq \sup_{j \in A_m} P(y_t = 0 \quad \text{i.o.} | \tau_m = j) \sum_{j \in A_m} P(\tau_m = j) \\
&= \sup_{j \in A_m} P(y_t = 0 \quad \text{i.o.} | \tau_m = j) \quad \text{(by equation (9))} \\
&\leq \sup_{j \in A_m} P(\exists t \geq j, y_t = 0 | \tau_m = j) \\
&\leq \sup_{j \in A_m} P(\varepsilon_{j+1} < -\beta_0 m \quad \text{or} \quad \varepsilon_{j+2} < -\beta_0 (m+1) \quad \text{or} \quad \varepsilon_{j+3} < -\beta_0 (m+2) \ldots | \tau_m = j) \\
&\leq \sum_{j=m}^{\infty} P(\varepsilon_t \leq -\beta_0 j)
\end{aligned}$$

where the last inequality follows by the independence of the ε_t, countable subadditivity of the probability measure and because $\{\tau_m = j\} \in \sigma(y_0, \varepsilon_1, \ldots, \varepsilon_{j-1}, \varepsilon_j)$ where $\sigma(y_0, \ldots, \varepsilon_{j-1}, \varepsilon_j)$ is the sub-sigma field generated by $(y_0, \varepsilon_1, \ldots, \varepsilon_{j-1}, \varepsilon_j)$. Now,

$$\begin{aligned}
&\sum_{j=m}^{\infty} P(\varepsilon_t \leq -\beta_0 j) \\
&= \sum_{j=m}^{\infty} \sum_{k=j}^{\infty} P(-\beta_0(k+1) < \varepsilon_t \leq -\beta_0 k) \\
&\leq \sum_{k=m}^{\infty} \sum_{j=1}^{k} \int_{-\beta_0(k+1)}^{-\beta_0 k} dF_\varepsilon(x) \\
&\leq \sum_{k=m}^{\infty} k \int_{-\beta_0(k+1)}^{-\beta_0 k} (|x|/(\beta_0 k)) dF_\varepsilon(x) \\
&\leq \beta_0^{-1} \int_{-\infty}^{-\beta_0 m} |x| dF_\varepsilon(x),
\end{aligned}$$

and the last expression converges to 0 as $m \to \infty$ if $E[|\varepsilon_t| I(\varepsilon_t < 0)] < \infty$, which completes the proof of Theorem 1. □

References

[1] Frederiksen, A., B. Honoré and L. Hu, 2007, Discrete Time Duration Models with Group-level Heterogeneity, Journal of Econometrics, forthcoming.

[2] Wooldridge, J. M, 2002, Econometric Analsysis of Cross Section and Panel Data. (MIT Press, Cambridge, Massachussets).