Economics Letters 121 (2013) 74–78

Contents lists available at ScienceDirect

Economics Letters

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

Partial identification in binary response models with nonignorable nonresponses

Tadao Hoshino *,[1]

Graduate School of Information Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo, Japan

HIGHLIGHTS

- We propose an estimation method for semiparametric binary response models with nonignorable nonresponses.
- The parameter of interest is partially identifiable without relying on restrictive distributional assumptions.
- Our estimation method, which is based on the special regressor approach, is easy to implement.
- The proposed estimator is consistent in the Hausdorff metric.

ARTICLE INFO

Article history:
Received 18 May 2013
Received in revised form
11 July 2013
Accepted 13 July 2013
Available online 18 July 2013

JEL classification:
C13
C14
C25

Keywords:
Semiparametric binary response models
Nonignorable nonresponses
Special regressor approach
Partial identification

ABSTRACT

This study investigates the identification of parameters in semiparametric binary response models of the form $y = 1(x'\beta + v + \varepsilon > 0)$ when there are nonignorable nonresponses. We propose an estimation procedure for the identified set, the set of parameters that are observationally indistinguishable from the true value β, based on the special regressor approach of Lewbel (2000). We show that the estimator for the identified set is consistent in the Hausdorff metric.

1. Introduction

In the field of survey and interview data analysis, nonresponses and missing data are common and often unavoidable. In many empirical studies, models are estimated using only a subsample of the complete data. These results are valid under the *missing-at-random* (MAR) assumption, which implies that missing data is ignorable. However, Manski (2003), for example, points out that such an assumption is untestable, so nonresponses are in general nonignorable. Indeed, the MAR assumption does not hold in many empirical settings, yielding biased estimates under such an assumption.

In this paper, we consider the estimation of binary response models with nonignorable missing response data without relying on the MAR assumption. Consider the following binary response model:

$$y = 1(x'\beta + v + \varepsilon > 0), \tag{1}$$

where y is a binary outcome, x is a $k \times 1$ vector of observed regressors, β is a $k \times 1$ vector of parameters to be estimated, ε is an unobserved error term, and v is a scalar random variable whose coefficient is normalized to 1 for identification. This paper extends (1) to the case in which y is not observable to econometricians with a positive probability less than 1. When an individual's response is not observable, it is in general impossible to infer whether the "potential" response is 1 or 0. In other words, what we can "observe" is only a random set Y defined by

$$Y = \begin{cases} \{y\} & \text{if } d = 1 \\ \{1, 0\} & \text{if } d = 0. \end{cases} \tag{2}$$

In the equation above, d is an indicator representing the observability of y. A formal definition of a random set is given in the next section.

* Tel.: +81 0 357342651.
E-mail addresses: hoshino.t.ai@m.titech.ac.jp, tadao0029@gmail.com.

[1] Research Fellow (PD), Japan Society for the Promotion of Science.

0165-1765/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.econlet.2013.07.009

When it is not credible to assume the MAR assumption, an alternative often used in the literature is a sample selection model (Heckman, 1979). In order to use such a model, we need to introduce additional structural and distributional assumptions on the relationship between the observability d, and its explanatory variables. The parameter estimates are generally not consistent when these assumptions are not met, but it is often very difficult to obtain a correct model specification. Thus, in this paper, we do not impose such structural and distributional assumptions on the observability of response data. In addition, we do not assume any parametric form for the distribution of the error term ε. Under this setup, this paper considers estimating the set of all observationally equivalent values of the parameter β. We call this set the identified set, and denote it by Θ_β. For a semiparametric binary response model, a straightforward estimator of the identified set, Θ_β, would be the maximum likelihood estimator. However, applying the maximum likelihood estimator is often problematic in terms of computational burden and the assumption that ε is independent of (x, v). To overcome these problems, this paper suggests using the method proposed by Lewbel (2000). If there are no missing responses in the population, i.e., $\mathrm{P}(d = 0) = 0$, and v can be used as a *special regressor*, Lewbel (2000) shows that β can be estimated by a linear regression of $[y - 1(v > 0)]/f(v|x)$ onto x, i.e.,

$$\beta = \mathrm{E}(xx')^{-1}\mathrm{E}\left(x\frac{y - 1(v > 0)}{f(v|x)}\right). \tag{3}$$

At the cost of assuming the existence of a special regressor, unlike the other estimators of semiparametric binary response models proposed by, for example, Klein and Spady (1993) and Ichimura (1993), the estimator in (3) does not require restrictive conditions on the error term such as statistical independence or single-index sufficiency. In addition, by utilizing the special regressor approach, the computation of the identified set Θ_β can be greatly simplified as compared with use of the maximum likelihood estimator.[2]

The remainder of this paper is organized as follows. In Sections 2 and 3, we describe the estimation and inference procedure for our model, respectively. In Section 4, we propose a method to determine the sign of the coefficient of the special regressor when it is not known a priori. In Section 5, we introduce an assumption called the *stigma-affecting response*, which is reasonable to assume in some empirical situations, and can improve the bound. Finally, in Section 6, we present the conclusion.

2. Consistent estimation of the identified set

First, let us introduce the following assumptions.

Assumption A. 1. The conditional distribution of v given x is absolutely continuous with respect to a Lebesgue measure with nondegenerate Radon–Nikodym conditional density $f(v|x)$. 2. The conditional distribution of ε given x is independent of v for all $(v, x) \in \mathrm{supp}(v, x)$. 3. (a) The conditional distribution of v given x has a support $[L, U]$ for some constants L and U such that $-\infty \leq L < 0 < U \leq \infty$; and (b) $\mathrm{supp}(-x'\beta - \varepsilon) \subseteq [L, U]$. 4. (a) $\mathrm{E}(x\varepsilon) = 0$; and (b) $\mathrm{E}(xx')$ exists and is nonsingular.

Assumption A1–3 characterize the special regressor v. The conditional independence condition in A2 is much weaker than the statistical independence condition. The large support condition in A3 implies that the probability of observing $y = 1$ approaches 0 (1) if v becomes sufficiently small (large). This condition can be relaxed by replacing it with the tail symmetry condition (for details, see Magnac and Maurin, 2007). Assumption A4(a) excludes the case where x is endogenous. If a set of suitable instrumental variables exists such that $\mathrm{E}(z\varepsilon) = 0$, this condition can be relaxed.

Now, we consider the partial identification of the parameters. Let us introduce some terms and their definitions in the field of *random set theory* (for a comprehensive review, see, e.g., Li et al., 2010, Molchanov, 2005). Random set theory provides very useful tools to analyze a certain class of partially identified models, as in Beresteanu and Molinari (2008) and Beresteanu et al. (2012). Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. Throughout this paper, we assume that the probability space is nonatomic. Let $K(\mathbb{R}^k)$ be the family of all nonempty closed subsets of $\mathbb{R}^k$.

Definition 1 (*Random Set*)**.** A set-valued mapping $F : \Omega \to K(\mathbb{R}^k)$ is called a random set if, for each open subset O in $\mathbb{R}^k$, $F^{-1}(O) := \{\omega \in \Omega : F(\omega) \cap O \neq \emptyset\} \in \mathcal{A}$.[3]

Definition 2 (*Selection*)**.** An $\mathbb{R}^k$-valued function $f : \Omega \to \mathbb{R}^k$ is called a selection for a random set $F : \Omega \to K(\mathbb{R}^k)$ if $f(\omega) \in F(\omega)$ for all $\omega \in \Omega$.

Let $\mathcal{S}(F)$ be a selection set in $L^1[\Omega; \mathbb{R}^k]$ for a random set F, where $L^1[\Omega; \mathbb{R}^k]$ is the space of measurable functions $f : \Omega \to \mathbb{R}^k$ such that $\int_\Omega |f|\mathrm{d}\mu$ is finite, i.e., $\mathcal{S}(F) := \{f \in L^1[\Omega; \mathbb{R}^k] : f(\omega) \in F(\omega) \text{ for all } \omega \in \Omega\}$.

Definition 3 (*Aumann Integral of a Random Set*)**.** For each random set F, the Aumann integral of F, denoted by $\mathbb{E}(F)$, is defined by $\mathbb{E}(F) = \left\{\int_\Omega f\mathrm{d}\mu : f \in \mathcal{S}(F)\right\}$.

Assumption B. 1. The random variables (Y, v, x) are defined on a nonatomic probability space $(\Omega, \mathcal{A}, \mu)$. 2. Any $y^* \in \mathcal{S}(Y)$ is admissible for the true y. 3. Let

$$y_U = \begin{cases} y & \text{if } d = 1 \\ 1 & \text{if } d = 0 \end{cases} \quad \text{and} \quad y_L = \begin{cases} y & \text{if } d = 1 \\ 0 & \text{if } d = 0. \end{cases}$$

$x_j\frac{y_U - 1(v>0)}{f(v|x)}$ and $x_j\frac{y_L - 1(v>0)}{f(v|x)}$, $j = 1, \ldots, k$, are random variables in $L^1[\Omega; \mathbb{R}]$.

The nonatomicity Assumption B1 is introduced in order to simplify the estimation of the identified set, and it is not too restrictive because the appropriate probability space for a sequence of i.i.d. random elements is nonatomic (see Beresteanu et al., 2012). Assumption B2 excludes the case, for example, in which $\mathrm{supp}(-x'\beta - \varepsilon)$ is known to researchers. If it were known and the observed value of v were larger (smaller) than its upper (lower) boundary, we could set y to 1 (0) regardless of the observability of y, yielding a smaller admissible set than $\mathcal{S}(Y)$. Define

$$G(\omega) := x(\omega)Y^*(\omega), \quad \text{where } Y^*(\omega) := \frac{Y - 1(v > 0)}{f(v|x)}(\omega).$$

Now, we characterize the population-identified set Θ_β as follows.

Proposition 1. *Suppose that Assumptions A and B hold. Then, the identified set for β is given by*

$$\Theta_\beta = \mathrm{E}(xx')^{-1}\mathbb{E}(G). \tag{4}$$

Further, the set in (4) is equivalent to

$$\mathrm{E}(xx')^{-1}\mathbb{E}(\mathrm{co}G), \tag{5}$$

where, for a set A, co$A$ is a convex hull of A.

[2] This paper is not the first to investigate identification and estimation in incomplete binary response models based on the special regressor approach. Magnac and Maurin (2008) consider a binary response model in which the special regressor v is either discrete or measured within intervals.

[3] In general, we can consider a mapping $F : \Omega \to K(\mathcal{X})$ with $\mathcal{X}$ being a general metric space. For the purpose of this study, it suffices to consider the case where $\mathcal{X} = \mathbb{R}^k$.