Efficient estimation of binary choice models under symmetry

Songnian Chen*

Department of Economics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, People’s Republic of China

Received 1 July 1997; received in revised form 1 June 1999

Abstract

This paper proposes a semiparametric maximum likelihood estimator for both the intercept and slope parameters in a binary choice model under symmetry and index restrictions. The estimator attains the semiparametric efficiency bound in Cosslett (1987) under the symmetry and independence restrictions. Compared with the estimator of Klein and Spady (1993), which attains the semiparametric efficiency bound in Chamberlain (1986), and Cosslett (1987) under the independence restriction, we show that there are possible efficiency gains in estimating the slope parameters by imposing the additional symmetry restriction. A small Monte Carlo study is carried out to illustrate the usefulness of our estimator. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C31; C34

Keywords: Binary choice; Symmetry; Semiparametric estimation; Efficiency bound

1. Introduction

In this paper we consider the estimation of the binary choice model defined by

\[ d = 1\{v(x, \theta_0) > u\}, \]

*Tel.: + 852-2358-7602; fax: + 852-2358-2084.
E-mail address: snchen@ust.hk (S. Chen)
where \(1\{A\}\) is the usual indicator function of the event \(A\), \(x\) is a vector of exogenous variables, \(v(x, \theta_0)\) is a given parametric function of \(x\) with an unknown parameter vector \(\theta_0\), and \(u\) is an unobservable error term. In most applications \(v(x, \theta_0)\) is assumed to be a linear function, then the model (the linear binary choice model hereafter) becomes

\[
d = 1\{x_0 + x_0 > u\}
\]

with \(\theta_0 = (x_0, \beta_0)'\), where \(x_0\) and \(\beta_0\) are the intercept and slope parameters, respectively.

For the binary choice model, traditionally, the logit and probit estimation methods are the most popular approaches by restricting the error distribution to parametric families. However, misspecification of the parametric distributions, in general, will result in inconsistent estimates for likelihood-based approaches. In addition, specific functional forms for the error distribution cannot usually be justified by economic theory. This consideration has motivated recent interest in estimating the binary choice model by semiparametric methods which do not assume a parametric distribution for \(u\). Most attention has been focused on the linear binary choice model (see Horowitz, 1993, for a survey) in semiparametric estimation literature. Under various weak distributional assumptions on the error term \(u\), a number of semiparametric estimators have been proposed for \(\theta_0\) (Ahn et al., 1996; Cavanagh and Sherman, 1998; Chen and Lee, 1998; Cosslett, 1983; Han, 1987; Härdle and Stoker, 1989; Horowitz and Härdle, 1996; Ichimura, 1993; Klein and Spady, 1993; Powell et al., 1989; Sherman, 1993), among others, for estimating the slope parameters with no location restriction on \(u\), and Chen, 1998, 1999a; Horowitz, 1992; Lewbel, 1997; Manski, 1985 for estimating the intercept and/or slope parameter with location restrictions imposed on \(u\). Cosslett (1987) considered semiparametric efficiency bounds for the parameters in the binary choice model under the assumption that the error term is independent of the regressors and various location restrictions. The estimators by Chen and Lee (1998) and Klein and Spady (1993), attain the efficiency bound of Cosslett (1987) (and Chamberlain, 1986) for the slope parameter under the independence restriction. Cosslett (1987) further pointed out that under an additional symmetry restriction, not only is the location parameter identified, but there is possible efficiency gain for estimating the slope parameter. Therefore, exploiting the symmetry restriction could potentially lead to more efficient estimator for the slope parameter than the estimator of Klein and Spady (1993).

Consistent estimation of the intercept term is also important for the binary choice model (see Chen, 1998, 1999a and Lewbel, 1997 for detailed discussions). In the context of the sample selection model, by imposing symmetry and index restrictions, Chen (1999b) recently considered \(\sqrt{n}\)-consistent estimation of both the intercept and slope parameters for the outcome equation without the usual exclusion restriction (see Chamberlain, 1986; Powell, 1989, etc.). However,
Chen’s (1999b) approach requires an initial \( \sqrt{n} \)-consistent estimator for the intercept term of the binary choice selection equation.

In this paper, we consider joint estimation of the intercept and slope parameters under index and symmetry restrictions. Following Klein and Spady (1993), we propose a semiparametric maximum likelihood estimator with both the index and symmetry restrictions taken into account in formulating the semiparametric likelihood function. The resulting estimator attains the efficiency bound of Cosslett (1987) under the independence and symmetry restrictions.

The article is organized as follows. Section 2 introduces our estimator and considers its large sample properties. In Section 2, we also compare Cosslett’s efficiency bounds under the independence restriction with and without the symmetry restriction, and discuss possible efficiency gain by imposing the symmetry restriction. In Section 3 a small Monte Carlo study is carried out to assess the finite sample performance of the estimator. Section 4 concludes.

### 2. Motivation and the estimator

For the above binary choice model, it is well known that some scale and location normalization is necessary in order to identify \( \theta_0 \). In parametric models, such as the probit and logit models, scale normalization is accomplished by setting the variance of \( u \) to one. In the semiparametric literature for the linear binary choice model, two common scale normalization schemes are to set either \( |\beta_0| = 1 \) or the coefficient of one component of \( \beta_0 \) to one. Location restrictions are imposed on various locational measures, for example, the conditional mean or median, of \( u \) directly. In this paper, the location and scale restrictions are imposed by setting \( v(x, \theta_0) = \alpha_0 + v_1(x, \beta_0) \) with \( v_1(x, \beta_0) = v_{1a}(x_1) + v_{1b}(x_2, \beta_0) \), and the error term \( u \) being symmetrically distributed conditional on \( x \).

Traditionally, the binary choice model is estimated by the maximum likelihood method, most notably for the logit and probit models. For a random sample of size \( n \), maximum likelihood estimator is defined by maximizing the log-likelihood function

\[
\ln L_n(\theta) = \sum_{i=1}^{n} \left[ d_i \ln F_x(v(x_i, \theta)) + (1 - d_i) \ln(1 - F_x(v(x_i, \theta))) \right],
\]

where the error term \( u \) is assumed to have distribution \( F_x(\cdot) \) given \( x \). However, misspecification of the parametric distribution function, in general, will lead to inconsistent estimates. Consequently, various semiparametric estimators have been proposed without a parametric specification for the error distribution. In particular, under the index restriction that the error distribution depends on \( x \) only through \( v_1(x, \beta_0) \), Klein and Spady (1993) proposed a semiparametric
maximum likelihood estimator for \( \beta_0 \) by maximizing

\[
\ln L_{\text{ks}}(\beta) = \sum_{i=1}^{n} \tau_{ni}[d_i \ln F_n(v_1(x_i, \beta)) + (1 - d_i) \ln(1 - F_n(v_1(x_i, \beta)))],
\]

where \( F_n \) is a nonparametric estimator for the unknown distribution of \( u - z_0 \), and \( \tau_{ni} \) are some trimming functions adopted for technical convenience. Since no location restriction is imposed in their case, they only provide an estimator for \( \beta_0 \). Under the condition that the error term is independent of the regressors, they further show that their estimator attains Cosslett’s (1987) (and Chamberlain’s, 1986) semiparametric efficiency bound.

In this paper, we consider joint estimation of \((z_0, \beta_0)\) under a location restriction in the form of conditional symmetry and an index restriction stronger than the one in Klein and Spady (1993). More specifically, we assume that the conditional density of \( u \) given \( x, f(u | x) \) is an even function of \( u \), and depends on \( x \) only through \( v^2(x, \theta_0) \); that is, \( f(-u | x) = f(u | x) \) and \( f(u | x) = f(u | v^2(x, \theta_0)) \). Following Klein and Spady (1993), we propose a semiparametric maximum likelihood estimator; moreover, we take into account the symmetry as well as the index restriction in formulating the semiparametric likelihood function. Under our assumptions, the location parameter \( z_0 \) can be consistently estimated; furthermore, there are possible efficiency gains in estimating \( \beta_0 \) by imposing the symmetry restriction.

To formulate the semiparametric likelihood function, let \( g(t, \theta) \) denote the density of \( v(x, \theta) \) evaluated at \( t \), \( P(t, \theta) = E(d | v(x, \theta) = t) \), \( P_0(t, \theta) = E(1 - d | v(x, \theta) = t) \), \( g_1(t, \theta) = P(t, \theta)g(t, \theta) \) and \( g_0(t, \theta) = P_0(t, \theta)g(t, \theta) = g(t, \theta) - g_1(t, \theta) \) with \( P(t) = P(t, \theta_0) \) and \( g(t) = g(t, \theta_0) \). Notice that under the symmetry and index restrictions, it can be easily shown that

\[ P(t) = P_0(-t) \]

namely,

\[ E(d | v(x, \theta_0) = t) = E((1 - d) | v(x, \theta_0) = -t) \]

which, in turn, suggests that under our conditions \( G_n(t, \theta_0) \) is a natural nonparametric kernel estimator for \( P(t) \), where \( G_n(t, \theta) \) is defined as the solution to the following problem\(^1\):

\[
\min_{\mu} \sum_{i=1}^{n} \left[ (d_i - \mu)^2 H\left( \frac{v(x_i, \theta) - t}{a_n} \right) + (1 - d_i - \mu)^2 H\left( \frac{v(x_i, \theta) + t}{a_n} \right) \right],
\]

\(^1\) I thank one referee for pointing this out.
where $H$ is a kernel function and $a_n$ is a sequence of bandwidth converging to zero as the sample size increases to infinity, that is,

$$G_n(t, \theta) = \frac{\sum_{i=1}^n [d_i H(v(x_i, \theta) - t)/a_n] + (1 - d_i)H((v(x_i, \theta) + t)/a_n)}{\sum_{i=1}^n [H((v(x_i, \theta) - t)/a_n) + H((v(x_i, \theta) + t)/a_n)].}$$

It is clear that $G_n(t, \theta)$ locally approximates (for $\theta$ in a neighborhood of $\theta_0$) $G_n(t, \theta_0)$. Also, $g_1(t, \theta)$, $g_0(t, \theta)$, and $g(t, \theta)$ can be estimated nonparametrically by $g_{n1}(t, \theta)$, $g_{n0}(t, \theta)$, and $g_n(t, \theta)$, respectively, where

$$g_{n1}(t, \theta) = \frac{1}{na_n} \sum_{i=1}^n d_i H\left(\frac{v(x_i, \theta) - t}{a_n}\right),$$

$$g_{n0}(t, \theta) = \frac{1}{na_n} \sum_{i=1}^n (1 - d_i)H\left(\frac{v(x_i, \theta) - t}{a_n}\right),$$

and $g_n(t, \theta) = g_{n1}(t, \theta) + g_{n0}(t, \theta)$. Hence, $G_n(t, \theta)$ can be rewritten as

$$G_n(t, \theta) = \frac{g_{n1}(t, \theta) + g_{n0}(-t, \theta)}{g_n(t, \theta) + g_n(-t, \theta)}.$$ 

These nonparametric estimators will provide the basis for constructing the semiparametric likelihood function.

Following Klein and Spady (1993) we introduce probability and likelihood trimmings to deal with technical difficulties arising from small estimated densities. For the probability trimming, let

$$\delta_{n1}(t, \theta) \equiv a_n^5 [e^{z_1}/(1 + e^{z_1})] \quad z_1 = [(a_n^{e^{z_1}} - g_{n1}(t, \theta))/a_n^{e^{z_1}}],$$

$$\delta_{n0}(t, \theta) \equiv a_n^5 [e^{z_0}/(1 + e^{z_0})] \quad z_0 = [(a_n^{e^{z_0}} - g_{n0}(t, \theta))/a_n^{e^{z_0}}],$$

$$\delta_n(t, \theta) \equiv a_n^5 [e^{z}/(1 + e^{z})] \quad z = [(a_n^{e^{z}} - g_{n1}(t, \theta))/a_n^{e^{z}}],$$

then we define the estimated probability function as

$$\hat{G}_n(t, \theta) = \frac{g_{n1}(t, \theta) + g_{n0}(-t, \theta) + \delta_{n1}(t, \theta) + \delta_{n0}(-t, \theta)}{g_n(t, \theta) + g_n(-t, \theta) + \delta_n(t, \theta) + \delta_n(-t, \theta)}.$$ 

We now consider the likelihood trimming. Let $\hat{\theta}_p$ be a preliminary consistent estimate for $\theta_0$ for which $|\hat{\theta}_p - \theta_0| = O_p(n^{-1/3})$. The estimators mentioned above satisfy this condition. Then the likelihood trimming function is defined as

$$\hat{\tau}_i(\theta) = \tau_i(\hat{\theta}_p)\hat{\tau}_{oi}(\theta),$$

where

$$\hat{\tau}_i(\theta) = \tau(g_{n1}(v(x_i, \theta), \theta), \varepsilon)\tau(g_{n1}(v(x_i, \theta), \theta), \varepsilon),$$

$$\hat{\tau}_{oi}(\theta) = \tau(g_{n0}(v(x_i, \theta), \theta), \varepsilon)\tau(g_{n0}(v(x_i, \theta), \theta), \varepsilon),$$

and

$$\tau(s, \varepsilon) \equiv \{1 + \exp[(a_n^{e^{z}} - s)/a_n^{e^{z}}]\}^{-1}.$$
with \( \varepsilon'' < \varepsilon' < 1 \). As in Klein and Spady (1993), the likelihood trimming is more severe than the probability trimming. Finally, we define our semiparametric maximum likelihood estimator, \( \hat{\theta} \), by maximizing

\[
\hat{Q}_n(\theta, \hat{\theta}) = \sum \left( \hat{\theta}_i (\hat{\theta}_p)_i / 2 \right) \ln \left[ \hat{G}_n^2(\theta(x_i, \theta), \theta) \right] \\
+ (1 - d_i) \ln \left[ (1 - \hat{G}_n(\theta(x_i, \theta), \theta))^2 \right].
\]  

(1)

Let \( D^r_m(z) \) denote the \( r \)th-order partial of \( m(z) \) with respect to \( z \) and \( D^0_m(z) = m(z) \). Let \( c \) denote a generic positive constant. We make the following assumptions.

**Assumption 1.** The observations \((d_i, x_i), i = 1, 2, \ldots, n\) are i.i.d. The conditional density of \( u, f(u \mid x) \), depends on \( x \) only through \( v^2(x, \theta_0) \), and symmetric around 0, namely, \( f(u \mid x) = f(u \mid v(x, \theta_0)) = f(u \mid v^2(x, \theta_0)) \) and \( f(-u \mid x) = f(u \mid x) \); and \( f(u \mid v(x, \theta_0) = t) \) is four times continuously differentiable with respect to \( (u, t) \) and its derivatives are uniformly bounded.

**Assumption 2.** The parameter vector \( \theta \) lies in a compact parameter space, \( \Theta \). The true value of \( \theta_0 \) is in the interior of \( \Theta \).

**Assumption 3.** There exist \( P, \bar{P} \) that do not depend on \( x \) such that

\[
0 < P \leq E(d \mid x) \leq \bar{P} < 1.
\]

**Assumption 4.** The index function \( v(x, \theta) \) is smooth in that for \( \theta \) in a neighborhood of \( \theta_0 \) and all \( s \):

\[
\{ | D^r_0 v(s; \theta)|, | \partial D^r_0 v(s; \theta)/\partial x | \} < c \quad (r = 0, 1, 2, 3, 4).
\]

Also, the conditional density of \( x_1 \) given \( x_2, f_{x_1 \mid x_2}(x_1) \) is supported on \([a, b]\) and smooth in that for all \( x \)

\[
| D^r_s f_{x_1 \mid x_2}(s) | < c \quad (r = 0, 1, 2, 3, 4; \ s \in [a, b]).
\]

**Assumption 5.** The kernel function \( H(v) \) is a symmetric function that integrates to one, four times continuously differentiable with bounded derivatives. Also, \( H(v) \) is a higher-order kernel in that \( \int v^2 H(v) dv = 0 \), and \( \int v^4 H(v) dv < \infty \).

**Assumption 6.** The bandwidth sequence \( \{a_n\} \) is chosen such that \( n^{-1/(6 + 2 \varepsilon)} < a_n < n^{-1/(8 - \varepsilon)} \).

**Assumption 7.** The model is identified in that for almost all \( x \),

\[
G(v(x, \theta_\star), \theta_\star) = G(v(x, \theta_0), \theta_0) \quad \text{implies} \quad \theta_\star = \theta_0,
\]
where
\[ G(v(x, \theta), \theta) = \frac{g_1(v(x, \theta), \theta) + g_0(-v(x, \theta), \theta)}{g(v(x, \theta), \theta) + g(-v(x, \theta), \theta)}. \]

Assumption 1 describes the model and the data. The symmetry restriction is a location normalization. Compared with the usual index restriction in the literature that \( f(u | x) = f(u | v(x, \theta_0)) \) (see, e.g. Ichimura, 1993; Klein and Spady, 1993), our index restriction is slightly stronger; however, in both cases, the unknown form of heteroscedasticity allowed is very limited. Assumptions 2–6 are similar to those in Klein and Spady (1993) (see Klein and Spady, 1993 for detailed discussions); in particular, a higher-order bias reducing kernel is used here. However, following Klein and Spady (1993), we can also adopt an adaptive kernel function with local smoothing that has similar bias reduction properties. Specifically, we set the bandwidth to be variable and data dependent. For sample size \( n \), let \( \hat{I}_j = (1/n_a L_n) \sum_{i=1}^n H(v(x_i, \theta) - t)/a_{np} \) be a preliminary density estimator with bandwidth \( a_{np} \). For \( 0 < v' < 1 \), set the bandwidth for observation \( j \) to \( a_{nj} = a_n \hat{v}(\theta) \hat{L}_j \), where \( \hat{v}(\theta) \) is the sample standard deviation of \( v(x, \theta) \),
\[ \hat{L}_j = \{[\hat{v}_j + (1 - \hat{v}_p)/m]\} \]
with
\[ \hat{v}_p = \tau(\hat{v}_j, a_{np}') \]
and \( m \) being the geometric mean of the \( \hat{v}_j \). As pointed out by Klein and Spady (1993), the estimated probabilities using local smoothing are always positive, whereas those based on higher-order kernels can be negative. Assumption 7 is an identification condition. Note that \( G(v(x, \theta_0), \theta_0) = P(v(x, \theta_0)) \) under our index and symmetry restrictions. Klein and Spady (1993) provided two sets of models for which the slope parameters are identified, allowing both monotonic and nonmonotonic models. We first consider the identification of the slope parameter. For a fixed \( \theta = (\alpha, \beta)' \), if
\[ G(v(x, \theta), \theta) = G(v(x, \theta_0), \theta_0) = P(v(x, \theta_0)) \]
for almost all \( x \), then define \( G_x(x, \beta) = G(x + v_1(x, \beta), (\alpha, \beta)') \), which can be viewed as a function of \( v_1(x, \beta) \), then by following the arguments of Klein and Spady (1993) we can show that \( \beta = \beta_0 \) if the assumptions in Theorems 1 or 2 of Klein and Spady (1993) are satisfied. Once the slope parameter is identified, we only need to show that
\[ G(v(x, (\alpha, \beta_0)'), (\alpha, \beta_0)') = G(v(x, \theta_0), \theta_0) \]
for almost all \( x \) implies \( \alpha = \alpha_0 \). It is straightforward to show that under our index and symmetry restrictions
\[ G(v(x, (\alpha, \beta_0)'), (\alpha, \beta_0)') = g_{w1}(x, \alpha)P(v(x, \theta_0)) \]
\[ + g_{w2}(x, \alpha)P(v(x, \theta_0 + 2(\alpha - \alpha_0))), \tag{2} \]
where
\[
g_{w1}(x, z) = \frac{g(v(x, (z, \beta_0')), (z, \beta_0'))}{g(v(x, (z, \beta_0')'), (z, \beta_0')) + g(-v(x, (z, \beta_0')'), (z, \beta_0'))} = \frac{g(v(x, \theta_0))}{g(v(x, \theta_0)) + g(-v(x, \theta_0) + 2z_0 - 2z)}
\]

and
\[
g_{w2}(x, z) = \frac{g(-v(x, (z, \beta_0')'), (z, \beta_0'))}{g(v(x, (z, \beta_0')'), (z, \beta_0')) + g(-v(x, (z, \beta_0')'), (z, \beta_0'))} = \frac{g(-v(x, \theta_0) + 2z_0 - 2z)}{g(v(x, \theta_0) + g(-v(x, \theta_0) + 2z_0 - 2z)}
\]

with \(g_{w1}(x, z) + g_{w2}(x, z) = 1\). Thus (2) is equivalent to
\[
g_{w2}(x, z)[P(v(x, \theta_0) + 2(z - z_0)) - P(v(x, \theta_0))] = 0.
\] (3)

When the support of \(v(x, \theta_0)\) is large enough so that there exists an interval \([a_v, b_v]\), we have \(g(t)g(-t + 2z_0 - 2z) > 0\) for \(t \in [a_v, b_v]\) and \((z, \beta_0) \in \Theta\), then identification of \(z_0\) follows if
\[
P(t) = P(t + 2(z - z_0))
\] (4)

for \(t \in [a_v, b_v]\), implies \(z = z_0\), which, in turn, obviously holds for monotonic models. Even many nonmonotonic models, (4) typically ensures the identification of \(z_0\); for example, identification of \(z_0\) follows easily from (4) when \(P(t) > 0.5\) if and only if \(t > 0\), and \(dP(t)/dt\) changes sign only once for \(t > 0\).

**Theorem 1.** Under Assumptions 1–7, the estimator \(\hat{\theta}\) is consistent and asymptotically normal,
\[
\sqrt{n}(\hat{\theta} - \theta_0) \overset{D}{\rightarrow} N(0, \Sigma),
\]
where
\[
\Sigma = \mathbb{E}\left\{\frac{\partial G}{\partial \theta} \frac{\partial G}{\partial \theta'} \left[\frac{1}{P(v)(1 - P(v))}\right]\right\}^{-1}
\]

with \(v = v(x, \theta_0)\) and \(\partial G/\partial \theta = [\partial G(v(x, \theta_0), \theta_0)]/\partial \theta\).

**Proof.** See the appendix. \(\square\)

We now turn to the issue of efficiency. Klein and Spady (1993) have shown that their semiparametric maximum likelihood estimator attains the asymptotic efficiency bound of Chamberlain (1986) and Cosslett (1987) associated with the
independence restriction. We will show that by also taking the symmetry restriction into account, our estimator attains the asymptotic efficiency bound of Cosslett (1987) associated with both the independence and symmetry restrictions.

Lemma 2. For the binary choice model under consideration here, the asymptotic covariance of \( \hat{\theta} \) achieves the efficiency bound as specified in Cosslett (1987) under the condition that the error is independent of the regressors, and symmetrically distributed around the origin.

Following the arguments of Lemma 2 of Chen and Lee (1998) and Klein and Spady (1993), we can show that under the independence and symmetry restrictions

\[
\frac{\partial}{\partial \theta} \left[ \mathbb{E}(d \mid v(x_i, \theta)) = v(x, \theta) \right] |_{\theta = \theta_0} = f(v) \left[ \frac{\partial v}{\partial \theta} - \mathbb{E} \left( \frac{\partial v}{\partial \theta} \right) \right],
\]

and

\[
\frac{\partial}{\partial \theta} \left[ \mathbb{E}(d \mid v(x_i, \theta)) = -v(x, \theta) \right] |_{\theta = \theta_0} = -f(v) \left[ \frac{\partial v}{\partial \theta} + \mathbb{E} \left( \frac{\partial v}{\partial \theta} \right) - v \right],
\]

where \( \partial v/\partial \theta = [\partial v(x, \theta_0)]/\partial \theta \). Therefore,

\[
\frac{\partial G}{\partial \theta} = f(v) \left( \frac{\partial v}{\partial \theta} - K(v) \right),
\]

where

\[
K(v) = \frac{1}{g(v) + g(-v)} \left[ g(v) \mathbb{E} \left( \frac{\partial v}{\partial \theta} \right) - g(-v) \mathbb{E} \left( \frac{\partial v}{\partial \theta} \right) - v \right],
\]

as specified in Cosslett (1987). Therefore,

\[
\Sigma^{-1} = \mathbb{E} \left\{ \frac{f^2(v)}{P(v)(1 - P(v))} \left( \frac{\partial v}{\partial \theta} - K(v) \right) \left( \frac{\partial v}{\partial \theta} - K(v) \right) \right\}. \tag{5}
\]

Then Lemma 2 follows by comparing (5) and Eq. (3.27) in Cosslett (1987).

Since the estimator by Klein and Spady \( \hat{\beta}_{ks} \) attains the efficiency bound under the independence restriction, while our estimator \( \hat{\theta} = (\hat{\alpha}, \hat{\beta}) \) attains the efficiency bound under both the independence and symmetry restrictions, it is of interest to examine whether the additional symmetry restriction leads to any efficiency gain in estimating \( \beta_0 \). From Cosslett (1987), the efficiency bound for \( \beta_0 \) under the independence restriction is

\[
I_1 = \mathbb{E} \left\{ \frac{f^2(v)}{P(v)(1 - P(v))} \left( \frac{\partial v}{\partial \beta} - \mathbb{E} \left( \frac{\partial v}{\partial \beta} \right) \right) \left( \frac{\partial v}{\partial \beta} - \mathbb{E} \left( \frac{\partial v}{\partial \beta} \right) \right) \right\},
\]
while the corresponding efficiency bound for $\beta_0$ under the independence and symmetry restrictions is $((I^{-1}_{2(\beta)})^{-1})$, where $(I^{-1}_{2(\beta)})$ is the lower right block of $I^{-1}_{2}$, with

$$I_2 = \mathbb{E}\left\{ \frac{f^2(v)}{P(v)(1 - P(v))} \left( \frac{\partial v}{\partial \theta} - K(v) \right) \left( \frac{\partial v}{\partial \theta} - K(v) \right) \right\}. $$

By some algebraic manipulation, we can show that

$$((I^{-1}_{2(\beta)})^{-1}) - I_1 = \left\langle \mathbb{E}\left( \frac{\partial v}{\partial \beta} \right), \mathbb{E}\left( \frac{\partial v}{\partial \beta} \right) \right\rangle - K(v) \right\rangle - \left\langle \mathbb{E}\left( \frac{\partial v}{\partial \beta} \right), g^* \right\rangle \langle g^*, g^* \rangle^{-1} \times \left\langle g^*, \mathbb{E}\left( \frac{\partial v}{\partial \beta} \right) \right\rangle - K(v),$$

where $g^* = [2g(-v)]/[g(v) + g(-v)]$, and $\langle \cdot, \cdot \rangle$ denotes an inner product such that

$$\langle g_1, g_2 \rangle = \mathbb{E}\left\{ \frac{f^2(v)}{P(v)(1 - P(v))} g_1 g_2 \right\}. $$

Consequently $((I^{-1}_{2(\beta)})^{-1}) - I_1$, can be viewed as the corresponding norm of the projection residual of $E[v \mid \beta] - K(v)$ onto $g^*$, it is thus semi-definite, as expected. When $v(x, \theta)$ is linear, and $x$ has a multivariate normal distribution, we have $E(x \mid v) = a + bv$ and $E(x \mid -v) = a - bv$, thus

$$E\left( \frac{\partial v}{\partial \beta} \right) - K(v) = \frac{g(-v)}{g(v) + g(-v)} \left( E\left( \frac{\partial v}{\partial \beta} \right) - v \right) = cg^*. $$

Therefore, in this special case $((I^{-1}_{2(\beta)})^{-1}) = I_1$, and there will be no efficiency gain in imposing the symmetry restriction. In fact, the semiparametric efficiency bound for the slope parameter under the independence restriction in this case agrees with the parametric bound, as noted in Cosslett (1987). In general, however,

$$\frac{g(-v)}{g(v) + g(-v)} \left( E\left( \frac{\partial v}{\partial \beta} \right) - v \right)$$

is of full rank, and there is indeed efficiency gain in estimating the slope parameter by imposing the additional symmetry restriction.
3. A small Monte Carlo study

In this section we perform some Monte Carlo simulations to investigate the finite sample performance of our estimator. The data is generated according to the following model:

\[ d = 1\{z_0 + x_1 + \beta_0 x_2 + u > 0\} \]

with \((z_0, \beta_0) = (0, 1)\). Four different designs are constructed by varying the distributions of the regressors and error term.

Notice that given the estimator for \(\beta_0, \hat{\beta}_{ks}\), of Klein and Spady (1993), \(z_0\) can also be estimated by \(\hat{z}_{1c}\), defined by maximizing \(\hat{Q}_n(x, \hat{\beta}_{ks}, \hat{\tau})\) in (1) with respect to \(\tau\). Here we consider the finite sample performance of our estimator, \((\hat{z}, \hat{\beta})\), the probit estimator \((\hat{z}_{pb}, \hat{\beta}_{pb})\), \(\hat{z}_{1c}\) and \(\hat{\beta}_{ks}\) defined above. We examine the efficiency loss relative to a correct parametric specification and efficiency gain in imposing symmetry and resulting joint estimation.

The results from 300 replications for each design are presented with sample sizes of 100 and 300. For each estimator under consideration, we report the mean value (Mean), the standard deviation (SD), and the root mean square error (RMSE). As in Klein and Spady (1993), we report results for the semiparametric estimators obtained without probability or likelihood trimming since estimates obtained are not sensitive to trimming. Also, as in Klein and Spady (1993), adaptive local smoothing with standard normal kernel is used for the semiparametric estimators, with local smoothing parameters defined without any trimming. The nonstochastic window width \(a_n\) and the corresponding pilot window component \(a_{np}\) are set to \(n^{-1/7}\) to satisfy the restriction specified above.

Tables 1 and 2 report the results for two homoscedastic designs in which \(u\) is drawn from \(N(0, 1)\) independent of \((x_1, x_2)\). In Homoscedastic Design I,

**Table 1**

Homoscedastic design I

<table>
<thead>
<tr>
<th>(n)</th>
<th>Estimator</th>
<th>Mean</th>
<th>SD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>(\hat{z}_{pb})</td>
<td>0.000</td>
<td>0.149</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td>(\hat{\beta}_{pb})</td>
<td>1.026</td>
<td>0.238</td>
<td>0.239</td>
</tr>
<tr>
<td></td>
<td>(\hat{z}_{ce})</td>
<td>-0.003</td>
<td>0.165</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>(\hat{z}_{ks})</td>
<td>1.037</td>
<td>0.283</td>
<td>0.285</td>
</tr>
<tr>
<td></td>
<td>(\hat{z}_{1})</td>
<td>-0.005</td>
<td>0.160</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>(\hat{z}_{2})</td>
<td>1.030</td>
<td>0.248</td>
<td>0.250</td>
</tr>
<tr>
<td>300</td>
<td>(\hat{z}_{pb})</td>
<td>0.005</td>
<td>0.091</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>(\hat{\beta}_{pb})</td>
<td>1.000</td>
<td>0.131</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>(\hat{z}_{ce})</td>
<td>0.003</td>
<td>0.095</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>(\hat{z}_{ks})</td>
<td>1.007</td>
<td>0.160</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>(\hat{z}_{1})</td>
<td>0.003</td>
<td>0.095</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>(\hat{z}_{2})</td>
<td>1.002</td>
<td>0.138</td>
<td>0.138</td>
</tr>
</tbody>
</table>
both $x_1$ and $x_2$ are drawn from N(0, 1) independent of each other, while $x_1$ is still drawn from N(0, 1) and $x_2$ is drawn independently from a standardized chi-squared distribution with 3 degrees of freedom with zero mean and unit variance in Homoscedastic Design II. The probit estimator, as expected, performs best in both designs, while our estimator is competitive and compared favorably with $\beta_{ks}$, even though earlier discussions on efficiency suggests that both the estimator of Klein and Spady (1993) and our estimator attain the parametric efficiency bound when $(x_1, x_2)$ is jointly normal. In both cases, $\hat{\beta}_{c1}$ is also quite competitive compared with our estimator for the intercept.

Tables 3 and 4 report the results for two heteroscedastic designs in which $u = 0.7u^*(1 + 0.5v + 0.25v^2)$ for $v = z_0 + x_1 + \beta_0 x_2$ with $u^*$ drawn from

### Table 2
Homoscedastic design II

<table>
<thead>
<tr>
<th>$n$</th>
<th>Estimator</th>
<th>Mean</th>
<th>SD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>$\hat{\beta}_{ph}$</td>
<td>-0.004</td>
<td>0.148</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{ph}$</td>
<td>0.995</td>
<td>0.234</td>
<td>0.234</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_c$</td>
<td>-0.007</td>
<td>0.169</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{ks}$</td>
<td>1.017</td>
<td>0.335</td>
<td>0.335</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
<td>-0.012</td>
<td>0.160</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_2$</td>
<td>0.977</td>
<td>0.240</td>
<td>0.241</td>
</tr>
<tr>
<td>300</td>
<td>$\hat{\beta}_{ph}$</td>
<td>0.009</td>
<td>0.087</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{ph}$</td>
<td>1.008</td>
<td>0.139</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_c$</td>
<td>0.004</td>
<td>0.094</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{ks}$</td>
<td>1.016</td>
<td>0.169</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
<td>0.002</td>
<td>0.092</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_2$</td>
<td>0.998</td>
<td>0.140</td>
<td>0.140</td>
</tr>
</tbody>
</table>

### Table 3
Heteroscedastic design I

<table>
<thead>
<tr>
<th>$n$</th>
<th>Estimator</th>
<th>Mean</th>
<th>SD</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>$\hat{\beta}_{ph}$</td>
<td>-0.004</td>
<td>0.251</td>
<td>0.251</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{ph}$</td>
<td>1.003</td>
<td>0.339</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_c$</td>
<td>-0.010</td>
<td>0.241</td>
<td>0.241</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{ks}$</td>
<td>1.007</td>
<td>0.308</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
<td>-0.012</td>
<td>0.240</td>
<td>0.240</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_2$</td>
<td>1.004</td>
<td>0.302</td>
<td>0.302</td>
</tr>
<tr>
<td>300</td>
<td>$\hat{\beta}_{ph}$</td>
<td>0.009</td>
<td>0.164</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{ph}$</td>
<td>0.981</td>
<td>0.231</td>
<td>0.232</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_c$</td>
<td>0.002</td>
<td>0.133</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_{ks}$</td>
<td>1.002</td>
<td>0.217</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_1$</td>
<td>0.001</td>
<td>0.130</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}_2$</td>
<td>0.996</td>
<td>0.193</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>$\hat{\delta}_b$</td>
<td>Mean</td>
<td>SD</td>
<td>RMSE</td>
</tr>
<tr>
<td>100</td>
<td>$\hat{\delta}_b$</td>
<td>$-0.056$</td>
<td>$0.235$</td>
<td>$0.241$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_b$</td>
<td>$0.910$</td>
<td>$0.366$</td>
<td>$0.377$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_c$</td>
<td>$-0.009$</td>
<td>$0.223$</td>
<td>$0.223$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_{ks}$</td>
<td>$0.987$</td>
<td>$0.336$</td>
<td>$0.336$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_1$</td>
<td>$-0.014$</td>
<td>$0.216$</td>
<td>$0.217$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_2$</td>
<td>$0.978$</td>
<td>$0.311$</td>
<td>$0.312$</td>
</tr>
<tr>
<td>300</td>
<td>$\hat{\delta}_b$</td>
<td>$0.083$</td>
<td>$0.164$</td>
<td>$0.184$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_b$</td>
<td>$0.904$</td>
<td>$0.220$</td>
<td>$0.240$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_c$</td>
<td>$0.000$</td>
<td>$0.136$</td>
<td>$0.136$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_{ks}$</td>
<td>$1.014$</td>
<td>$0.193$</td>
<td>$0.194$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_1$</td>
<td>$0.003$</td>
<td>$0.137$</td>
<td>$0.137$</td>
</tr>
<tr>
<td></td>
<td>$\hat{\delta}_2$</td>
<td>$1.010$</td>
<td>$0.190$</td>
<td>$0.190$</td>
</tr>
</tbody>
</table>

N(0, 1) independent of $(x_1, x_2)$. In Heteroscedastic Design I, $(x_1, x_2)$ is drawn as in Homoscedastic Design I, and in Heteroscedastic Design II, $(x_1, x_2)$ is drawn as in Homoscedastic Design II. In Heteroscedastic Design I, the probit estimator still has small bias despite misspecification, but its bias increases significantly in Heteroscedastic Design II. In both cases, the probit estimator has the largest RMSE, whereas all the semiparametric estimators perform well.

4. Conclusions

In this paper we have proposed a semiparametric maximum likelihood estimator for both the intercept and slope parameters in a binary choice model under symmetry and index restrictions. The estimator is consistent and asymptotically normal. Furthermore, it attains the semiparametric efficiency bound in Cosslett (1987) under the symmetry and independence restrictions. A small Monte Carlo study is carried out to illustrate the usefulness of our estimator. Compared with the estimator of Klein and Spady (1993), which attains the semiparametric efficiency bound in Chamberlain (1986) and Cosslett (1987) under the independence restriction, we show that there are possible efficiency gains in imposing the additional symmetry restriction. At the same time, however, our approach depends crucially on the validity of the symmetry restriction. The symmetry requirement is a testable restriction. One apparent approach is to use Hausman test by comparing our estimator and the estimator of Klein and Spady (1993). Alternatively, conditional moment tests (Newey, 1985) can be used since the conditional moment restriction $E(d|x) = G(v(x, \theta_0), \theta_0)$ holds almost surely only when the symmetry restriction is satisfied.
Recently, Chamberlain (1986) derived the semiparametric efficiency bound for the slope parameters for the binary choice sample selection model under the independence restriction. Various semiparametric estimators have been proposed for the slope parameters. In particular, the estimators by Ai (1997) and Chen and Lee (1998) attain the efficiency bound. In view of the results obtained in the paper, there might be possible efficiency gains in imposing an additional symmetry restriction. This is a topic for future research.

Acknowledgements

I would like to thank Bo Honore, Roger Klein, Lung-Fei Lee and Jim Powell and two anonymous referees for their helpful comments. I am also grateful to Roger Klein for providing Gauss codes.

Appendix. Proof of Theorem 1

Theorem 1 is proved by following Chen and Lee (1998), Klein and Spady (1993) and Lee (1994). Consistency follows from arguments similar to those in Klein and Spady (1993) by showing that the semiparametric likelihood function converges uniformly in $\theta$ to a function uniquely maximized at $\theta_0$. To prove asymptotic normality, we first consider a Taylor expansion to obtain

$$\sqrt{n} (\hat{\theta} - \theta_0) = \left[ - \frac{\partial^2 \hat{Q}(\theta_0, \hat{\tau})}{\partial \theta \partial \theta'} \right]^{-1} \sqrt{n} \frac{\partial \hat{Q}(\theta_0, \hat{\tau})}{\partial \theta} \quad (A.1)$$

where $\theta_0^+$ lies between $\hat{\theta}$ and $\theta_0$. Similar to Lemma 5 of Klein and Spady (1993), we can show that

$$- \frac{\partial^2 \hat{Q}(\theta_0, \hat{\tau})}{\partial \theta \partial \theta'} = E \left[ \frac{\partial G}{\partial \theta} \frac{\partial G}{\partial \theta'} \frac{G}{P(v)(1 - P(v))} \right] + o_p(1).$$

For the gradient term in (A.1), similar to Lemma 3 of Klein and Spady (1993), we have the following Taylor expansion for the likelihood trimming term

$$\hat{\tau}_i(\hat{\theta}_p, \epsilon^*) = \hat{\tau}_i(\theta_0, \epsilon^*) + L_m(\theta_0, \epsilon^*)(\hat{\theta}_p - \theta_0) + \frac{1}{2}(\hat{\theta}_p - \theta_0)^T Q_m(\theta_0^+, \epsilon^*)(\hat{\theta}_p - \theta_0),$$

where $L_m(\theta, \epsilon^*)$ and $Q_m(\theta, \epsilon^*)$ represent the first- and second-order partial derivative functions and $\theta_0^+$ lies between $\hat{\theta}_p$ and $\theta_0$. Thus, we obtain

$$\sqrt{n} \frac{\partial \hat{Q}(\theta_0, \hat{\tau})}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{\tau}_i(\theta_0, \epsilon^*) \frac{d_i - G_i}{G_i(1 - G_i)} \frac{\partial G_i}{\partial \theta}$$

$$= S_{n1} + S_{n2}(\hat{\theta}_p - \theta_0) + \sqrt{n}(\hat{\theta}_p - \theta_0) S_{n3}(\hat{\theta}_p - \theta_0).$$
where

\[
S_{n1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{d_i - \hat{G}_i}{G_i(1 - G_i)} \frac{\partial \hat{G}_i}{\partial \theta},
\]

\[
S_{n2} = \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} L_{ni}(\theta_0, \epsilon^n) \frac{d_i - \hat{G}_i}{G_i(1 - G_i)} \frac{\partial \hat{G}_i}{\partial \theta} \right]
\]

and

\[
S_{n3} = \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Q_{ni}(\theta_2^*, \epsilon^n) \frac{d_i - \hat{G}_i}{G_i(1 - G_i)} \frac{\partial \hat{G}_i}{\partial \theta} \right]
\]

with \( \hat{G}_i = \hat{G}_n(v(x_i, \theta_0), \theta_0) \) and \( \frac{\partial \hat{G}_i}{\partial \theta} = [\partial \hat{G}_n(v(x_i, \theta_0), \theta_0)]/\partial \theta \). First, write \( S_{n1} \) as

\[
S_{n1} = S_{n10} + S_{n11} + S_{n12},
\]

where

\[
S_{n10} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tau_i(\theta_0) \frac{d_i - P(v_i)}{P(v_i)(1 - P(v_i))} \frac{\partial G(v(x_i, \theta_0, \theta_0)}{\partial \theta},
\]

\[
S_{n11} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (d_i - P(v_i)) \left[ \frac{\hat{G}_i}{G_i(1 - G_i)} \frac{\partial \hat{G}_i}{\partial \theta} - \frac{\tau_i(\theta_0)}{P(v_i)(1 - P(v_i))} \frac{\partial G(v(x_i, \theta_0, \theta_0)}{\partial \theta} \right]
\]

and

\[
S_{n12} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tau_{1i}(\theta_0, \epsilon^n) \frac{\hat{G}_i - P(v_i)}{G_i(1 - G_i)} \frac{\partial \hat{G}_i}{\partial \theta}
\]

with \( \tau_i(\theta) = \tau_{1i}(\theta)\tau_{0i}(\theta) \), for

\[
\tau_{1i}(\theta) = \tau(g_1(v(x_i, \theta), \theta), \epsilon^n)\tau(g_1(-v(x_i, \theta), \theta), \epsilon^n),
\]

and

\[
\tau_{0i}(\theta) = \tau(g_0(v(x_i, \theta), \theta), \epsilon^n)\tau(g_0(-v(x_i, \theta), \theta), \epsilon^n),
\]

Following the arguments in Lee (1994, pp. 381–384), which involves Taylor expansion and repeated applications of Propositions 4 and 6 of Lee (1994), we can show that \( S_{n11} = o_p(1) \). For \( S_{n12} \), by applying Taylor expansion and U-statistic projections as in Lee (1994) and Chen and Lee (1998), we can
show that

\[ S_{n12} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tau_i(\theta_0) \frac{\hat{G}_i - P(v_i)}{P(v_i)(1 - P(v_i))} \frac{\partial G(v(x_i, \theta_0), \theta_0)}{\partial \theta} + o_p(1) \]

\[ = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tau_i(\theta_0) g(v_i) g(-v_i)(q(v_i) - q(-v_i))(d_i - P(v_i)) + o_p(1), \]

where

\[ q(v) = \frac{f(v)}{F(v)(1 - F(v))} \left[ g(v) + g(-v) \right] \left[ E\left( \frac{\partial \hat{v}}{\partial \theta} \right) + E\left( \frac{\partial \hat{v}}{\partial \theta} \right) - v \right]. \]

Note that \( q(v) = q(-v) \), thus \( S_{n12} = o_p(1) \). By similar arguments, we can show that

\[ n^{-1/3} S_{n2} = o_p(1) \]

hence, \( S_{n2}(\hat{\theta}_p - \theta_0) = o_p(1) \). Finally, we turn to \( S_{n3} \). Following the arguments of Lemma 5 in Klein and Spady (1993), we can show that

\[ 1 - \frac{n}{\sqrt{n}} Q_{n1}(\theta^+, \theta^-) \frac{d_i - \hat{G}_i}{\hat{G}_i(1 - \hat{G}_i)} \frac{\partial \hat{G}_i}{\partial \theta} = o_p(1). \]

Therefore, \( \sqrt{n}(\hat{\theta}_p - \theta_0)^{-1} S_{n3}(\hat{\theta}_p - \theta_0) = o_p(1) \). Combining the above results, we have

\[ \sqrt{n}(\hat{\theta} - \theta_0) = \left[ E\left( \frac{\partial G \partial G}{\partial \theta \partial \theta} \frac{1}{P(v)(1 - P(v))} \right) \right]^{-1} \]

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tau_i(\theta_0) \frac{d_i - P(v_i)}{P(v_i)(1 - P(v_i))} \frac{\partial G(v(x_i, \theta_0), \theta_0)}{\partial \theta} + o_p(1). \]

Then Theorem 1 follows by applying a central limit theorem (Serfling, 1980, p. 32). □

References

Chen, S., 1998. Rank estimation of a location parameter in the binary choice model. Hong Kong University of Science and Technology.


