Econometric Theory, **15**, 1999, 79–98. Printed in the United States of America.

SEMIPARAMETRIC ESTIMATION OF A LOCATION PARAMETER IN THE BINARY CHOICE MODEL

SONGNIAN CHEN
Hong Kong University of Science and Technology

This paper considers the estimation of a location parameter in the binary choice model with some weak distributional assumptions imposed on the error term in the latent regression model. Two estimators are proposed here, both of which are two-step estimators; in the first step, the slope parameters are consistently estimated by existing methods; in the second step, the location parameter is consistently estimated based on a moment condition. The estimators are shown to be consistent and asymptotically normal. A small Monte Carlo study illustrates the usefulness of the estimators. We also point out that the location and slope parameters can be estimated simultaneously.

1. INTRODUCTION

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

$$d = \begin{cases} 1 & \text{if } x\beta_0 - \alpha_0 - v > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where d is the indicator of the response (or participation) variable, x is a vector of exogenous variables, $\alpha_0 \in R$ and $\beta_0 \in R^q$ are unknown parameters, and the error term v is assumed to satisfy a monotonic index and symmetry condition.

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. For the binary choice model, 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 can not usually be justified by economic theory. Here we consider estimating the binary choice model without assuming a parametric distribution for v. As a result, our estimators will be robust to the misspecification of the error distribution. It is well known that some scale normalization is necessary to identify the intercept

I thank Jianqing Fan, Bo Honoré, Roger Klein, Lung-Fei Lee, James Powell, two anonymous referees, and the co-editor Joel Horowitz for helpful comments. Address correspondence to: Songnian Chen, Department of Economics, the Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, People's Republic of China; e-mail: snchen@ust.hk.

© 1999 Cambridge University Press 0266-4666/99 $9.50

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

and slope parameters. We adopt a common normalization scheme by setting $\beta_{01} = 1$, where β_{01} is the first component of β_0. Abusing notation slightly, we still use (α_0, β_0) to denote the normalized parameters. There have been several semiparametric estimators proposed for (α_0, β_0) in the literature. Ahn, Ichimura, and Powell (1996), Cosslett (1983), Han (1987), Horowitz and Härdle (1996), Ichimura (1993), Klein and Spady (1993), Powell, Stock, and Stoker (1989), and Sherman (1993), among others, considered estimating the slope parameter β_0 under a single-index restriction. The intercept term α_0 is not identified under the single-index restriction because no centering assumption is made about the distribution of v. Under an independence and mean restriction Lewbel (1997) considered estimating the intercept term α_0, whereas Manski (1985) and Horowitz (1992) considered estimating both the slope and intercept terms under a conditional median restriction.

The estimators for the intercept term in Manski (1985) and Horowitz (1992) converge at a rate slower than the usual parametric rate because they are based on a median restriction. In estimating α_0, Lewbel (1997) imposed homoskedasticity and a strong restriction on the support of the linear index $x\beta_0$.[1] We will focus on root-n (n is the sample size) consistent estimation of the intercept term α_0 under a symmetry and monotonic index assumption that strengthens both the index and conditional median restrictions. We allow certain unknown heteroskedasticity.

By obtaining root-n consistent estimators for α_0, our results will complement the existing semiparametric estimation literature, thus making semiparametric estimators for the binary choice model fully comparable to their parametric counterparts. Model specification tests like the Hausman-type test can then be based on the slope parameters and the intercept term, giving rise to the possibility that the resulting tests could be more powerful than similar ones based on the slope parameters alone.

Estimation of location parameters is important in many empirical applications. Estimation of the intercept term in the binary choice model is an important step in obtaining a consistent estimator for the intercept term in the censored sample selection model (see, e.g., Chen, 1996; Chen and Lee, 1996), which itself has important economic implications in assessing the impact on earnings of job training and union status, among other applications. Lewbel (1997) noted the importance of estimation of the location parameter in the binary choice model in determining the distribution of reservation prices in consumer demand analysis.

A useful prediction rule[2] for the response variable d is (see, e.g., Greene, 1993) as follows:

$$\hat{d} = 1 \quad \text{if } x\hat{\beta} - \hat{\alpha} > 0,$$

where $(\hat{\alpha}, \hat{\beta})$ is some estimate for (α_0, β_0). As a result, estimation of α_0 is needed for this purpose.[3] Such a prediction rule is also used in constructing some goodness-of-fit tests for discrete choice models (Amemiya, 1981; Maddala, 1983).

In the next section we shall propose two root-n consistent estimators for the intercept α_0. Cosslett (1987) derived the semiparametric efficiency bound for the parameters in the binary choice model under the assumption that the error term is

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

independent of the regressors and possesses a constant median. It turns out the efficiency bound for the location parameter is infinite, which implies that regular root-n consistent estimators for α_0 do not exist (Chamberlain, 1986). Cosslett also showed that when the error term also has a symmetric distribution around a known constant, then the information bound for the location parameter will become finite, giving rise to the possibility of root-n estimation of the intercept term α_0; in fact, given an initial root-n consistent estimator for (α_0, β_0), Chen (1997) proposed a one-step semiparametric scoring estimator that achieves Cosslett's efficiency bound. The estimators proposed in this paper can be used as initial estimators for Chen's efficient estimator. Finally, we also briefly discuss a procedure for estimating both the slope and intercept terms simultaneously.

2. THE ESTIMATORS

Recall the binary choice model

$$d = 1\{x\beta_0 - \alpha_0 - v > 0\}, \tag{2}$$

where for simplicity, β_0 is taken to be the normalized parameter vector (for detailed discussions, see Cosslett, 1983; Manski, 1985; Ichimura, 1993). Let $z = x\beta_0$. Let $F(t, \beta) = E(d \mid x\beta = t)$, for any possible value of β, and $F(t) = F(t, \beta_0)$.[4] Here we assume that the error term v satisfies a monotonic index and symmetry condition. More specifically, the conditional density of v, $f(v \mid x)$, depends on x only through $|\alpha_0 - z|$, the magnitude of the linear index $x\beta_0 - \alpha_0$, and symmetric around 0, i.e., $f(v \mid x) = f(v \mid |\alpha_0 - z|)$ and $f(-v \mid x) = f(v \mid x)$; in addition, we assume that $F(t)$ is strictly increasing in t for $t \in R$. Therefore, we have

$$F(t) + F(2\alpha_0 - t) = \int_{-\infty}^{(t-\alpha_0)} f(s \mid |\alpha_0 - t|)\, ds + \int_{-\infty}^{(\alpha_0 - t)} f(s \mid |\alpha_0 - t|)\, ds = 1 \tag{3}$$

for any t. Thus we could solve for α_0 from equation (3) if F and β_0 were unknown. However, for a random sample of size n and any possible value of β, $F(t, \beta)$ can be estimated by the local linear regression[5] (see Fan and Gijbels, 1995) that minimizes

$$\sum_{i=1}^{n} (d_i - \eta_0 - \eta_1(x, \beta - t))^2 K\left(\frac{x_i\beta - t}{h_1}\right) \tag{4}$$

with respect to η_0 and η_1, where K is a symmetric density function and h_1 is a positive bandwidth. Let $F_n(t, \beta)$ be the estimate for the local intercept term in the local linear regression (4); $F(t)$ can be estimated by $F_n(t, \hat{\beta})$, where $\hat{\beta}$ is a consistent estimator for β_0.

Now we consider estimating α_0 given some first-step estimator for β_0. Two estimators are proposed here based on equation (3). First, note that if there exist observations i, j such that

$$F_i + F_j = 1,$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

where $F_i = F(z_i)$ and $F_j = F(z_j)$, then, by the strict monotonicity, we have

$$z_i + z_j = 2\alpha_0,$$

hence,

$$\alpha_0 = \frac{\sum_{i=j}(z_i + z_j)1(F_i + F_j = 1)}{2\sum_{i=j}1(F_i + F_j = 1)}. \tag{5}$$

This equality cannot be implemented directly even if $F(\cdot)$ were known because there may not exist pairs of observations satisfying $F_i + F_j = 1$; furthermore, z_i's are not observable because they depend on unknown parameters β_0. This approach can still be implemented if F is smooth enough and there exists a consistent estimator $\hat{\beta}$ for β_0. In this case, we have

$$\hat{z}_i + \hat{z}_j \approx 2\alpha_0,$$

where $\hat{z}_i = x_i\hat{\beta}$ and $\hat{z}_j = x_j\hat{\beta}$, if

$$F(\hat{z}_i) + F(\hat{z}_j) \approx 1.$$

An infeasible estimator for α_0 can then be defined as

$$\tilde{\alpha}_1 = \frac{\sum_{i=j}(\hat{z}_i + \hat{z}_j)k\left(\frac{F(\hat{z}_i) + F(\hat{z}_j) - 1}{h_2}\right)w(\hat{z}_i)w(\hat{z}_j)}{2\sum_{i=j}k\left(\frac{F(\hat{z}_i) + F(\hat{z}_j) - 1}{h_2}\right)w(\hat{z}_i)w(\hat{z}_j)},$$

where $k(\cdot)$ is a univariate kernel function declining to zero as its argument increases in magnitude and h_2 is a bandwidth sequence converging to zero; thus, the kernel weight $k((F(\hat{z}_i) + F(\hat{z}_j) - 1)/h_2)$ gives declining weight to pairs with large values of $|F(\hat{z}_i) + F(\hat{z}_j) - 1|$, and $w(\cdot)$ is a trimming function that is chosen to equal to zero when $F(t)$ cannot be accurately estimated. Our first estimator, $\hat{\alpha}_1$, is defined by replacing $F(\cdot)$ with $F_n(\cdot, \hat{\beta})$ in $\tilde{\alpha}_1$,

$$\hat{\alpha}_1 = \frac{A_n(\hat{\beta}, 0)}{B_n(\hat{\beta})},$$

where

$$A_n(\beta, \alpha) = \frac{1}{n(n-1)h_2}\sum_{i=j}(x_i\beta + x_j\beta - 2\alpha) \times k\left(\frac{F_n(x_i\beta, \beta) + F_n(x_j\beta, \beta) - 1}{h_2}\right)w(x_i\beta)w(x_j\beta)$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

and

$$B_n(\beta) = \frac{2}{n(n-1)h_2} \sum_{i=j} k\left(\frac{F_n(x_i\beta,\beta) + F_n(x_j\beta,\beta) - 1}{h_2} \right) w(x_i\beta) w(x_j\beta).$$

We now turn to the second estimator. For a fixed t, α_0 solves the equation

$$F(t) + F(2\alpha - t) = 1. \tag{6}$$

Then, for each $\hat{z}_i, i = 1,2 \ldots n$, define $\hat{\alpha}_{2i}$ as a solution to

$$F_n(\hat{z}_i, \hat{\beta}) + F_n(2\alpha - \hat{z}_i, \hat{\beta}) = 1,$$

a sample analog to equation (6) when $t = \hat{z}_i$, i.e., $\hat{\alpha}_{2i}$ satisfies[6]

$$F_n(\hat{z}_i, \hat{\beta}) + F_n(2\hat{\alpha}_{2i} - \hat{z}_i, \hat{\beta}) = 1.$$

Consequently, the second estimator for α_0 is defined as

$$\hat{\alpha}_2 = \frac{\frac{1}{n} \sum_{i=1}^{n} \hat{\alpha}_{2i} w(\hat{z}_i)}{\frac{1}{n} \sum_{i=1}^{n} w(\hat{z}_i)},$$

where $w(\hat{z}_i)$ is a trimming factor.

Now we consider large sample properties of the estimators. We make the following assumptions.

Assumption 1. The vectors (x_i, d_i) are independent and identically distributed across i, with finite fourth-order moments for each component of x_i.

Assumption 2. The conditional density of v, $f(v|x)$, depends on x only through $|\alpha_0 - z|$, and symmetric around 0, i.e., $f(v|x) = f(v||\alpha_0 - z|)$ and $f(-v|x) = f(v|x)$. In addition, $F(t)$ is strictly increasing in t.

Assumption 3. The preliminary estimator $\hat{\beta}$ of β_0 is root-n consistent and has the following asymptotic linear representation:

$$\hat{\beta} = \beta_0 + \frac{1}{n} \sum \psi_i + o_p(n^{-1/2})$$

for some $\psi_i = \psi(d_i, x_i)$ such that $E\psi(d_i, x_i) = 0$ and $E\|\psi(d_i, x_i)\|^2 < \infty$.

Assumption 4. The kernel function $K(\cdot)$ is a symmetric density with a bounded support. It is twice continuously differentiable.

Assumption 5. The kernel function $k(\cdot)$ has a bounded support. It is s times continuously differentiable, $\int k(u)\, du = 1$ and $\int u^l k(u)\, du = 0$ for $l = 1,2,\ldots s$.

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

Assumption 6.

(i) The support of the distribution of x is not contained in any proper linear subspace of R^q.
(ii) For almost every $\tilde{x} = (x_2, \ldots, x_q)'$, the distribution of x_1 conditional on $\tilde{x}$ has everywhere positive density with respect to Lebesgue measure. As before, $\beta_{01} = 1$ by scale normalization.

Let $p(t, \beta)$ denote the density of $x\beta$ evaluated at t. The term $\lambda(\cdot)$ denotes the inverse of $F(\cdot)$ and $g(\cdot)$ denotes the density of $F(z)$. We choose bounded sets $S = [s_1, s_2]$ and $S_{\varepsilon_0} = [s_1 + \varepsilon_0, s_2 - \varepsilon_0]$ for a sufficiently small positive ε_0.

Assumption 7.

(i) $0 \leq w(t) \leq 1, w(t) = 0$ for $t \notin S$, and $w(t) = 1$ for $t \in S_b = (b1, b2)$. For any $z \in S_b$, then $z \in S_{\varepsilon_0}$ and $2\alpha_0 - z \in S_{\varepsilon_0}$.
(ii) $p(t, \beta_0)$ and $f(t - \alpha_0 | x)$ are uniformly bounded away from zero for $t \in S$.

Assumption 8.

(i) $F(t, \beta)p(t, \beta)$ and $p(t, \beta)$ are uniformly bounded and twice continuously differentiable in t and β, and the trimming function $w(\cdot)$ is continuously differentiable.
(ii) $\lambda(F)$, $g(F)$, and $w(t)$ are s times continuously differentiable, and the partial derivatives are continuous and uniformly bounded for $t \in S$ and $F \in [F(s_1), F(s_2)]$.

Assumption 9. The bandwidth sequences satisfy

(i) $nh_2^s \to 0$, $nh_1^8 h_2^{-6} \to 0$, $h_2^{-2} h_1 \to 0$, and $nh_2^4 h_1^3 \ln^2 n \to \infty$, as $n \to \infty$.
(ii) $nh_1^6 \to 0$ and $nh_1^4 \ln^2 n \to \infty$.

Assumptions 1 and 2 describe the model and the data. Note that certain unknown heteroskedasticity is allowed. Several existing estimators for the slope parameters for β_0 (among others, Ahn et al., 1996; Han, 1987; Horowitz and Härdle, 1996; Ichimura, 1993; Klein and Spady, 1993; Powell et al., 1989; Sherman, 1993) satisfy Assumption 3. Assumptions 4 and 5 describe the kernel functions; note that the second-step kernel function k in defining the first estimator $\hat{\alpha}_1$ is a "higher order" bias reduction kernel, which, together with the rate of convergence condition for the bandwidth parameters in Assumption 9, ensures that $\hat{\alpha}_1$ is asymptotically unbiased. Assumption 6 is used for identifying the slope parameters by Manski (1985) under a conditional quantile restriction, which is certainly sufficient under our symmetry condition. Ahn et al. (1996) provided a similar set of conditions also including the presence of a continuously distributed component x_1 for identifying the slope parameters under a monotone single-index restriction; but their conditions do not require the support of x_1 to be the real line. The identification conditions of Ahn et al. (1996) for identifying the slope parameters also apply here because we impose a stronger restriction in our setting. Assumption 7 essentially is an identification condition for the intercept term; it requires a portion of the population with participation probability below 0.5 and

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

another portion with participation probability above 0.5 because the equality $F(z) + F(2\alpha_0 - z) = 1$ holds for $z \in S_b$. Mainly this assumption is made to avoid the situation that the supports of v and $x\beta_0$ are too small to identify the location parameter. A sufficient condition is that the supports of both $x\beta_0$ and the error term v are the whole line. Assumption 8 is some boundedness and smoothness conditions, which can be justified by more primitive conditions on the distributions of the variables in the model (for some discussions on similar conditions, see Lee, 1994, Sherman, 1994). Assumption 9 describes the rate of convergence for the chosen bandwidths; notice that the kernel bandwidth h_2 goes to zero more slowly than h_1.

THEOREM 1. *If Assumptions* 1–9 *hold, then* $\hat{\alpha}_1$ *is consistent and asymptotically normal*,

$$\sqrt{n}(\hat{\alpha}_1 - \alpha_0) \xrightarrow{d} N(0, \Sigma_1),$$

where $\Sigma_1 = B^{-2}[D_{1\xi\xi} + \Omega_1 D_{1\psi\xi} + D_{1\xi\psi}\Omega_1' + \Omega_1 D_{\psi\psi}\Omega_1']$ *for* $D_{1\xi\xi} = E\xi_{1i}\xi_{1i}'$, $D_{\psi\psi} = E\psi_i\psi_i'$, $D_{i\xi\psi} = E\xi_{1i}\psi_i' = D_{1\psi\xi}'$, $B = 2Eg(1 - F_i)w(z_i)w(2\alpha_0 - z_j)$, *and* $\Omega_1 = A_1 + A_2$ *with*

$$A_1 = -E\left(\frac{\partial F(x_i\beta_0, \beta_0)}{\partial \beta'} + E\left(\frac{\partial F(x\beta_0, \beta_0)}{\partial \beta'}\middle| z = 2\alpha_0 - z_i\right)\right)$$
$$\times g(1 - F_i)\lambda'(1 - F_i)w(z_i)w(2\alpha_0 - z_i),$$

$$A_2 = Eg(1 - F_i)w(z_i)w(2\alpha_0 - z_j)(x_i + E(x|z = 2\alpha_0 - z_i)),$$

and

$$\xi_{1i} = -2(d_i - F_i)\lambda'(1 - F_i)w(z_i)w(2\alpha_0 - z_i)g(1 - F_i).$$

Proof. See the Appendix.

THEOREM 2. *If Assumptions* 1–4, 6–8, *and* 9(ii) *hold, then* $\hat{\alpha}_2$ *is consistent and asymptotically normal*,

$$\sqrt{n}(\hat{\alpha}_2 - \alpha_0) \xrightarrow{d} N(0, \Sigma_2),$$

where

$$\Sigma_2 = W^{-2}[D_{2\xi\xi} + \Omega_2 D_{2\psi\xi} + D_{2\xi\psi}\Omega_2' + \Omega_2 D_{\psi\psi}\Omega_2']$$

with

$$\Omega_2 = -E\frac{1}{2f(\alpha_0 - z_i|x_i)}\left(\frac{\partial F(x_i\beta_0, \beta_0)}{\partial \beta'} + \frac{\partial F(2\alpha_0 - x_i\beta_0, \beta_0)}{\partial \beta'}\right)w(z_i)$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

and

$$\xi_{2i} = -\frac{1}{2f(\alpha_0 - z_i|x_i)p(z_i)}(d_i - F_i)[w(z_i)p(z_i) + w(2\alpha_0 - z_i)p(2\alpha_0 - z_i)]$$

for $D_{2\xi\xi} = E\xi_{2i}\xi_{2i}'$, $D_{\psi\psi} = E\psi_i\psi_i'$, *and* $D_{2\xi\psi} = E\xi_{2i}\psi_i' = D_{2\psi\xi}'$, $W = Ew(z_i)$.

Proof. See the Appendix.

Consistent estimators for the asymptotic covariance matrices are needed to conduct large sample inferences. Consider, for example, estimating Σ_1. A consistent estimator of Σ_1 can be easily obtained if we can find consistent estimators of $A_1, A_2, B, D_{1\xi\xi}, D_{1\xi\psi}$, and $D_{\psi\psi}$. Estimation of A_1, A_2, and B is straightforward; following the proof of Theorem 1, we can show that $A_{n1}(\hat{\beta}, \hat{\alpha}), A_{n2}(\hat{\beta})$, and $B_n(\hat{\beta})$ are consistent for A_1, A_2, and B, respectively, where

$$A_{n1}(\beta,\alpha) = \frac{1}{n(n-1)}\sum_{i=j}(x_i\beta + x_j\beta - 2\alpha)\left(\frac{\partial F_n(x_i\beta,\beta)}{\partial\beta} + \frac{\partial F_n(x_j\beta,\beta)}{\partial\beta}\right) \times \frac{1}{h_2^2}k'\left(\frac{F_n(x_i\beta,\beta) + F_n(x_j\beta,\beta) - 1}{h_2}\right)w(x_i\beta)w(x_j\beta)$$

and

$$A_{n2}(\beta) = \frac{1}{n(n-1)}\sum_{i=j}(x_i + x_j) \times \frac{1}{h_2}k\left(\frac{F_n(x_i\beta,\beta) + F_n(x_j\beta,\beta) - 1}{h_2}\right)w(x_i\beta)w(x_j\beta).$$

Consistent estimation of the matrices $D_{1\xi\xi}, D_{1\xi\psi}$, and $D_{\psi\psi}$ requires suitable estimates of the components $\{\xi_{1i}\}$ and $\{\psi_i\}$ in the asymptotically linear representation in (A.3). Specifically, it is useful to assume that $\{\hat{\xi}_{1ni}\}$ and $\{\psi_{ni}\}$ exist such that

$$\frac{1}{n}\sum|\xi_{1i} - \hat{\xi}_{1ni}|^2 = o_p(1) \tag{7}$$

and

$$\frac{1}{n}\sum|\psi_{1i} - \hat{\psi}_{1ni}|^2 = o_p(1).$$

One example of $\{\hat{\psi}_{ni}\}$ is suggested in Powell et al. (1989). An analogous sequence of $\{\hat{\xi}_{1ni}\}$ can be constructed as

$$\hat{\xi}_{1ni} = 2(d_i - F_n(\hat{z}_i,\hat{\beta}))\frac{1}{(n-1)h_2^2} \times \sum_{j=i}(\hat{z}_i + \hat{z}_j - 2\hat{\alpha})k'\left(\frac{F_n(\hat{z}_i,\hat{\beta}) + F_n(\hat{z}_j,\hat{\beta}) - 1}{h_2}\right)w(\hat{z}_i)w(\hat{z}_j).$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

By following the proof of Theorem 1, it is straightforward to show that (7) is satisfied. With the sequences of $\{\hat{\xi}_{1ni}\}$ and $\{\psi_{ni}\}$ given, the remaining components of Σ_1, namely, $D_{1\xi\xi}$, $D_{1\xi\psi}$, and $D_{\psi\psi}$, can be consistently estimated as in Powell (1989). Define

$$\hat{D}_{1\xi\psi} = \frac{1}{n}\sum_{i=1}^{n} \hat{\xi}_{1ni}\hat{\psi}'_{ni}$$

with analogous definition of $\hat{D}_{1\xi\xi}$, $\hat{D}_{1\psi\xi}$, and $\hat{D}_{\psi\psi}$, which can be shown to be consistent for $D_{1\xi\xi}$, $D_{1\xi\psi}$, and $D_{\psi\psi}$, respectively.

Now we consider estimating Σ_2. First, we can show that

$$\hat{\Omega}_2 = -\frac{1}{n}\sum_{i=1}^{n}\left[2\,\frac{\partial F_n(\hat{z}_i,\hat{\beta})}{\partial t}\right]^{-1}\left(\frac{\partial F_n(x_i\hat{\beta},\hat{\beta})}{\partial \beta'} + \frac{\partial F_n(2\hat{\alpha}-x_i\hat{\beta},\hat{\beta})}{\partial \beta'}\right)w(\hat{z}_i)$$

and $W_n = 1/n\sum_{i=1}^{n} w(\hat{z}_i)$ are consistent for Ω_2 and W. Next, define

$$\hat{\xi}_{2ni} = -\left[2\,\frac{\partial F_n(\hat{z}_i,\hat{\beta})}{\partial t}\,p_n(\hat{z}_i)\right]^{-1}(d_i - F_n(\hat{z}_i,\hat{\beta}))$$
$$\times\,[w(\hat{z}_i)p_n(\hat{z}_i) + w(2\hat{\alpha}-\hat{z}_i)p_n(2\hat{\alpha}-\hat{z}_i)]$$

with $p_n(\hat{z}_i)$ a consistent estimate for the density of z_i, for example, by a kernel nonparametric estimate. By following the approach in estimating Σ_1, it is not difficult to construct a similar consistent estimator for Σ_2.

For the two estimators proposed previously, $\hat{\alpha}_1$ has a closed form given a first-step estimator $\hat{\beta}$, a combination of which with the noniterative procedures for estimating β_0 by Ahn et al. (1996), Horowitz and Hardle (1996), and Powell et al. (1989) will be very attractive computationally because it avoids nonlinear optimization or solving nonlinear equations. The second estimator, $\hat{\alpha}_2$, is computationally more intensive, involving a one-dimensional grid search. However, $\hat{\alpha}_2$ involves only one nonparametric approximation compared with two approximations in constructing $\hat{\alpha}_1$; thus $\hat{\alpha}_2$ will be useful in cases where $\hat{\alpha}_1$ is found to be sensitive to the choice of the second bandwidth h_2.

Our estimators are based on equation (3), which suggests that the performance of our estimators will be adversely affected when $P(z \in S_b)$ is small, namely, when most individuals have participation probabilities above (or below) 0.5. The estimator proposed by Lewbel requires that the support of the error term v be contained in that of z; therefore his estimator is likely to have large bias when the participation probability is bounded away from zero or one.

3. A MONTE CARLO STUDY

In this section we present a small Monte Carlo study to illustrate the usefulness of the proposed estimator. The data are generated according to the following model:

$$d = 1\{x + \alpha_0 + v > 0\},$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

with $\alpha_0 = 1$. Notice that for simplicity, we are treating the case that β_0 is assumed to be known. Four different designs are constructed by varying the distributions of the error term v and the regressor x.

Here we consider the finite-sample performance of our estimators, $\hat{\alpha}_1$ and $\hat{\alpha}_2$, along with Lewbel's estimator (1997), $\hat{\alpha}_l$. The results from 300 replications from 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). We adopt the quadratic kernel in estimating the function F

$$K(u) = 0.9375I(|u| < 1)(1 - u^2)^2$$

as in Lewbel (1997). The same kernel[7] is used as the second-step kernel $k(\cdot)$ in constructing $\hat{\alpha}_2$. The bandwidth selection for $\hat{\alpha}_l$ follows the rule suggested in Lewbel (1997),[8] which yields $h_l = 0.630$ for $n = 100$ and $h_l = 0.406$ for $n = 300$. Like Lewbel, we also report the results with one-half and double h_l for α_l. It is difficult to choose bandwidths with rigorous theoretical justifications. The bandwidth for estimating F is chosen by the generalized cross-validation. Also, the estimator $\hat{\alpha}_2$ is not very sensitive to the choice of the bandwidth[9] h_2; we report the results with $h_2 = 0.1$, 0.05, and 0.025. The sample is discarded and redrawn if $\min_i F_n(\hat{z}_i, \hat{\beta}) > 0.5$ or $\max_i F_n(\hat{z}_i, \hat{\beta}) < 0.5$ because the discarded sample is not informative for the intercept term under our symmetry condition. For the trimming function $w(\cdot)$, its smoothness is not important in practice. Let $F_{n*} = \max\{0.025, \min_i F_n(\hat{z}_i, \hat{\beta})\}$, $F^{n*} = \min\{0.975, \max_i F_n(\hat{z}_i, \hat{\beta})\}$, and $F_{n\alpha} = \min\{0.5 - F_{n*}, F^{n*} - 0.5\}$; then we set $w(\cdot) = 1$ if $F_n(\hat{z}_i, \hat{\beta}) \in (0.5 - F_{n\alpha}, 0.5 + F_{n\alpha})$, and 0 otherwise; namely, we select the observations falling in the largest interval symmetric around 0.5 within $(0.025, 0.975)$ in terms of $F_n(\hat{z}_i, \hat{\beta})$.[10]

Table 1 reports the simulation results for the case in which both x and v are drawn from $N(0,1)$ independent of each other (homoskedastic design I). In this case, all estimators perform well; $\hat{\alpha}_l$ has the smallest RMSE, whereas our estimators, $\hat{\alpha}_1$ and $\hat{\alpha}_2$, have smaller biases. As expected, performances improve when the sample size increases from 100 to 300. Note that in this design, only about 15% of the observations have participation probability below 0.5 whereas that of the remaining 85% of the observations exceeds 0.5. The performance of our estimators is adversely affected by this unbalanced design.

In Table 2 we report the results when x is drawn from $N(-1,1)$ and v from $N(0,1)$ with x and v independent of each other (homoskedastic design II). In this design, about 50% of the observations have participation probability below 0.5, and the remaining half have participation probability above 0.5. Consequently, the relative performance of our estimators improves significantly.

In Table 3 we consider a design with a heteroskedastic error; the error term $v = v_1 \cdot (|1 + x|^{0.3} + 0.25)$ where v_1 and x are drawn from independent $N(0,1)$ (heteroskedastic design). Note that the homoskedasticity required for $\hat{\alpha}_l$ is violated whereas the symmetry and monotonic index assumption for our estimators are

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

TABLE 1. Homoskedastic design I

n	Estimators	Mean	SD	RMSE
100	$\hat{\alpha}_1, h_2 = 0.100$	0.973	0.230	0.231
	$\hat{\alpha}_1, h_2 = 0.050$	0.972	0.251	0.252
	$\hat{\alpha}_1, h_2 = 0.025$	0.967	0.268	0.269
	$\hat{\alpha}_2$	0.959	0.223	0.226
	$\hat{\alpha}_l, h_l = 0.315$	0.911	0.169	0.191
	$\hat{\alpha}_l, h_l = 0.630$	0.937	0.175	0.186
	$\hat{\alpha}_l, h_l = 1.260$	0.958	0.176	0.180
300	$\hat{\alpha}_1, h_2 = 0.100$	0.986	0.132	0.132
	$\hat{\alpha}_1 h_2 = 0.050$	0.991	0.134	0.134
	$\hat{\alpha}_1, h_2 = 0.025$	0.992	0.135	0.135
	$\hat{\alpha}_2$	0.972	0.135	0.138
	$\hat{\alpha}_l, h_l = 0.203$	0.956	0.118	0.126
	$\hat{\alpha}_l, h_l = 0.406$	0.969	0.123	0.127
	$\hat{\alpha}_l, h_l = 0.812$	0.974	0.119	0.122

still satisfied. Consequently, $\hat{\alpha}_l$ have large biases when our estimators still perform satisfactorily.

The results with a different homoskedastic design are reported in Table 4 (homoskedastic design III); in this case x is drawn from $N(0,1)$ conditional on $|x| < 1.6$,

TABLE 2. Homoskedastic design II

n	Estimators	Mean	SD	RMSE
100	$\hat{\alpha}_1, h_2 = 0.100$	0.996	0.144	0.143
	$\hat{\alpha}_1, h_2 = 0.050$	0.997	0.145	0.145
	$\hat{\alpha}_1, h_2 = 0.025$	0.997	0.145	0.145
	$\hat{\alpha}_2$	0.997	0.143	0.143
	$\hat{\alpha}_l, h_l = 0.315$	0.975	0.165	0.166
	$\hat{\alpha}_l, h_l = 0.630$	0.999	0.169	0.169
	$\hat{\alpha}_l, h_l = 1.260$	1.065	0.177	0.188
300	$\hat{\alpha}_1, h_2 = 0.100$	1.001	0.079	0.079
	$\hat{\alpha}_1, h_2 = 0.050$	1.001	0.080	0.080
	$\hat{\alpha}_1, h_2 = 0.025$	1.001	0.080	0.080
	$\hat{\alpha}_2$	1.002	0.080	0.079
	$\hat{\alpha}_l, h_l = 0.203$	0.996	0.098	0.098
	$\hat{\alpha}_l, h_l = 0.406$	1.008	0.098	0.098
	$\hat{\alpha}_l, h_l = 0.812$	1.034	0.098	0.103

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

TABLE 3. Heteroskedastic design

n	Estimators	Mean	SD	RMSE
100	$\hat{\alpha}_1, h_2 = 0.100$	0.956	0.285	0.288
	$\hat{\alpha}_1, h_2 = 0.050$	0.957	0.297	0.300
	$\hat{\alpha}_1, h_2 = 0.025$	0.949	0.316	0.319
	$\hat{\alpha}_2$	0.936	0.281	0.288
	$\hat{\alpha}_l, h_l = 0.315$	0.799	0.203	0.286
	$\hat{\alpha}_l, h_l = 0.630$	0.824	0.221	0.283
	$\hat{\alpha}_l, h_l = 1.260$	0.841	0.222	0.273
300	$\hat{\alpha}_1, h_2 = 0.100$	0.983	0.168	0.168
	$\hat{\alpha}_1, h_2 = 0.050$	0.990	0.171	0.171
	$\hat{\alpha}_1, h_2 = 0.025$	0.992	0.171	0.171
	$\hat{\alpha}_2$	0.964	0.180	0.183
	$\hat{\alpha}_l, h_l = 0.203$	0.858	0.143	0.201
	$\hat{\alpha}_l, h_l = 0.406$	0.878	0.154	0.196
	$\hat{\alpha}_l, h_l = 0.812$	0.882	0.153	0.193

with v drawn from $N(0,1)$ independent of x. Because the consistency of $\hat{\alpha}_l$ requires that the linear index have heavier tails than that of the error term, which does not hold here, Lewbel's procedure incurs substantial biases, with small variances though.[11] Our estimators still perform well in this case.

TABLE 4. Homoskedastic design III

n	Estimators	Mean	SD	RMSE
100	$\hat{\alpha}_1, h_2 = 0.100$	0.932	0.222	0.231
	$\hat{\alpha}_1, h_2 = 0.050$	0.929	0.248	0.258
	$\hat{\alpha}_1, h_2 = 0.025$	0.917	0.273	0.285
	$\hat{\alpha}_2$	0.927	0.228	0.239
	$\hat{\alpha}_l, h_l = 0.315$	0.792	0.122	0.241
	$\hat{\alpha}_l, h_l = 0.630$	0.820	0.126	0.220
	$\hat{\alpha}_l, h_l = 1.260$	0.866	0.136	0.191
300	$\hat{\alpha}_1, h_2 = 0.100$	1.006	0.175	0.175
	$\hat{\alpha}_1, h_2 = 0.050$	1.010	0.177	0.177
	$\hat{\alpha}_1, h_2 = 0.025$	1.011	0.177	0.177
	$\hat{\alpha}_2$	1.001	0.185	0.185
	$\hat{\alpha}_l, h_l = 0.203$	0.828	0.074	0.187
	$\hat{\alpha}_l, h_l = 0.406$	0.842	0.076	0.176
	$\hat{\alpha}_l, h_l = 0.812$	0.863	0.079	0.158

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

4. EXTENSION AND CONCLUSION

In this paper we considered semiparametric estimation of a location parameter in the binary choice model under a symmetry and monotonic index restriction. Two estimators are proposed here, both of which are consistent and asymptotically normal. A small Monte Carlo study illustrates the usefulness of the estimators.

Both of the estimators are two-step estimators based on equation (3), given a first-step consistent estimator $\hat{\beta}$ for β_0; in fact equation (3) also suggests that (α_0, β_0) can be jointly estimated. Let $x\beta_0 = x_1 - x_2\delta_0$, i.e., the coefficient of x_1 has been normalized to one. For observations (i, j) if

$$F(z_i) + F(z_j) = 1$$

then

$$x_{1i} + x_{1j} = 2\alpha_0 + (x_{2i} + x_{2j})\delta_0.$$

Following the idea behind the first estimator $\hat{\alpha}_1$, we can propose the following estimator for $\theta_0 = (\alpha_0, \delta_0')'$:

$$\hat{\theta} = S_{n2}^{-1} S_{n1},$$

where

$$S_{n2} = \frac{1}{n(n-1)} \sum_{i=j} (x_i^* + x_j^*)'(x_i^* + x_j^*) k\left(\frac{F_n(x_i) + F_n(x_j) - 1}{h_2}\right) w^*(x_i) w^*(x_j)$$

and

$$S_{n2} = \frac{1}{n(n-1)} \sum_{i=j} (x_i^* + x_j^*)'(x_{1i} + x_{1j}) k\left(\frac{F_n(x_i) + F_n(x_j) - 1}{h_2}\right) w^*(x_i) w^*(x_j),$$

where $F_n(\cdot)$ denotes a nonparametric estimate for $F(\cdot)$, $x_i^* = -1, x_i)$, for $i = 1, 2, \ldots, n$, and $w^*(\cdot)$ is an exogenous trimming function. One undesirable feature associated with this estimator is that high dimensional nonparametric estimation is involved in constructing $F_n(\cdot)$. Note that $\hat{\theta}$ reduces to $\hat{\alpha}_1$ when β_0 is known. Analysis of $\hat{\theta}$ will be similar to that of $\hat{\alpha}_1$ and is not considered here.

NOTES

1. Actually, Lewbel proposed a procedure estimating the intercept α_0 when β_0 is known, whereas he conjectured that his results carry over to the situation where β_0 must be estimated.

2. Note that $x\hat{\beta} - \hat{\alpha} > 0$ is equivalent to $F(x\hat{\beta} - \hat{\alpha} + \alpha_0) > 0.5$ where $F(\cdot)$ is the participation probability function to be defined later. An alternative prediction rule without estimating α_0 is to set $\hat{d} = 1$ if $\hat{F}(x\hat{\beta}) > 0.5$ where $\hat{F}(\cdot)$ is a nonparametric estimator for $F(\cdot)$; however, with a root-n consistent estimator $\hat{\alpha}$, $F(x\hat{\beta} - \hat{\alpha} + \alpha_0)$ converges faster than $\hat{F}(x\hat{\beta})$ for a given x. Furthermore, the prediction rule based on estimated linear indices is easier to implement.

3. In fact, Manski's estimator (Manski, 1985) is devised by maximizing the number of correct predictions based on this prediction rule.

4. When v is independent of x, then $F(t)$ is the accumulative distribution function of $\alpha_0 + v$.

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

5. Nonparametric estimation based on kernel fitting or series approximation also works here. However, in our context, in finite samples local linear fitting is likely to better capture the mononicity of the regression function than the local constant fitting based on ordinary kernel estimation when positive kernels are used.

6. It will be shown later that $\hat{\alpha}'_{2i}s$ are well defined in large samples.

7. A fourth-order kernel $K_4(v) = 1.1\exp(-v^2/2) - 0.1\exp(-v^2/22)/\sqrt{11}$ is also experimented with; the estimates based on the higher order kernel have similar RMSE's but are slightly more sensitive to the choice of the bandwidth.

8. Which is based on Powell and Stoker (1996).

9. Note that probability functions range from 0 to 1; thus rescaling is not needed.

10. A smoothed version of this trimming scheme can be easily incorporated in the definition of our two estimators, only at the expense of more complicated notations.

11. The small variances associated with Lewbel's estimator here are not unexpected because the estimated density function of x enters as the denominator in his approach whereas in this design x has larger densities in its support than in the previous designs. Note that when x has a finite support smaller than that of v, Lewbel's estimator essentially corresponds to a different parameter than the intercept term.

REFERENCES

Ahn, H., H. Ichimura, & J.L. Powell (1996) Simple Estimators for Monotone Index Models. Manuscript, Department of Economics, University of California at Berkeley.

Amemiya, T. (1981) Qualitative response models: A survey. *Journal of Economic Literature* 18, 1483–1536.

Chamberlain, G. (1986) Asymptotic efficiency in semiparametric models with censoring. *Journal of Econometrics* 32, 189–218.

Chen, S. (1996) Distribution-free estimation of the random-coefficient dummy endogenous variable model. *Journal of Econometrics*, forthcoming.

Chen, S. (1997) Efficient Estimation of a Binary Choice Model. Manuscript.

Chen, S. & L.F. Lee (1996) Semiparametric Estimation of Conditional Expectation Sample Selection Models. Manuscript, The Hong Kong University of Science and Technology.

Cosslett, S.R. (1983) Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica* 51, 765–782.

Cosslett, S.R. (1987) Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. *Econometrica* 55, 559–587.

Fan, J. & I. Gijbels (1995) *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Greene, W. (1993) *Econometric Analysis.* New York: Macmillan.

Han, A.K. (1987) Nonparametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics* 35, 303–316.

Horowitz, J.L. (1992) A smooth maximum score estimator for the binary response model. *Econometrica* 60, 505–531.

Horowitz, J.L. & W. Härdle (1996) Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association* 91, 1632–1640.

Ichimura, H. (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* 58, 71–120.

Klein, R.W. & R.S. Spady (1993) An efficient semiparametric estimator of the binary response model. *Econometrica* 61, 387–421.

Lee, L.F. (1994) Semiparametric instrumental variable estimation of simultaneous equation sample selection models. *Journal of Econometrics* 63, 341–388.

Lewbel, A. (1997) Semiparametric estimation of location and other discrete choice moments. *Econometric Theory* 13, 32–51.

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

Maddala, G.S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.
Manski, C.F. (1985) Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics* 27, 313–333.
Newey, K.N. (1994) Kernel estimation of partial means and a general variance estimator. *Econometric Theory* 10, 233–253.
Powell, J.L. (1989) Semiparametric Estimation of Bivariate Latent Variable Models. Manuscript.
Powell, J.L. & T.M. Stoker (1996) Optimal bandwidth choice for density weighted averages. *Journal of Econometrics* 75, 291–316.
Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of weighted average derivatives. *Econometrica* 57, 1403–1430.
Sherman, R.P. (1993) The limiting distribution of the maximum rank correlation estimator. *Econometrica* 61, 123–138.
Sherman, R.P. (1994) U-processes in the analysis of a generalized semiparametric regression estimator. *Econometric Theory* 11, 372–395.

APPENDIX

The following two lemmas present some results for rates of convergence and asymptotic linear representations for local linear estimates. These results are useful to prove the main theorems.

LEMMA 1. *Under Assumptions* 1–4, 6–8, *and* 9(ii), *for* $t \in S$, *and* β *belonging to a neighborhood of* β_0, *we have*

$$\sup_{t,\beta} |F_n(t,\beta) - F(t,\beta)| = O_p(h_1^2 + (nh_1)^{-1/2} \ln n),$$

$$\sup_{t,\beta} \left| \frac{\partial F_n(t,\beta)}{\partial \beta} - \frac{\partial F(t,\beta)}{\partial \beta} \right| = O_p(h_1 + (nh_1^3)^{-1/2} \ln n),$$

$$\sup_{t,\beta} \left| \frac{\partial F_n(t,\beta)}{\partial t} - \frac{\partial F(t,\beta)}{\partial t} \right| = O_p(h_1 + (nh_1^3)^{-1/2} \ln n),$$

$$\sup_{\beta} \left| \frac{\partial F_n(x\beta,\beta)}{\partial \beta} - \frac{\partial F(x\beta,\beta)}{\partial \beta} \right| = O_p(h_1 + (nh_1^3)^{-1/2} \ln n),$$

and

$$\sup_{t} |1 - F_n(t,\hat{\beta}) - F_n(2\alpha_0 - t,\hat{\beta})| = O_p(h_1^2 + (nh_1)^{-1/2} \ln n).$$

Proof. We need to consider the bias and variance terms for the local linear estimates. We can deduce the rates of convergence for the variance terms following the approach in Newey (Lemma B.1, 1994). The rates of convergence for the bias terms can be derived as in Fan and Gijbels (1995, Sect. 3.7). The last equality also relies on the root-n consistency of $\hat{\beta}$ and the symmetry and monotonic index assumption. ■

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

Similarly, we can prove the following lemma.

LEMMA 2. *Under Assumptions* 1, 4, 7, 8, *and* 9(ii),

$$F_n(t,\beta_0) - F(t) = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{p(t)}(d_i - F(t))K\left(\frac{z_i - t}{h_1}\right) + O_p(h_1^2 + (nh_1)^{-1}\ln^2 n)$$

uniformly in $t \in S$.

Proof of Theorem 1. From the definition of $\hat{\alpha}_1$, we have

$$\sqrt{n}(\hat{\alpha}_1 - \alpha_0) = \frac{\sqrt{n}A_n(\hat{\beta},\alpha_0)}{B_n(\hat{\beta})}.$$

We will show that $B_n(\hat{\beta})$ converges to B in probability and $\sqrt{n}A_n(\hat{\beta})$ is asymptotically normal. First, from Assumptions 1–9 and Lemma 1 and from Lemma 3.1 of Powell et al. (1989) we can show that

$$\begin{aligned}
B_n(\hat{\beta}) &= \frac{2}{n(n-1)}\sum_{i=j}\frac{1}{h_2}k\left(\frac{F_i + F_j - 1}{h_2}\right)w(z_i)w(z_j) \\
&\quad + h_2^{-2}O_p\left(\sup_t|F_n(t,\hat{\beta}) - F(t,\hat{\beta})|\right) + O_p(h_2^{-2})(\hat{\beta} - \beta_0) \\
&= 2\lim E\frac{1}{h_2}k\left(\frac{F_i + F_j - 1}{h_2}\right)w(z_i)w(z_j) + o_p(1) \\
&= 2Eg(1 - F_i)w(\lambda(F_i))w(\lambda(1 - F_i)) + o_p(1) \\
&= B + o_p(1). \qquad \textbf{(A.1)}
\end{aligned}$$

Next, we write $\sqrt{n}A_n(\hat{\beta})$ as

$$\sqrt{n}A_n(\hat{\beta}) = A_{n0} + (A_{n1}(\bar{\beta},\alpha_0) + A_{n2}(\bar{\beta}) + A_{n3}(\bar{\beta},\alpha_0))\sqrt{n}(\hat{\beta} - \beta_0),$$

where

$$\begin{aligned}
A_{n0} &= \frac{1}{\sqrt{n}(n-1)}\sum_{i=j}(z_i + z_j - 2\alpha_0) \\
&\quad \times \frac{1}{h_2}k\left(\frac{F_n(z_i,\beta_0) + F_n(z_j,\beta_0) - 1}{h_2}\right)w(z_i)w(z_j), \\
A_{n3}(\beta,\alpha) &= \frac{1}{n(n-1)}\sum_{i=j}(x_i\beta + x_j\beta - 2\alpha)\frac{1}{h_2}k\left(\frac{F_n(x_i\beta,\beta) + F_n(x_j\beta,\beta) - 1}{h_2}\right) \\
&\quad \times (w'(x_i\beta)w(x_j\beta)x_i + w(x_j\beta)w'(x_j\beta)x_j),
\end{aligned}$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

where $\bar{\beta}$ lies on the line segment of β_0 and $\hat{\beta}$ and $w'(\cdot)$ is the derivative of $w(\cdot)$. Similar to (A.1), we can show that $A_{n1}(\bar{\beta}, \alpha_0) = A_1 + o_p(1)$, $A_{n2}(\bar{\beta}) = A_2 + o_p(1)$, and $A_{n3}(\bar{\beta}, \alpha_0) = o_p(1)$. Finally, we consider A_{n0},

$$A_{n0} = A_{n01} + A_{n02} + o_p(1),$$

where

$$A_{n01} = \frac{1}{\sqrt{n}(n-1)} \sum_{i=j} (z_i + z_j - 2\alpha_0) \frac{1}{h_2} k\left(\frac{F_i + F_j - 1}{h_2}\right) w(z_i) w(z_j)$$

and

$$\begin{aligned} A_{n02} = {} & \frac{1}{\sqrt{n}(n-1)} \sum_{i=j} (z_i + z_j - 2\alpha_0) \frac{1}{h_2^2} k'\left(\frac{F_i + F_j - 1}{h_2}\right) w(z_i) w(z_j) \\ & \times (F_n(z_i, \beta_0) + F_n(z_j, \beta_0) - F_i - F_j) + n^{1/2} h_2^{-3} O_p\left(\sup_t |F_n(t) - F(t)|^2\right) \\ = {} & \frac{2}{\sqrt{n}(n-1)(n-2)} \sum_{i=j=l} q_{ijl} + o_p(1) \end{aligned}$$

with

$$q_{ijl} = \frac{1}{h_2^2 h_1 p(z_i)} (z_i + z_j - 2\alpha_0)(d_l - F_l) k'\left(\frac{F_i + F_j - 1}{h_2}\right) K\left(\frac{z_l - z_i}{h_1}\right) w(z_i) w(z_j)$$

by Lemmas 1 and 2. Again by a U-statistic projection (Powell et al., 1989, Lemma 3.1) we obtain that $A_{n01} = o_p(1)$, and

$$A_{n02} = \frac{2}{\sqrt{n}} \sum_{l=1}^{n} E(q_{ijl}|l) + o_p(1) = \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \xi_{1i} + o_p(1) \tag{A.2}$$

because $E(q_{ijl}|i) = E(q_{ijl}|j) = Eq_{ijl} = 0$. Therefore

$$\begin{aligned} \sqrt{n}(\hat{\alpha}_1 - \alpha_0) &= B^{-1} \left[(A_1 + A_2)\sqrt{n}(\hat{\beta} - \beta_0) + \frac{1}{\sqrt{n}} \sum_{l=1}^{n} \xi_{1i} \right] + o_p(1) \\ &= B^{-1} \frac{1}{\sqrt{n}} \sum_{l=1}^{n} (\xi_{1i} + \Omega_1 \psi_{1i}) + o_p(1). \end{aligned} \tag{A.3}$$

Theorem 1 follows easily. ■

Proof of Theorem 2. From Lemma 1, with probability close to one, there exists N, for $n > N$, such that

$$\sup_{z \in S_b} |[F_n(z, \hat{\beta}) + F_n(2\alpha_0 - z + \varepsilon_0, \hat{\beta})] - [F(z) - F(2\alpha_0 - z + \varepsilon_0)]| < \varepsilon_0 f_0/2$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

holds, where $f_0 = \min_{t\in S} f(t-\alpha_0|x)$. From the symmetry assumption, we have

$$\min_{z\in S_b}[F(z)+F(2\alpha_0-z+\varepsilon_0)-1] = \min_{z\in S_b}[F(2\alpha_0-z+\varepsilon_0)-F(2\alpha_0-z)] > \varepsilon_0 f_0,$$

thus

$$\min_{z\in S_b}[F_n(z,\hat{\beta})+F_n(2\alpha_0-z+\varepsilon_0,\hat{\beta})-1] > \varepsilon_0 f_0/2.$$

Similarly, we can show that

$$\max_{z\in S_b}[F_n(z,\hat{\beta})+F_n(2\alpha_0-z-\varepsilon_0,\hat{\beta})-1] < -\varepsilon_0 f_0/2.$$

Therefore, with probability close to one, for $\hat{z}_i \in S_b$, there exists $\hat{\alpha}_{2i}$ such that $2\hat{\alpha}_{2i}-\hat{z}_i \in S$,

$$F_n(2\hat{\alpha}_{2i}-\hat{z}_i,\hat{\beta}) = 1-F_n(\hat{z}_i,\hat{\beta}), \tag{A.4}$$

and

$$\sup_i|\hat{\alpha}_{2i}-\alpha_0| < \varepsilon_0/4.$$

Notice that from equation (A.4), we have

$$F_n(2\alpha_0-\hat{z}_i,\hat{\beta}) + 2\frac{\partial F_n}{\partial t}(2\bar{\alpha}_{2i}-\hat{z}_i,\hat{\beta})(\hat{\alpha}_{2i}-\alpha_0) = 1-F_n(\hat{z}_i,\hat{\beta}),$$

where $\hat{\alpha}_{2i}$ lies between $\hat{\alpha}_{2i}$ and α_0, thus

$$[\hat{\alpha}_{2i}-\alpha_0] = \left[-2\frac{\partial F_n}{\partial t}(2\bar{\alpha}_{2i}-\hat{z}_i,\hat{\beta})\right]^{-1}[F_n(\hat{z}_i,\hat{\beta})+F_n(2\alpha_0-\hat{z}_i,\hat{\beta})-1].$$

From Lemma 1, we have

$$\begin{aligned}\sup_i|\hat{\alpha}_{2i}-\alpha_0| &= O_p\Big(\sup_{t\in S}|1-F_n(t,\hat{\beta})-F_n(2\alpha_0-t,\hat{\beta})|\Big)\\ &= O_p((nh_1)^{-1/2}\ln n + h_1^2).\end{aligned}$$

We now derive the asymptotic distribution for $\hat{\alpha}_2$. From the definition of $\hat{\alpha}_2$, we obtain

$$\sqrt{n}(\hat{\alpha}_2-\alpha_0) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\alpha}_{2i}-\alpha_0)w(\hat{z}_i)}{\frac{1}{n}\sum_{i=1}^{n}w(\hat{z}_i)}$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

First we look at the numerator. From Lemma 1, we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\alpha}_{2i}-\alpha_0)w(\hat{z}_i)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[-2\frac{\partial F_n}{\partial t}(2\bar{\alpha}_{2i}-\hat{z}_i,\hat{\beta})\right]^{-1}[F_n(\hat{z}_i,\hat{\beta})+F_n(2\alpha_0-\hat{z}_i,\hat{\beta})-1]w(\hat{z}_i)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[-2\frac{\partial F}{\partial t}(2\alpha_0-\hat{z}_i,\hat{\beta})\right]^{-1}[F_n(\hat{z}_i,\hat{\beta})+F_n(2\alpha_0-\hat{z}_i,\hat{\beta})-1]w(\hat{z}_i)\\
&\qquad+\sqrt{n}\left\{\sup_{t\in S,|\beta-\beta_0|<\varepsilon_0}\left|\frac{\partial F_n(t,\beta)}{\partial t}-\frac{\partial F(t,\beta)}{\partial t}\right|+\sup_i|\hat{\alpha}_{2i}-\alpha_0|\right\}\\
&\qquad\times\sup_{t\in S}|1-F_n(t,\hat{\beta})-F_n(2\alpha_0-t,\hat{\beta})|w(t)+o_p(1)\\
&\quad=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[-2\frac{\partial F}{\partial t}(2\alpha_0-\hat{z}_i,\hat{\beta})\right]^{-1}[F_n(\hat{z}_i,\hat{\beta})+F_n(2\alpha_0-\hat{z}_i,\hat{\beta})-1]w(\hat{z}_i)\\
&\qquad+\sqrt{n}O_p[(h_1+(nh_1^3)^{-1/2}\ln n)(h_1^2+(nh_1)^{-1/2}\ln n)]\\
&\quad=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{2f(\alpha_0-z_i|x_i)}[F_n(\hat{z}_i,\hat{\beta})+F_n(2\alpha_0-\hat{z}_i,\hat{\beta})-1]w(\hat{z}_i)+o_p(1)\\
&\quad=C_{1n}+(C_{2n}+C_{3n})\sqrt{n}(\hat{\beta}-\beta_0)+o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
C_{1n}&=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{2f(\alpha_0-z_i|x_i)}(F_n(z_i,\beta_0)+F_n(2\alpha_0-z_i,\beta_0)-1]w(z_i),\\
C_{2n}&=-\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2f(\alpha_0-z_i|x_i)}\left(\frac{\partial F_n(x_i\beta_0,\beta_0)}{\partial\beta'}+\frac{\partial F_n(2\alpha_0-x_i\beta_0,\beta_0)}{\partial\beta'}\right)w(z_i)\\
&=-E\frac{1}{2f(\alpha_0-z_i|x_i)}\left(\frac{\partial F(x_i\beta_0,\beta_0)}{\partial\beta'}+\frac{\partial F(2\alpha_0-x_i\beta_0,\beta_0)}{\partial\beta'}\right)w(z_i)+o_p(1)\\
&=\Omega_2+o_p(1),
\end{aligned}$$

and

$$C_{3n}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2f(\alpha_0-z_i|x_i)}[1-F_n(z_i,\beta_0)-F_n(2\alpha_0-z_i,\beta_0)]\frac{\partial w(x_i\beta_0)}{\partial\beta'}=o_p(1).$$

We write C_{1n} as $C_{1n}=C_{11n}+C_{12n}$, where

$$C_{11n}=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{2f(\alpha_0-z_i|x_i)}[F_n(z_i)-F_i]w(z_i)$$

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press

and

$$C_{12n} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{2f(\alpha_0 - z_i|x_i)} [F_n(2\alpha_0 - z_i) - F(2\alpha_0 - z_i)]w(z_i).$$

Similar to Lemma 2, we can show that

$$C_{11n} = -\frac{1}{\sqrt{n}(n-1)} \sum_{j=i}^{n} \frac{1}{2f(\alpha_0 - z_i|x_i)p(z_i)} (d_j - F_j) \frac{1}{h_1} K\left(\frac{z_j - z_i}{h_1}\right) w(z_i) + o_p(1)$$

$$= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{2f(\alpha_0 - z_i|x_i)} (d_i - F_i)w(z_i) + o_p(1)$$

and

$$C_{12n} = -\frac{1}{\sqrt{n}(n-1)} \sum_{j=i}^{n} \frac{1}{2f(\alpha_0 - z_i|x_i)p(2\alpha_0 - z_i)h} (d_j - F_j)K$$

$$\times \left(\frac{z_j + z_i - 2\alpha}{h}\right) w(z_i) + o_p(1)$$

$$= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{2f(\alpha_0 - z_i|x_i)p(z_i)} (d_i - F_i)p(2\alpha_0 - z_i)w(2\alpha_0 - z_i) + o_p(1).$$

Hence

$$C_{1n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{2i} + o_p(1).$$

Also we can show that

$$\frac{1}{n} \sum_{i=1}^{n} w(\hat{z}_i) = Ew(z_i) + o_p(1).$$

Consequently

$$\sqrt{n}(\hat{\alpha}_2 - \alpha_0) = W^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{2i} + \Omega_2 \sqrt{n}(\hat{\beta} - \beta_0) \right]. \tag{A.5}$$

Theorem 2 follows easily. ■

https://doi.org/10.1017/S0266466699151041 Published online by Cambridge University Press