ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 78, No. 1 (January, 2010), 159–168

BINARY RESPONSE MODELS FOR PANEL DATA: IDENTIFICATION AND INFORMATION

GARY CHAMBERLAIN
Harvard University, Cambridge, MA 02138, U.S.A.

Econometrica, Vol. 78, No. 1 (January, 2010), 159–168

BINARY RESPONSE MODELS FOR PANEL DATA: IDENTIFICATION AND INFORMATION

BY GARY CHAMBERLAIN[1]

This paper considers a panel data model for predicting a binary outcome. The conditional probability of a positive response is obtained by evaluating a given distribution function (F) at a linear combination of the predictor variables. One of the predictor variables is unobserved. It is a random effect that varies across individuals but is constant over time. The semiparametric aspect is that the conditional distribution of the random effect, given the predictor variables, is unrestricted.

This paper has two results. If the support of the observed predictor variables is bounded, then identification is possible only in the logistic case. Even if the support is unbounded, so that (from Manski (1987)) identification holds quite generally, the information bound is zero unless F is logistic. Hence consistent estimation at the standard pn rate is possible only in the logistic case.

KEYWORDS: Panel data, binary response, correlated random effects, identification, information bound.

1. INTRODUCTION

THIS PAPER CONSIDERS a panel data model for predicting a binary outcome. The conditional probability of a positive response is obtained by evaluating a given distribution function (F) at a linear combination of the predictor variables. One of the predictor variables is unobserved. It is a random effect that varies across individuals but is constant over time. The semiparametric aspect is that the conditional distribution of the random effect, given the predictor variables, is unrestricted.

When the distribution function F is logistic, Rasch's (1960, 1961) conditional likelihood approach can be used to obtain a consistent estimator. Andersen (1970) examined the properties of this estimator. See Chamberlain (1984) for a review and additional results.

Manski (1987) showed that consistent estimation is possible without specifying a functional form for F. Furthermore, the form of F can be allowed to depend on the predictor variables in a time-invariant way. Identification does, however, require an unbounded support for at least one of the predictor variables.

This paper has two results. If the support of the observed predictor variables is bounded, then identification is possible only in the logistic case. Even if the support is unbounded, so that (from Manski (1987)) identification holds quite generally, the information bound is zero unless F is logistic. Hence consistent estimation at the standard $\sqrt{n}$ rate is possible only in the logistic case.

[1]Financial support was provided by the National Science Foundation.

 DOI: 10.3982/ECTA8419

2. IDENTIFICATION

The random vector $(y_{i1}, y_{i2}, x'_{i1}, x'_{i2}, c_i)$ is independently and identically distributed (i.i.d.) for $i = 1, \ldots, n$. We observe $z'_i = (y_{i1}, y_{i2}, x'_{i1}, x'_{i2})$; the latent variable c_i is not observed. The binary variable $y_{it} = 0$ or 1 and $x'_i \equiv (x'_{i1}, x'_{i2})$ has support $X \subset \mathcal{R}^J \times \mathcal{R}^J$. We assume that

$$\text{Prob}(y_{it} = 1 | x_i, c_i) = F(\alpha_0 d_{it} + \beta'_0 x_{it} + c_i) \qquad (t = 1, 2),$$

where $d_{it} = 1$ if $t = 2$ and $= 0$ otherwise. The distribution function F is given as part of the prior specification; it is strictly increasing on the whole line with a bounded, continuous derivative, and with $\lim_{s \to -\infty} F(s) = 0$ and $\lim_{s \to \infty} F(s) = 1$. Furthermore, y_{i1} and y_{i2} are independent conditional on x_i, c_i. The parameter space $\Theta = \Theta_1 \times \Theta_2$, where Θ_1 is an open subset of $\mathcal{R}$, Θ_2 is an open subset of $\mathcal{R}^J$, and $\theta'_0 \equiv (\alpha_0, \beta'_0) \in \Theta$. We assume that Θ_2 contains all $\beta \neq 0$ with $|\beta|$ sufficiently small.

Define

$$p(x, c, \theta) = \begin{pmatrix} [1 - F(\beta' x_1 + c)][1 - F(\alpha + \beta' x_2 + c)] \\ [1 - F(\beta' x_1 + c)] F(\alpha + \beta' x_2 + c) \\ F(\beta' x_1 + c)[1 - F(\alpha + \beta' x_2 + c)] \end{pmatrix}.$$

Let $\mathcal{G}$ consist of the mappings from X into the space of probability measures on $\mathcal{R}$. We let G_x denote G evaluated at x for $G \in \mathcal{G}$. We shall say that identification fails at θ_0 if

$$\int p(x, c, \theta_0) G_{0x}(dc) = \int p(x, c, \theta^*) G^*_x(dc)$$

for all $x \in X$, where G_0 and $G^* \in \mathcal{G}$, $\theta^* \in \Theta$, and $\theta^* \neq \theta_0$. Then (θ_0, G_0) and (θ^*, G^*) give the same conditional distribution for (y_1, y_2) given x.

The distribution F is logistic if

$$F(s) = \exp(\phi_1 + \phi_2 s)/[1 + \exp(\phi_1 + \phi_2 s)]$$

for some $\phi_1, \phi_2 \in \mathcal{R}$.

THEOREM 1: *If X is bounded, then identification fails for all θ_0 in some open subset of Θ if F is not logistic.*

PROOF: Let $\text{co}\, p(x, \mathcal{R}, \theta)$ denote the convex hull of the set $\{p(x, c, \theta) : c \in \mathcal{R}\}$. Suppose that for some $\alpha \in \Theta_1$, this convex hull contains an open ball B (in $\mathcal{R}^3$) when $\beta = 0$. Then for any θ_0 and $\theta^* \in \Theta$ sufficiently close to $(\alpha, 0)$,

$$\text{co}\, p(x, \mathcal{R}, \theta_0) \cap \text{co}\, p(x, \mathcal{R}, \theta^*)$$

is nonempty for all $x \in X$. Then for each $x \in X$, there are probability measures G_{0x} and G_x^* such that

$$\int p(x, c, \theta_0) G_{0x}(dc) = \int p(x, c, \theta^*) G_x^*(dc).$$

Hence θ_0 is not identified unless the dimension of $\mathrm{co}\, p(x, \mathcal{R}, (\alpha, 0))$ is 2 for all $\alpha \in \Theta_1$. In that case, for each $\alpha \in \Theta_1$, there exist scalars $\psi_1, \dots, \psi_4$ (not all zero) such that

$$\begin{aligned} &\psi_1[1 - F(c)][1 - F(\alpha + c)] + \psi_2[1 - F(c)]F(\alpha + c) \\ &\quad + \psi_3 F(c)[1 - F(\alpha + c)] = \psi_4 \end{aligned}$$

for all $c \in \mathcal{R}$. Letting $c \to \infty$ gives $\psi_4 = 0$; letting $c \to -\infty$ gives $\psi_1 = 0$. Hence, with $Q \equiv F/(1 - F)$, we have

$$\psi_2 Q(\alpha + c) + \psi_3 Q(c) = 0$$

and so

$$Q(\alpha + c) = Q(\alpha)Q(c)/Q(0)$$

for all $\alpha \in \Theta_1$ and all $c \in \mathcal{R}$. The only positive, continuously differentiable solution to this form of Cauchy's equation is

$$Q(s) = \exp(\phi_1 + \phi_2 s);$$

then the result follows from $F = Q/(1 + Q)$. *Q.E.D.*

Manski's (1987) model is specified as (dropping the i subscripts)

$$y_t = \begin{cases} 1, & \text{if } \theta_0' w_t + c + u_t \geq 0, \\ 0, & \text{otherwise} \end{cases} \qquad (t = 1, 2),$$

$$u_1 | w_1, w_2, c \overset{d}{=} u_2 | w_1, w_2, c.$$

The key restriction here is that the latent error, u_t, should have an identical distribution in both periods, conditional on the predictor variables w_1, w_2, c. (Also the support of the distribution of u_t conditional on w_1, w_2, c is assumed to be $\mathcal{R}$.) This allows for a certain kind of heteroskedasticity, but does not permit, for example, the conditional distribution of u_t to be more sensitive to w_t than to w_s ($s, t = 1, 2; s \neq t$).

Our model, with $w_t' = (d_t, x_t')$, imposes additional restrictions. We require u_1, u_2 to be independent of w_1, w_2, c and to be i.i.d. over time with a known distribution ($-u_t \overset{d}{=} F$).

Manski's result is that θ_0 is identified up to scale if a component of $w_2 - w_1$ (with a nonzero coefficient) has positive Lebesgue density on the whole line, conditional on the other components of $w_2 - w_1$ and on c. (In addition, the support of the $K \times 1$ vector $w_2 - w_1$ should not be contained in a proper linear subspace of $\mathcal{R}^K$.) Our scale normalization is built in to the given specification for the u_t distribution, but we can make comparisons by considering ratios of coefficients. The proof of Theorem 1 shows that a logistic F is necessary for such ratios to be identified. It is the bounded support for the predictor variables (w_1, w_2) that accounts for the difference in results.

In the next section, we shall assume that x_i has positive Lebesgue density on all of $\mathcal{R}^{2J}$. Then identification is possible in general, but we shall see that the information bound is zero except for the logistic case.

3. INFORMATION

Our semiparametric information bound is based on considering the least favorable parametric subfamily. This idea is owing to Stein (1956) and has been developed by Levit (1975), Begun, Hall, Huang, and Wellner (1983), and Pfanzagl (1982).

The observations consist of independent and identically distributed (i.i.d.) random vectors $z_1, \ldots, z_n$ with values in Z, a subset of a Euclidean space. The distribution of z_1 has positive density $f(z; \theta_0, g_0)$ with respect to a σ-finite measure μ. The parametric component θ_0 is an element of Θ, an open subset of $\mathcal{R}^K$. The nonparametric component g_0 is an element of Γ, an infinite-dimensional set. A path λ through g_0 is a mapping from an open interval $(c, d) \subset \mathcal{R}$ into Γ, where $\lambda(\delta_0) = g_0$ for a unique $\delta_0 \in (c, d)$. The path λ is used to construct a parametric likelihood function

$$f_\lambda(z; \theta, \delta) = f(z; \theta, \lambda(\delta)).$$

Let $\gamma' = (\theta', \delta)$ and $\gamma_0' = (\theta_0', \delta_0)$. Then we have mean-square differentiability at γ_0 if

$$f_\lambda^{1/2}(z; \gamma) - f_\lambda^{1/2}(z; \gamma_0) = \sum_{j=1}^{K+1} \psi_{\lambda j}(z)(\gamma_j - \gamma_{0j}) + r(z; \gamma),$$

where

$$\int r^2(z; \gamma)\mu(dz)/|\gamma - \gamma_0|^2 \to 0$$

as $\gamma \to \gamma_0$. If the mean-square differential exists and if the partial derivatives exist almost everywhere with respect to μ (a.e. μ), then

$$\psi_{\lambda j}(z) = \frac{1}{2} f_\lambda^{-1/2}(z; \gamma_0)\, \partial f_\lambda(z; \gamma_0)/\partial \gamma_j \qquad \text{(a.e. } \mu).$$

The partial information for θ_{0k} $(k = 1, \dots, K)$ in the path λ is

$$I_{\lambda,k} = 4 \inf_{\alpha \in \mathcal{R}^{K+1}:\, \alpha_k = 1} \int \left(\sum_{j=1}^{K+1} \alpha_j \psi_{\lambda j}(z) \right)^2 \mu(dz).$$

Given a set Λ of paths, we define

$$I_{\Lambda,k} = \inf_{\lambda \in \Lambda} I_{\lambda,k}.$$

We let $I_\Lambda = 0$ denote that $I_{\Lambda,k} = 0$ for $k = 1, \dots, K$. In that case, no component of θ_0 can be estimated at a $\sqrt{n}$ rate; see Chamberlain (1986, Theorem 2).

In the panel data model from Section 2, we shall assume that the conditional distribution of c given x has a density g with respect to Lebesgue measure. Then the likelihood is based on the density (with respect to μ)

$$f(z; \theta, g) = \int A(z, c, \theta) g(c, x)\, dc,$$

where

$$A(z, c, \theta) = \prod_{t=1}^{2} F(\alpha d_t + \beta' x_t + c)^{y_t} \cdot [1 - F(\alpha d_t + \beta' x_t + c)]^{(1-y_t)}.$$

The measure μ is defined on $Z = Y \times X$, where $Y = \{(0,0), (0,1), (1,0), (1,1)\}$ and X is now $\mathcal{R}^J \times \mathcal{R}^J$. If $B_1 \subset Y$ and B_2 is a Borel subset of X, then

$$\mu(B_1 \times B_2) = \tau(B_1) v(B_2),$$

where τ is the counting measure on Y and the measure v gives the probability distribution of $x_i \equiv (x_{i1}, x_{i2})$.

We shall assume that v has positive Lebesgue density on all of $\mathcal{R}^{2J}$. In addition, we shall assume that $\beta_{0J} \neq 0$ for $\beta_0 \in \Theta_2$.

We shall use the following specification for Γ:

DEFINITION 1: Γ consists of all measurable functions $g : \mathcal{R} \times X \to \mathcal{R}$ such that (i) $\inf_{(c,x) \in B} g(c, x) > 0$ for any compact subset B of $\mathcal{R} \times X$; (ii) for all $x \in X$, $\int_{-\infty}^{\infty} g(c, x)\, dc = 1$; (iii) for each $z \in Z$ and $\theta \in \Theta$, there exists a neighborhood $C \subset \Theta$ of θ and a measurable function $q_1 : \mathcal{R} \to \mathcal{R}$ such that $\int q_1(c) g(c, x)\, dc < \infty$ and

$$|\partial A(z, c, \eta)/\partial \theta| \leq q_1(c) \quad \text{for all } \eta \in C;$$

(iv) for each $\theta \in \Theta$, there is a neighborhood $C \subset \Theta$ of θ and a function $q_2 : Z \to \mathcal{R}$ such that $\int q_2(z)\mu(dz) < \infty$ and

$$\left[\int |\partial A(z, c, \eta)/\partial\theta| g(c, x)\, dc\right]^2 \Big/ \int A(z, c, \eta) g(c, x)\, dc \leq q_2(z) \quad \text{for all } \eta \in C.$$

Part (iii) of Definition 1 ensures that f is continuously differentiable with respect to θ for each $z \in Z$; part (iv) ensures that the mean-square derivative of $f^{1/2}$ with respect to θ exists.

We shall work with the following set of paths:

DEFINITION 2: Λ consists of the paths

$$\lambda(\delta) = g_0[1 + (\delta - \delta_0)h],$$

where $g_0 \in \Gamma$ and $h : \mathcal{R} \times X \to \mathcal{R}$ is a bounded, measurable function with

$$\int g_0(c, x)h(c, x)\, dc = 0 \quad \text{for all } x \in X.$$

THEOREM 2: *$I_\Lambda = 0$ for all θ_0 in Θ if F is not logistic.*

PROOF: It is straightforward to check that $\lambda(\delta) \in \Gamma$ for δ sufficiently close to δ_0. Define the parametric likelihood function $f_\lambda(z; \theta, \delta) = f(z; \theta, \lambda(\delta))$. Now let $\gamma' = (\theta', \delta)$ and $\gamma_0' = (\theta_0', \delta_0)$, and apply the mean-value theorem to obtain

$$f_\lambda^{1/2}(z; \gamma) - f_\lambda^{1/2}(z; \gamma_0) = \frac{\partial f_\lambda^{1/2}(z; \gamma_0)}{\partial \gamma'}(\gamma - \gamma_0) + r(z; \gamma),$$

where

$$r(z; \gamma) = \left[\frac{\partial f_\lambda^{1/2}(z; \tilde{\gamma})}{\partial \gamma} - \frac{\partial f_\lambda^{1/2}(z; \gamma_0)}{\partial \gamma}\right]'(\gamma - \gamma_0)$$

and $\tilde{\gamma}$ is on the line segment joining γ and γ_0:

$$\frac{r^2(z; \gamma)}{|\gamma - \gamma_0|^2} \leq \left|\frac{\partial f_\lambda^{1/2}(z; \tilde{\gamma})}{\partial \gamma} - \frac{\partial f_\lambda^{1/2}(z; \gamma_0)}{\partial \gamma}\right|^2 \to 0$$

as $\gamma \to \gamma_0$ (a.e. μ) since the partial derivatives are continuous in γ. Then the dominated convergence theorem implies that

$$\int r^2(z; \gamma)\mu(dz)/|\gamma - \gamma_0|^2 \to 0$$

and so the mean-square differentiability condition is satisfied.

Now we need to show that if F is not logistic, then given $k \in \{1, \ldots, K\}$ and given $\varepsilon > 0$, there is a path $\lambda \in \Lambda$ such that

$$(1)\qquad 4\int\left(\frac{\partial f_\lambda^{1/2}(z;\gamma_0)}{\partial\theta_k} - \frac{\partial f_\lambda^{1/2}(z;\gamma_0)}{\partial\delta}\right)^2 \mu(dz)$$
$$= \sum_y \int f^{-1}(z;\theta_0, g_0)$$
$$\times\left(\frac{\partial f(z;\theta_0,g_0)}{\partial\theta_k} - \int A(z,c,\theta_0)g_0(c,x)h(c,x)\,dc\right)^2 v(dx)$$
$$< \varepsilon.$$

Since $\int f^{-1}(z;\theta_0,g_0)[\partial f(z;\theta_0,g_0)/\partial\theta_k]^2 v(dx) < \infty$, we can choose an $\varepsilon' > 0$ such that (1) is satisfied if there is a compact subset B of $\mathcal{R}^{2J}$ with $v(B) > 1-\varepsilon'$, $h(c,x) = 0$ for $x \notin B$, and

$$(2)\qquad \sum_y \int_B f^{-1}(z;\theta_0,g_0)$$
$$\times\left(\frac{\partial f(z;\theta_0,g_0)}{\partial\theta_k} - \int A(z,c,\theta_0)g_0(c,x)h(c,x)\,dc\right)^2 v(dx) < \varepsilon'.$$

Since $f(z;\theta_0,g_0)$ is bounded away from 0 for $x \in B$, there is an $\varepsilon'' > 0$ such that (2) is satisfied if there is a bounded, measurable function $m : \mathcal{R} \times B \to \mathcal{R}$ such that for all $x \in B : m(c,x) = 0$ for $|c|$ sufficiently large, $\int m(c,x)\,dc = 0$ and

$$(3)\qquad \left[\sum_y\left(\frac{\partial f(z;\theta_0,g_0)}{\partial\theta_k} - \int A(z,c,\theta_0)m(c,x)\,dc\right)^2\right]^{1/2} < \varepsilon''.$$

Then we set

$$h(c,x) = 1(x \in B)m(c,x)/g_0(c,x).$$

($1(\cdot)$ is the indicator function that equals 1 if the condition is satisfied and equals 0 otherwise.)

Let $r(x)$ denote the 4×1 vector whose elements are $\partial f(z;\theta_0,g_0)/\partial\theta_k$ for $y = (0,0), (0,1), (1,0), (1,1)$. Note that $l'r(x) = 0$, where l is a 4×1 vector of 1's. Define

$$a(x,c,\theta) = \begin{pmatrix} [1-F(\beta'x_1+c)][1-F(\alpha+\beta'x_2+c)] \\ [1-F(\beta'x_1+c)]F(\alpha+\beta'x_2+c) \\ F(\beta'x_1+c)[1-F(\alpha+\beta'x_2+c)] \\ F(\beta'x_1+c)F(\alpha+\beta'x_2+c) \end{pmatrix}.$$

Then (3) can be written as

$$(3') \qquad \left| r(x) - \int a(x, c, \theta_0) m(c, x)\, dc \right| < \varepsilon''.$$

Suppose that for all $x \in \mathcal{R}^{2J}$ except for a set with v-probability zero, there exist points $c_j(x) \in \mathcal{R}$ $(j = 1, \ldots, 4)$ with

$$\big[a(x, c_1(x), \theta_0), \ldots, a(x, c_4(x), \theta_0)\big]$$

nonsingular. Then for each such x, there is a neighborhood C_x of x such that

$$\big[a(x', c_1(x), \theta_0), \ldots, a(x', c_4(x), \theta_0)\big]$$

is nonsingular for all x' in the closure of C_x. The C_x provide an open cover of a compact set B with $v(B) > 1 - \varepsilon'$. Hence there is a finite subcover, and we can partition B into Borel subsets $D_1, \ldots, D_m$ and choose the c_j to be simple (hence measurable) functions of the form $c_j(x) = \sum_{k=1}^{m} \kappa_{jk} 1(x \in D_k)$. Furthermore, we can choose the c_j such that

$$H(x) = \big[a(x, c_1(x), \theta_0), \ldots, a(x, c_4(x), \theta_0)\big]$$

has its determinant bounded away from zero for $x \in B$.

Define $b(x) = H^{-1}(x) r(x)$. Since $l'H(x) = l'$, we have

$$l'b(x) = l'H(x)b(x) = l'r(x) = 0.$$

Then (3′) can be written as

$$(3'') \qquad \left| \sum_{j=1}^{4} a(x, c_j(x), \theta_0) b_j(x) - \int a(x, c, \theta_0) m(c, x)\, dc \right| < \varepsilon''.$$

Set

$$m(c, x) = \sum_{j=1}^{4} 1\big(|c - c_j(x)| < \delta\big) b_j(x)/(2\delta).$$

Then m is bounded and measurable, $m(c, x) = 0$ for $|c|$ sufficiently large,

$$\int m(c, x)\, dc = \sum_{j=1}^{4} b_j(x) = 0,$$

and (3″) is satisfied if $\delta > 0$ is sufficiently small.

We conclude that $I_A = 0$ unless, for all x in a set S with positive Lebesgue (outer) measure, $\{a(x, c, \theta_0) : c \in \mathcal{R}\}$ lies in a proper linear subspace of $\mathcal{R}^4$.

Then for each such x, there exists a nonzero $\psi \in \mathcal{R}^4$ such that $\psi' a(x, c, \theta_0) = 0$ for all $c \in \mathcal{R}$; that is,

$$\begin{aligned}&\psi_1[1 - F(\beta_0' x_1 + c)][1 - F(\alpha_0 + \beta_0' x_2 + c)]\\&\quad + \psi_2[1 - F(\beta_0' x_1 + c)]F(\alpha_0 + \beta_0' x_2 + c)\\&\quad + \psi_3 F(\beta_0' x_1 + c)[1 - F(\alpha_0 + \beta_0' x_2 + c)]\\&\quad + \psi_4 F(\beta_0' x_1 + c)F(\alpha_0 + \beta_0' x_2 + c) = 0.\end{aligned}$$

Taking the limit as $c \to \infty$ gives $\psi_4 = 0$, and letting $c \to -\infty$ gives $\psi_1 = 0$. Hence, with $Q \equiv F/(1 - F)$, we have

$$\psi_2 Q(\alpha_0 + \beta_0' x_2 + c) + \psi_3 Q(\beta_0' x_1 + c) = 0$$

and so

$$Q(\alpha_0 + \beta_0' x_2 + c) = Q(\alpha_0 + \beta_0' x_2) Q(\beta_0' x_1 + c)/Q(\beta_0' x_1). \tag{4}$$

Then (4) holds for all x in the closure of S. Define $M(s) = \log Q(s)$ and $\dot{M}(s) = dM(s)/ds$, and take the partial derivative with respect to the Jth component of x_2:

$$\dot{M}(\alpha_0 + \beta_0' x_2 + c)\beta_{0J} = \dot{M}(\alpha_0 + \beta_0' x_2)\beta_{0J}$$

for all $c \in \mathcal{R}$. Hence $M(s) = \phi_1 + \phi_2 s$, and the result follows from $F = \exp(M)/[1 + \exp(M)]$. *Q.E.D.*

If $0 \notin \Theta_1$, then we can reparameterize in terms of $\tilde{\theta} = (\alpha, \beta/\alpha)$. Then we can apply the proof of Theorem 2 to show that the information bound for $\tilde{\beta}_0 \equiv \beta_0/\alpha_0$ is 0 unless F is logistic. Hence, even though a consistent estimator of $\tilde{\beta}_0$ is available from Manski (1987) (under the additional assumption that the parameter space bounds $|\beta_{0J}|/(|\alpha_0| + |\beta_0|)$ away from 0), estimation at the standard $\sqrt{n}$ rate is possible only in the logistic case.

REFERENCES

ANDERSEN, E. B. (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society, Ser. B*, 32, 283–301. [159]

BEGUN, J., W. HALL, W. HUANG, AND J. WELLNER (1983): "Information and Asymptotic Efficiency in Parametric–Nonparametric Models," *The Annals of Statistics*, 11, 432–452. [162]

CHAMBERLAIN, G. (1984): "Panel Data," in *Handbook of Econometrics*, Vol. II, ed. by Z. Griliches and M. Intriligator. Amsterdam: North-Holland, Chapter 22. [159]

——— (1986): "Asymptotic Efficiency in Semiparametric Models With Censoring," *Journal of Econometrics*, 32, 189–218. [163]

LEVIT, B. (1975): "On the Efficiency of a Class of Nonparametric Estimates," *Theory of Probability and Its Applications*, 20, 723–740. [162]

MANSKI, C. (1987): "Semiparametric Analysis of Random Effects Linear Models From Binary Panel Data," *Econometrica*, 55, 357–362. [159,161,167]

PFANZAGL, J. (1982): *Contributions to a General Asymptotic Statistical Theory*. Lecture Notes in Statistics, Vol. 13. New York: Springer-Verlag. [162]

RASCH, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Denmarks Paedagogiske Institute. [159]

——— (1961): "On General Laws and the Meaning of Measurement in Psychology," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4. Berkeley: University of California Press. [159]

STEIN, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley: University of California Press. [162]

Dept. of Economics, Harvard University, 123 Littauer Center, 1875 Cambridge Street, Cambridge, MA 02138, U.S.A.; gary_chamberlain@harvard.edu.

Manuscript received February, 2009; final revision received August, 2009.