W I

SSRI Workshop series 8512

Social Systems Research Institute

University of Wisconsin – Madison.

SEMIPARAMETRIC ANALYSIS
OF RANDOM EFFECTS LINEAR
MODELS FROM BINARY
PANEL DATA

Charles F. Manski

8512

This research was supported under National Science Foundation Grant SES-8319335 and by a grant from the University of Wisconsin Graduate School. Jim Powell has provided very useful comments.

October 1985

ABSTRACT

Andersen (1970) considered the problem of inference on random effects linear models from binary response panel data. He showed that inference is possible if the disturbances for each panel member are known to be white noise with the logistic distribution and if the observed explanatory variables vary over time. A conditional maximum likelihood estimator consistently estimates the model parameters up to scale. The present note shows that inference remains possible if the disturbances for each panel member are known only to be time-stationary with unbounded support and if the explanatory variables vary enough over time. A conditional version of the maximum score estimator (Manski, 1975, 1985) consistently estimates the model parameters up to scale.

Andersen(1970) considered the problem of inference on random effects linear models from binary response panel data. He showed that inference is possible if the disturbances for each panel member are known to be white noise with the logistic distribution and if the observed explanatory variables vary over time. Nothing need be known about the distribution of effects. Andersen proved that a conditional maximum likelihood estimator consistently estimates the model parameters up to scale. For a review of this and related results, see Chamberlain(1984).

The present note shows that inference remains possible if the disturbances for each panel member are known only to be time-stationary with unbounded support and if the explanatory variables vary enough over time. The note proves that a conditional version of the maximum score estimator (Manski,1975,1985) consistently estimates the model parameters up to scale.

Section 1 sets out assumptions and notation. Section 2 proves identification under the assumptions. Section 3 develops a consistent estimator.

1. Assumptions

It suffices to consider the case where two observations are available for each person. Thus, let $[(y_t, x_t, u_t; t=0,1), c]$ be a random vector. Here, y_t is the scalar response variable in period t, x_t is the corresponding K-vector of observed explanatory variables, and u_t is the unobserved scalar disturbance. The random variable c is the unobserved time invariant person-specific effect. The random effects linear model has the form

(1) $$y_t = x_t\beta + c + u_t, \qquad t = 0,1,$$

where $\beta \in R^K$ is a parameter. Define the binary indicator z_t such that $z_t=1$ if $y_t \geq 0$ and $z_t=0$ otherwise. The binary response panel data problem is to combine observations on $(z_t,x_t;t=0,1)$ with prior information so as to learn about β.

To specify the prior information assumed in this paper, let F denote the population distribution of $[(y_t,x_t,u_t;t=0,1),c]$. Let $u\equiv(u_0,u_1)$, $x\equiv(x_0,x_1)$, $z\equiv(z_0,z_1)$. Let $F_{c|x}$ denote the distribution of c conditional on x and let $F_{u|xc}$ denote the distribution of u conditional on (x,c).

Following the literature, we impose no restrictions on $F_{c|x}$ but do presume prior information about $F_{u|xc}$. In particular, we maintain

<u>Assumption 1(Disturbances)</u>: (a) $F_{u_1|xc} = F_{u_0|xc}$, all (x,c).
(b) The support of $F_{u_0|xc}$ is dense in R^1, all (x,c).∎

Part (a) of Assumption 1 says that u_t is stationary, conditional on (x,c). Equivalently, u_t is stationary conditional on the identity of the panel member. This restriction is critical to our analysis. Part (b) is a regularity condition. Its purpose is to guarantee that for all c, the event $z_1 \neq z_0$ occurs with positive probability. We could accomodate disturbances with bounded support if we were to assume that the support of c is bounded. Note that Assumption 1 places no restriction on the form of serial dependence between u_0 and u_1. Nor does it restrict the manner in which $F_{u|xc}$ may vary with (x,c).

Let $\|\ \|$ be a norm on R^K and let $\beta^* = \beta/\|\beta\|$ denote the normalized parameter vector. We shall show in Section 2 that Assumption 1 identifies $\beta*$ provided that the explanatory variables x_t vary sufficiently over time. Let $w = x_1 - x_0$ and let F_w denote the distribution of w. The following condition on F_w will suffice.

<u>Assumption 2(Explanatory Variables):</u> (a) The support of F_w is not contained in any proper linear subspace of R^K.

(b) There exists at least one $k \in [1,2,3,...,K]$ such that $\beta_k \neq 0$ and such that, for almost every value of $\tilde{w} = (w_1, w_2, ..., w_{k-1}, w_{k+1}, ..., w_K)$, the scalar random variable w_k has everywhere positive Lebesgue density conditional on $\tilde{w}$. Without loss of generality, let $k=K$.∎

Assumption 2 has the same form as Manski(1985), Assumption 2. Part (a) is the familiar full-rank condition. It prevents a global failure of identification. If Assumption 1 were strengthened to be the white noise logistic assumption of Andersen, Part (a) would suffice to identify $\beta*$. In our semiparametric setting, Part (a) bounds $\beta*$ but does not identify it. Part (b) prevents a local failure of identification. Part (b) is a substantive restriction. It implies that wb has everywhere positive density for all b such that $b_K \neq 0$.

Finally, we need to state the sampling assumption. This is

<u>Assumption 3(Sampling):</u> A sample of N independent realizations are drawn from F. For each $n=1,...,N$, $(z_{tn}, x_{tn}; t=0,1)$ is observed.∎

Actually, the assumption of random sampling from F is much stronger than necessary. Versions of the consistency result of Section 3 can be proved for sufficiently regular i.n.i.d. and dependent sampling processes.

2. Identification

Let $E(z|x)$ denote the expectation of the observable binary indicators conditional on the observable explanatory variables. Assume that $E(z|x)$ is known for a set of x values having F_x-probability one. Then Assumptions 1 and 2 identify β^*. The following Lemma is the key.

Lemma 1: Let Assumption 1 hold. Then

$$
\begin{aligned}
&x_1\beta > x_0\beta \Leftrightarrow E(z_1|x) > E(z_0|x)\\
(2)\qquad &x_1\beta = x_0\beta \Leftrightarrow E(z_1|x) = E(z_0|x)\\
&x_1\beta < x_0\beta \Leftrightarrow E(z_1|x) < E(z_0|x).\blacksquare
\end{aligned}
$$

Proof: For all (x,c,t), $E(z_t|x,c) = P(y_t \geq 0|x,c)$. In general,

$$P(y_1 \geq 0|x,c) = \int_{-x_1\beta-c}^{\infty} dF_{u_1|xc}.$$

$$P(y_0 \geq 0|x,c) = \int_{-x_0\beta-c}^{\infty} dF_{u_0|xc}.$$

It follows from this and from Assumption 1 that for all c,

$$x_1\beta \gtreqless x_0\beta \Leftrightarrow P(y_1 \geq 0|x,c) \gtreqless P(y_0 \geq 0|x,c).$$

Equivalently, for all c,

$$x_1\beta \gtreqless x_0\beta \Leftrightarrow E(z_1|x,c) \gtreqless E(z_0|x,c).$$

The result now follows immediately.

Q.E.D.

Lemma 1 relates the parameter β to the observable (z,x). Rewriting (2) as

$$
\begin{aligned}
&w\beta > 0 \Leftrightarrow E(z_1-z_0|x) > 0\\
(2')\qquad &w\beta = 0 \Leftrightarrow E(z_1-z_0|x) = 0\\
&w\beta < 0 \Leftrightarrow E(z_1-z_0|x) < 0,
\end{aligned}
$$

we see that Assumption 1 implies the same form of relationship as was shown in Manski(1985), equation (1) to follow from a linear median regression assumption on cross-section data. This makes it natural to ask whether the present panel data problem has a median regression interpretation. In fact it does.

Let $M(z_1-z_0|x,z_1\neq z_0)$ denote the median of z_1-z_0 conditional on x and on the event $z_1\neq z_0$. I am grateful to Jim Powell for help in showing that Lemma 1 has the following Corollary.

<u>Corollary</u>: Let Assumption 1 hold. Then

(3) $M(z_1-z_0|x,z_1\neq z_0) = \text{sgn}(w\beta)$.∎

Proof: The distribution of z_1-z_0 conditional on x and on the event $z_1\neq z_0$ is Bernoulli with

$$P(z_1-z_0=1|x,z_1\neq z_0) = P(z_1=1,z_0=0|x)/P(z_1\neq z_0\ |x)$$

$$P(z_1-z_0=-1|x,z_1\neq z_0) = P(z_1=0,z_0=1|x)/P(z_1\neq z_0\ |x).$$

It follows that

$$M(z_1-z_0|x,z_1\neq z_0) = \text{sgn}[P(z_1=1,z_0=0|x)-P(z_1=0,z_0=1|x)].$$

But

$$P(z_1=1,z_0=0|x) = P(z_1=1|x) - P(z_1=1,z_0=1|x)$$

$$P(z_1=0,z_0=1|x) = P(z_0=1|x) - P(z_1=1,z_0=1|x).$$

Hence,

$$M(z_1-z_0|x,z_1\neq z_0) = \text{sgn}[P(z_1=1|x)-P(z_0=1|x)].$$

By Lemma 1,

$$\text{sgn}[P(z_1=1|x)-P(z_0=1|x)] = \text{sgn}(w\beta).$$

Q.E.D.

Now consider $b \in R^K$, $b \neq \beta$. Lemma 1 distinguishes b from β if there exists a set of w values having positive F_w-probability such that (2′) does not hold when b is substituted for β. In this case, we say that β is identified relative to b. Formally, let

$$(4) \qquad W_b \equiv [w \in R^K: \text{sgn}(wb) \neq \text{sgn}(w\beta)].$$

Then β is identified relative to b if

$$(5) \qquad R(b) \equiv \int_{W_b} dF_w > 0.$$

Clearly the scale of β is not identified. Under Assumption 2, the normalized parameter β^* is identified. This was shown in Manski (1985), Lemma 2 and is restated below.

<u>Lemma 2</u>: Let Assumption 2 hold. Then $R(b) > 0$ for all $b \in R^K$ such that $b/\|b\| \neq \beta^*$.■

3. Consistent Estimation

Our development of a consistent estimator for $\beta*$ follows closely the approach yielding the maximum score estimator. The main idea is to find a function of (z,x,b) whose expectation over (z,x) is maximized uniquely at $\beta*$. We then propose maximization of the sample analog of this expectation. Lemma 3 provides the desired function.

Lemma 3: Let Assumptions 1 and 2 hold. Define

$$(6) \qquad H(b) \equiv E[\mathrm{sgn}(wb)(z_1-z_0)].$$

Then $H(\beta*) > H(b)$ for all $b \in R^K$ such that $b/\|b\| \neq \beta*$.■

Proof: For all $b \in R^K$,

$$H(\beta*)-H(b) = E[\{\mathrm{sgn}(w\beta)-\mathrm{sgn}(wb)\}(z_1-z_0)] = 2\int_{W_b} \mathrm{sgn}(w\beta)E[(z_1-z_0)|w]dF_w.$$

Given Assumption 1, Lemma 1 implies that for all w,

$$\mathrm{sgn}(w\beta)E[(z_1-z_0)|w] = \left|E[(z_1-z_0)|w]\right|.$$

Therefore,

$$H(\beta*)-H(b) = 2\int_{W_b} \left|E[(z_1-z_0)|w]\right| dF_w \geq 0.$$

Under Assumptions 1 and 2, $E[(z_1-z_0)|w] \neq 0$ for almost all w. It now follows from Lemma 2 that $H(\beta*)-H(b) > 0$ whenever $b/\|b\| \neq \beta*$.

Q.E.D.

Now consider estimating $\beta*$ by maximizing the sample analog of $H(*)$, namely the sample average function

$$(7) \qquad H_N(b) \equiv \frac{1}{N}\sum_{n=1}^{N} \mathrm{sgn}(w_n b)(z_{1n}-z_{0n}).$$

Observe that the behavior of $H_N(*)$ is unaffected by removing observations having $z_1=z_0$. Comparison of (7) with Manski(1985), equation (5) shows that the estimator maximizing $H_N(*)$ is maximum score applied to the observations having $z_1\neq z_0$. Thus, we have derived a conditional maximum score estimator. The conditioning event $z_1\neq z_0$ is the same as that used by Andersen to form his conditional maximum likelihood estimator.

Consistency of the proposed estimator follows from Lemma 3, from the fact that $H_N(*)$ behaves like $H(*)$ as the sample size increases, and from the fact that $H(*)$ is smooth as a function of b. The consistency theorem stated below imposes Assumptions 1-3 plus the minor requirement that the parameter space B be bounded away from $b_K=0$. Proof of the theorem parallels the proof of Manski(1985), Theorem 1, where consistency of the maximum score estimator is shown.

Theorem: Let Assumptions 1, 2, and 3 hold. Let there exist a known $\eta>0$ such that $|\beta_K|/\|\beta\| \geq \eta$. Define $B \equiv [b\in R^K: \|b\|=1 \cap |b_K|\geq\eta]$. Let $N\to\infty$. Then the estimator maximizing the criterion function $H_N(*)$ over B is strongly consistent for β^*.■

Proof: By construction, B is compact and $\beta^*\in B$. Lemma 3 implies that on B, $H(*)$ has its unique maximum at β^*. Manski(1985) Lemma 4 shows that as $N\to\infty$, $H_N(*)$ converges to $H(*)$ uniformly on B, almost surely. By Manski(1985) Lemma 5, H is continuous on B. These properties imply strong consistency.

Q.E.D.

References

Andersen, E.(1970), "Asymptotic Properties of Conditional Maximum Likelihood Estimators", Journal of the Royal Statistical Society, Series B, 32, 283-301.

Chamberlain, G.(1984), "Panel Data", Chapter 22 of Handbook of Econometrics, Volume II, Edited by Z. Griliches and M. Intriligator, Amsterdam:North-Holland.

Manski, C.(1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice", Journal of Econometrics, 3, 205-228.

Manski, C.(1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", Journal of Econometrics, 27, 313-333.