A Root-N Consistent Semiparametric Estimator for Fixed Effect Binary Response Panel Data

Lee, M.J.

Publication date:
1996

entER
for
nomic Research

Discussion paper

[redacted]

Tilburg University

Center
for
Economic Research

No. 9664

A ROOT-N CONSISTENT SEMIPARAMETRIC ESTIMATOR FOR FIXED EFFECT BINARY RESPONSE PANEL DATA

By Myoung-jae Lee

July 1996

ISSN 0924-7815

A ROOT-N CONSISTENT SEMIPARAMETRIC ESTIMATOR

FOR FIXED EFFECT BINARY RESPONSE PANEL DATA

(July 24, 1996)

Myoung-jae Lee

Department of Econometrics

Tilburg University

P.O.Box 90153

5000 LE Tilburg

The Netherlands

E-mail: M.J.Lee@KUB.NL

We propose a $\sqrt{N}$-consistent estimator for binary response panel data where the individual specific effect may be correlated with the regressors. The estimator is asymptotically normal with a simple variance matrix. The estimator does not require specifying the error term distribution nor depends on a smoothing parameter as in other nonparametric methods; it, however, puts restrictions on the relationship between the level of the regression function and its increment over time. Extensions to fixed effect ordered response panel data are considered.

Key words: fixed effect, panel data, binary response, single index, semiparametrics, ordered response.

1. Introduction.

Consider a "fixed-effect" binary response panel data $(y_{it},x_{it}')'$, where $y_{it}=1[y_{it}^*>0]$ with $1[A]=1$ if the event A holds and 0 otherwise,

$$(1.1)\qquad y_{it}^* = x_{it}'\beta + \alpha_i + u_{it}, \qquad i=1\ldots N, \quad t=1,2,$$

x_{it} is a k×1 regressor vector and α_i is an unobserved unit-specific effect possibly correlated with $x_i\equiv(x_{i1}',x_{i2}')'$. Chamberlain (1980) proposed a parametric estimator for β under the independence of $u_i\equiv(u_{i1},u_{i2})'$ from x_i and u_{i1} and u_{i2} being iid Logistic. Manski (1987) proposed a semiparametric estimator ("panel maximum score" estimator (PMSE)) for β without specifying the distribution of the error terms under the following restrictions: with $x_{i2}-x_{1i} \equiv \Delta x_i = (\Delta x_{i1}\ldots\Delta x_{ik})'$,

(1.2) $u_{i1}|(\alpha_i,x_i')$ and $u_{i2}|(\alpha_i,x_i')$ follow the same distribution, and the support of $u_{i1}|(\alpha_i,x_i')$ is R^1 for a.e. $(\alpha_i,x_i')'$.

(1.3) $E\{(\Delta x_i-E\Delta x_i)(\Delta x_i-E\Delta x_i)\}'$ is p.d., and there is a regressor, say the kth, such that $\beta_k\neq 0$ and $\Delta x_{ik}|(\Delta x_{i1}\ldots\Delta x_{i,k-1})$ has an everywhere positive Lebesque density for a.e. $(\Delta x_{i1}\ldots\Delta x_{i,k-1})$.

Charlier et al. (1995) and Kyriazidou (1994) smooth PMSE as Horowitz (1992) smooths the (cross-section) maximum score estimator (MSE) of Manski (1985). MSE is $N^{1/3}$-consistent with no practical asymptotic distribution (Kim and Pollard (1990)); PMSE is likely to have a similar asymptotic behavior. The smoothed PMSE is asymptotically normal, but still short of being $\sqrt{N}$-consistent and depends on a smoothing parameter.

In this paper, we propose a $\sqrt{N}$-consistent semiparametric estimator for (1.1). In the semiparametric econometric literature, it has been difficult to extend various $\sqrt{N}$-consistent estimators for single equation models to fixed-effect panel, with the only exception being Honoré (1992) for panel Tobit models. In addition to being useful for (1.1) itself, our estimator can be used also as a first stage estimator for panel sample selection models with a fixed effect in the selection equation, the kind of model analyzed in Kyriazidou (1994). There however is a price to pay: the estimator restrains the relationship between the level of the regression function and its increment over time; also certain restriction on the relation between α and x is necessary.

In Section 2, the estimator is introduced. In Section 3, the identification and consistency are shown. In Section 4, its asymptotic distribution is derived. In Section 5, efficiency and other interesting aspects of the distribution are discussed. In Section 6, two extensions are considered: (1.1) with more than two waves, and fixed effect ordered response panel data. Finally, in Section 7, conclusions are drawn.

2. The Estimator.

Define

(2.1) $$\Delta y_i \equiv y_{i2}-y_{i1} \qquad \text{(and recall } \Delta x_i \equiv x_{i2}-x_{i1}\text{)}.$$

PMSE is obtained by maximizing

(2.2) $$(1/N)\Sigma_i \text{sgn}(\Delta x_i{}'b)\cdot\Delta y_i$$

with respect to (wrt) b, and its consistency is based upon

(2.3) $$\text{Med}(\Delta y_i|x_i,\Delta y_i\neq 0) = \text{sgn}(\Delta x_i{}'\beta),$$

where Med(y|A) is the median of y conditional on A, and sgn(B)=1 if B>0, 0 if B=0 and -1 if B<0. Note that Δy_i given $\Delta y_i \neq 0$ can take only ±1.

Our estimator is obtained by maximizing

$$(2.4)\qquad \{N(N-1)\}^{-1}\Sigma_{i\neq j}\mathrm{sgn}(\Delta x_i{}'b-\Delta x_j{}'b)\cdot(\Delta y_i-\Delta y_j)\cdot\Delta y_i{}^2\Delta y_j{}^2$$

$$(2.5)\qquad = \binom{N}{2}^{-1}\Sigma_{i<j,\Delta y_i\neq\Delta y_j,\Delta y_i\neq 0,\Delta y_j\neq 0}\ \mathrm{sgn}(\Delta x_i{}'b-\Delta x_j{}'b)\cdot(\Delta y_i-\Delta y_j)$$

wrt b, and its consistency is based on the following key equation

$$(2.6)\qquad \mathrm{Med}(\Delta y_i-\Delta y_j\ |\Delta x_i,\Delta x_j,\Delta y_i\neq\Delta y_j,\Delta y_i\neq 0,\Delta y_j\neq 0) = 2\cdot\mathrm{sgn}(\Delta x_i{}'\beta-\Delta x_j{}'\beta).$$

The step from (2.3) to (2.6) is not trivial, for median is not a linear functional; even if so, we would get $\mathrm{sgn}(\Delta x_i{}'\beta)-\mathrm{sgn}(\Delta x_j{}'\beta)$, not the right-hand side of (2.6).

Although $\Delta y_i-\Delta y_j$ can take five values $(0,\pm1,\pm2)$, the conditioning event in (2.6) excludes the middle three values $(0,\pm1)$ to make $\Delta y_i-\Delta y_j$ binary, which is one of the key ideas behind (2.6). The main impetus for the $\sqrt{N}$-consistency of our estimator is the double sums where one sum turns the non-smooth optimand into the smooth function

$$(2.7)\qquad E\{\mathrm{sgn}(\Delta x'b-\Delta x_j{}'b)\cdot(\Delta y-\Delta y_j)\cdot\Delta y^2\Delta y_j{}^2|\Delta x_j,\Delta y_j\}.$$

Then the usual extremum estimator theory is applicable to $(1/\sqrt{N})\Sigma_j(2.7)$; this is not exactly true, for the form of $E\{\cdot|\Delta x_j,\Delta y_j\}$ is a nuisance parameter (Newey and McFadden (1994,p.2200-2202)), resulting in variance four times larger than that of the maximizer of $(1/\sqrt{N})\Sigma_j(2.7)$ as shown in Theorem 2 below. Interpersonal comparison in the double sums provides more information than that used in PMSE, leading to the $\sqrt{N}$-consistency.

For a latent linear cross-section model

$$(2.8)\qquad z_i{}^* = w_i{}'\beta + u_i,$$

Han (1987)'s "maximum rank correlation estimator" (RCE) maximizes

$$(2.9)\qquad \{N(N-1)\}^{-1}\Sigma_{i\neq j}1[w_i{}'b>w_j{}'b]\cdot 1[z_i>z_j]$$

wrt b under the independence of u and x where $z=\tau(z^*)$ with τ being a non-decreasing transformation. Sherman (1993) proved the asymptotic normality of RCE and derived the variance, despite the apparent non-smoothness of the maximand.

The summand in (2.4) takes ±2 depending on whether the sign of $\Delta x_i'b-\Delta x_j'b$ agrees with that of $\Delta y_i-\Delta y_j$ given $\{\Delta y_i\neq 0, \Delta y_j\neq 0\}$. Adding 2 to the summand and dividing it by 4, the summand takes 1 or 0. Then maximizing (2.4) is equivalent to maximizing

$$(2.10) \qquad \{N(N-1)\}^{-1}\Sigma_{i\neq j}1[\Delta x_i'b>\Delta x_j'b]\cdot 1[\Delta y_i>\Delta y_j]\cdot \Delta y_i^2\Delta y_j^2.$$

From this, our estimator may be viewed as a panel data analog of RCE, but the analogy holds only for binary response. For any non-decreasing transformation $\tau(\cdot)$ for which RCE is good, (2.10) is not applicable since $\Delta y=\tau(y_2^*)-\tau(y_1^*)\neq\tau(\Delta y^*)$ in general; the appearance of $1[\Delta y_i>\Delta y_j]$ in (2.10) is an artifact due to the event $\{\Delta y_i\neq\Delta y_j, \Delta y_i\neq 0, \Delta y_j\neq 0\}$.

3. Identification and Consistency.

In this section, we state assumptions and discuss identification of our estimator from which its consistency easily follows. Since β is identified only up to a scale, normalize β by setting $\beta_k=1$, which requires $\beta_k\neq 0$ a priori; from now on, we use

$$\beta = (\beta_1\ldots\beta_{k-1},1)' \equiv (\beta_c',1)', \qquad b = (b_1\ldots b_{k-1},1)' \equiv (b_c',1)'.$$

In practice, one has to try both $\beta_k=1$ and $\beta_k=-1$, for $\text{sgn}(\beta_k)$ is unknown. But since $\text{sgn}(\beta_k)$ is estimated at a faster rate than $\sqrt{N}$, the remainder of the paper is not affected by omitting the case $\beta_k=-1$; see Horowitz (1992) for this point.

Corresponding to β_c, define Δx_{ic} as the components of Δx_i other than the last one Δx_{ik}. Time-invariant variables (e.g. intercept) drop out in Δx_i, and time-variant but "unit-invariant" variables (e.g. time dummies and macro-economic shocks common to all units) drop out in $\Delta x_i - \Delta x_j$, which can reduce the dimension of the parameters further below k-1; to simplify the notation, however, we will ignore this. Let $|H|^2 \equiv \Sigma_i \Sigma_j h_{ij}^2$ for a matrix $H=[h_{ij}]$.

Assumption 1:

(a) The parameter space B is a compact subset of R^{k-1}.

(b) $(y_{i1}, y_{i2}, x_{i1}', x_{i2}')'$, i=1...N, are iid.

(c) (1.3).

(d) $(u_1, u_2)|\alpha$ is independent of $(x_1', x_2')|\alpha$ and the support of $(u_1, u_2)|\alpha$ is R^2 for a.e. α.

(e) $P(\Delta y=1|\Delta x'\beta)$ is an increasing function of $\Delta x'\beta$ and $P(\Delta y=-1|\Delta x'\beta)$ is a decreasing function of $\Delta x'\beta$ for a.e. $\Delta x'\beta$.

(f) "Index increment sufficiency" of $(x_1'\beta, \alpha)$:

$$P(x_1'\beta, \alpha|\Delta x) = P(x_1'\beta, \alpha|\Delta x'\beta) \qquad \text{for a.e. } \Delta x$$

where $P(\cdot|\cdot)$ denotes the conditional distributions.

(a) and (b) are self-explanatory. The second part of (c) involving $\Delta x_{ik}|(\cdot)$ is in fact stronger than necessary, for what we need is

(c)' the kth component of $\Delta x_i - \Delta x_j$ has an everywhere positive Lebesque density conditional on the other components of $\Delta x_i - \Delta x_j$ a.e.

To see this (and to simplify notations), let

$$(3.1) \qquad \Delta x_i \equiv w_i = (w_{i1} \ldots w_{ik})' \equiv (w_{ic}', w_{ik})', \qquad \Delta y_i \equiv z_i$$

where $w_{ic}(\equiv \Delta x_{ic})$ is the first k-1 components of w_i, and observe

$$(3.2) \qquad f_{(w_{ik}-w_{jk})|w_{ic},w_{jc}}(a_o) = \int f_{w_k|w_{ic}}(a_o+w_{jk}) \cdot f_{w_k|w_{jc}}(w_{jk}) dw_{jk}$$

where $f_{a|b}(a_o)$ denote a conditional density of $a|b$ at $a=a_o$. From (3.2), we can see that the second part of (c) implies (c)', but not necessarily the other way around; the second part of (c) is however more primitive.

The assumption (d) is weaker than (1.2) in that (d) does not restrict the marginal distributions, while (d) is stronger than (1.2) in that it requires the conditional independence. In PMSE, only one person i matters, which makes it possible to prove (2.3) by conditioning on x_i and α_i. For our estimator, however, two persons with different x and α appear, so we cannot condition on $x_i, \alpha_i, x_j, \alpha_j$ in proving (2.6); (d) is used to leave only $\Delta x_i'\beta$ and $\Delta x_j'\beta$ in the conditioning set. (d) allows correlation between u_1 and u_2 as well as between (u_1, u_2) and α.

The conditional independence in (d) does not necessarily imply the unconditional independence; e.g. u_1, u_2, x_1 and x_2 may be independently drawn from $N(0, 1+\alpha^2)$. This example shows that u_1 and u_2 can have heteroskedasticity of an unknown form, so long as the conditional variance depends only on α. Considering that RCE requires independence between the error term and the regressor, (d) is the natural analog for panel data, and looks difficult to relax. Finally, (d) does not allow the lagged dependent variable as a regressor, for having y_1 in x_2 implies a correlation between u_1 and x_2.

As for Assumption (e), observe that

$$(3.3) \qquad P(\Delta y=1 | x_1'\beta, \Delta x'\beta, \alpha) = P(u_1 < -x_1'\beta-\alpha, u_2 > -x_1'\beta-\Delta x'\beta-\alpha \; | x_1'\beta, \Delta x'\beta, \alpha)$$

is an increasing function of $\Delta x'\beta$ for a given $x_1'\beta$ and α. Integrate this wrt $(x_1'\beta, \alpha) | \Delta x'\beta$ to get the "weighted average" $P(\Delta y=1 | \Delta x'\beta)$:

$$(3.4) \qquad \int P(\Delta y=1 | x_1'\beta, \Delta x'\beta, \alpha) \cdot f_{x_1'\beta | (\alpha, \Delta x'\beta)}(x_1'\beta) \cdot F_{\alpha | \Delta x'\beta}(d\alpha) \cdot d(x_1'\beta)$$

where the weight is the product of $f_{x_1'\beta | (\alpha, \Delta x'\beta)}$ and $F_{\alpha | \Delta x'\beta}$, a conditional density for $x_1'\beta | (\alpha, \Delta x'\beta)$ and conditional distribution for

$\alpha|\Delta x'\beta$ respectively. Since α is time-invariant and $\Delta x'\beta$ is an increment over time, α can be independent of $\Delta x'\beta$, in which case (3.4) becomes

(3.5) $$\int P(\Delta y=1|x_1'\beta,\Delta x'\beta,\alpha)\cdot f_{x_1'\beta|(\alpha,\Delta x'\beta)}(x_1'\beta)\cdot F_\alpha(d\alpha)\cdot d(x_1'\beta).$$

Thus one overly sufficient condition for Assumption (e) is

(e)' α is independent of $\Delta x'\beta$, and $f_{x_1'\beta|(\alpha,\Delta x'\beta)}$ is a non-decreasing function of $\Delta x'\beta$ for a.e. $(\alpha,\Delta x'\beta)'$.

Note that $f_{x_1'\beta|(\alpha,\Delta x'\beta)}$ is in fact allowed to be decreasing in $\Delta x'\beta$ for some $\Delta x'\beta$ so long as $f_{x_1'\beta|(\alpha,\Delta x'\beta)}$ is increasing in $\Delta x'\beta$ for some other $\Delta x'\beta$ such that the weighted average $P(\Delta y=1|\Delta x'\beta)$ remains increasing in $\Delta x'\beta$. A similar argument holds for $P(\Delta y=-1|\Delta x'\beta)$.

To understand the second part of (e)' better, consider a labor force participation decision where x consists of age, age^2, yearly job-experience, the other household income, marital-status (dummy), and the number of dependents. Assume that unemployment is voluntary to call $x_1'\beta$ "desire to work" at time 1. If a higher $\Delta x'\beta$ implies the lower desire to work for a.e. α, then (e)' fails. If a higher $\Delta x'\beta$ implies the higher desire to work for a.e. α, then (e)' holds. In mixed-cases, (e) depends on which case dominates in the weighted average (3.4). If $x_1'\beta$ and $x_2'\beta$ are iid for a.e. α, then $x_1'\beta$ and $\Delta x'\beta=x_2'\beta-x_1'\beta$ are negatively related, leading to a failure of (e)'. But iid is an exception rather than a rule for micro panel data.

The following is sufficient for (e)' (and so (e)):

(e)" α is independent of $\Delta x'\beta$, and each component of Δx is independent of its level at t=1 for a.e. α,

which can be easier to check than (e)'. Recall the labor example ahead.

First, Δage=1 and it is independent of the age satisfying (e)", but Δage=1 violates (1.3). Second, $\Delta age^2=(age_1+1)^2-age_1^2=2\cdot age_1+1$; the increment is positively related to the level, which is good for (e) and (e)'. Third, if income increases by 5%, then income=$0.05\cdot income_1$; the increment is positively related to the level (if log(income) is used, the increment would be independent of the level). Fourth, Δjob (increment in job experience) is 0 or 1; Δjob is likely to be positively correlated with the level (seniority). Fifth, for marital status dummy (m=1 if married), Δm is negatively related with m_1 when $\Delta m \neq 0$, for $\Delta m=1$ (-1) is associated with $m_1=0$ (1), but $P(\Delta m=0)$ is far greater than $P(\Delta m \neq 0)$ so that Δm may weigh little for $\Delta x'\beta$. Sixth, as for the number of dependents d, the relation between Δd and d_1 is not clear. If x includes only those variables whose increment is independent of its level at t=1 for a.e. α, then (e)" holds.

Assumption (f) is used to prove Lemma 1 below: $P(\Delta y=1|\Delta x_c,\Delta x'\beta) = P(\Delta y=1|\Delta x'\beta)$ for a.e. Δx. Without Lemma 1, $\Delta x'\beta$ would not be the only determinant for Δy: our estimator would not be then valid. Since α is time-invariant and Δx is increment over time, we may assume that α is independent of Δx, as we argued analogously for Assumption (e) ahead. The following (f)' is sufficient for (f) and easier to grasp:

(f)' α is independent of Δx, and (integrating out α in (f)),

$$P(x_1'\beta|\Delta x) \ (=P(x_1'\beta|\Delta x_c,\Delta x'\beta)) = P(x_1'\beta|\Delta x'\beta) \quad \text{for a.e. } \Delta x.$$

In the labor example, (f)' dictates that, given the increment in the desire to work, no other increments should give information for the level of desire at t=1. First, since Δage=1 always, Δage does not give any information for $x_1'\beta$; but as mentioned ahead, age violates (1.3). Second, $\Delta age^2=2\cdot age_1+1$ shows age_1 and this may give information about

$x_1'\beta$, if the desire to work depends on age. Third, if Δincome=5% for everybody, then Δincome does not give any information for $x_1'\beta$. Fourth, it is not clear whether Δjob gives any information on $x_1'\beta$; one could argue for both negative and positive relations with $x_1'\beta$. Fifth, if $\Delta m=1$ or $\Delta d=1$, then this may imply relatively higher $x_1'\beta$ for males and lower $x_1'\beta$ for females. Overall, this example illustrates that, while some of Δx_c may be related to $x_1'\beta$ (so violating (f)'), it is not clear whether the relations are positive or negative, in which case (f)' may not be too farfetched.

The proofs for the following two lemmas are in the appendix.

Lemma 1: Under Assumption 1, Δy is independent of Δx_c given $\Delta x'\beta$:

$$P(\Delta y \mid \Delta x_c, \Delta x'\beta) = P(\Delta y \mid \Delta x'\beta), \quad \text{a.e. } \Delta x.$$

The lemma is a panel analog of index (increment) sufficiency: with $\Delta x'\beta(=w'\beta)$ given, $\Delta x_c(=w_c)$ does not provide any new information for $\Delta y(=z)$. In cross-section models as (2.8) with the same notation w and z, index sufficiency is an assumption, while Lemma 1 is a derived result, for $1[\Delta y=1]$ depends on $x_1'\beta$ and $x_2'\beta$ (and others) not just on $\Delta x'\beta$.

Lemma 2: Under Assumption 1, (2.6) holds.

In essence, (2.6) "identifies" $(\Delta x_i-\Delta x_j)'\beta$. To separate β from $\Delta x_i-\Delta x_j$, $P(X_{id})>0$ is sufficient where

$$X_{id} \equiv \{(x_i', x_j'): \mathrm{sgn}(\Delta x_i'\beta-\Delta x_j'\beta)\neq\mathrm{sgn}(\Delta x_i'b-\Delta x_j'b) \text{ for any } b\neq\beta, b\in B\}.$$

$P(X_{id})>0$ can be proven under Assumption 1, following Manski (1985). With the identification, the a.s. consistency follows from the uniform convergence of U-processes in Sherman (1994) and the continuity of the population version of (2.4). We state this as a theorem:

Theorem 1: Under Assumption 1, β is identified, and the estimator maximizing (2.4) is a.s. consistent for β.

4. The Asymptotic Distribution.

The asymptotic distribution is derived following Sherman (1993) who builds on the techniques in Pollard (1984) and Pakes and Pollard (1989); see also Arcones et al. (1994). It may be done as well following Honoré and Powell (1994) with the technique of Huber (1967).

The key expression is (recall the definitions in (3.1))

$$\tau(z_j,w_j,b_c) \equiv E_{i|j}\{\mathrm{sgn}(w_i'b-w_j'b)\cdot(z_i-z_j)\cdot z_i^2 z_j^2\}, \quad i\neq j, \tag{4.1}$$

where $i|j$ means that the integral is wrt (z_i,w_i') conditional on (z_j,w_j'). Denote the gradient of $\tau(z_j,w_j,b_c)$ wrt b_c as $\nabla_1\tau(z_j,w_j,b_c)$. Denote the second order derivative matrix wrt b_c as $\nabla_2\tau(z_j,w_j,b_c)$. We need the following Assumption 2; near the end of this section, we provide a more primitive Assumption 3 that implies Assumption 2.

Assumption 2:

(a) $\nabla_2\tau(z_j,w_j,b_c)$ exists for a.e. $(z_j,w_j')'$ in a neighborhood N_β of β_c.

(b) There is an integrable function $M(z_j,w_j)$ such that for all $b_c\in N_\beta$

$$|\nabla_2\tau(z_j,w_j,b_c)-\nabla_2\tau(z_j,w_j,\beta_c)| \le M(z_j,w_j)\cdot|b_c-\beta_c|.$$

(c) $E|\nabla_1\tau(z_j,w_j,\beta_c)|^2<\infty$.

(d) $E\nabla_2\tau(z_j,w_j,\beta_c)$ is n.d.

Let $E^{-1}(\lambda)\equiv\{E(\lambda)\}^{-1}$. Denoting the estimator as b_{cN} and omitting z_j,

w_j and β_c in τ, Sherman (1993) proves (for RCE)

(4.2) $$\sqrt{N}(b_{cN}-\beta_c) =^d N(0,\ 2^2\cdot E^{-1}\nabla_2\tau\cdot E(\nabla_1\tau\nabla_1\tau')\cdot E^{-1}\nabla_2\tau\ )$$

under Assumption 2, consistency of β_c, the a.s. continuity of the $\psi(b) \equiv 1[w_i'b>w_j'b]\cdot 1[z_i>z_j]$ wrt the product measure for "i and j", and that $\{\psi(b), b\in B\}$ is a "Euclidean class" in the sense of Pakes and Pollard (1989). Since the Euclidean property easily holds for our estimator and Assumption 1 and 2 are sufficient for the other conditions, (4.2) can be used for our estimator. 2^2 in (4.2) is due to the order of the U-process (2.4) being 2; if the order were 3, we would get 3^2.

Deriving $\nabla_1\tau$ and $\nabla_2\tau$ in the appendix, we obtain

Theorem 2: Under the assumptions 1 and 2, $\sqrt{N}(b_{cN}-\beta_c)$ is asymptotically normal with the variance matrix $4\cdot E^{-1}\nabla_2\tau\cdot E(\nabla_1\tau\nabla_1\tau')\cdot E^{-1}\nabla_2\tau$,

$$E\nabla_2\tau = -8\cdot E[\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot g_o(w'\beta)\cdot\eta(w'\beta)\ ],$$

$$E(\nabla_1\tau\nabla_1\tau') = 16\cdot E[\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot g_o(w'\beta)^2$$
$$\cdot P(z=-1|w'\beta)P(z=1|w'\beta)\cdot\{P(z=-1|w'\beta)+P(z=1|w'\beta)\}\ ]$$

where $g_o(\cdot)$ is a Lebesque density of $w'\beta$,

(4.3) $$\eta(w'\beta) = P'(z=1|w'\beta)P(z=-1|w'\beta) - P'(z=-1|w'\beta)P(z=1|w'\beta),$$

and $P'(\cdot|w'\beta)\equiv dP(z=1|w'\beta)/d(w'\beta)$ which exists as shown in the proof for the sufficiency of Assumption 3 for Assumption 2.

$E\nabla_2\tau$ is n.d. because, due to Assumption 1(e),

(4.4) $$P'(z=1|w'\beta) > 0 > P'(z=-1|w'\beta).$$

The variance of RCE for the binary response model $z=1[z^*>0]$ with z^* in (2.8) has the same format as in (4.2) with (Sherman (1993,p.134))

$$(4.5)\quad E\nabla_2\tau = -4\cdot E[\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot g_o(w'\beta)\cdot P'(z=1|w'\beta)\ ],$$

$$E(\nabla_1\tau\nabla_1\tau') = 16\cdot E[\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot g_o(w'\beta)^2 \cdot P(z=0|w'\beta)P(z=1|w'\beta)\cdot\{P(z=0|w'\beta)+P(z=1|w'\beta)\}\ ].$$

These are almost the same as those of our estimator except at four aspects. First, $z=0$ occurs in RCE while $z=-1$ occurs in our case. Second, $P(z=0|w'\beta)+P(z=1|w'\beta)=1$ for RCE, while $P(z=-1|w'\beta)+P(z=1|w'\beta)<1$ for our estimator, which is a "normalizing factor" since only $z=\pm1$ is used for our estimator. Third, $P'(z=1|w'\beta)$ appears in RCE instead of $\eta(w'\beta)$. To see $\eta(w'\beta)=P'(z=1|w'\beta)$ for the binary case, replace $z=-1$ with $z=0$ to get

$$(4.6)\qquad \eta(w'\beta) = P'(z=1|w'\beta)P(z=0|w'\beta) - P'(z=0|w'\beta)P(z=1|w'\beta)$$
$$= P'(z=1|w'\beta)\{P(z=0|w'\beta)+P(z=1|w'\beta)\} = P'(z=1|w'\beta)$$

due to $P'(z=1|w'\beta)=-P'(z=0|w'\beta)$ in the binary response. Fourth, $E\nabla_2\tau$ in (4.5) has 4 while $E\nabla_2\tau$ in Theorem 2 has $8=4\times2$; the number 4 in RCE is an error which should be 8 (Sherman agrees with this error (personal communication)). Having 2 multiplied into 4 in $E\nabla_2\tau$ of Theorem 2 gets rids of 2^2 in (4.2). In Example 3.3 of Arcones et al. (1994,p.1472) where 3^2 appears for a third order U-process, 3^2 is cancelled by 3 in $E\nabla_2\tau$, which supports our argument.

In $E(\nabla_1\tau\nabla_1\tau')$ of Theorem 2, it looks as if the variance becomes larger as $P(z=-1|w'\beta)+P(z=1|w'\beta)$ goes up, which is counter-intuitive for $z=\pm1$ is the informative part of z. To see that this is not the case, define

$$(4.7)\qquad P_1=P(z=1|w'\beta),\quad P_{-1}=P(z=-1|w'\beta),\quad P_1'=P'(z=1|w'\beta),$$
$$P_{-1}'=P'(z=-1|w'\beta),\quad P_{\pm1}=P_1+P_{-1}.$$

Then it is shown in the appendix that

(4.8) $\eta(w'\beta) = [\partial \ln\{(P_1/P_{\pm1})/(P_{-1}/P_{\pm1})\}/\partial(w'\beta)]\cdot(P_{-1}/P_{\pm1})\cdot(P_1/P_{\pm1})P_{\pm1}^{\,2}$.

Also rewrite $P_1P_{-1}P_{\pm1}$ in $E(\nabla_1\tau\nabla_1\tau')$ as $(P_1/P_{\pm1})(P_{-1}/P_{\pm1})\cdot P_{\pm1}^{\,3}$. Substituting this and (4.8) into the variance of our estimator (i.e., replacing P_1 and P_{-1} with the normalized versions $P_1/P_{\pm1}$ and $P_{-1}/P_{\pm1}$ respectively), overall we get $P_{\pm1}^{-1}$ in the variance heuristically, implying that a larger $P_{\pm1}$ can lead to a smaller variance.

The variance matrix can be estimated by plugging in nonparametric estimates for $E(w_c|w'\beta)$, $g_o(w'\beta)$, $P(z=1|w'\beta)$, $P(z=-1|w'\beta)$, $P'(z=1|w'\beta)$ and $P'(z=-1|w'\beta)$ and invoking theorems in Andrews (1995) as done in Lee (1996) for instance. But this requires many regularity conditions. A theoretically easier way is using numerical derivatives twice to estimate $\nabla_1\tau$ and $\nabla_2\tau$, because one sum in the double sums already plays the role of a nonparametric estimate for the smooth summand in (2.7). Only new condition needed is the speed of the bandwidth parameter ε in the numerical derivatives: $N^{1/2}\varepsilon\to\infty$ for $\nabla_1\tau$ and $N^{1/4}\varepsilon\to\infty$ for $\nabla_2\tau$; see Section 7 of Sherman (1993).

To be specific, for a given i, estimate $\nabla_1\tau(z_i,w_i,\beta_c)$ with $\nabla_{1N}\tau(z_i,w_i,b_{cN})$ which is a numerical first derivative vector of

(4.9) $$(1/N)\Sigma_{j=1,\,j\neq i}^{N}\mathrm{sgn}(\Delta x_i'b_{cN}-\Delta x_j'b_{cN})(\Delta y_i-\Delta y_j)\Delta y_i^{\,2}\Delta y_j^{\,2},$$

and then estimate $E(\nabla_1\tau\nabla_1\tau')$ with

(4.10) $$(1/N)\Sigma_{i=1}^{N}\ \nabla_{1N}\tau(z_i,w_i,b_{cN})\cdot\nabla_{1N}\tau(z_i,w_i,b_{cN})'.$$

Denoting a numerical second order derivative matrix for (4.9) as $\nabla_{2N}\tau(z_i,w_i,b_{cN})$, $E\nabla_2\tau$ can be estimated by $(1/N)\Sigma_i\nabla_{2N}\tau(z_i,w_i,b_{cN})$.

Although we proposed numerical derivatives, this does not necessarily mean that they will work better in practice than a variance estimate with conditional means replaced by kernel estimates. Note that both idea require smoothing parameters. For numerical derivatives, it is hard to choose a proper one unless the derivatives tend to get "stabilized" as the smoothing parameter gets smaller. On the other hand, for the kernel-based method, since the conditioning variable is one-dimensional, one can choose a bandwidth easily using graphs, which can be an important advantage in practice.

The following Assumption 3 implies Assumption 2; the proof is in the appendix.

Assumption 3: There exists a Lebesque density $h(w_k|w_c,z)$ for $w_k|(w_c',z)$ which is everywhere positive for a.e. $(w_c',z)'$ and satisfies:

(a) $h(w_k|w_c,z)$, $\partial h(w_k|w_c,z)/\partial w_k$, $\partial^2 h(w_k|w_c,z)/\partial w_k^2$ are respectively bounded by M_0, M_1 and M_2 a.e. (w_c',z); M_0, M_1, M_2 are constants not depending on (w',z).

(b) $|\partial h(w_k|w_c,z)/\partial w_k|<h(w_k|w_c,z)\cdot\zeta_1$ where ζ_1 is a constant not depending on (w',z).

(c) $|\partial h^2(w_k|w_c,z)/\partial w_k^2|<h(w_k|w_c,z)\cdot\zeta_2$ where ζ_2 is a constant not depending on (w',z).

(d) $E|w|^3<\infty$.

What we actually need for Assumption 2 is various differentiability and Lipschitz continuity conditions wrt $w'\beta$ for $P(z|w'\beta)$ and a density $g(w'\beta|w_c)$ for $w'\beta|w_c$. But because $1[z=1]$ depends on $x_1'\beta$ and $x_2'\beta$ (and on others), assumptions put directly on $P(z|w'\beta)$ can be difficult to check. Instead, observe the following three facts. First, (4.11) $$P(z|w'\beta) = f(w'\beta|z)\cdot P(z)/g_o(w'\beta)$$

where $f(w'\beta|z)$ is a density for $w'\beta|z$. Second,

(4.12) $$h(t-w_c'\beta_c|w_c,z) = \gamma(t|w_c,z)$$

where $\gamma(w'\beta|w_c,z)$ is a density for $w'\beta|(w_c',z)$. Third, $f(w'\beta|z)$, $g(w'\beta|w_c)$ and $g_o(w'\beta)$ are derived from $\gamma(w'\beta|w_c,z)$ (thus from h) by integrating out some variables. This explains why we choose to impose assumptions on h. The everywhere positivity of h is not necessary; it is sufficient for $f(w'\beta|z)$ to be so a.e. z, but we chose the former to be coherent. (b) and (c) are used only once in proving a Lipschitz continuity of $P'(z|w'\beta)$ which is used for Assumption 2(b).

5. More on Asymptotic Distribution.

Chamberlain (1992) shows that,

(5.1) if u_1 and u_2 are iid given $(x',\alpha)'$
and if the support of x is all of R^{2k},

then only logistic distributions for the distribution of $u_1|(x,\alpha)$ are allowed for any $\sqrt{N}$-consistent "pure fixed-effect" estimator for β, which makes the conditional logit of Chamberlain (1980) attractive. For this, we make the following three remarks, the last of which is further discussed (in relation to efficiency of our estimator) later.

First, Assumption 1(e) and (f) restrict the relation between α and x somewhat, deviating from the pure fixed-effect assumed by Chamberlain (1992). This seems to give enough information for our $\sqrt{N}$-consistency despite the unknown distribution of (u_1,u_2). Thus, Chamberlain (1992)'s and our results are not contradictory under (5.1).

Second, Chamberlain (1992) does not address the case of u_1 and u_2 given (x,α) being dependent. With the lagged dependent variable not allowed in the regressor, serial correlation is more likely than not, and u_1 and u_2 are highly unlikely to be logistic in this case. Although our estimator does not allow for the lagged dependent variable, it does allow for serial correlation that is excluded by (5.1).

Third, consider maximizing a conditional log-likelihood function:

$$(5.2) \qquad (1/N)\Sigma_i\{\ 1[z_i=-1]\cdot \ln(P_{-1|\pm}) + 1[z_i=1]\cdot \ln(P_{1|\pm})\ \}$$

which includes the conditional logit as a special case, where (recall the notations in (4.7))

$$P_{-1|\pm}\equiv P(z=-1|z\neq 0,w_i'b) = P_{-1}/P_{\pm 1}, \quad P_{1|\pm}\equiv P(z=1|z\neq 0,w_i'b) = P_1/P_{\pm 1}.$$

In the appendix, it is shown that the variance matrix for the conditional maximum likelihood estimator (MLE) is

$$(5.3) \qquad E^{-1}[\ ww'\cdot\eta(w'\beta)^2\cdot(P_{-1}P_1P_{\pm 1})^{-1}\ ].$$

Later, we show heuristically that this variance with w replaced by $w_c-E(w_c|w'\beta)$ can be attained by taking one-step from our estimator. That is, even if (5.1) holds, asymptotically there is not much to lose using our estimator, if Assumption 1 and 2 hold.

As well known, the asymptotic variance matrix of an estimator b_N maximizing an objective function $Q_N(b)\equiv(1/N)\Sigma_i q(\vartheta_i,b)$ where ϑ_i are iid has the form $H^{-1}GH^{-1}$ where

$$G\equiv E(q_b q_b'), \quad q_b\equiv\partial q(\beta)/\partial b, \quad H\equiv -E(q_{bb'}), \quad q_{bb'}\equiv\partial q_b/\partial b'.$$

If H=G as in MLE, then the remaining H^{-1} is often an efficiency bound for some models. One intuitive way to achieve the simplification from $H^{-1}GH^{-1}$ to H^{-1} is maximizing $(1/N)\Sigma_i\omega_i q(\vartheta_i,b)$ where ω_i is a weight to be chosen such that $-E^{-1}(\omega\cdot q_{bb'})=E(\omega^2 q_b q_b')$. The weighting is just "taking one step" from the first estimator, and an elementary example is a weighted least squares estimator.

If we apply this intuition to our estimator, we get the asymptotic variance matrix (with $\omega = \{g_o(w'\beta)\cdot P_{-1}P_1P_{\pm1}\}^{-1}\cdot\eta(w'\beta)$)

$$(5.4) \qquad E^{-1}[\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot\eta(w'\beta)^2\cdot(P_{-1}P_1P_{\pm1})^{-1}\ ];$$

compare this to (5.3).

For the cross-section binary response model (2.8), using (4.6) and replacing z=-1 with z=0, (5.4) becomes (note that $P_{\pm1}=1$ now)

$$(5.5) \qquad E^{-1}[\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot(P_1')^2\cdot(P_0P_1)^{-1}\ ],$$

which is the semiparametric efficiency bound (under independence between the error term and the regressors) derived by Chamberlain (1986) and Cosslett (1987); Klein and Spady (1993) show an estimator attaining this bound. This suggests that (5.4) may be an efficiency bound for some model with (1.1). In fact, it is not difficult to show that (5.5) is attained also by taking one step (weighting) from RCE.

Taking one step from our estimator to get a new estimator with the variance (5.4) may improve efficiency in view of the above discussion, but this will ruin one nice feature of our estimator: since the weight ω depends $g_o(w'\beta)$ and $\eta(\omega'\beta)$, estimating ω requires smoothing, and hence the one-step "improved" estimator will depend on a bandwidth. This was the main reason not to pursue this line of extension seriously; but it will be still interesting to find out whether (5.4) is indeed the semiparametric efficiency bound for (1.1) under Assumption 1 and 2.

6. Extensions to Multiple Waves and Fixed Effect Ordered Response.

We consider two ways to extend (1.1) for more than two periods ("waves"). One is using a minimum distance estimation idea after each adjacent two periods are used. For instance, if t=1,2,3, then β can be estimated by b_{12} using t=1,2 and by b_{23} using t=2,3. Since the population parameters for b_{12} and b_{13} should be equal, this leads to a minimum distance estimation; see Lee (1995) to see how these can be implemented in practice. The other way is to follow Charlier et al (1995): maximize

$$(6.1)\qquad \Sigma_{s<t}\Sigma_{i\neq j}\mathrm{sgn}(\Delta x_{ist}'b-\Delta x_{jst}'b)\cdot(\Delta y_{ist}-\Delta y_{jst})\cdot\Delta y_{ist}^{2}\Delta y_{jst}^{2}$$

where s runs 1 to T (the last period), t runs 2 to T and

$$(6.2)\qquad \Delta x_{ist} \equiv x_{it}-x_{is}, \quad \text{and } \Delta y_{ist} \equiv y_{it}-y_{is}.$$

If the panel is unbalanced, then we can attach a product of dummy variables, say $d_{ist}d_{jst}$, to the summand in (6.1), where d_{ist}=1 if i is observed at s and t, and 0 otherwise.

We consider two ways to extend our estimator to fixed effect ordered response panel data. One is a minimum distance estimation after the ordered response has been collapsed into binary responses around different thresholds. But this requires getting the influence functions that depend on smoothing parameters, and the estimate itself will depend on smoothing parameters. The other, explained in the following, avoids this problem and is simple in deriving its asymptotic variance.

Suppose that y_{it} takes three values 0, 1, and 2, and y_{it}=2 if $y_{it}^{*}>\gamma>0$, where γ is a (un)known threshold. There are two kinds of adjacent moves in y_{it}:

$$0 \text{ to } 1, \text{ or } 1 \text{ to } 0 \quad \text{and} \quad 1 \text{ to } 2, \text{ or } 2 \text{ to } 1.$$

Define indicator functions for these two as d_{01i} and d_{12i} respectively. Maximize

$$(6.3) \qquad \{N(N-1)\}^{-1}\Sigma_{i\neq j}\text{sgn}(\Delta x_i'b-\Delta x_j'b)\cdot(\Delta y_i-\Delta y_j)\cdot(d_{01i}d_{01j}+d_{12i}d_{12j})$$

wrt b. If there are M categories, then this estimator can be extended straightforwardly by considering all adjacent moves (0 to 1, 1 to 2, ... M-2 to M-1); namely, there will be M-1 terms in the maximand.

The estimator for (6.3) is $\sqrt{N}$-consistent and asymptotically normal. Due to no overlap between the moves from 0 to 1 and 1 to 2, the variance matrix is simple: it is $4\cdot E^{-1}\nabla_2\tau\cdot E(\nabla_1\tau\nabla_1\tau')\cdot E^{-1}\nabla_2\tau$, where

$$E\nabla_2\tau = -8\cdot E[\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot g_o(w'\beta)\cdot\eta(w'\beta)\ ],$$

$$E(\nabla_1\tau\nabla_1\tau') = 16\cdot E(\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot g_o(w'\beta)^2$$
$$\cdot[P(1 \text{ to } 0|w'\beta)\cdot P(0 \text{ to } 1|w'\beta)\cdot\{P(1 \text{ to } 0|w'\beta)+P(0 \text{ to } 1|w'\beta)\}$$
$$+ P(2 \text{ to } 1|w'\beta)\cdot P(1 \text{ to } 2|w'\beta)\cdot\{P(2 \text{ to } 1|w'\beta)+P(1 \text{ to } 2|w'\beta)\}]\ ),$$

$$\eta(w'\beta) = P'(0 \text{ to } 1|w'\beta)\cdot P(1 \text{ to } 0|w'\beta) - P'(1 \text{ to } 0|w'\beta)\cdot P(0 \text{ to } 1|w'\beta)$$
$$+ P'(1 \text{ to } 2|w'\beta)\cdot P(2 \text{ to } 1|w'\beta) - P'(2 \text{ to } 1|w'\beta)\cdot P(1 \text{ to } 2|w'\beta).$$

It may be possible to design a more efficient estimator because two types of information are not used in (6.3). One is non-adjacent moves such as 0 to 2. The other is the information on the thresholds. For instance, when the thresholds are equally spaced in the M category case, $\Delta x'\beta$ for a move 0 to M-1 should be larger than that for a move 0 to 1; Lee (1992) considers other intermediate possibilities for thresholds. It seems difficult to accommodate these pieces of information in our framework which is critically built on binary response.

7 Conclusions.

We proposed a semiparametric $\sqrt{N}$-consistent estimator for fixed effect binary response panel data which was then extended to fixed effect ordered response. The estimator does not depend on a smoothing parameter and is asymptotically normal with a simple variance matrix. The estimator, however, imposes restrictions on the relationship between the level of regression function and its increment, and on the relationship between the individual-specific term and regressors.

Appendix

Proof of Lemma 1 and Lemma 2.

(x_1', x_2') is one-to-one to $(x_{1c}', x_1'\beta, x_{2c}', x_2'\beta)$ which is in turn one-to-one to $(x_{1c}', x_1'\beta, \Delta x_c', \Delta x'\beta)$ where the index i is omitted,

$$x_{tc} = (x_{t1} \ldots x_{t,k-1})', \quad t=1,2 \qquad \text{and} \qquad \Delta x_c = x_{2c} - x_{1c}.$$

Observe

(a.1) $P(\Delta y=1|x,\alpha) = P(u_1<-x_1'\beta-\alpha, u_2>-x_1'\beta-\Delta x'\beta-\alpha|x_{1c}', x_1'\beta, \Delta x_c', \Delta x'\beta, \alpha)$.

Under the independence of $(u_1,u_2)|\alpha$ from $(x_1',x_2')|\alpha$, this becomes

(a.2) $$P(u_1<-x_1'\beta-\alpha,\ u_2>-x_1'\beta-\Delta x'\beta-\alpha\ |x_1'\beta, \Delta x'\beta, \alpha).$$

Integrate (a.2) wrt $(x_1'\beta,\alpha)|(\Delta x_c, \Delta x'\beta)$ and invoke Assumption 1(f) to get Lemma 1.

Turning to Lemma 2, on $C\equiv\{\Delta x_i, \Delta x_j, \Delta y_i\neq 0, \Delta y_j\neq 0, \Delta y_i\neq\Delta y_j\}$, Δy_i can take only ±1 and $\Delta y_i-\Delta y_j$ can take only ±2. We need to prove

(a.3) $$P(\Delta y_i>\Delta y_j|C) >(<)\ P(\Delta y_i<\Delta y_j|C) \quad \text{iff} \quad \Delta x_i'\beta >(<)\ \Delta x_j'\beta.$$

$P(\Delta y_i>\Delta y_j|C)$ is

(a.4) $$P(\Delta y_i=1, \Delta y_j=-1\ |\Delta x_i, \Delta x_j, \Delta y_i\neq 0, \Delta y_j\neq 0, \Delta y_i\neq\Delta y_j)$$

$$= P(\Delta y_i=1, \Delta y_j=-1, \Delta y_i\neq\Delta y_j\ |\Delta x_i, \Delta x_j, \Delta y_i\neq 0, \Delta y_j\neq 0)/\pi$$

$$= P(\Delta y_i=1, \Delta y_j=-1\ |\Delta x_i, \Delta x_j, \Delta y_i\neq 0, \Delta y_j\neq 0)/\pi$$

(a.5) $$= P(\Delta y_i=1|\Delta x_i, \Delta y_i\neq 0)\cdot P(\Delta y_j=-1|\Delta x_j, \Delta y_j\neq 0)/\pi$$

where

$$\pi = P(\Delta y_i\neq\Delta y_j\ |\Delta x_i, \Delta x_j, \Delta y_i\neq 0, \Delta y_j\neq 0).$$

Likewise,

(a.6) $$P(\Delta y_i<\Delta y_j|C) = P(\Delta y_i=-1|\Delta x_i, \Delta y_i\neq 0)\cdot P(\Delta y_j=1|\Delta x_j, \Delta y_j\neq 0)/\pi.$$

Since π appears in both (a.5) and (a.6), it can be ignored. Also instead of comparing (a.5) and (a.6), we can compare

(a.7) $\quad P(\Delta y_i=1|\Delta x_i)\cdot P(\Delta y_j=-1|\Delta x_j)$ vs. $P(\Delta y_i=-1|\Delta x_i)\cdot P(\Delta y_j=1|\Delta x_j)$

due to $P(\Delta y_i=1|\Delta x_i,\Delta y_i\neq 0) = P(\Delta y_i=1|\Delta x_i)/P(\Delta y_i\neq 0|\Delta x_i)$ and the appearance of $P(\Delta y_i\neq 0|\Delta x_i)$ in both terms in (a.7). Furthermore, due to Lemma 1, we can compare as well

(a.8) $P(\Delta y_i=1|\Delta x_i'\beta)\cdot P(\Delta y_j=-1|\Delta x_j'\beta)$ vs. $P(\Delta y_i=-1|\Delta x_i'\beta)\cdot P(\Delta y_j=1|\Delta x_j'\beta)$.

The two terms in (a.8) are equal whenever $\Delta x_i'\beta=\Delta x_j'\beta$. In (a.2), $P(\Delta y=1\ |x_1'\beta,\Delta x'\beta,\alpha)$ is increasing in $\Delta x'\beta$ for all given $x_1'\beta$ and α. Either combining this with Assumption 1(e)' or simply invoking Assumption 1(e), $P(\Delta y=1|\Delta x'\beta)$ is increasing in $\Delta x'\beta$. Likewise,

(a.9) $\quad P(\Delta y=-1|\Delta x'\beta) = P(u_1>-x_1'\beta-\alpha,\ u_2<-x_1'\beta-\Delta x'\beta-\alpha\ |\Delta x'\beta)$

which is decreasing in $\Delta x'\beta$. This proves Lemma 2.

Proof for Theorem 2.

Using $w_k=w'\beta-w_c'\beta_c$, rewrite $\tau(z_j,w_j,b_c)$ as

(b.1) $$E_{w|j}[\ 1[w_c'(b_c-\beta_c)+w'\beta>w_{jc}'(b_c-\beta_c)+w_j'\beta]\cdot E(z_j^2z-z_jz^2|w,z_j)$$
$$-1[w_c'(b_c-\beta_c)+w'\beta<w_{jc}'(b_c-\beta_c)+w_j'\beta]\cdot E(z_j^2z-z_jz^2|w,z_j)\ ];$$

note that $(z-z_j)\cdot z^2z_j^2 = z^3z_j^2-z_j^3z^2 = z_j^2z-z_jz^2$. Recalling that $g(\cdot|\cdot)$ is the density of $w'\beta|w_c$, $\tau(z_j,w_j,b_c)$ is

(b.2) $$E_{w_c|j}[\ \int_{(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta}^{\infty} E(z_j^2z-z_jz^2|w'\beta,z_j)\cdot g(w'\beta|w_c)d(w'\beta)$$
$$-\int_{-\infty}^{(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta} E(z_j^2z-z_jz^2|w'\beta,z_j)\cdot g(w'\beta|w_c)d(w'\beta)\ ]$$

where $\cdot|j$ means that z_j and w_j are fixed. Differentiating this wrt b_c,

(b.3) $$\nabla_1\tau(z_j,w_j,b_c) = 2\cdot E_{w_c|j}[\ E\{z_j^2z-z_jz^2|(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta,z_j\}$$

$$\cdot g\{(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta)|w_c\}\cdot(w_c-w_{jc})\ ].$$

With $b_c=\beta_c$,

$$\nabla_1\tau(z_j,w_j,\beta_c) = 2\cdot E_{w_c|j}[E(z_j^2z-z_jz^2|w'\beta=w_j'\beta,z_j)\cdot g(w_j'\beta|w_c)(w_c-w_{jc})]$$

(b.4) $$= 2\cdot E_{w_c|j}[\ \{E(z|w_j'\beta)z_j^2-z_jE(z^2|w_j'\beta)\}\cdot g(w_j'\beta|w_c)w_c\ ]$$

(b.5) $$-2\cdot E_{w_c|j}[\ \{E(z|w_j'\beta)z_j^2-z_jE(z^2|w_j'\beta)\}\cdot g(w_j'\beta|w_c)w_{jc}\ ].$$

The term $w_j'\beta$ in $E(\cdot)$ and $g(\cdot)$ takes the place of $w'\beta$ and thus should be regarded as $w'\beta=w_j'\beta$ here and below.

Due to $E_{w_c|j}g(w_j'\beta|w_c) = g_o(w_j'\beta)$, rewrite (b.5) as

(b.6) $$-2\cdot\{E(z|w_j'\beta)z_j^2-z_jE(z^2|w_j'\beta)\}\cdot g_o(w_j'\beta)\cdot w_{jc}.$$

Examine (b.4). Multiplying and dividing by $g_o(w_j'\beta)$, the first term is

$$= 2z_j^2g_o(w_j'\beta)\cdot E(z|w_j'\beta)\cdot\int w_cg(w_j'\beta|w_c)dF(w_c)/g_o(w_j'\beta)$$

$$= 2z_j^2g_o(w_j'\beta)\cdot E(z|w_j'\beta)\int w_cdF(w_c|w_j'\beta) = 2z_j^2g_o(w_j'\beta)E(z|w_j'\beta)E(w_c|w_j'\beta)$$

where $F(w_c)$ and $F(w_c|w'\beta)$ denote respectively the distribution of w_c and $w_c|w'\beta$. Doing analogously for the second term of (b.4), (b.4) becomes

(b.7) $$2\cdot\{E(z|w_j'\beta)z_j^2-z_jE(z^2|w_j'\beta)\}\cdot g_o(w_j'\beta)\cdot E(w_c|w_j'\beta).$$

Thus, putting (b.6) to (b.7) together,

(b.8) $$\nabla_1\tau(z_j,w_j,\beta_c) = -2\cdot\{w_{jc}-E(w_{jc}|w_j'\beta)\}\cdot g_o(w_j'\beta)$$
$$\cdot\{z_j^2E(z|w_j'\beta) - z_jE(z^2|w_j'\beta)\}.$$

Note that

(b.9) $$E(z|w'\beta)=P(z=1|w'\beta)-P(z=-1|w'\beta) \quad\text{and}\quad E(z^2|w'\beta)=P(z=\pm1|w'\beta).$$

From $\nabla_1\tau$, (now drop the subscript j and denote $\{E(\lambda)\}^2$ as $E^2(\lambda)$)

$$E(\nabla_1\tau\nabla_1\tau') = E[\ 4\{g_o(w'\beta)\}^2\cdot\{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'$$

$$\cdot\{z^2E^2(z|w'\beta) + z^2E^2(z^2|w'\beta) - 2zE(z|w'\beta)E(z^2|w'\beta)\}\ ]$$

for z takes only 0 and ±1. Define W, a function of w_c and $w'\beta$, as

$$W \equiv 4\{g_o(w'\beta)\}^2\{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'.$$

Then

$$E(\nabla_1\tau\nabla_1\tau'|w'\beta) = E(z^2W|w'\beta)\cdot E^2(z|w'\beta) + E(z^2W|w'\beta)\cdot E^2(z^2|w'\beta)$$

$$-2E(zW|w'\beta)\cdot E(z|w'\beta)\cdot E(z^2|w'\beta)$$

$$= E(W|w'\beta)\cdot E(z^2|w'\beta)\cdot E^2(z|w'\beta) + E(W|w'\beta)\cdot E^3(z^2|w'\beta)$$

$$- 2E(W|w'\beta)\cdot E(z^2|w'\beta)\cdot E^2(z|w'\beta) \qquad \text{due to Lemma 1}$$

$$= E(W|w'\beta)\cdot\{E^3(z^2|w'\beta) - E(z^2|w'\beta)\cdot E^2(z|w'\beta)\}.$$

The term $\{\cdots\}$ is (recall (b.9))

$$(P_1+P_{-1})^3 - (P_1+P_{-1})\cdot(P_1-P_{-1})^2 = 4(P_1+P_{-1})P_1P_{-1}.$$

Hence,

$$E(\nabla_1\tau\nabla_1\tau') = E\{\ E(\nabla_1\tau\nabla_1\tau'|w'\beta)\ \} = E\{\ E(W|w'\beta)\cdot 4(P_1+P_{-1})P_1P_{-1}\}$$

$$= E[\ E\{W\cdot 4(P_1+P_{-1})P_1P_{-1}|w'\beta\}\ ] = E[\ W\cdot 4(P_1+P_{-1})P_1P_{-1}\ ]$$

(b.10) $$= 16\cdot E[\ \{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\cdot\{g_o(w'\beta)\}^2$$

$$\cdot P(z=-1|w'\beta)P(z=1|w'\beta)\cdot\{P(z=-1|w'\beta)+P(z=1|w'\beta)\}\ ].$$

Turning to $E\nabla_2\tau$, observe that

$\nabla_2\tau(z_j,w_j,b_c) =$

$$-2\cdot E_{w_c|j}[\ \partial E\{z_j{}^2z-z_jz^2|\ w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta,z_j\}/\partial(w'\beta)$$

$$\cdot g\{(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta|w_c\}\cdot(w_c-w_{jc})(w_c-w_{jc})'$$

$$+ E\{z_j{}^2z-z_jz^2|\ w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta,z_j\}$$

$$\cdot\partial g\{w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta|w_c\}/\partial(w'\beta)\cdot(w_c-w_{jc})(w_c-w_{jc})'\ ].$$

With $b_c=\beta_c$,

(b.11) $\nabla_2\tau(z_j, w_j, \beta_c) =$

$$-2E_{w_c|j}[\ \partial E(z_j^2 z - z_j z^2 | w'\beta = w_j'\beta, z_j)/\partial(w'\beta)\cdot g(w_j'\beta|w_c)\cdot(w_c - w_{jc})(w_c - w_{jc})'$$

$$+\ E(z_j^2 z - z_j z^2 | w_j'\beta, z_j)\cdot\partial g(w'\beta = w_j'\beta|w_c)/\partial(w'\beta)\cdot(w_c - w_{jc})(w_c - w_{jc})'\ ].$$

The second term can be ignored as far as $E\nabla_2\tau$ is concerned for it becomes zero due to $E(z_j^2 z - z_j z^2 | w_j'\beta, z_j)$ and Lemma 1. As for the first term of (b.11), it is

(b.12) $$-2\cdot\partial E(z_j^2 z - z_j z^2 | w'\beta = w_j'\beta, z_j)/\partial(w'\beta)$$

$$\cdot E_{w_c|j}\{g(w_j'\beta|w_c)\cdot(w_c - w_{jc})(w_c - w_{jc})'\}.$$

Taking $E_{z_j|w_j'\beta}$ on (b.12), the second term that depends only on $w_j'\beta$ can be taken out of the expectation. As for the first term, we get

(b.13) $$-2\cdot E(z_j^2|w_j'\beta)\cdot\partial E(z|w'\beta = w_j'\beta)/\partial(w'\beta)$$

$$+\ 2\cdot E(z_j|w_j'\beta)\cdot\partial E(z^2|w'\beta = w_j'\beta)/\partial(w'\beta)$$

(b.14) $$= -2\cdot(P_1 + P_{-1})\cdot(P_1' - P_{-1}') + 2\cdot(P_1 - P_{-1})\cdot(P_1' + P_{-1}')$$

$$= -2P_1P_1' + 2P_1P_{-1}' - 2P_{-1}P_1' + 2P_{-1}P_{-1}' + 2P_1P_1' - 2P_{-1}P_1' + 2P_1P_{-1}' - 2P_{-1}P_{-1}'$$

$$= -4\cdot(P_{-1}P_1' - P_1P_{-1}') = -4\eta(w_j'\beta).$$

Hence

(b.15) $$E_{z_j|w_j'\beta}\{(b.12)\} = -4\eta(w_j'\beta)\cdot E_{w_c|j}\{g(w_j'\beta|w_c)\cdot(w_c - w_{jc})(w_c - w_{jc})'\}.$$

Rewrite $E_{w_c|j}\{\ \cdot\}$ in (b.15) as

(b.16) $$E_{w_c|w_j'\beta}\{g(w_j'\beta|w_c)\cdot(w_cw_c' - w_{jc}w_c' - w_cw_{jc}' + w_{jc}w_{jc}')\}$$

which has four terms. The first term of (b.16) is

(b.17) $$\int w_cw_c'g(w_j'\beta|w_c)\cdot F(dw_c)$$

$$= g_o(w_j'\beta)\cdot\int w_cw_c'g(w_j'\beta|w_c)\cdot F(dw_c)/g_o(w_j'\beta)$$

$$= g_o(w_j'\beta)\cdot\int w_c w_c' F(dw_c|w_j'\beta) = g_o(w_j'\beta)\cdot E(w_c w_c'|w_j'\beta).$$

Analogously, the last three terms in (b.16) are respectively

$$\text{(b.18)} \qquad -g_o(w_j'\beta)\cdot w_{jc}\cdot E(w_c'|w_j'\beta), \qquad -g_o(w_j'\beta)\cdot E(w_c|w_j'\beta)w_{jc}',$$

$$g_o(w_j'\beta)\cdot w_{jc}w_{jc}'.$$

Insert (b.17) and (b.18) into (b.15) and take $E_{w_j'\beta}$ on (b.15) to get

$$\text{(b.19)} \quad E\nabla_2\tau = -8\cdot E[\ \eta(w'\beta)g_o(w'\beta)\cdot\{E(w_c w_c'|w'\beta)-E(w_c|w'\beta)E(w_c'|w'\beta)\}\ ];$$

note that 4 in (b.15) is replaced by 8 at this step.

As the final step, we show that (b.19) is the same as $E\nabla_2\tau$ in Theorem 2 which is

$$\text{(b.20)} \qquad -8\cdot E[\ \eta(w'\beta)g_o(w'\beta)\cdot\{w_c-E(w_c|w'\beta)\}\{w_c-E(w_c|w'\beta)\}'\ ]$$

$$=-8E[\eta(w'\beta)g_o(w'\beta)\{w_c w_c'-E(w_c|w'\beta)w_c'-w_c E(w_c'|w'\beta)+E(w_c|w'\beta)E(w_c'|w'\beta)]$$

$$= -8\cdot E[\ \eta(w'\beta)g_o(w'\beta)\cdot\{E(w_c w_c'|w'\beta)-E(w_c|w'\beta)E(w_c'|w'\beta)\}\ ].\blacksquare$$

Proof for the sufficiency of Assumption 3 for Assumption 2.

Assumption 2(d) was already proven. 2(c) is implied by 3(a). For 2(a), it is enough for the following to be bounded for any $b_c \in N_\beta$:

$$\text{(c.1)} \qquad \partial P\{z=1|\ w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta\}/\partial(w'\beta).$$

Before we proceed, recall (4.12) implying the followings. First, $g(w'\beta|w_c)$, $g_o(w'\beta)$ and $f(w'\beta|z)$ are bounded by M_0 when h is bounded by M_0; the same can be said for the first derivatives $\partial h/\partial w_k$, $g'\equiv\partial g/\partial(w'\beta)$, $g_o'\equiv\partial g_o/\partial(w'\beta)$, $f'\equiv\partial f/\partial(w'\beta)$ and the bound M_1 for $\partial h/\partial w_k$. Second, $g(w'\beta|w_c)$, $g_o(w'\beta)$ and $f(w'\beta|z)$ are Lipschitz continuous wrt $w'\beta$ with the Lipschitz constant M_1, when $h(w_k|w_c,z)$ is so wrt w_k. Third, since

$$\text{(c.2)} \quad g(w'\beta|w_c) = \int\gamma(w'\beta|w_c,z)F(dz), \quad g_o(w'\beta) = \int\gamma(w'\beta|w_c,z)F(dw_c,dz),$$

reversing the order of integration and differentiation, γ, g and g_o are twice differentiable wrt $w'\beta$. Fourth, because $\partial h(w_k|w_c,z)/\partial w_k$ is Lipschitz-continuous wrt w_k with the Lipschitz-constant M_2, g', f' and g_o' are so as well wrt $w'\beta$.

Take log on (4.11) and differentiate wrt $w'\beta$ to get

$$\text{(c.3)} \qquad P'(z|w'\beta) = P(z|w'\beta) \cdot\{ f'(w'\beta|z)/f(w'\beta|z) - g_o'(w'\beta)/g_o(w'\beta) \}.$$

Under 3(b),

$$|f'(w'\beta|z)| \le \int|\partial\gamma(w'\beta|w_c,z)/\partial(w'\beta)|F(dw_c) \le \zeta_1\cdot\int h(w_k|w_c,z)F(dw_c)$$

$$\Rightarrow \quad |f'(w'\beta|z)|/f(w'\beta|z) \le \zeta_1 \quad \text{for all z.}$$

Also

$$|g_o'(w'\beta)| \le \Sigma_{q=-1}^{1}|f'(w'\beta|z=q)|P(z=q) \le \zeta_1 g_o(w'\beta)$$

$$\Rightarrow \quad |g_o'(w'\beta)|/g_o(w'\beta) \le \zeta_1.$$

Therefore

$$\text{(c.4)} \qquad |P'(z|w'\beta)| \le 2\zeta_1 \text{ for a.e. } w'\beta,$$

proving that (c.1) is bounded for any $b_c \in N_\beta$.

Before turning to 2(b), we establish a Lipschitz continuity of $P'(z|w'\beta)$ wrt $w'\beta$. Differentiate $P'(z|w'\beta)$ in (c.3) wrt $w'\beta$ to get

$$P'(z|w'\beta)\cdot\{f'(w'\beta|z)/f(w'\beta|z) - g_o'(w'\beta)/g_o(w'\beta)\}$$

$$+ P(z|w'\beta)\cdot\partial\{ f'(w'\beta|z)/f(w'\beta|z) - g_o'(w'\beta)/g_o(w'\beta) \}/\partial(w'\beta)$$

$$\le 4\zeta_1^{\,2} + 2\zeta_1\cdot|\partial\{ f'(w'\beta|z)/f(w'\beta|z) - g_o'(w'\beta)/g_o(w'\beta) \}/\partial(w'\beta)|.$$

Focus on the last term, which is (use " to denote second derivatives)

$$f''(w'\beta|z)/f(w'\beta|z) - \{f'(w'\beta|z)/f(w'\beta|z)\}^2$$

$$- [\; g_o''(w'\beta)/g_o(w'\beta) - \{g_o'(w'\beta)/g_o(w'\beta)\}^2 \;].$$

Under 3(c),

$$|f''(w'\beta|z)| \le \int|\partial^2\gamma(w'\beta|w_c,z)/\partial(w'\beta)^2|F(dw_c) \le \zeta_2\cdot\int h(w_k|w_c,z)F(dw_c)$$

$$\Rightarrow \quad |f''(w'\beta|z)|/f(w'\beta|z) \le \zeta_2 \quad \text{for all } z.$$

Also

$$|g_o''(w'\beta)| \le \Sigma_{q=-1}^{1} |f''(w'\beta|z=q)|P(z=q) \le \zeta_2 g_o(w'\beta).$$

$$\Rightarrow \quad |g_o''(w'\beta)|/g_o(w'\beta) \le \zeta_2.$$

Therefore

$$\text{(c.5)} \qquad |P''(z|w'\beta)| \le 4\zeta_1^2 + 4\zeta_1(\zeta_2+\zeta_1^2) \qquad \text{for a.e. } w'\beta.$$

This implies that $P'(z|w'\beta)$ is Lipschitz continuous wrt $w'\beta$.

Turning to 2(b), rewrite $\nabla_2\tau(z_j,w_j,b_c)-\nabla_2\tau(z_j,w_j,\beta_c)$ as

$$\text{(c.6)} \quad -2\cdot E_{w_c|j}[\ \partial E\{z^2z_j-z_j^2z|\ w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta,z_j\}/\partial(w'\beta)$$
$$\cdot g\{(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta)|w_c\}\cdot(w_c-w_{jc})(w_c-w_{jc})'\ ]$$

$$+2\cdot E_{w_c|j}[\ \partial E\{z^2z_j-z_j^2z|\ w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta,z_j\}/\partial(w'\beta)$$
$$\cdot g\{w_j'\beta|w_c\}\cdot(w_c-w_{jc})(w_c-w_{jc})'\ ]$$

$$\text{(c.7)} \quad -2\cdot E_{w_c|j}[\ \partial E\{z^2z_j-z_j^2z|\ w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta,z_j\}/\partial(w'\beta)$$
$$\cdot g\{w_j'\beta|w_c\}\cdot(w_c-w_{jc})(w_c-w_{jc})'\ ]$$

$$+2\cdot E_{w_c|j}[\ \partial E\{z^2z_j-z_j^2z|\ w'\beta=w_j'\beta,z_j\}/\partial(w'\beta)$$
$$\cdot g\{w_j'\beta|w_c\}\cdot(w_c-w_{jc})(w_c-w_{jc})'\ ]$$

$$\text{(c.8)} \quad +\ 2\cdot E_{w_c|j}[\ E\{z^2z_j-z_j^2z|\ w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta,z_j\}$$
$$\cdot\partial g\{w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta)|w_c\}/\partial(w'\beta)\cdot(w_c-w_{jc})(w_c-w_{jc})'\ ]$$

$$-\ 2\cdot E_{w_c|j}[\ E\{z^2z_j-z_j^2z|\ w'\beta=w_j'\beta,z_j\}$$
$$\cdot\partial g\{w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta)|w_c\}/\partial(w'\beta)\cdot(w_c-w_{jc})(w_c-w_{jc})'\ ]$$

$$\text{(c.9)} \quad +\ 2\cdot E_{w_c|j}[\ E\{z^2z_j-z_j^2z|\ w'\beta=w_j'\beta,z_j\}$$
$$\cdot\partial g\{w'\beta=(w_{jc}-w_c)'(b_c-\beta_c)+w_j'\beta)|w_c\}/\partial(w'\beta)\cdot(w_c-w_{jc})(w_c-w_{jc})'\ ]$$

$$- 2\cdot E_{w_c|j}[\ E\{z^2 z_j - z_j^2 z|\ w'\beta = w_j'\beta, z_j\}$$

$$\cdot \partial g\{w_j'\beta|w_c\}/\partial(w'\beta)\cdot(w_c - w_{jc})(w_c - w_{jc})'\ ].$$

$|(c.6)|$ to $|(c.9)|$ are less than respectively

(c.10) $$12\cdot\zeta_1 M_1\cdot E_{w_c|j}|w_c - w_{jc}|^3\cdot|b_c - \beta_c|,$$

(c.11) $$2\cdot\{4\zeta_1^2 + 4\zeta_1(\zeta_2 + \zeta_1^2)\}\cdot M_0\cdot E_{w_c|j}|w_c - w_{jc}|^3\cdot|b_c - \beta_c|,$$

(c.12) $$4\zeta_1 M_1\cdot E_{w_c|j}|w_c - w_{jc}|^3\cdot|b_c - \beta_c|,$$

(c.13) $$6\cdot M_2\cdot E_{w_c|j}|w_c - w_{jc}|^3\cdot|b_c - \beta_c|,$$

which proves 2(b). ■

Proof for (4.8).

Rewrite $\eta = P_1'P_{-1} - P_{-1}'P_1$ as

(d.1) $$\{(P_1'/P_{\pm1})(P_{-1}/P_{\pm1}) - (P_{-1}'/P_{\pm1})(P_1/P_{\pm1})\}\cdot P_{\pm1}^2.$$

Observe that, with $t \equiv w'\beta$,

(d.2) $$\partial(P_1/P_{\pm1})/\partial t = P_1'/P_{\pm1} - (P_1/P_{\pm1})\cdot\{(P_1' + P_{-1}')/P_{\pm1}\},$$

(d.3) $$\partial(P_{-1}/P_{\pm1})/\partial t = P_{-1}'/P_{\pm1} - (P_{-1}/P_{\pm1})\cdot\{(P_1' + P_{-1}')/P_{\pm1}\}.$$

Multiplying (d.2) by $P_{-1}/P_{\pm1}$ and (d.3) by $P_1/P_{\pm1}$, and taking the difference, the second terms in (d.2) and (d.3) are eliminated, and

$$\eta = [\{\partial(P_1/P_{\pm1})/\partial t\}\cdot(P_{-1}/P_{\pm1}) - \{\partial(P_{-1}/P_{\pm1})/\partial t\}\cdot(P_1/P_{\pm1})]\cdot P_{\pm1}^2.$$

$$= [\ \{\partial\ln(P_1/P_{\pm1})/\partial t\}\cdot(P_1/P_{\pm1})P_{-1}/P_{\pm1}$$

$$-\{\partial\ln(P_{-1}/P_{\pm1})/\partial t\}\cdot(P_{-1}/P_{\pm1})P_1/P_{\pm1}\ ]\cdot P_{\pm1}^2$$

(d.4) $$= \{\partial\ln(P_1/P_{-1})/\partial t\}\cdot(P_{-1}/P_{\pm1})\cdot(P_1/P_{\pm1})\cdot P_{\pm1}^2.$$ ■

Proof for (5.3).

The score vector is

$$w\cdot\{\ 1[z=-1]\cdot P_{-1|\pm}'/P_{-1|\pm} + 1[z=1]\cdot P_{1|\pm}'/P_{1|\pm}\ \}. \tag{e.1}$$

Observe

$$P_{1|\pm}' = \partial(P_1/P_{\pm1})/\partial b = (P_1'P_{-1}-P_1P_{-1}')/P_{\pm1}^2 \tag{e.2}$$

$$P_{-1|\pm}' = \partial(P_{-1}/P_{\pm1})/\partial b = (P_{-1}'P_1-P_{-1}P_1')/P_{\pm1}^2 = -P_{1|\pm}'.$$

Using this, $P_{-1|\pm}=P_{-1}/P_{\pm1}$ and $P_{1|\pm}=P_1/P_{\pm1}$, the score vector is

$$w\cdot\{\ -1[z=-1]\cdot P_{-1}^{-1} + 1[z=1]\cdot P_1^{-1}\ \}\cdot(P_1'P_{-1}-P_1P_{-1}')/P_{\pm1} \tag{e.3}$$

$$= w\cdot\{\ -1[z=-1]\cdot P_{-1}^{-1} + 1[z=1]\cdot P_1^{-1}\ \}\cdot\eta/P_{\pm1}.$$

Thus, the outer product is

$$ww'\cdot\{\ 1[z=-1]\cdot P_{-1}^{-2} + 1[z=1]\cdot P_1^{-2}\ \}\cdot\eta^2/P_{\pm1}^2. \tag{e.4}$$

Take $E(\cdot|w)$ to get

$$ww'\cdot\{P_{-1}^{-1}+P_1^{-1}\}\cdot\eta^2/P_{\pm1}^2 = ww'\cdot\{P_{\pm1}/(P_{-1}P_1)\}\cdot\eta^2/P_{\pm1}^2 \tag{e.5}$$

$$= ww'\cdot\eta^2\cdot(P_{-1}P_1P_{\pm1})^{-1}.$$

Take $E_w(\cdot)$ to get (5.3).■

References

Arcones, M.A., Z.Chen and E.Giné, 1994, Estimators related to U-processes with applications to multivariate Medians: asymptotic normality, Annals of Statistics 22, 1460-1477.

Andrews, D.W.K., 1995, Nonparametric kernel estimation for semiparametric models, Econometric Theory 11, 560-596.

Chamberlain, G., 1980, Analysis of covariance with qualitative data, Review of Economic Studies 47, 225-238.

Chamberlain, G., 1986, Asymptotic efficiency in semiparametric models with censoring, Journal of Econometrics 32, 189-218.

Chamberlain, G., 1992, Binary response models for panel data: identification and information, Harvard Univ., unpublished paper.

Charlier, E., B.Melenberg and A.van Soest, 1995, A smoothed maximum score estimator for the binary choice panel data model with an application to labor force participation, Statistica Neerlandica 49, 324-343.

Cosslett, S.R., 1987, Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models, Econometrica 55, 559-585.

Han, A.K., 1987, Nonparametric analysis of a generalized regression model, Journal of Econometrics 35, 303-316.

Honoré, B.E., 1992, Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects, Econometrica 60, 533-565.

Honoré, B.E. and J.L.Powell, 1994, Pairwise difference estimators of censored and truncated regression models, Journal of Econometrics 64, 241-278.

Horowitz, J., 1992, A smoothed maximum score estimator for the binary response model, Econometrica 60, 505-531.

Huber, P., 1967, The behavior of maximum likelihood estimates under nonstandard conditions, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA.

Kim, J.K. and D.Pollard, 1990, Cube-root asymptotics, Annals of Statistics 18, 191-219.

Klein, R.W. and R.H.Spady, 1993, An efficient semiparametric estimator for binary response models, Econometrica 61, 387-421.

Kyriazidou, E., 1994, Estimation of a panel data sample selection model, University of Chicago, unpublished paper.

Lee, M.J., 1992, Median regression for ordered discrete response, Journal of Econometrics 51, 59-77.

Lee, M.J., 1995, Semiparametric estimation of simultaneous equations with limited dependent variables: a case study of female labor supply, Journal of Applied Econometrics 10, 187-200.

Lee, M.J., 1996, Nonparametric two-stage estimation of simultaneous equations with limited endogenous regressors, Econometric Theory 12, 305-330.

Manski, C.F., 1985, Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator, Journal of Econometrics 27, 313-333.

Manski, C.F., 1987, Semiparametric analysis of random effects linear models from binary panel data, Econometrica 55, 357-362.

Newey, W.K., and D.McFadden, 1994, Large sample estimation and hypothesis testing, in Handbook of Econometrics IV, North-Holland.

Pakes, A. and D.Pollard, 1989, Simulation and the asymptotics of optimization estimators, Econometrica 57, 1027-1057.

Pollard, D., 1984, Convergence of stochastic processes, Springer.

Sherman, R.P., 1993, The limiting distribution of the maximum rank correlation estimator, Econometrica 61, 123-137.

Sherman, R.P., 1994, Maximal inequalities for degenerate U-processes with applications to optimization estimators, Annals of Statistics 22, 439-459.

No.	Author(s)	Title
9599	H. Bloemen	The Relation between Wealth and Labour Market Transitions: an Empirical Study for the Netherlands
95100	J. Blanc and L. Lenzini	Analysis of Commmunication Systems with Timed Token Protocols using the Power-Series Algorithm
95101	R. Beetsma and L. Bovenberg	The Interaction of Fiscal and Monetary Policy in a Monetary Union: Balancing Credibility and Flexibility
95102	P. de Bijl	Aftermarkets: The Monopoly Case
95103	F. Kumah	Unanticipated Money and the Demand for Foreign Assets - A Rational Expectations Approach
95104	M. Vázquez-Brage, A. van den Nouweland, I. García-Jurado	Owen's Coalitional Value and Aircraft Landing Fees
95105	Y. Kwan and G. Chow	Estimating Economic Effects of the Great Leap Forward and the Cultural Revolution in China
95106	P. Verheyen	The missing Link in Budget Models of Nonprofit Institutions; Two Practical Dutch Applications
95107	J. Miller	Should we Offer the Unemployment Places on Labour Market Programmes With the Intention That They Reject Them?
95108	C. van Raalte and H. Webers	Statial Competition with Intermediated matching
95109	W. Verkooijen, J. Plasmans and H. Daniels	Long-Run Exchange Rate Determination: A Neutral Network Study
95110	E. van der Heijden, J. Nelissen, J. Potters and H. Verbon	Transfers and Reciprocity in Overlapping-Generations Experiments
95111	A. van den Elzen and D. Talman	An Algorithmic approach towards the Tracing Procedure of Harsanyi and Selten
95112	R. Beetsma and L. Bovenberg	Does Monetary Unification Lead to Excessive Debt Accumulation?
95113	H. Keuzenkamp	Keynes and the Logic of Econometric Method
95114	E. Charlier, B. Melenberg and A. van Soest	Estimation of a Censored Regression Panel Data Model Using Conditional Moment Restrictions Efficiently
95115	N. G. Noorderhaven, B. Nooteboom and H. Berger	Exploring Determinants of Perceived Interfirm Dependence in Industrial Supplier Relations

No.	Author(s)	Title
95116	F. Kleibergen and H. Hoek	Bayesian Analysis of ARMA models using Noninformative Priors
95117	J. Lemmen and S. Eijffinger	The Fundamental Determiniants of Financial Integration in the European Union
95118	B. Meijboom and J. Rongen	Clustering, Logistics, and Spatial Economics
95119	A. de Jong, F. de Roon and C. Veld	An Empirical Analysis of the Hedging Effectiveness of Currency Futures
95120	J. Miller	The Effects of Labour Market Policies When There is a Loss of Skill During Unemployment
95121	S. Eijffinger, M. Hoeberichts and E. Schaling	Optimal Conservativeness in the Rogoff (1985) Model: A Graphical and Closed-Form Solution
95122	W. Ploberger and H. Bierens	Asymptotic Power of the Integrated Conditional Moment Test Against Global and Large Local Alternatives
95123	H. Bierens	Nonparametric Cointegration Analysis
95124	H. Bierens and W. Ploberger	Asymptotic Theory of Integrated Conditional Moment Tests
95125	E. van Damme	Equilibrium Selection in Team Games
95126	J. Potters and F. van Winden	Comparative Statics of a Signaling Game: An Experimental Study
9601	U. Gneezy	Probability Judgements in Multi-Stage Problems: Experimental Evidence of Systematic Biases
9602	C. Fernández and M. Steel	On Bayesian Inference under Sampling from Scale Mixtures of Normals
9603	J. Osiewalski and M. Steel	Numerical Tools for the Bayesian Analysis of Stochastic Frontier Models
9604	J. Potters and J. Wit	Bets and Bids: Favorite-Longshot Bias and Winner's Curse
9605	H. Gremmen and J. Potters	Assessing the Efficacy of Gaming in Economics Educating
9606	J. Potters and F. van Winden	The Performance of Professionals and Students in an Experimantal Study of Lobbying
9607	J.Kleijnen, B. Bettonvil and W. van Groenendaal	Validation of Simulation Models: Regression Analysis Revisited
9608	C. Fershtman and N. Gandal	The Effect of the Arab Boycott on Israel: The Automobile Market
9609	H. Uhlig	Bayesian Vector Autoregressions with Stochastic Volatility

No.	Author(s)	Title
9610	G. Hendrikse	Organizational Change and Vested Interests
9611	F. Janssen, R. Heuts and T. de Kok	On the (R,s,Q) Inventory Model when Demand is Modelled as a compound Bernoulli Process
9612	G. Fiestras-Janeiro P. Borm and F. van Megen	Protective Behaviour in Games
9613	F. van Megen, G. Facchini, P. Borm and S. Tijs	Strong Nash Equilibria and the Potential Maximizer
9614	J. Miller	Do Labour Market Programmes Necessarily Crowd out Regular Employment? - A Matching Model Analysis
9615	H. Huizinga	Unemployment Benefits and Redistributive Taxation in the Presence of Labor Quality Externalities
9616	G. Asheim and M. Dufwenberg	Admissibility and Common Knowlegde
9617	C. Fernández, J. Osiewalski and M. Steel	On the Use of Panel Data in Bayesian Stochastic Frontier Models
9618	H. Huizinga	Intrafirm Information Transfers and Wages
9619	B. Melenberg and B. Werker	On the Pricing of Options in Incomplete Markets
9620	F. Kleibergen	Reduced Rank Regression using Generalized Method of Moments Estimators
9621	F. Janssen, R. Heuts and T. de Kok	The Value of Information in an (R,s,Q) Inventory Model
9622	F. Groot, C. Withagen and A. de Zeeuw	Strong Time-Consistency in the Cartel-Versus-Fringe Model
9623	R. Euwals and A. van Soest	Desired and Actual Labour Supply of Unmarried Men and Women in the Netherlands
9624	M. Khanman, M. Perry and P.J. Reny	An Ex-Post Envy-Free and Efficient Allocation Mechanism: Imperfect Information Without Common Priors
9625	C. Eaves, G. van der Laan D. Talman and Z. Yang	Balanced Simplices on Polytopes
9626	H. Bloemen and A. Kalwij	Female Employment and Timing of Births Decisions: a Multiple State Transition Model
9627	P. Bolton	Strategic Experimentation: a Revision
9628	C. Fershtman and Y. Weiss	Social Rewards, Externalities and Stable Preferences

No.	Author(s)	Title
9629	P.M. Kort, G. Feichtinger R.F. Hartl and J.L. Haunschmied	Optimal Enforcement Policies (Crackdowns) on a Drug Market
9630	C. Fershtman and A. de Zeeuw	Tradeable Emission Permits in Oligopoly
9631	A. Cukierman	The Economics of Central Banking
9632	A. Cukierman	Targeting Monetary Aggregates and Inflation in Europe
9633	F. de Roon, C. Veld and J. Wei	A Study on the Efficiency of the Market for Dutch Long Term Call Options
9634	B. van Aarle, L. Bovenberg M. Raith	Money, Fiscal Deficits and Government Debt in a Monetary Union
9635	G. van der Laan, D. Talman and Z. Yang	Existence of and Equilibrium in a Competitive Economy with Indivisibilities and Money
9636	I. Brouwer, J. van der Put C. Veld	Contrarian Investment Strategies in a European Context
9637	M. Berg, F. van der Duyn Schouten and J. Jansen	Optimal Service Policies to Remote Customers with Delay-Limits
9638	F.C. Drost and C.A.J. Klaassen	Efficient Estimation in Semiparametric GARCH Models
9639	A.G. de Kok and F.B.S.L.P. Janssen	Demand Management in Multi-Stage Distribution Chain
9640	D.A. Freund, T.J. Kniesner A.T. LoSasso	How Managed Care Affects Medicaid Utilization A Synthetic Difference-in-Differences Zero-Inflated Count Model
9641	M.J. Lee	Instrumental Variable Estimation For Linear Panel Data Models
9642	W. Härdle, E. Mammen and M. Müller	Testing Parametric versus Semiparametric Modelling in Generalized Linear Models
9643	C. Dustmann, N. Rajah and A. van Soest	Part-Time Work, School Success and School Leaving
9644	S.C.W. Eijffinger and M. Hoeberichts	The Trade Off Between Central Bank Independence and Conservativeness
9645	R. Sarin and P. Wakker	A Single-Stage Approach to Anscombe and Aumann's Expected Utility

No.	Author(s)	Title
9646	J.P. Ziliak and T.J. Kniesner	The Importance of Sample Attrition in Life Cycle Labor Supply Estimation
9647	P.M. Kort	Optimal R&D Investments of the Firm
9648	M.P. Berg	Performance Comparisons for Maintained Items
9649	H. Uhlig and Y. Xu	Effort and the Cycle: Cyclical Implications of Efficiency Wages
9650	M. Slikker and A. van den Nouweland	Communication Situations with a Hierarchical Player Partition
9651	H.L.F. de Groot	The Struggle for Rents in a Schumpeterian Economy
9652	R.M. de Jong and J. Davidson	Consistency of Kernel Estimators of heteroscedastic and Autocorrelated Covariance Matrices
9653	J. Suijs, A. De Waegenaere and P. Borm	Stochastic Cooperative Games in Insurance and Reinsurance
9654	A.N. Banerjee and J.R. Magnus	Testing the Sensitivity of OLS when the Variance Matrix is (Partially) Unknown
9655	A. Kalwij	Estimating the Economic Return to Schooling on the basis of Panel Data
9656	M. Lind and F. van Megen	Order Based Cost Allocation Rules
9657	A. van Soest, P. Fontein and Rob Euwals	Earnings Capacity and Labour Market Participation
9658	C. Fernández and M.F.J. Steel	On Bayesian Modelling of Fat Tails and Skewness
9659	R. Sarin and P. Wakker	Revealed Likelihood and Knightian Uncertainty
9660	J.R. Magnus and J. Durbin	A Classical Problem in Linear Regression or How to Estimate the Mean of a Univariate Normal Distribution with Known Variance
9661	U. Gneezy and J. Potters	An Experiment on Risk Taking and Evaluation Periods
9662	H.J. Bierens	Nonparametric Nonlinear Co-Trending Analysis, with an Application to Interest and Inflation in the U.S.
9663	J.P.C. Blanc	Optimization of Periodic Polling Systems with Non-Preemptive, Time-Limited Service
9664	M.J. Lee	A Root-N Consistent Semiparametric Estimator for Fixed Effect Binary Response Panel Data