Econometric Theory, **17**, 2001, 913–932. Printed in the United States of America.

THE INFORMATION BOUND OF A DYNAMIC PANEL LOGIT MODEL WITH FIXED EFFECTS

JINYONG HAHN
Brown University

In this paper, I calculate the semiparametric information bound in two dynamic panel data logit models with individual specific effects. In such a model without any other regressors, it is well known that the conditional maximum likelihood estimator yields a $\sqrt{n}$-consistent estimator. In the case where the model includes strictly exogenous continuous regressors, Honoré and Kyriazidou (2000, *Econometrica* 68, 839–874) suggest a consistent estimator whose rate of convergence is slower than $\sqrt{n}$. Information bounds calculated in this paper suggest that the conditional maximum likelihood estimator is *not* efficient for models without any other regressor and that $\sqrt{n}$-consistent estimation is infeasible in more general models.

I. INTRODUCTION

Discrete choice models may be the most widely investigated among a general class of nonlinear models. In the panel context, it is often of interest to allow for individual effects to avoid the potential omitted variable bias. When the distribution of individual effects is allowed to depend in an arbitrary way on a set of strictly exogenous regressors (fixed effects approach), the problem becomes a semiparametric problem. Consistent and/or efficient estimation of the parametric component in discrete choice panel models with fixed effects in this semiparametric context is not completely understood yet, although there exist a few well-known identification results. For example, among the class of panel data models where all the regressors are strictly exogenous, Manski (1987) shows that identification is possible if one of the regressors has the entire real line as the support, whereas Chamberlain (1992) has shown that identification is infeasible except in the logit model if the support of the regressors is bounded. Chamberlain (1993) also discusses the condition under which identification fails in the case where the set of regressors includes a lagged dependent variable and only two discrete dependent variables are observed ($T = 2$).

In this paper, I examine whether the insight of Chamberlain (1993) generalizes to the $T = 3$ case by considering panel logit models where the set of ex-

Helpful comments by the co-editor and anonymous referees are acknowledged. Address correspondence to: Jinyong Hahn, Department of Economics, Brown University, Box B, Providence, RI 02912, USA.

© 2001 Cambridge University Press 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

planatory variables includes a one period lagged dependent variable. In Section 2, I consider the case where the lagged dependent variable is the only explanatory variable. It is well known that the conditional maximum likelihood (CMLE) estimator is $\sqrt{n}$-consistent for such a model. See Andersen (1970) and Chamberlain (1985). It is suggested that the conditional maximum likelihood estimator is *not* semiparametrically efficient by showing that the semiparametric asymptotic variance bound is strictly smaller than the asymptotic variance of the CMLE.[1] In Section 3, I consider an extended model whose regressors include time dummies, and I argue that $\sqrt{n}$-consistent estimation is infeasible by showing that the semiparametric information bound is singular. I argue that such result suggests that substantial improvement over the Honoré and Kyriazidou (2000) estimator is unlikely.

2. BASIC MODEL

In this section, I consider a simple dynamic panel logit model with fixed effects and examine whether the CMLE is semiparametrically efficient. Consider the model where

$$\Pr[y_{i1}=1|u_i]=\frac{\exp(u_i)}{1+\exp(u_i)},$$

$$\Pr[y_{i2}=1|u_i,y_{i2}]=\frac{\exp(u_i+\beta y_{i1})}{1+\exp(u_i+\beta y_{i1})},$$

$$\Pr[y_{i3}=1|u_i,y_{i1},y_{i2}]=\frac{\exp(u_i+\beta y_{i2})}{1+\exp(u_i+\beta y_{i2})}, \tag{1}$$

where y_{it} denotes binary dependent variables and u_i denotes fixed effects. I assume that the distribution of u is nonparametrically specified. I argue that the asymptotic variance of the CMLE is strictly larger than the asymptotic variance bound for β. This suggests that the CMLE is *not* efficient in the semiparametric sense. Although it is well known that the CMLE is $\sqrt{n}$-consistent, the finding that potentially there exists an additional source of information is new to my knowledge.

2.1. Conditional Maximum Likelihood Estimation

I first discuss, very briefly, how the CMLE of Andersen (1970) is implemented. It is easy to see that the conditional density of $(y_{i2},y_{i1})=(1,0)$ given $(y_{i3},y_{i1}+y_{i2})=(1,1)$ equals

$$\frac{\exp(\beta)}{1+\exp(\beta)}. \tag{2}$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

Let

$$w_i \equiv \begin{cases} 1 & \text{if} \quad (y_{i2}, y_{i1}) = (1,0) \\ 0 & \text{if} \quad (y_{i2}, y_{i1}) = (0,1), \end{cases} \qquad A_i \equiv 1(y_{i2} + y_{i1} = 1, y_{i3} = 1),$$

where $1(\cdot)$ denotes the standard indicator function. Given independent and identically distributed (i.i.d.) observations (y_{i3}, y_{i2}, y_{i1}) $i = 1, \ldots, n$, the CMLE maximizes

$$\sum_{i=1}^{n} A_i \cdot \left[w_i \log \frac{\exp(\beta)}{1 + \exp(\beta)} + (1 - w_i) \log \frac{1}{1 + \exp(\beta)} \right]$$

$$= \sum_{i=1}^{n} A_i \cdot [w_i \beta - \log(1 + \exp(\beta))].$$

It is straightforward to show that the CMLE is $\sqrt{n}$-consistent and asymptotically normal. Therefore, the information for β is nonzero.

2.2. Main Result

Now, I argue that the CMLE is *not* semiparametrically efficient. Note that the density $\mathcal{L}^B$ of (y_3, y_2, y_1) is equal to

$$\int p^B(y_3, y_2, y_1; v, \beta) f(v)\, dv,$$

where

$$p^B(y_3, y_2, y_1; v, \beta) \equiv \frac{\exp(y_1 \cdot v)}{1 + \exp(v)} \frac{\exp(y_2 \cdot (v + \beta \cdot y_1))}{1 + \exp(v + \beta \cdot y_1)} \frac{\exp(y_3 \cdot (v + \beta \cdot y_2))}{1 + \exp(v + \beta \cdot y_2)}$$

$$= \exp[(y_1 + y_2)(v - \ln(1 + e^{v+\beta}) + \ln(1 + e^{v}))$$

$$+ y_3 v + y_2 \beta y_1 + y_3 \beta y_2 - 3 \ln(1 + e^{v})] \qquad \textbf{(3)}$$

denotes the density of (y_3, y_2, y_1) given $u = v$ and $f(\cdot)$ denotes the density of u. (I omit the i subscript whenever obvious.)

Observe that $(y_3, y_2 + y_1)$ is sufficient for u as can be seen in (3). Such an observation suggests that the nonparametric tangent space takes the form

$$\mathcal{T}^B \equiv \{E[k(u)|y_3, y_2 + y_1] : E[k(u)] = 0\},$$

which is formally established by Lemma 1 in the Appendix. Let

$$\ell^B \equiv \frac{\partial \log \mathcal{L}^B}{\partial \beta} = E\left[\frac{\partial \log p^B(y_3, y_2, y_1; u, \beta)}{\partial \beta} \middle| y_3, y_2, y_1 \right].$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

The efficient score for β is the residual $\tilde{\ell}^B$ when ℓ^B is projected on the nonparametric tangent space $\mathcal{T}^B$:

$$\tilde{\ell}^B \equiv \ell^B - \text{Proj}[\ell^B|\mathcal{T}^B].$$

Here, $\text{Proj}[\cdot|\mathcal{T}]$ denotes the projection (in the L_2 sense) on some arbitrary space $\mathcal{T}$. The asymptotic variance bound for β is then equal to $(E[(\tilde{\ell}^B)^2])^{-1}$.[2]

Observe that $\mathcal{T}^B$ is a subset of the set $\mathcal{T}_0$ of zero-mean square-integrable functions of $(y_3, y_2 + y_1)$. Letting $\mathcal{C}^B$ denote the orthogonal complement of $\mathcal{T}^B$ in $\mathcal{T}_0$, we may write $\text{Proj}[\ell^B|\mathcal{T}_0] = \text{Proj}[\ell^B|\mathcal{T}^B] + \text{Proj}[\ell^B|\mathcal{C}^B]$. Because $\mathcal{T}_0$ is the space of zero-mean square-integrable functions of $(y_3, y_2 + y_1)$, we also have

$$\text{Proj}[\ell^B|\mathcal{T}_0] = E[\ell^B|y_3, y_2 + y_1],$$

from which we obtain

$$E[\ell^B|y_3, y_2 + y_1] = \text{Proj}[\ell^B|\mathcal{T}^B] + \text{Proj}[\ell^B|\mathcal{C}^B] \quad \textbf{(4)}$$

and

$$\tilde{\ell}^B = (\ell^B - E[\ell^B|y_3, y_2 + y_1]) + \text{Proj}[\ell^B|\mathcal{C}^B].$$

Observe that $(\ell^B - \text{Proj}[\ell^B|\mathcal{T}_0])$ and $\text{Proj}[\ell^B|\mathcal{C}^B]$ are orthogonal to each other by construction. It follows that

$$E[(\tilde{\ell}^B)^2] = E[(\ell^B - E[\ell^B|y_3, y_2 + y_1])^2] + E[(\text{Proj}[\ell^B|\mathcal{C}^B])^2].$$

It is not difficult to show that the asymptotic variance of CMLE is equal to $(E[(\ell^B - E[\ell^B|y_3, y_2 + y_1])^2])^{-1}$. If $E[(\text{Proj}[\ell^B|\mathcal{C}^B])^2] > 0$, then the asymptotic variance bound would be strictly smaller than the asymptotic variance of CMLE. Therefore, the CMLE would be semiparametrically efficient if and only if

$$E[(\text{Proj}[\ell^B|\mathcal{C}^B])^2] = 0. \quad \textbf{(5)}$$

Comparing (4) and (5), we can conclude that CMLE is semiparametrically efficient if and only if

$$E[\ell^B|y_3, y_2 + y_1] = \text{Proj}[\ell^B|\mathcal{T}^B].$$

Obviously, this would happen if and only if there exists some $K(u)$ with $E[K(u)] = 0$ such that

$$E[\ell^B|y_3, y_2 + y_1] = E[K(u)|y_3, y_2 + y_1].$$

I establish that it is *not* the case.

THEOREM 1. *There does not exist any* $K(u)$ *with* $E[K(u)] = 0$ *such that*

$$E[\ell^B|y_3, y_2 + y_1] = E[K(u)|y_3, y_2 + y_1] \quad \textit{a.e.}$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

Although Theorem 1 seems to suggest that the CMLE is semiparametrically inefficient, it is not yet clear whether there indeed exists an additional source of information. I do not yet know how to construct an estimator with asymptotic variance smaller than that of CMLE. It could simply be due to my inability, or it could be due to a more fundamental problem. Ritov and Bickel (1990) found that there are cases where semiparametric efficiency bounds are totally misleading. It is not yet clear whether the model (1) belongs to such pathological cases.

2.3. Proof of Theorem 1

I will tentatively suppose that there exists $K(u)$ such that

$$E[K(u)] = 0, \qquad E[K(u)|y_3, y_2 + y_1] = E[\ell^B|y_3, y_2 + y_1] \quad \forall (y_3, y_2 + y_1) \qquad \textbf{(6)}$$

and then obtain a contradiction.

2.3.1. Intuition. Write $z \equiv \exp(u)$ and $b \equiv \exp(\beta)$ for convenience. Using Bayes' theorem, it can be established that the condition $E[K(u)|y_3, y_2 + y_1] = E[\ell^B|y_3, y_2 + y_1]$ is equivalent to

$$E_u\left[K(u)\cdot \frac{b^{y_2+y_3} z^{y_1+y_2+y_3}}{(1+z)^{3-y_1-y_2}\cdot(1+bz)^{y_1+y_2}}\right] = E_u\left[\frac{\partial p^B(y_3, y_2, y_1; u, \beta)}{\partial \beta}\right], \qquad \textbf{(7)}$$

where $E_u[\cdot]$ denotes the expectation taken with respect to u keeping (y_3, y_2, y_1) fixed.[3] Observe that (6) in fact consists of seven restrictions including six equations from every possible combination of $(y_3, y_2 + y_1)$. It is shown subsequently that the expression $E_u[K(u)\cdot b^{y_2+y_3} z^{y_1+y_2+y_3}/((1+z)^{3-y_1-y_2})\cdot(1+bz)^{y_1+y_2})]$ on the left-hand side (LHS) of (7) can be written as a linear combination of six expectations $E[K(u)z^5/(1+z)^3(1+bz)^2], \ldots, E[K(u)z/(1+z)^3(1+bz)^2]$, and $E[K(u)/(1+z)^3(1+bz)^2]$.

The proof is based on the observation that (6) consists of seven linear restrictions involving only six "parameters" and the intuition that a system of linear equations usually does not have any solution if the number of parameters is smaller than that of the equations.

2.3.2. First Step: Alternative Characterization of (6). I will first provide an alternative characterization of (6). Without loss of generality, we may reparametrize $K(u)$ such that $K(u) \equiv (1+z)^3(1+bz)^2 g(z)$.[4] Equation (6) can now be written as

$$b^2\cdot E[g(z)z^5] + (2b + 3b^2)\cdot E[g(z)z^4] + (1 + 6b + 3b^2)\cdot E[g(z)z^3] + (3 + 6b + b^2)\cdot E[g(z)z^2] + (3 + 2b)\cdot E[g(z)z] + E[g(z)] = 0 \qquad \textbf{(8)}$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

TABLE 1. Alternative characterization of (9)

(0,0)	$b^2 \cdot E[g(z)z^2] + 2b \cdot E[g(z)z] + E[g(z)] = 0$
(0,1)	$b \cdot E[g(z)z^3] + (1+b) \cdot E[g(z)z^2] + E[g(z)z] = -E[bz^2/(1+bz)^2(1+z)^2]$
(0,2)	$E[g(z)z^4] + 2 \cdot E[g(z)z^3] + E[g(z)z^2] = E[z^2/(1+bz)^2(1+z)] - E[2bz^3/(1+bz)^3(1+z)]$
(1,0)	$b^2 \cdot E[g(z)z^3] + 2b \cdot E[g(z)z^2] + E[g(z)z] = 0$
(1,1)	$b \cdot E[g(z)z^4] + (1+b) \cdot E[g(z)z^3] + E[g(z)z^2] = b/(1+b)E[z^2/(1+bz)(1+z)^2] - E[bz^3/(1+bz)^2(1+z)^2]$
(1,2)	$E[g(z)z^5] + 2 \cdot E[g(z)z^4] + E[g(z)z^3] = E[z^3/(1+bz)^2(1+z)] - E[2bz^4/(1+bz)^3(1+z)]$

Note: The first column denotes the six possible combinations of $(y_3, y_1 + y_2)$. The equalities in the second column are obtained by applying the expression (7) to each combination.

and

$$E_u[b^{y_2+y_3} z^{y_1+y_2+y_3} (1+z)^{y_1+y_2} (1+bz)^{2-y_1-y_2} g(u)] = E_u\left[\frac{\partial p^B(y_3, y_2, y_1; u, \beta)}{\partial \beta}\right] \quad \forall (y_3, y_2 + y_1). \tag{9}$$

Table 1 provides a concrete summary of (9) for every $(y_3, y_2 + y_1)$ combination. It is convenient to rewrite equation (8) and the six equalities in matrix notations as

$$\Omega^B \underset{6\times 1}{\varsigma_g} = r^B, \tag{10}$$

where

$$\underset{6\times 1}{\varsigma_g} \equiv (E[g(z)z^5], E[g(z)z^4], \ldots, E[g(z)z], E[g(z)])' \tag{11}$$

and Ω^B and r^B are appropriate 7×6 and 7×1 matrices.[5]

2.3.3. Second Step: Contradiction. I now show that equation (10) cannot hold, thereby obtaining a contradiction. Because Ω^B is a 7×6 matrix, the rows of Ω^B cannot be linearly independent. It can be shown that[6]

$$(1, -1, -2, -b, -1, -1-b, -b^2)\Omega^B = 0,$$

which, along with $\Omega^B \varsigma_g = r^B$, implies that

$$(1, -1, -2, -b, -1, -1-b, -b^2) r^B = 0.$$

But, in fact, we have

$$(1, -1, -2, -b, -1, -1-b, -b^2) r^B = E\left[\frac{b^2 z^3}{(1+bz)^2(1+z)}\right] > 0.$$

We therefore have a contradiction. ■

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

3. TIME DUMMIES

In this section, I consider a model with time dummies:

$$\Pr[y_1 = 1|u] = \frac{\exp(u)}{1+\exp(u)},$$
$$\Pr[y_2 = 1|u, y_1] = \frac{\exp(u+\gamma_2+\beta y_1)}{1+\exp(u+\gamma_2+\beta y_1)},$$
$$\Pr[y_3 = 1|u, y_1, y_2] = \frac{\exp(u+\gamma_3+\beta y_2)}{1+\exp(u+\gamma_3+\beta y_2)}. \tag{12}$$

I show that the partial information bound for the time dependence parameter is equal to zero.

It is not clear whether β is identified or not. I could not find any identification result in the literature. I could not formally derive it, either. Even then, the semiparametric information bound calculation for this case is interesting because of its implication with models with strictly exogenous variables. Honoré and Kyriazidou (2000) recently considered a model of the form

$$\Pr[y_t = 1|u, x_1, \ldots, x_T, y_1, \ldots, y_{t-1}] = \frac{\exp(u+\beta y_{t-1}+\gamma' x_t)}{1+\exp(u+\beta y_{t-1}+\gamma' x_t)},$$

where $(x_1, \ldots, x_T)$ denotes the set of strictly exogenous variables. Based on the intuition that the CMLE is applicable to the set where $\{x_2 = x_3\}$, they developed a consistent estimator for β. Not surprisingly, the rate of convergence of their estimator is slower than $\sqrt{n}$ if the set $\{x_2 = x_3\}$ has probability zero. It is therefore of interest to examine if more information is available from the set $\{x_2 \neq x_3\}$. The result in this section suggests that the information for β is concentrated on the set $\{x_2 = x_3\}$. It is shown subsequently that if the true values of γ_2 and γ_3 are different from each other, the partial information for β is equal to zero. It follows that the partial information for β in the set $\{\gamma' x_2 \neq \gamma' x_3\}$ is zero, and therefore, substantial improvement over the Honoré and Kyriazidou (2000) estimator is unlikely.

3.1. Main Result

Note that the density $\mathcal{L}^D$ of (y_1, y_2, y_3) is equal to

$$\int p(y_1, y_2, y_3; v, \beta, \gamma_2, \gamma_3) \cdot f(v)\, dv,$$

where

$$p(y_1, y_2, y_3; v, \beta, \gamma_2, \gamma_3) \equiv \frac{\exp(y_1 \cdot v)}{1+\exp(v)} \frac{\exp(y_2 \cdot (v+\gamma_2+\beta \cdot y_1))}{1+\exp(v+\gamma_2+\beta \cdot y_1)} \frac{\exp(y_3 \cdot (v+\gamma_3+\beta \cdot y_2))}{1+\exp(v+\gamma_3+\beta \cdot y_2)} \tag{13}$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

and $f(\cdot)$ denotes the density of u. By the same arguments as in the previous section, we obtain

$$\ell_\beta^D \equiv \frac{\partial \log \mathcal{L}^D}{\partial \beta} = E\left[\left.\frac{\partial \log p}{\partial \beta}\right| y_1, y_2, y_3\right],$$

$$\ell_2^D \equiv \frac{\partial \log \mathcal{L}^D}{\partial \gamma_2} = E\left[\left.\frac{\partial \log p}{\partial \gamma_2}\right| y_1, y_2, y_3\right],$$

$$\ell_3^D \equiv \frac{\partial \log \mathcal{L}^D}{\partial \gamma_3} = E\left[\left.\frac{\partial \log p}{\partial \gamma_3}\right| y_1, y_2, y_3\right]$$

and the nonparametric tangent space

$$\mathcal{T}^D \equiv \{E[k(u)|y_1, y_2, y_3] : E[k(u)] = 0\}.$$

The efficient score for β is the residual when $(\ell_\beta^D, \ell_2^D, \ell_3^D)$ is projected on the nonparametric tangent space. Let $\tilde{\ell}_\beta^D, \tilde{\ell}_2^D, \tilde{\ell}_3^D$ denote the residuals when $\ell_\beta^D, \ell_2^D, \ell_3^D$ are projected on the nonparametric tangent space $\mathcal{T}^D$. The partial information for β can then be written as $\min_{\theta_2, \theta_2} E[(\tilde{\ell}_\beta^D + \theta_2 \cdot \tilde{\ell}_2^D + \theta_3 \cdot \tilde{\ell}_3^D)^2]$.[7]

THEOREM 2. *Assume that $\gamma_2 \neq 0, \gamma_3 \neq 0, \gamma_2 \neq \gamma_3$. Then, there exists a distribution of u such that the semiparametric information bound for β is zero.*[8] *Here, (γ_2, γ_3) and the distribution of u refer to the true value of (γ_2, γ_3) and the distribution of u.*

Theorem 2 does not rule out the possibility that there may exist a distribution of u such that the semiparametric information bound is positive, which seems to imply that there may exist a $\sqrt{n}$-consistent estimator for some distribution of u. Without identification of the fixed effects distribution, it seems impossible to make a case-by-case judgment of the particular data generating process faced. For all practical purposes, Theorem 2 implies that there cannot exist a $\sqrt{n}$-consistent estimator. The semiparametric information bound as defined in Bickel, Klaassen, Ritov, and Wellner (1993), e.g., is computed at the true distribution, and therefore, the bound in Theorem 2 refers to the one computed at the true distribution of u. On the other hand, a statement that some estimator is $\sqrt{n}$-consistent would be usually understood as the statement that the estimator converges at the speed of $\sqrt{n}$ for any arbitrary distribution of u. Such semantic convention suggests that the "semiparametric information bound" in this case be defined as the infimum of semiparametric information bounds, where the infimum is taken with respect to the distribution of u.[9] If the semiparametric information bound and the rate of consistency are understood in such practical terms, Theorem 2 implies that there cannot exist a $\sqrt{n}$-consistent estimator.

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

3.2. Proof of Theorem 2

The information bound would be equal to zero if there existed some real numbers θ_2, and θ_3 such that

$$\tilde{\ell}_\beta^D + \theta_2 \cdot \tilde{\ell}_2^D + \theta_3 \cdot \tilde{\ell}_3^D = 0 \quad \text{a.e.} \tag{14}$$

3.2.1. Intuition. Because $\tilde{\ell}_\beta^D + \theta_2 \cdot \tilde{\ell}_2^D + \theta_3 \cdot \tilde{\ell}_3^D$ is the residual of the projection of $\ell_\beta^D + \theta_2 \cdot \ell_2^D + \theta_3 \cdot \ell_3^D$ on $\mathcal{T}^D$, we would have (14) if

$$\ell_\beta^D + \theta_2 \cdot \ell_2^D + \theta_3 \cdot \ell_3^D = E[k(u)|y_1, y_2, y_3]$$

almost everywhere, i.e., for every (y_1, y_2, y_3) combination.[10] Using Bayes' theorem, we may rewrite the preceding equality as

$$E_u\left[\frac{\partial p}{\partial \beta} + \theta_2 \cdot \frac{\partial p}{\partial \gamma_2} + \theta_3 \cdot \frac{\partial p}{\partial \gamma_3}\right] = E_u[k(u) \cdot p(y_1, y_2, y_3; u, \beta, \gamma_2, \gamma_3)]. \tag{15}$$

For convenience, let $g(\cdot)$ be defined such that $k(u) \equiv g(z)(1 + z)(1 + cz) \times (1 + dz)(1 + bcz)(1 + bdz)$, where I write $c \equiv \exp(\gamma_2)$, $d \equiv \exp(\gamma_3)$, $z \equiv \exp(u)$. We can then write (15) as

$$E_u[g(z)(1 + z)(1 + cz)(1 + dz)(1 + bcz)(1 + bdz)p(y_1, y_2, y_3; u, \beta, \gamma_2, \gamma_3)]$$
$$= E_u\left[\frac{\partial p}{\partial \beta} + \theta_2 \cdot \frac{\partial p}{\partial \gamma_2} + \theta_3 \cdot \frac{\partial p}{\partial \gamma_3}\right] \quad \forall (y_1, y_2, y_3). \tag{16}$$

Subsequently it is shown that the LHS of (16) can be written as a linear combination of six expectations $E[K(u)z^5/(1 + z)^3(1 + bz)^2], \ldots, E[K(u)z/(1 + z)^3(1 + bz)^2]$, and $E[K(u)/(1 + z)^3(1 + bz)^2]$. The proof is based on the observation that (16) consists of eight linear equations involving eight "parameters" including θ_2, θ_3, and the six expectations and on the intuition that such system of linear equations usually has a solution.

3.2.2. First Step: Alternative Characterization of (16). With some tedious algebra, it can be shown that (16) holds if and only if the eight equalities summarized in Table 2 hold. It is convenient to rewrite the eight equalities in matrix notation as

$$\Xi^D \varsigma_g = E_u[P_\beta^D] + \theta_2 \cdot E_u[P_2^D] + \theta_3 \cdot E_u[P_3^D], \tag{17}$$

where $\Xi^D, P_\beta^D, P_2^D, P_3^D$ are appropriate 8×6, 8×1, 8×1, 8×1 matrices[11] and ς_g is as defined in (11).

3.2.3. Second Step: Solving for (θ_2, θ_3). I will first solve for row vectors λ satisfying equality

$$\lambda \Xi^D = 0. \tag{18}$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

TABLE 2. Alternative characterization of (16)

(0,0,0)	$cb^2d\cdot E_u[z^2g(z)] + (db + cb)\cdot E_u[zg(z)] + E_u[g(z)]$ $= E_u[\partial p(0,0,0;u,\beta,\gamma_2,\gamma_3)/\partial\beta] + \theta_2\cdot E_u[\partial p(0,0,0;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_2]$ $+ \theta_3\cdot E_u[\partial p(0,0,0;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_3]$
(0,0,1)	$d^2cb^2\cdot E_u[z^3g(z)] + (d^2b + cdb)\cdot E_u[z^2g(z)] + d\cdot E_u[zg(z)]$ $= E_u[\partial p(0,0,1;u,\beta,\gamma_2,\gamma_3)/\partial\beta] + \theta_2\cdot E_u[\partial p(0,0,1;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_2]$ $+ \theta_3\cdot E_u[\partial p(0,0,1;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_3]$
(0,1,0)	$c^2db\cdot E_u[z^3g(z)] + (c^2b + cd)\cdot E_u[z^2g(z)] + c\cdot E_u[zg(z)]$ $= E_u[\partial p(0,1,0;u,\beta,\gamma_2,\gamma_3)/\partial\beta] + \theta_2\cdot E_u[\partial p(0,1,0;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_2]$ $+ \theta_3\cdot E_u[\partial p(0,1,0;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_3]$
(0,1,1)	$c^2d^2b^2\cdot E_u[z^4g(z)] + (dc^2b^2 + cd^2b)\cdot E_u[z^3g(z)] + cdb\cdot E_u[z^2g(z)]$ $= E_u[\partial p(0,1,1;u,\beta,\gamma_2,\gamma_3)/\partial\beta] + \theta_2\cdot E_u[\partial p(0,1,1;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_2]$ $+ \theta_3\cdot E_u[\partial p(0,1,1;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_3]$
(1,0,0)	$cdb\cdot E_u[z^3g(z)] + (db + c)\cdot E_u[z^2g(z)] + E_u[zg(z)]$ $= E_u[\partial p(1,0,0;u,\beta,\gamma_2,\gamma_3)/\partial\beta] + \theta_2\cdot E_u[\partial p(1,0,0;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_2]$ $+ \theta_3\cdot E_u[\partial p(1,0,0;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_3]$
(1,0,1)	$cd^2b\cdot E_u[z^4g(z)] + (d^2b + cd)\cdot E_u[z^3g(z)] + d\cdot E_u[z^2g(z)]$ $= E_u[\partial p(1,0,1;u,\beta,\gamma_2,\gamma_3)/\partial\beta] + \theta_2\cdot E_u[\partial p(1,0,1;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_2]$ $+ \theta_3\cdot E_u[\partial p(1,0,1;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_3]$
(1,1,0)	$c^2db\cdot E_u[z^4g(z)] + (cdb + c^2b)\cdot E_u[z^3g(z)] + cb\cdot E_u[z^2g(z)]$ $= E_u[\partial p(1,1,0;u,\beta,\gamma_2,\gamma_3)/\partial\beta] + \theta_2\cdot E_u[\partial p(1,1,0;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_2]$ $+ \theta_3\cdot E_u[\partial p(1,1,0;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_3]$
(1,1,1)	$c^2d^2b^2\cdot E_u[z^5g(z)] + (d^2cb^2 + dc^2b^2)\cdot E_u[z^4g(z)] + cb^2d\cdot E_u[z^3g(z)]$ $= E_u[\partial p(1,1,1;u,\beta,\gamma_2,\gamma_3)/\partial\beta] + \theta_2\cdot E_u[\partial p(1,1,1;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_2]$ $+ \theta_3\cdot E_u[\partial p(1,1,1;u,\beta,\gamma_2,\gamma_3)/\partial\gamma_3]$

Note: The first column denotes the eight possible combinations of (y_1, y_2, y_3).

Lemma 2 in the Appendix shows that there are two linearly independent row vectors solving (18), say, λ_1^D and λ_2^D. Let $\Lambda^D \equiv ((\lambda_1^D)', (\lambda_2^D)')'$. Premultiplying (17) by Λ^D, we obtain

$$\Lambda^D E[P_\beta^D] + \left[\underbrace{\Lambda^D E[P_2^D]}_{2\times 1}, \underbrace{\Lambda^D E[P_3^D]}_{2\times 1}\right]\begin{bmatrix}\theta_2\\ \theta_3\end{bmatrix} = 0.$$

Lemma 3 in the Appendix shows that there exists a distribution of u with density bounded away from zero on an open interval, say, $\mathcal{O}$, such that the 2×2 matrix $[\Lambda^D E[P_2^D], \Lambda^D E[P_3^D]]$ is nonsingular. Therefore, (17) is satisfied by $(\theta_2, \theta_3) = (\theta_2^*, \theta_3^*)$, where

$$\begin{bmatrix}\theta_2^*\\ \theta_3^*\end{bmatrix} \equiv -[\Lambda^D E[P_2^D], \Lambda^D E[P_3^D]]^{-1}\Lambda^D E[P_\beta^D].$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

It remains to show that (17) is satisfied for some g. Lemma 6 in the Appendix establishes that such g exists if u has a continuous distribution with density bounded away from zero on an open interval. ■

4. SUMMARY

In this paper, I considered two dynamic panel logit models with fixed effects for $T = 3$. It was suggested that CMLE is not semiparametrically efficient when the model does not include any time dummies. It was also shown that the semiparametric information bound is equal to zero when the model includes time dummies. Both results are restrictive because they are concerned only with $T = 3$ cases. I do not yet know how they may generalize to $T > 3$ cases, although the proof methodology employed in this paper suggests that such generalization would involve very technical arguments.

NOTES

1. In the previous version of this paper, I incorrectly stated that the CMLE is semiparametrically efficient.

2. See, e.g., Bickel et al. (1993) for more complete discussion of the semiparametric efficiency bound.

3. $E_u[\cdot]$ is an integral with respect to the *marginal* distribution of u, whereas $E[\cdot|y_3, y_2, y_1]$ is an integral with respect to the *conditional* distribution of u. Evaluation of both sides of the expression (9) becomes substantially easier with the introduction of the notation $E_u[\cdot]$.

4. By writing $K(u) = (1 + z)^3(1 + bz)^2 g(z)$, it becomes substantially easier to recognize that the expression $E_u[K(u)b^{y_2+y_3}z^{y_1+y_2+y_3}/((1 + z)^{3-y_1-y_2}(1 + bz)^{y_1+y_2})]$ on the LHS of (7) can be written as a linear combination of six expectations $E[K(u)z^5/(1 + z)^3(1 + bz)^2], \ldots, E[K(u)z/(1 + z)^3(1 + bz)^2]$, and $E[K(u)/(1 + z)^3(1 + bz)^2]$.

5. See Definitions 1 and 2 in the Appendix.

6. The vector $(1, -1, -2, -b, -1, -1 - b, -b^2)'$ can be found as a kernel of the transpose of Ω^B.

7. See Chamberlain (1986) for a variational characterization of the partial semiparametric information bound.

8. The assumption that $\gamma_2 \neq 0, \gamma_3 \neq 0$, and $\gamma_2 \neq \gamma_3$ is harmless. After all, if $\gamma_2 = \gamma_3 = 0$, we get back to the basic model (1). The assumption that $\gamma_2 \neq \gamma_3$ is also harmless. If $\gamma_2 = \gamma_3$, then we may use the CMLE approach adopted by Honoré and Kyriazidou (2000).

9. Chamberlain (1992) implicitly relies on such definition of the information bound.

10. If such θ's exist, the zero mean restriction on k is automatically satisfied because

$$E[k(u)] = E[\ell_\beta^D] + \theta_2 \cdot E[\ell_2^D] + \theta_3 \cdot E[\ell_3^D] = 0 + \theta_2 \cdot 0 + \theta_3 \cdot 0 = 0.$$

11. See Definitions 3–6.

REFERENCES

Andersen, E. (1970) Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, Series B* 32, 283–301.

Bickel, P., C. Klaassen, Y. Ritov, & J. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

Chamberlain, G. (1985) Heterogeneity, omitted variable bias, and duration dependence. In J.J. Heckman & B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*. New York: Cambridge University Press.
Chamberlain, G. (1986) Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics* 32, 189–218.
Chamberlain, G. (1992) Binary Response Models for Panel Data: Identification and Information. Manuscript.
Chamberlain, G. (1993) Feedback in Panel Data Models. Seminar notes.
Honoré, B. & E. Kyriazidou (2000) Panel data discrete choice models with lagged dependent variables. *Econometrica* 68, 839–874.
Manski, C. (1987) Semiparametric analysis of random effects linear models from binary response data. *Econometrica* 55, 357–362.
Ritov, Y. & P. Bickel (1990) Achieving information bounds in non- and semiparametric models. *Annals of Statistics* 18, 925–938.

TECHNICAL APPENDIX

A.1. Notation.

Definition 1.

$$\Omega^B \equiv \begin{bmatrix} b^2 & 2b+3b^2 & 1+6b+3b^2 & 3+6b+b^2 & 3+2b & 1 \\ 0 & 0 & 0 & b^2 & 2b & 1 \\ 0 & 0 & b & 1+b & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & b^2 & 2b & 1 & 0 \\ 0 & b & 1+b & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Definition 2.

$$r^B \equiv \begin{bmatrix} 0 \\ 0 \\ -E\left[\dfrac{bz^2}{(1+bz)^2(1+z)^2}\right] \\ E\left[\dfrac{z^2}{(1+bz)^2(1+z)}\right] - E\left[\dfrac{2bz^3}{(1+bz)^3(1+z)}\right] \\ 0 \\ \dfrac{b}{1+b}E\left[\dfrac{z^2}{(1+bz)(1+z)^2}\right] - E\left[\dfrac{bz^3}{(1+bz)^2(1+z)^2}\right] \\ E\left[\dfrac{z^3}{(1+bz)^2(1+z)}\right] - E\left[\dfrac{2bz^4}{(1+bz)^3(1+z)}\right] \end{bmatrix}.$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

Definition 3.

$$\Xi^D \equiv \begin{bmatrix} 0 & 0 & 0 & cb^2d & db+cb & 1 \\ 0 & 0 & d^2cb^2 & d^2b+cdb & d & 0 \\ 0 & 0 & c^2db & c^2b+cd & c & 0 \\ 0 & c^2d^2b^2 & dc^2b^2+cd^2b & cdb & 0 & 0 \\ 0 & 0 & cdb & db+c & 1 & 0 \\ 0 & cd^2b & d^2b+cd & d & 0 & 0 \\ 0 & c^2db & cdb+c^2b & cb & 0 & 0 \\ c^2d^2b^2 & d^2cb^2+dc^2b^2 & cb^2d & 0 & 0 & 0 \end{bmatrix}.$$

Definition 4.

$$P_\beta^D \equiv \begin{bmatrix} 0 \\ 0 \\ -\dfrac{z^2dcb}{(1+z)(1+zc)(1+zdb)^2} \\ \left(\dfrac{z^2cd}{(1+z)(1+zc)(1+zdb)} - \dfrac{bz^3cd^2}{(1+z)(1+zc)(1+zdb)^2}\right)b \\ -\dfrac{z^2cb}{(1+z)(1+zcb)^2(1+zd)} \\ -\dfrac{z^3dcb}{(1+z)(1+zcb)^2(1+zd)} \\ \left(\dfrac{z^2c}{(1+z)(1+zcb)(1+zdb)} - \dfrac{z^3c^2b}{(1+z)(1+zcb)^2(1+zdb)} - \dfrac{z^3dcb}{(1+z)(1+zcb)(1+zdb)^2}\right)b \\ \left(2\dfrac{z^3dcb}{(1+z)(1+zcb)(1+zdb)} - \dfrac{z^4c^2b^2d}{(1+z)(1+zcb)^2(1+zdb)} - \dfrac{z^4cb^2d^2}{(1+z)(1+zcb)(1+zdb)^2}\right)b \end{bmatrix}.$$

Definition 5.

$$P_2^D \equiv \begin{bmatrix} -\dfrac{zc}{(1+z)(1+zc)^2(1+zd)} \\ -\dfrac{z^2dc}{(1+z)(1+zc)^2(1+zd)} \\ \left(\dfrac{z}{(1+z)(1+zc)(1+zdb)} - \dfrac{z^2c}{(1+z)(1+zc)^2(1+zdb)}\right)c \\ \left(\dfrac{z^2db}{(1+z)(1+zc)(1+zdb)} - \dfrac{z^3cdb}{(1+z)(1+zc)^2(1+zdb)}\right)c \\ -\dfrac{z^2cb}{(1+z)(1+zcb)^2(1+zd)} \\ -\dfrac{z^3cdb}{(1+z)(1+zcb)^2(1+zd)} \\ \left(\dfrac{z^2b}{(1+z)(1+zcb)(1+zdb)} - \dfrac{z^3cb^2}{(1+z)(1+zcb)^2(1+zdb)}\right)c \\ \left(\dfrac{z^3b^2d}{(1+z)(1+zcb)(1+zdb)} - \dfrac{z^4cb^3d}{(1+z)(1+zcb)^2(1+zdb)}\right)c \end{bmatrix}.$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

Definition 6.

$$P_3^D \equiv \begin{bmatrix} -\dfrac{zd}{(1+z)(1+zc)(1+zd)^2} \\ \left(\dfrac{z}{(1+z)(1+zc)(1+zd)} - \dfrac{z^2d}{(1+z)(1+zc)(1+zd)^2}\right)d \\ -\dfrac{z^2dcb}{(1+z)(1+zc)(1+zdb)^2} \\ \left(\dfrac{z^2cb}{(1+z)(1+zc)(1+zdb)} - \dfrac{z^3cdb^2}{(1+z)(1+zc)(1+zdb)^2}\right)d \\ -\dfrac{z^2d}{(1+z)(1+zcb)(1+zd)^2} \\ \left(\dfrac{z^2}{(1+z)(1+zcb)(1+zd)} - \dfrac{z^3d}{(1+z)(1+zcb)(1+zd)^2}\right)d \\ -\dfrac{z^3cb^2d}{(1+z)(1+zcb)(1+zdb)^2} \\ \left(\dfrac{z^3cb^2}{(1+z)(1+zcb)(1+zdb)} - \dfrac{z^4cb^3d}{(1+z)(1+zcb)(1+zdb)^2}\right)d \end{bmatrix}.$$

A.2. Lemmas.

LEMMA 1. *The nonparametric tangent space of the model (1) is equal to*

$$\mathcal{T}^B = \{E[k(u)|y_3, y_2 + y_1] : E[k(u)] = 0\}.$$

Proof. Consider a one-dimensional parametric subfamily $f(\cdot, \zeta)$ of the density of u indexed by ζ such that $f(\cdot, \zeta_0) = f(\cdot)$. Then the density of (y_3, y_2, y_1) of the parametric submodel is given by

$$\mathcal{L}^B(\beta, \zeta) = \int p^B(y_3, y_2, y_1; v, \beta) f(v, \zeta)\, dv.$$

Observe that

$$\left.\frac{\partial \log \mathcal{L}^B}{\partial \zeta}\right|_{\zeta=\zeta_0} = \frac{\int k(v) \cdot p^B(y_3, y_2, y_1; v, \beta) f(v)\, dv}{\int p^B(y_3, y_2, y_1; v, \beta) f(v)\, dv} = E[k(u)|y_3, y_2, y_1],$$

where

$$k(v) = \frac{\partial}{\partial \zeta} \log f(v, \zeta)|_{\zeta=\zeta_0}.$$

(For a more rigorous justification of the second equality, see, e.g., Bickel et al., 1993, Proposition A.5.5.) Therefore, it easily follows that the nonparametric tangent space is equal to $\mathcal{T}^B \equiv \{E[k(u)|y_3, y_2, y_1] : E[k(u)] = 0\}$. The conclusion follows from the observation that $(y_3, y_2 + y_1)$ is sufficient for u as can be seen in (3). ■

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

LEMMA 2. *If $-c + db \neq 0$, then a solution x to (18) is an element of the subspace spanned by*

$$\lambda_1^D = \left(0 \quad \frac{c(-d+c)}{d(-c+db)} \quad 1 \quad 0 \quad -\frac{d(cb-c)}{-c+db} \quad -\frac{cb(-c+d)}{-c+db} \quad -\frac{db(-d+c)}{-c+db} \quad 0\right)$$

and

$$\lambda_2^D = \left(0 \quad -\frac{b(-d+c)}{-c+db} \quad 0 \quad 1 \quad -\frac{d(-cb+db)}{-c+db} \quad -\frac{cb(db-d)}{-c+db} \quad -\frac{db(-c+d)}{-c+db} \quad 0\right).$$

If $-c + db = 0$, then it is an element of the subspace spanned by

$$\lambda_1^D = \left(0 \quad -b \quad 1 \quad 1 \quad 0 \quad -cb-c \quad \frac{c}{b} \quad 0\right)$$

and

$$\lambda_2^D = \left(0 \quad -\frac{b}{c} \quad 0 \quad 0 \quad 1 \quad -b \quad 1 \quad 0\right).$$

LEMMA 3. *There exists a distribution of u with density bounded away from zero on an open interval such that the 2×2 matrix $[\Lambda^D E[P_2^D], \Lambda^D E[P_3^D]]$ is nonsingular.*

Proof. The proof is based on inspection of the determinant, computed by MAPLE. Suppose first that $-c + db \neq 0$, in which case we can show using Lemma 2 that

$$\Lambda^D E[P_\beta^D] = \begin{bmatrix} \dfrac{c^2db(-d+c)(b-1)m_3}{-c+db} \\ c^2d^2b^2m_4 + \dfrac{cdb(b^2d^2 - c^2b + cdb - cd)m_3}{-c+db} + cdbm_2 \end{bmatrix},$$

$$\Lambda^D E[P_2^D] = \begin{bmatrix} -\dfrac{c^2d^2b^2(-d+c)m_4}{-c+db} - \dfrac{cdb(c^2b - cdb - d^2b + c^2)m_3}{-c+db} \\ + \dfrac{cb(cdb + d^2b - c^2 - 2cd + d^2)m_2}{-c+db} + cm_1 \\ \dfrac{d^3b^2c^2(b-1)m_4}{-c+db} + \dfrac{cb^2d^2(d+c)(b-1)m_3}{-c+db} + \dfrac{cd^2b(b-1)m_2}{-c+db} \end{bmatrix},$$

and

$$\Lambda^D E[P_3^D] = \begin{bmatrix} \dfrac{c^2d^2b^2(-d+c)m_4}{-c+db} + \dfrac{cb^2d(2c+d)(-d+c)m_3}{-c+db} \\ - \dfrac{c(b^2d^2 - c^2b - 2cdb + d^2b + cd)m_2}{-c+db} + \dfrac{c(-d+c)m_1}{-c+db} \\ -\dfrac{c^2d^2b^2(-d+c)m_4}{-c+db} - \dfrac{cdb(cb^2d - 2d^2b - cdb + c^2b + cd)m_3}{-c+db} \\ - \dfrac{db(-d+c)(cb+c+db)m_2}{-c+db} - \dfrac{db(-d+c)m_1}{-c+db} \end{bmatrix},$$

where

$$m_j \equiv m_j(b,c,d), \qquad j = 1,2,3,4$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

and

$$m_j(b,c,d) \equiv E\left[\frac{z^j}{(1+z)(1+cz)(1+dz)(1+bcz)(1+bdz)}\right]. \tag{A.1}$$

Therefore, such θ's exist if the matrix $[\Lambda^D E[P_2^D], \Lambda^D E[P_3^D]]$ is nonsingular. It can be shown that the determinant of $[\Lambda^D E[P_2^D], \Lambda^D E[P_3^D]]$ is equal to $-cdb/(-c+db)\cdot\Delta^D$, where Δ^D equals

$$\begin{aligned}
&bm_2^2c^3 - bd^3m_2^2 - d^2m_2m_1 - b^2d^3m_2^2 + b^2m_2^2c^3 + c^2m_1m_2 + cm_1^2 - dm_1^2 \\
&\quad - b^2d^4c^2m_4m_3 - b^2d^4cm_4m_2 - b^2d^4m_3m_2 - b^2d^4cm_3^2 - 4b^2d^3cm_2m_3 \\
&\quad - 2b^2d^3m_3^2c^2 - b^2d^3c^2m_2m_4 - 2b^2d^2m_2^2c + b^2d^2c^4m_4m_3 + 3b^2d^2c^3m_2m_4 \\
&\quad + b^2dc^2m_2^2 + b^2dm_3c^2m_1 + b^2dc^4m_4m_2 + 4b^2dm_2c^3m_3 + b^2dm_3^2c^4 \\
&\quad + b^2m_3c^4m_2 - bd^3cm_4m_1 - 2bd^3cm_2m_3 - bd^3c^2m_2m_4 - bd^3m_3m_1 \\
&\quad - 2bd^2m_2m_1 + bd^2m_3^2c^3 - 3bd^2cm_1m_3 + 2bdc^2m_2^2 + bdm_3c^2m_1 + bdc^3m_4m_1 \\
&\quad + 2bdm_2c^3m_3 + bc^3m_1m_3 + 2bc^2m_1m_2 - b^3d^4c^3m_4^2 - 2b^3d^4c^2m_4m_3 \\
&\quad - b^3d^4cm_3^2 - b^3d^3m_3^2c^2 - b^3d^3cm_2m_3 + b^3d^3c^4m_4^2 - b^3d^3c^2m_2m_4 \\
&\quad + 2b^3d^2m_3^2c^3 + 2b^3d^2c^4m_4m_3 + b^3dm_2c^3m_3 + b^3dm_3^2c^4 - d^2m_2^2c \\
&\quad + dm_3c^2m_1,
\end{aligned} \tag{A.2}$$

which is nonzero for some distribution of u as shown in Lemma 4. We therefore have (17) satisfied for some θ_2 and θ_3, say, $\theta_2^*(b,c,d)$ and $\theta_3^*(b,c,d)$.

Now consider the case where

$$-c + db = 0.$$

We can then show that the same argument repeats for

$$\Lambda^D = \begin{bmatrix} 0 & -b & 1 & 1 & 0 & -cb-c & \frac{c}{b} & 0 \\ 0 & -\frac{b}{c} & 0 & 0 & 1 & -b & 1 & 0 \end{bmatrix},$$

$$\Lambda^D E[P_\beta^D] = \begin{bmatrix} c^4m_4 + \frac{c^3(b+1)m_3}{b} + c^2m_2 \\ -c^2(b-1)m_3 \end{bmatrix},$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

$$[\Lambda^D E[P_2^D], \Lambda^D E[P_3^D]]$$

$$= \begin{bmatrix} \frac{c^4(b+1)m_4}{b} + \frac{c^3(b^2+3b+1)m_3}{b} + \frac{c^2(b+1)^2 m_2}{b} + cm_1 & -\frac{c^4 m_4}{b} - c^3(b+2)m_3 - c^2(b+2)m_2 - cm_1 \\ c^3 m_4 + c^2(b+1)m_3 + cm_2 & -c^3 m_4 - c^2(2b+1)m_3 - c(b+2)m_2 - m_1 \end{bmatrix},$$

and the determinant of $[\Lambda^D E[P_2^D], \Lambda^D E[P_3^D]]$ equal to $-c/b$ times

$$\begin{aligned} & cm_2 m_1 + c^2 m_3 m_1 + c^3 m_4 m_1 + 3c^2 m_2^2 b + 3c^2 m_2^2 b^2 + 3c^4 m_3^2 b + 4c^4 m_3^2 b^2 \\ & + c^4 m_3^2 b^3 + 2c^4 m_4 m_2 + c^6 m_4^2 b + 3c^4 m_4 b m_2 + c^4 m_4 b^2 m_2 + c^5 m_4 m_3 \\ & + 3c^5 m_4 b m_3 + 2c^5 m_4 b^2 m_3 + 6c^3 m_3 b^2 m_2 + 2c^3 m_3 b^3 m_2 + m_1^2 b + 2c^2 m_2^2 \\ & + c^4 m_3^2 + c^2 m_2^2 b^3 + 3c^3 m_3 m_2 + 7c^3 m_3 b m_2 + 3c m_2 b m_1 + 2c m_2 b^2 m_1 \\ & + 3c^2 m_3 b m_1 + 2c^2 m_3 b^2 m_1 + c^3 m_4 b m_1. \end{aligned}$$

As before, we can show that there exists a distribution of u with density bounded away from zero on an open interval such that this determinant is nonzero. ■

LEMMA 4. *There exists a distribution of z with density bounded away from zero on an open interval such that (A.2) is nonzero.*

Proof. The proof is based on the intuition that, if z is supported on a very small interval around x with density proportional to the denominator

$$(1+z)(1+cz)(1+dz)(1+bcz)(1+bdz)$$

in the definition (A.1) of m_j, then the determinant can be approximated by a polynomial function $h(x)$ where m_j's in (A.2) are replaced by x^j's. The polynomial function in x is obviously nonzero Lebesgue almost everywhere, and there would exist an x such that $h(x) \neq 0$.

Substituting $m_j = x^j$ in (A.2), we obtain $(d-c)x^2$ times

$$\begin{aligned} h(x) \equiv & d^3 b^3 c^3 x^6 + d^2 b^2 c^2 (c+d)(2b+1)\cdot x^5 \\ & + dbc(2c^2 b + c^2 b^2 + 5cdb + cd + 3cb^2 d + 2d^2 b + d^2 b^2)\cdot x^4 \\ & + b(c+d)(c^2 b + 4cdb + 3cd + cb^2 d + d^2 b)\cdot x^3 \\ & + (2c^2 b + c^2 b^2 + 5cdb + cd + 3cb^2 d + 2d^2 b + d^2 b^2)\cdot x^2 \\ & + (c+d)(2b+1)\cdot x \\ & + 1. \end{aligned}$$

Because $h(0) = 1$ and $h(\infty) = \infty$, and because $h(\cdot)$ is continuous on $(0,\infty)$, we can find some $\sigma \neq 0$ and $x^* > 0$ such that $h(x^*) = \sigma$.

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

Consider now a density of z that is proportional to

$$(1+z)(1+cz)(1+dz)(1+bcz)(1+bdz)$$

on $(x^*-\varepsilon, x^*+\varepsilon)$, and zero elsewhere. Note that

$$m_j(x^*,\varepsilon) \equiv C(\varepsilon)^{-1}\int_{(x^*-\varepsilon,x^*+\varepsilon)} \frac{z^j(1+z)(1+cz)(1+dz)(1+bcz)(1+bdz)}{(1+z)(1+cz)(1+dz)(1+bcz)(1+bdz)}\,dz$$

$$= C(\varepsilon)^{-1}\int_{(x^*-\varepsilon,x^*+\varepsilon)} z^j dz,$$

where

$$C(\varepsilon) = \int_{(x^*-\varepsilon,x^*+\varepsilon)} (1+z)(1+cz)(1+dz)(1+bcz)(1+bdz)\,dz.$$

Observe that

$$\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}C(\varepsilon) = (1+x^*)(1+cx^*)(1+dx^*)(1+bcx^*)(1+bdx^*)$$

and

$$\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_{(x^*-\varepsilon,x^*+\varepsilon)} z^j dz = (x^*)^j.$$

It therefore follows that

$$\lim_{\varepsilon\to 0} m_j(x^*,\varepsilon) = \frac{\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_{(x^*-\varepsilon,x^*+\varepsilon)} z^j dz}{\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}C(\varepsilon)}$$

$$= \frac{(x^*)^j}{(1+x^*)(1+cx^*)(1+dx^*)(1+bcx^*)(1+bdx^*)}.$$

Now, substitute $m_j = m_j(x^*,\varepsilon)$ in (A.2), and take the limit as $\varepsilon \to 0$. It is quite easy to see that the limit equals $(d-c)(x^*)^2$ times

$$\frac{h(x^*)}{[(1+x^*)(1+cx^*)(1+dx^*)(1+bcx^*)(1+bdx^*)]^2},$$

which is nonzero. Therefore, there exists $\varepsilon > 0$ sufficiently small such that (A.2) is nonzero. ■

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

LEMMA 5. *For any open interval, there exist a subset* $\{z_1,\ldots,z_6\} \subset \mathcal{I}$ *and some* $\varepsilon > 0$ *such that the determinant of the* 6×6 *matrix*

$$\begin{bmatrix} \frac{1}{2\varepsilon}\int_{z_1-\varepsilon}^{z_1+\varepsilon} z^5\,dz & \frac{1}{2\varepsilon}\int_{z_2-\varepsilon}^{z_2+\varepsilon} z^5\,dz & \cdots & \frac{1}{2\varepsilon}\int_{z_5-\varepsilon}^{z_5+\varepsilon} z^5\,dz & \frac{1}{2\varepsilon}\int_{z_6-\varepsilon}^{z_6+\varepsilon} z^5\,dz \\ \frac{1}{2\varepsilon}\int_{z_1-\varepsilon}^{z_1+\varepsilon} z^4\,dz & \frac{1}{2\varepsilon}\int_{z_2-\varepsilon}^{z_2+\varepsilon} z^4\,dz & \cdots & \frac{1}{2\varepsilon}\int_{z_5-\varepsilon}^{z_5+\varepsilon} z^4\,dz & \frac{1}{2\varepsilon}\int_{z_6-\varepsilon}^{z_6+\varepsilon} z^4\,dz \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{2\varepsilon}\int_{z_1-\varepsilon}^{z_1+\varepsilon} z\,dz & \frac{1}{2\varepsilon}\int_{z_2-\varepsilon}^{z_2+\varepsilon} z\,dz & \cdots & \frac{1}{2\varepsilon}\int_{z_5-\varepsilon}^{z_5+\varepsilon} z\,dz & \frac{1}{2\varepsilon}\int_{z_6-\varepsilon}^{z_6+\varepsilon} z\,dz \\ \frac{1}{2\varepsilon}\int_{z_1-\varepsilon}^{z_1+\varepsilon} 1\,dz & \frac{1}{2\varepsilon}\int_{z_2-\varepsilon}^{z_2+\varepsilon} 1\,dz & \cdots & \frac{1}{2\varepsilon}\int_{z_5-\varepsilon}^{z_5+\varepsilon} 1\,dz & \frac{1}{2\varepsilon}\int_{z_6-\varepsilon}^{z_6+\varepsilon} 1\,dz \end{bmatrix}$$

is nonzero.

Proof. There exists a subset $\{z_1,\ldots,z_6\} \subset \mathcal{I}$ such that the determinant of the 6×6 matrix

$$\begin{bmatrix} z_1^5 & z_2^5 & z_3^5 & z_4^5 & z_5^5 & z_6^5 \\ z_1^4 & z_2^4 & z_3^4 & z_4^4 & z_5^4 & z_6^4 \\ z_1^3 & z_2^3 & z_3^3 & z_4^3 & z_5^3 & z_6^3 \\ z_1^2 & z_2^2 & z_3^2 & z_4^2 & z_5^2 & z_6^2 \\ z_1 & z_2 & z_3 & z_4 & z_5 & z_6 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

is nonzero. This is because for any set $\{z_1,\ldots,z_5\} \subset \mathcal{I}$ with distinct elements, the determinant regarded as a real valued function of z_6 is polynomial in z_6 and is equal to zero for only a finite number of values of z_6. By continuity, the determinant would be nonzero even if every (k,j) element is replaced by $(1/2\varepsilon)\int_{z_j-\varepsilon}^{z_j+\varepsilon} z^{6-k}\,dz$ for $\varepsilon > 0$ near zero. ■

LEMMA 6. *Assume that the density of u is bounded away from zero in an open interval. Then, there exists some g such that*

$$\Xi^D \varsigma_g = E_u[P_\beta^D] + \theta_2^* \cdot E_u[P_2^D] + \theta_3^* \cdot E_u[P_3^D].$$

Proof. Let Υ denote a 6×8 matrix with full row rank, the six rows of which are orthogonal to λ_1 and λ_2. Multiplying both sides of (17) by Υ, we obtain

$$\Upsilon \Xi \varsigma_g = \Upsilon(E[P_\beta^D] + \theta_2^* \cdot E[P_2^D] + \theta_3^* \cdot E[P_3^D]).$$

Because $\Upsilon\Xi$ is nonsingular by construction, we should have

$$\varsigma_g = (\Upsilon\Xi)^{-1}\Upsilon(E[P_\beta^D] + \theta_2^* \cdot E[P_2^D] + \theta_3^* \cdot E[P_3^D]). \qquad \textbf{(A.3)}$$

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press

Let $\varphi(z)$ denote the density of z that is bounded away from zero on an open interval $\mathcal{O}$. Let $(z_1-\varepsilon, z_1+\varepsilon),\ldots,(z_6-\varepsilon, z_6+\varepsilon)$ denote some collection of nonoverlapping open intervals contained in $\mathcal{O}$. By Lemma 5, we can choose $z_1,\ldots,z_6$ and ε such that the matrix S with

$$S_{(k,j)} = \frac{1}{2\varepsilon}\int_{z_j-\varepsilon}^{z_j+\varepsilon} z^{6-k}dz$$

is nonsingular. Here, $S_{(k,j)}$ denotes the (k,j) element of S.

It is now shown that $g(\cdot)$ defined by

$$g(z) \equiv \sum_{j=1}^{6} \frac{w_j}{2\varepsilon\varphi(z)}\, 1(z_j-\varepsilon < z < z_j+\varepsilon)$$

satisfies (A.3) for some choice of $w_1,\ldots,w_6$. (Here, $1(\cdot)$ is the usual indicator function.) Note that we should have

$$E[g(z)z^k] = \sum_{j=1}^{6} w_j \cdot \frac{1}{2\varepsilon}\int_{z_j-\varepsilon}^{z_j+\varepsilon} z^{6-k}dz = \sum_{j=1}^{6} S_{(k,j)}w_j,$$

and therefore, we may write $\varsigma_g = Sw$, where $w = (w_1,\ldots,w_6)'$. If we take

$$w = S^{-1}(\Upsilon\Xi)^{-1}\Upsilon(E[P_\beta^D] + \theta_2^*\cdot E[P_2^D] + \theta_3^*\cdot E[P_3^D]),$$

the $g(\cdot)$ would satisfy (A.3). ■

https://doi.org/10.1017/S0266466601175031 Published online by Cambridge University Press