A note on sufficiency in binary panel models

Koen Jochmans[*] Thierry Magnac[†]
Sciences Po, Paris Toulouse School of Economics

This version: February 8, 2017

Abstract

Consider estimating the slope coefficients of a fixed-effect binary-choice model from two-period panel data. Two approaches to semiparametric estimation at the regular parametric rate have been proposed. One is based on a sufficiency requirement, the other is based on a conditional-median restriction. We show that, under standard assumptions, both conditions are equivalent.

Keywords: binary choice, fixed effects, panel data, regular estimation, sufficiency.

1 Introduction

A classic problem in panel data analysis is the estimation of the vector of slope coefficients, β, in fixed-effect linear models from binary response data on n observations.

In seminal work, Rasch (1960) constructed a conditional maximum-likelihood estimator for the fixed-effect logit model by building on a sufficiency argument. Chamberlain (2010) and Magnac (2004) have shown that sufficiency is necessary for estimation at the $n^{-1/2}$ rate to be possible in general.

[*]Department of Economics, Sciences Po, 28 rue des Saints Pères, 75007 Paris, France.

[†]Toulouse School of Economics, Université de Toulouse, 21 Allée de Brienne, 31000 Toulouse, France.

We thank the Co-Editor and two referees for comments. Jochmans gratefully acknowledges financial support from the European Research Council through Starting Grant n° 715787. Magnac gratefully acknowledges financial support from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 grant agreement n° 295298.

Manski (1987) proposed a maximum-score estimator of β. His estimator relies on a conditional median restriction and does not require sufficiency. However, it converges at the slow rate $n^{-1/3}$. Horowitz (1992) suggested smoothing the maximum-score criterion function and showed that, by doing so, the convergence rate can be improved, although the $n^{-1/2}$-rate remains unattainable. Lee (1999) has given an alternative conditional-median restriction and derived a $n^{-1/2}$-consistent maximum rank-correlation estimator of β. He provided sufficient conditions for this condition to hold that restrict the distribution of the fixed effects and the covariates. It can be shown that these restrictions involve the unknown parameter β through index-sufficiency requirements on the distribution of the covariates, and that these can severely restrict the values that β is allowed to take.

We reconsider the conditional-median restriction of Lee (1999) under standard assumptions and look for conditions that imply it to hold for any β. We find that imposing the conditional-median restriction is equivalent to requiring sufficiency.

2 Model and assumptions

Suppose that binary outcomes $y_i = (y_{i1}, y_{i2})$ relate to a set of observable covariates $x_i = (x_{i1}, x_{i2})$ through the threshold-crossing model

$$y_{i1} = 1\{x_{i1}\beta + \alpha_i \geq u_{i1}\}, \qquad y_{i2} = 1\{x_{i2}\beta + \alpha_i \geq u_{i2}\},$$

where $u_i = (u_{i1}, u_{i2})$ are latent disturbances, α_i is an unobserved effect, and β is a parameter vector of conformable dimension, say k.

The challenge is to construct an estimator of β from a random sample $\{(y_i, x_i), i = 1, \ldots, n\}$ that converges at the regular $n^{-1/2}$ rate.

Let $\Delta y_i = y_{i2} - y_{i1}$ and $\Delta x_i = x_{i2} - x_{i1}$. The following assumption will be maintained throughout.

Assumption 1 (Identification and regularity)**.**

(a) u_i *is independent of* (x_i, α_i).

(b) Δx_i *is not contained in a proper linear subspace of* $\mathcal{R}^k$.

(c) The first component of Δx_i *continuously varies over the whole real line* $\mathcal{R}$ *(for almost all values of the other components), and the first component of* β *is not equal to zero and normalized to one.*

(d) α_i *varies continuously over the whole real line* $\mathcal{R}$ *(for almost all values of* x_i*).*

(e) The distribution of u_i *admits a strictly positive, continuous, and bounded density function with respect to Lebesgue measure.*

Parts (a)–(c) collect sufficient conditions that ensure that β is (semiparametrically) identified while Parts (d)–(e) are conventional regularity conditions that allow the use of differential calculus (see Magnac 2004). In the sequel we will omit the 'almost surely' qualifier from all conditional statements.

Assumption 1 does not parametrize the distribution of u_i nor does it restrict the dependence between α_i and x_i. As such, our approach is semiparametric and we treat the α_i as fixed effects. This is to be contrasted with a random-effect approach, where the distribution of α_i given x_i (and the distribution of u_i) is parametrized (see, e.g., Chamberlain 1980). In such a case standard inference can be performed through the (marginal) likelihood. A middle ground would be to impose semiparametric restrictions on the dependence between α_i and x_i. For example, Honoré and Lewbel (2002) construct a $n^{-1/2}$-consistent estimator under the condition that one of the regressors is conditionally independent of the fixed effects and that this special regressor satisfies a large-support condition.

3 Conditions for regular estimation

Magnac (2004, Theorem 1) has shown that, under Assumption 1, the semiparametric efficiency bound for β is zero unless $y_{i1} + y_{i2}$ is a sufficient statistic for α_i. Sufficiency can be stated as follows.

Condition 1 (Sufficiency)**.** *There exists a real function* G*, independent of* α_i*, such that*

$$\Pr(\Delta y_i = 1 | x_i, \Delta y_i \neq 0, \alpha_i) = \Pr(\Delta y_i = 1 | x_i, \Delta y_i \neq 0) = G(\Delta x_i \beta)$$

for all $\alpha_i \in \mathcal{R}$*.*

Condition 1 states that data in first-differences follow a single-indexed binary-choice model. This yields a variety of estimators of β, such as semiparametric maximum likelihood (Klein and Spady 1993), that are $n^{-1/2}$-consistent under standard assumptions.

Magnac (2004, Theorem 3) derived conditions on the distributions of u_i and Δu_i that imply that Condition 1 holds.

On the other hand, Lee (1999) considered estimation of β based on a sign restriction. We write $\operatorname{med}(x)$ for the median of random variable x and let $\operatorname{sgn}(x) = 1\{x \geq 0\} - 1\{x < 0\}$.

Condition 2 (Median restriction)**.** *For any two observations i and j,*

$$\operatorname{med}\left(\frac{\Delta y_i - \Delta y_j}{2}\middle| x_i, x_j, \Delta y_i \neq 0, \Delta y_j \neq 0, \Delta y_i \neq \Delta y_j\right) = \operatorname{sgn}(\Delta x_i \beta - \Delta x_j \beta)$$

holds.

Condition 2 suggests a rank estimator for β. Conditions for this estimator to be $n^{-1/2}$-consistent are stated in Sherman (1993).

Lee (1999, Assumption 1) restricted the joint distribution of α_i, x_i, and $x_{i1}\beta, x_{i2}\beta$ to ensure that Condition 2 holds. Aside from these restrictions going against the fixed-effect approach, they do not hold uniformly in β, in general. The Appendix contains additional discussion and an example.

4 Equivalence

The main result of this paper is the equivalence of Conditions 1 and 2 as requirements for $n^{-1/2}$-consistent estimation of any β.

Theorem 1 (Equivalence)**.** *Let Assumption 1 hold. Then Condition 2 holds for any β and any joint distribution of (α_i, x_i) if and only if Condition 1 holds for any β and any joint distribution of (α_i, x_i).*

Proof. We start with two lemmas that are instrumental in showing Theorem 1. We will routinely make use of the fact that, for events A, B, and C,

$$\frac{\Pr(A|C)}{\Pr(B|C)} = \frac{\Pr(A)}{\Pr(B)}$$

if $A \subset C$ and $B \subset C$.

Lemma 1. *Condition 1 is equivalent to the existence of a continuously-differentiable, strictly-decreasing function c, independent of α_i, such that*

$$\frac{\Pr(\Delta y_i = -1|x_i, \alpha_i)}{\Pr(\Delta y_i = 1|x_i, \alpha_i)} = c(\Delta x_i \beta)$$

for all $\alpha_i \in \mathcal{R}$.

Proof. Conditional on $\Delta y_i \neq 0$ and on α_i, x_i, the random variable Δy_i is Bernoulli with success probability

$$\Pr(\Delta y_i = 1 | x_i, \Delta y_i \neq 0, \alpha_i) = \frac{1}{1 + \frac{\Pr(\Delta y_i = -1 | x_i, \alpha_i)}{\Pr(\Delta y_i = \ \ 1 | x_i, \alpha_i)}}.$$

Re-arranging this expression and enforcing Condition 1 shows that

$$\frac{\Pr(\Delta y_i = -1 | x_i, \alpha_i)}{\Pr(\Delta y_i = \ \ 1 | x_i, \alpha_i)} = \frac{1 - G(\Delta x_i \beta)}{G(\Delta x_i \beta)},$$

which is a function of $\Delta x_i \beta$ only. Monotonicity and differentiability of this function follow easily, as in Magnac (2004, Proof of Theorem 2). This completes the proof of Lemma 1. $\square$

Lemma 2. *Let*

$$\tilde{c}(x_i) = \frac{\Pr(\Delta y_i = -1 | x_i)}{\Pr(\Delta y_i = \ \ 1 | x_i)}.$$

Condition 2 is equivalent to the sign restriction

$$\operatorname{sgn}(\tilde{c}(x_j) - \tilde{c}(x_i)) = \operatorname{sgn}(\Delta x_i \beta - \Delta x_j \beta)$$

holding for any two observations i and j.

Proof. Conditional on $\Delta y_i \neq 0, \Delta y_j \neq 0, \Delta y_i \neq \Delta y_j$ (and the covariates),

$$\frac{\Delta y_i - \Delta y_j}{2} = \begin{cases} \ \ 1 & \text{if } \Delta y_i = 1 \text{ and } \Delta y_j = -1 \\ -1 & \text{if } \Delta y_j = 1 \text{ and } \Delta y_i = -1 \end{cases}.$$

Therefore, it is Bernoulli with success probability

$$\Pr(\Delta y_i = 1, \Delta y_j = -1 | x_i, x_j, \Delta y_i \neq 0, \Delta y_j \neq 0, \Delta y_i \neq \Delta y_j) = \frac{1}{1 + r(x_i, x_j)},$$

where

$$r(x_i, x_j) = \frac{\Pr(\Delta y_i = -1, \Delta y_j = \ \ 1 | x_i, x_j, \Delta y_i \neq 0, \Delta y_j \neq 0, \Delta y_i \neq \Delta y_j)}{\Pr(\Delta y_i = \ \ 1, \Delta y_j = -1 | x_i, x_j, \Delta y_i \neq 0, \Delta y_j \neq 0, \Delta y_i \neq \Delta y_j)}.$$

Note that

$$\operatorname{med}\left(\frac{\Delta y_i - \Delta y_j}{2} \middle| x_i, x_j, \Delta y_i \neq 0, \Delta y_j \neq 0, \Delta y_i \neq \Delta y_j \right) = \operatorname{sgn}\left(\frac{1}{1 + r(x_i, x_j)} - \frac{r(x_i, x_j)}{1 + r(x_i, x_j)} \right).$$

By the Bernoulli nature of the outcomes in the first step and random sampling of the observations in the second step, we have that

$$r(x_i, x_j) = \frac{\Pr(\Delta y_i = -1, \Delta y_j = \ \ 1 | x_i, x_j)}{\Pr(\Delta y_i = \ \ 1, \Delta y_j = -1 | x_i, x_j)} = \frac{\Pr(\Delta y_i = -1 | x_i)}{\Pr(\Delta y_i = \ \ 1 | x_i)} \frac{\Pr(\Delta y_j = \ \ 1 | x_j)}{\Pr(\Delta y_j = -1 | x_j)} = \frac{\tilde{c}(x_i)}{\tilde{c}(x_j)}.$$

Thus, Condition 2 can be written as

$$\operatorname{sgn}(\tilde{c}(x_j) - \tilde{c}(x_i)) = \operatorname{sgn}(\Delta x_i \beta - \Delta x_j \beta).$$

This completes the proof of Lemma 2. $\square$

We first establish that Condition 1 implies Condition 2. Armed with Lemmas 1 and 2 this is a simple task. First note that, because the function c is strictly decreasing by Lemma 1, Condition 1 implies that

$$\mathrm{sgn}(c(\Delta x_j\beta) - c(\Delta x_i\beta)) = \mathrm{sgn}(\Delta x_i\beta - \Delta x_j\beta).$$

Under Condition 1 we also have that

$$c(\Delta x_i\beta) = \frac{\Pr(\Delta y_i = -1|x_i, \alpha_i)}{\Pr(\Delta y_i = \ \ 1|x_i, \alpha_i)} = \frac{\Pr(\Delta y_i = -1|x_i)}{\Pr(\Delta y_i = \ \ 1|x_i)} = \tilde{c}(x_i).$$

Therefore,

$$\mathrm{sgn}(\tilde{c}(x_j) - \tilde{c}(x_i)) = \mathrm{sgn}(\Delta x_i\beta - \Delta x_j\beta).$$

By Lemma 2, this is Condition 2.

To see that Condition 2 implies Condition 1, first note that Assumption 1(a) gives

$$\frac{\Pr(\Delta y_i = -1|x_i, \alpha_i)}{\Pr(\Delta y_i = \ \ 1|x_i, \alpha_i)} = \frac{\Pr(u_{i1} \leq \gamma_i - \frac{1}{2}\Delta x_i\beta, u_{i2} > \gamma_i + \frac{1}{2}\Delta x_i\beta)}{\Pr(u_{i1} > \gamma_i - \frac{1}{2}\Delta x_i\beta, u_{i2} \leq \gamma_i + \frac{1}{2}\Delta x_i\beta)}$$

where we let $\gamma_i = \alpha_i + \frac{1}{2}(x_{i1} + x_{i2})\beta$. We may therefore write

$$\Pr(\Delta y_i = 1|x_i, \Delta y_i \neq 0, \alpha_i) = \tilde{G}(\Delta x_i\beta, \gamma_i)$$

for some function $\tilde{G}$. Hence,

$$\Pr(\Delta y_i = 1|x_i, \Delta y_i \neq 0) = \textstyle\int_{-\infty}^{+\infty} \tilde{G}(\Delta x_i\beta, \gamma)\, p(\gamma|x_i, \Delta y_i \neq 0)\, d\gamma,$$

where $p(\gamma_i|x_i, \Delta y_i \neq 0)$ denotes the density of γ_i given x_i and $\Delta y_i \neq 0$. Next, by Lemma 2, Condition 2 implies that

$$\begin{aligned}
\Delta x_i\beta = \Delta x_j\beta &\Longleftrightarrow \tilde{c}(x_i) = \tilde{c}(x_j) \\
&\Longleftrightarrow \frac{\Pr(\Delta y_i = -1|x_i)}{\Pr(\Delta y_i = \ \ 1|x_i)} = \frac{\Pr(\Delta y_j = -1|x_j)}{\Pr(\Delta y_j = \ \ 1|x_j)} \\
&\Longleftrightarrow \frac{\Pr(\Delta y_i = -1|x_i, \Delta y_i \neq 0)}{\Pr(\Delta y_i = \ \ 1|x_i, \Delta y_i \neq 0)} = \frac{\Pr(\Delta y_j = -1|x_j, \Delta y_j \neq 0)}{\Pr(\Delta y_j = \ \ 1|x_j, \Delta y_j \neq 0)} \\
&\Longleftrightarrow \Pr(\Delta y_i = 1|x_i, \Delta y_i \neq 0) = \Pr(\Delta y_j = 1|x_j, \Delta y_j \neq 0) \\
&\Longleftrightarrow \textstyle\int_{-\infty}^{+\infty} \tilde{G}(\Delta x_i\beta, \gamma)\, p(\gamma|x_i, \Delta y_i \neq 0)\, d\gamma = \int_{-\infty}^{+\infty} \tilde{G}(\Delta x_j\beta, \gamma)\, p(\gamma|x_j, \Delta y_j \neq 0)\, d\gamma
\end{aligned}$$

where the last step follows from the definition of $\tilde{G}$ above. Therefore, when $\Delta x_i\beta = \Delta x_j\beta = v$ (say), it must be that

$$\textstyle\int_{-\infty}^{+\infty} \tilde{G}(v, \gamma)\ \{p(\gamma|x_i, \Delta y_i \neq 0) - p(\gamma|x_j, \Delta y_j \neq 0)\}\ d\gamma = 0. \tag{1}$$

If the dependence between γ_i and x_i is unrestricted, this equality can only hold if $\tilde{G}(v,\gamma)$ is (almost surely) constant in γ. We have the following result, which is Condition 1 and concludes the proof of the theorem.

Lemma 3. *For all v and almost all γ_i (or α_i)*

$$\tilde{G}(\Delta x_i\beta,\gamma_i) = \Pr(\Delta y_i = 1|x_i, \Delta y_i \neq 0, \alpha_i) = \Pr(\Delta y_i = 1|x_i, \Delta y_i \neq 0) = G(\Delta x_i\beta)$$

for some function G.

Proof. First note that Assumption 1(a) implies that

$$\Pr(\Delta y_i \neq 0|x_i,\alpha_i) = \Pr(\Delta y_i = 1|x_i,\alpha_i) + \Pr(\Delta y_i = -1|x_i,\alpha_i) = h(\Delta x_i\beta,\gamma_i)$$

for some function h. This gives the factorization

$$\Pr(\Delta y_i = 1|x_i, \Delta y_i \neq 0) = \frac{\int_{-\infty}^{+\infty}\tilde{G}(\Delta x_i\beta,\gamma)\,h(\Delta x_i\beta,\gamma)\,p(\gamma|x_i)\,d\gamma}{\int_{-\infty}^{+\infty}h(\Delta x_i\beta,\gamma)\,p(\gamma|x_i)\,d\gamma},$$

where $p(\gamma_i|x_i)$ is the density of γ_i given x_i. Now fix x_i and v. Let $p_0(\gamma) = p(\gamma|x_i)$. By Assumption 1(c) there always exists an x_j for which (1) must hold. Let $p_1(\gamma) = p(\gamma|x_j)$. Then (1) can be written as

$$\frac{\int_{-\infty}^{+\infty}\tilde{G}(v,\gamma)\,h(v,\gamma)\,p_0(\gamma)\,d\gamma}{\int_{-\infty}^{+\infty}h(v,\gamma)\,p_0(\gamma)\,d\gamma} = \frac{\int_{-\infty}^{+\infty}\tilde{G}(v,\gamma)\,h(v,\gamma)\,p_1(\gamma)\,d\gamma}{\int_{-\infty}^{+\infty}h(v,\gamma)\,p_1(\gamma)\,d\gamma} \tag{2}$$

Because $p_1(\gamma)$ is unrestricted we may set

$$p_1(\gamma) = \begin{cases} p_0(\gamma)\,(1+\varepsilon) & \text{if } \gamma \in \mathcal{A} \\ p_0(\gamma)\,(1-\varepsilon') & \text{if } \gamma \notin \mathcal{A} \end{cases},$$

where

$$\mathcal{A} = \left\{\gamma \in \mathcal{R} : \tilde{G}(v,\gamma) \geq \overline{G}(v)\right\}, \qquad \overline{G}(v) = \frac{\int_{-\infty}^{+\infty}\tilde{G}(v,\gamma)\,h(v,\gamma)\,p_0(\gamma)\,d\gamma}{\int_{-\infty}^{+\infty}h(v,\gamma)\,p_0(\gamma)\,d\gamma},$$

and $(\varepsilon,\varepsilon') \in [0,1)^2$ can be chosen such that $\varepsilon+\varepsilon' \in (0,1)$. Note that $\Pr(\gamma \in \mathcal{A}) > 0$ because of Assumption 1(d). Furthermore, because $\int_{-\infty}^{+\infty} p_1(\gamma)\,d\gamma = 1$ we have $\Pr(\gamma \in \mathcal{A}) = \varepsilon'/(\varepsilon+\varepsilon')$ and $\Pr(\gamma \notin \mathcal{A}) = \varepsilon/(\varepsilon+\varepsilon')$. Also, as

$$\int_{-\infty}^{+\infty}h(v,\gamma)\,p_1(\gamma)\,d\gamma = (1+\varepsilon)\int_{\gamma\in\mathcal{A}}h(v,\gamma)\,p_0(\gamma)\,d\gamma + (1-\varepsilon')\int_{\gamma\notin\mathcal{A}}h(v,\gamma)\,p_0(\gamma)\,d\gamma,$$

we can write

$$\int_{-\infty}^{+\infty}h(v,\gamma)\,p_1(\gamma)\,d\gamma = ((1+\varepsilon)\,\lambda + (1-\varepsilon')\,(1-\lambda))\int_{-\infty}^{+\infty}h(v,\gamma)\,p_0(\gamma)\,d\gamma \tag{3}$$

for

$$\lambda = \frac{\int_{\gamma \in \mathcal{A}} h(v,\gamma)\, p_0(\gamma)\, d\gamma}{\int_{-\infty}^{+\infty} h(v,\gamma)\, p_0(\gamma)\, d\gamma} \in [0,1].$$

Because $h(v,\gamma) > 0$ and $p_0(\gamma) > 0$ for almost all γ and $\Pr(\gamma \in \mathcal{A}) > 0$ we have that $\lambda > 0$ and that $\lambda = 1$ if and only if $\Pr(\gamma \in \mathcal{A}) = 1$. Now, re-arranging (2) and using (3) gives

$$\begin{aligned} 0 = &\left(\frac{(\varepsilon + \varepsilon')\,(1-\lambda)}{(1+\varepsilon)\,\lambda + (1-\varepsilon')\,(1-\lambda)}\right) \frac{\int_{\gamma \in \mathcal{A}} \tilde{G}(v,\gamma)\, h(v,\gamma)\, p_0(\gamma)\, d\gamma}{\int_{-\infty}^{+\infty} h(v,\gamma)\, p_0(\gamma)\, d\gamma} \\ &- \left(\frac{(\varepsilon + \varepsilon')\,\lambda}{(1+\varepsilon)\,\lambda + (1-\varepsilon')\,(1-\lambda)}\right) \frac{\int_{\gamma \notin \mathcal{A}} \tilde{G}(v,\gamma)\, h(v,\gamma)\, p_0(\gamma)\, d\gamma}{\int_{-\infty}^{+\infty} h(v,\gamma)\, p_0(\gamma)\, d\gamma} \end{aligned} \tag{4}$$

while, by definition of the set $\mathcal{A}$, we have

$$\frac{\int_{\gamma \in \mathcal{A}} \tilde{G}(v,\gamma)\, h(v,\gamma)\, p_0(\gamma)\, d\gamma}{\int_{-\infty}^{+\infty} h(v,\gamma)\, p_0(\gamma)\, d\gamma} \geq \lambda\, \overline{G}(v), \qquad \frac{\int_{\gamma \notin \mathcal{A}} \tilde{G}(v,\gamma)\, h(v,\gamma)\, p_0(\gamma)\, d\gamma}{\int_{-\infty}^{+\infty} h(v,\gamma)\, p_0(\gamma)\, d\gamma} \leq (1-\lambda)\, \overline{G}(v), \tag{5}$$

with a strict inequality of the second expression if and only if $\lambda < 1$. Suppose that $\lambda < 1$. Then, combining (4) and (5) gives the inequality

$$\left(\frac{(\varepsilon + \varepsilon')\,(1-\lambda)\,\lambda}{(1+\varepsilon)\,\lambda + (1-\varepsilon')\,(1-\lambda)}\right) \overline{G}(v) < \left(\frac{(\varepsilon + \varepsilon')\,(1-\lambda)\,\lambda}{(1+\varepsilon)\,\lambda + (1-\varepsilon')\,(1-\lambda)}\right) \overline{G}(v),$$

which is a contradiction since $\varepsilon + \varepsilon' > 0$ and $\bar{G}(v) > 0$. Thus, we must have that $\lambda = 1$, and so $\Pr(\gamma \in \mathcal{A}) = 1$. Therefore, we have for any v

$$\Pr\left(G(v,\gamma) \geq \overline{G}(v)\right) = 1$$

and, by symmetry, for any v

$$\Pr\left(G(v,\gamma) \leq \overline{G}(v)\right) = 1.$$

Therefore for any v, $\tilde{G}(v,\gamma)$ is constant (almost surely) in γ and $\Delta y_i \neq 0$ is sufficient for γ_i. This completes the proof of the lemma. □

□

Appendix

The notation in Lee (1999) decomposes x into its continuously varying single component whose coefficient is equal to 1 and the remaining variables. We shall denote by a the first component and by z the remaining variables, so that $x = (a, z)$. We denote by θ the coefficient of z in $x\beta$ so that $\beta = (1, \theta)$, and omit the subscript i throughout.

Assumptions (g) and (h) of Lee (1999) can be written as

$$\begin{array}{lll} \text{(g)} & \alpha & \perp \Delta z \mid \Delta a + \theta \Delta z, \\ \text{(h)} & a_1 + \theta z_1 & \perp \Delta z \mid \Delta a + \theta \Delta z, \alpha \end{array}$$

in which, e.g., $\Delta z = z_2 - z_1$.

We first prove that these conditions imply an index sufficiency requirement on the distribution function of regressors. Second, we provide an example in which these conditions restrict the parameter of interest to only two possible values, except in non-generic cases.

Index sufficiency Denote by f the density with respect to some dominating measure and rewrite (h) as

$$f(a_1 + \theta z_1, \Delta z \mid \Delta a + \theta \Delta z, \alpha) = f(a_1 + \theta z_1 \mid \Delta a + \theta \Delta z, \alpha) f(\Delta z \mid \Delta a + \theta \Delta z, \alpha).$$

As Condition (g) can be written as

$$f(\Delta z \mid \Delta a + \theta \Delta z, \alpha) = f(\Delta z \mid \Delta a + \theta \Delta z),$$

we therefore have that

$$f(a_1 + \theta z_1, \Delta z \mid \Delta a + \theta \Delta z, \alpha) = f(a_1 + \theta z_1 \mid \Delta a + \theta \Delta z, \alpha) f(\Delta z \mid \Delta a + \theta \Delta z),$$

which we can multiply by $f(\alpha \mid \Delta a + \theta \Delta z)$ and integrate with respect to α to get

$$f(a_1 + \theta z_1, \Delta z \mid \Delta a + \theta \Delta z) = f(a_1 + \theta z_1 \mid \Delta a + \theta \Delta z) f(\Delta z \mid \Delta a + \theta \Delta z).$$

As this expression can be rewritten as

$$f(\Delta z \mid \Delta a + \theta \Delta z, a_1 + z_1 \theta) = f(\Delta z \mid \Delta a + \theta \Delta z),$$

Conditions (g) and (h) of Lee (1999) demand that

$$f(\Delta z \mid a_1 + z_1 \theta, a_2 + z_2 \theta) = f(\Delta z \mid \Delta a + \theta \Delta z, a_1 + z_1 \theta) = f(\Delta z \mid \Delta a + \theta \Delta z),$$

or in terms of the original variables, that

$$f(\Delta z \mid x_1 \beta, x_2 \beta) = f(\Delta z \mid \Delta x \beta),$$

This is an index sufficiency requirement on the data generating process of the regressors x that is driven by the parameter of interest, β.

Example To illustrate, suppose that z is a single dimensional regressor and that regressors are jointly normal with a restricted covariance matrix allowing for contemporaneous correlation only. Moreover,

$$\begin{pmatrix} a_1 \\ a_2 \\ z_1 \\ z_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{a_1} \\ \mu_{a_2} \\ \mu_{z_1} \\ \mu_{z_2} \end{pmatrix}, \begin{pmatrix} \sigma_{a_1}^2 & 0 & \sigma_{a_1 z_1} & 0 \\ 0 & \sigma_{a_2}^2 & 0 & \sigma_{a_2 z_2} \\ \sigma_{a_1 z_1} & 0 & \sigma_{z_1}^2 & 0 \\ 0 & \sigma_{a_2 z_2} & 0 & \sigma_{z_2}^2 \end{pmatrix}\right).$$

Then

$$\begin{pmatrix} \Delta z \\ x_1\beta \\ x_2\beta \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{12} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{13} & \Sigma_{23} & \Sigma_{33} \end{pmatrix}\right)$$

for

$$\begin{aligned}
\mu_1 &= \mu_{z_2} - \mu_{z_1} \\
\mu_2 &= \mu_{a_1} + \mu_{z_1}\theta \\
\mu_3 &= \mu_{a_2} + \mu_{z_2}\theta
\end{aligned}$$

and

$$\begin{aligned}
\Sigma_{11} &= \mathrm{var}(\Delta z) = \mathrm{var}(z_1) + \mathrm{var}(z_2) \\
\Sigma_{12} &= \mathrm{cov}(\Delta z, x_1\beta) = -\mathrm{cov}(z_1, a_1 + z_1\theta) \\
&= -\mathrm{cov}(a_1, z_1) - \theta\,\mathrm{var}(z_1) \\
&= -\sigma_{a_1 z_1} - \theta\sigma_{z_1}^2 \\
\Sigma_{13} &= \mathrm{cov}(\Delta z, x_2\beta) = \mathrm{cov}(z_2, a_2 + z_2\theta) \\
&= \mathrm{cov}(a_2, z_2) + \theta\,\mathrm{var}(z_2) \\
&= \sigma_{a_2 z_2} + \theta\sigma_{z_2}^2 \\
\Sigma_{22} &= \mathrm{var}(x_1\beta) = \mathrm{var}(a_1 + z_1\theta) \\
&= \mathrm{var}(a_1) + \theta^2\,\mathrm{var}(z_1) + \theta\,2\mathrm{cov}(a_1, z_1) \\
&= \sigma_{a_1}^2 + 2\theta\sigma_{a_1 z_1} + \theta^2\sigma_{z_1}^2 \\
\Sigma_{33} &= \mathrm{var}(x_2\beta) = \mathrm{var}(a_2 + z_2\theta) \\
&= \mathrm{var}(a_2) + \theta^2\,\mathrm{var}(z_2) + \theta\,2\mathrm{cov}(a_2, z_2) \\
&= \sigma_{a_2}^2 + 2\theta\sigma_{a_2 z_2} + \theta^2\sigma_{z_2}^2 \\
\Sigma_{23} &= \mathrm{cov}(x_1\beta, x_2\beta) = 0.
\end{aligned}$$

From standard results on the multivariate normal distribution we have that

$$\Delta z | x_1\beta, x_2\beta$$

is normal with constant variance and conditional mean function

$$m(x_1\beta, x_2\beta) = \mu_1 + \frac{(\Sigma_{13}\Sigma_{22} - \Sigma_{12}\Sigma_{23})(x_2\beta - \mu_3) - (\Sigma_{13}\Sigma_{23} - \Sigma_{12}\Sigma_{33})(x_1\beta - \mu_2)}{\Sigma_{22}\Sigma_{33} - \Sigma_{23}^2}.$$

To satisfy the condition of index sufficiency we need that

$$(\Sigma_{13}\Sigma_{22} - \Sigma_{12}\Sigma_{23}) = (\Sigma_{13}\Sigma_{23} - \Sigma_{12}\Sigma_{33}).$$

Plugging-in the expressions from above, this becomes

$$(\sigma_{a_2 z_2} + \theta\sigma_{z_2}^2)(\sigma_{a_1}^2 + 2\theta\sigma_{a_1 z_1} + \theta^2\sigma_{z_1}^2) = (\sigma_{a_1 z_1} + \theta\sigma_{z_1}^2)(\sigma_{a_2}^2 + 2\theta\sigma_{a_2 z_2} + \theta^2\sigma_{z_2}^2).$$

We can write this condition as the third-order polynomial equation (in θ)

$$C + B\theta + A\theta^2 + D\theta^3 = 0$$

with coefficients

$$\begin{aligned}
C &= \sigma_{a_1}^2\sigma_{a_2 z_2} - \sigma_{a_2}^2\sigma_{a_1 z_1} \\
B &= \sigma_{a_1}^2\sigma_{z_2}^2 + 2\sigma_{a_2 z_2}\sigma_{a_1 z_1} - \sigma_{a_2}^2\sigma_{z_1}^2 - 2\sigma_{a_2 z_2}\sigma_{a_1 z_1} \\
&= \sigma_{a_1}^2\sigma_{z_2}^2 - \sigma_{a_2}^2\sigma_{z_1}^2 \\
A &= \sigma_{a_1 z_1}\sigma_{z_2}^2 - \sigma_{a_2 z_2}\sigma_{z_1}^2 \\
D &= 0.
\end{aligned}$$

For $t = 1, 2$, let

$$\rho_t = \frac{\sigma_{a_t z_t}}{\sigma_{a_t}\sigma_{z_t}}, r_t = \frac{\sigma_{a_t}}{\sigma_{z_t}}.$$

Then

$$\begin{aligned}
\frac{C}{\sigma_{a_1}\sigma_{a_2}\sigma_{z_1}\sigma_{z_2}} &= \rho_2 r_1 - \rho_1 r_2 \\
\frac{B}{\sigma_{a_1}\sigma_{a_2}\sigma_{z_1}\sigma_{z_2}} &= \frac{r_1}{r_2} - \frac{r_2}{r_1} \\
\frac{A}{\sigma_{a_1}\sigma_{a_2}\sigma_{z_1}\sigma_{z_2}} &= \frac{\rho_1}{r_2} - \frac{\rho_2}{r_1}.
\end{aligned}$$

The polynomial condition therefore is

$$(\rho_2 r_1 - \rho_1 r_2) + \left(\frac{r_1}{r_2} - \frac{r_2}{r_1}\right)\theta + \left(\frac{\rho_1}{r_2} - \frac{\rho_2}{r_1}\right)\theta^2 = 0.$$

Note that the leading polynomial coefficient is equal to zero if and only if $\rho_1 r_1 = \rho_2 r_2$. This leads to three mutually-exclusive cases:

(i) The data are stationary, that is, $\rho_1 = \rho_2$ and $r_1 = r_2$. Then all polynomial coefficients are zero so that all values of θ satisfy Lee's restriction.

(ii) We have $\rho_1 r_1 = \rho_2 r_2$ but $r_1 \neq r_2$. Then the resulting linear equation admits one and only one solution in θ.

(iii) The leading polynomial coefficient is non-zero, so, $\rho_1 r_1 \neq \rho_2 r_2$. In this case the discriminant of the second-order polynomial equals

$$\begin{aligned}\Delta &= \left(\frac{r_1}{r_2} - \frac{r_2}{r_1}\right)^2 - 4\left(\frac{\rho_1}{r_2} - \frac{\rho_2}{r_1}\right)(\rho_2 r_1 - \rho_1 r_2) \\ &= \left(\frac{r_1}{r_2}\right)^2 + \left(\frac{r_2}{r_1}\right)^2 - 2 - 4\left(\rho_1\rho_2\left\{\frac{r_1}{r_2} + \frac{r_2}{r_1}\right\} - (\rho_1^2 + \rho_2^2)\right).\end{aligned}$$

Set $x = \frac{r_1}{r_2} \geq 0$ and write

$$\Delta(x) = x^2 + \frac{1}{x^2} - 2 - 4(\rho_1\rho_2(x + \frac{1}{x}) - (\rho_1^2 + \rho_2^2)),$$

which is smooth for $x > 0$. The derivative of Δ with respect to x equals

$$\begin{aligned}\Delta'(x) &= 2x - \frac{2}{x^3} - 4(\rho_1\rho_2(1 - \frac{1}{x^2})) \\ &= \frac{2}{x^3}(x^4 - 1) - 4\rho_1\rho_2\frac{1}{x^2}(x^2 - 1) \\ &= \frac{2}{x^3}(x^2 - 1)(x^2 + 1 - 2\rho_1\rho_2 x).\end{aligned}$$

Note that the Cauchy-Schwarz inequality implies that $x^2 + 1 - 2\rho_1\rho_2 x \geq 0$ so that, for $x \geq 0$,

$$\operatorname{sgn}(\Delta'(x)) = \operatorname{sgn}(x - 1).$$

Further, $\Delta(1) = 4(\rho_1 - \rho_2)^2$. Therefore, $\Delta(x)$ is always non-negative. Hence, in this case, the polynomial condition generically has two solutions in θ.

Conclusion Conditions (g) and (h) of Lee (1999) imply an index-sufficiency condition for the distribution function of regressors. In generic cases in a standard example, this condition is restrictive and is not verified by every possible value of the parameter of interest, θ, but only two.

References

Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225–238.

Chamberlain, G. (2010), "Binary Response Models for Panel Data: Identification and Information," *Econometrica*, 78, 159–168.

Honoré, B. E., and Lewbel, A. (2002), "Semiparametric Binary Choice Panel Data Models Without Strictly Exogeneous Regressors," *Econometrica*, 70, 2053–2063.

Horowitz, J. L. (1992), "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

Klein, R. W., and Spady, R. H. (1993), "An Efficient Semiparametric Estimator for Binary Choice Models," *Econometrica*, 61, 387–421.

Lee, M.-J. (1999), "A Root-N Consistent Semiparametric Estimator for Related-Effects Binary Response Panel Data," *Econometrica*, 67, 427–433.

Magnac, T. (2004), "Panel Binary Variables and Sufficiency: Generalizing Conditional Logit," *Econometrica*, 72, 1859–1876.

Manski, C. F. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–362.

Rasch, G. (1960), "Probabilistic Models for Some Intelligence and Attainment Tests,", Unpublished report, The Danish Institute of Educational Research, Copenhagen.

Sherman, R. P. (1993), "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137.