ORIGINAL ARTICLE

Identification in dynamic binary choice models

Gary Chamberlain[1]

© The Author(s) 2023

Abstract

This paper studies identification in a binary choice panel data model with choice probabilities depending on a lagged outcome, additional observed regressors and an unobserved unit-specific effect. It is shown that with two consecutive periods of data identification is not possible (in a neighborhood of zero), even in the logistic case.

Keywords Panel data · Binary choice · Feedback · Identification

JEL Classification C23 · C33

1 Foreword from the editors

In 1993, Gary Chamberlain released Harvard Institute of Economic Research (HIER) discussion paper no. 1656, titled "Feedback in Panel Data." The material in this paper appears to be a revision and extension of ideas that circulated at least as early as 1991, in a draft paper titled "Sequential Moment Conditions in Panel Data." Some of the material in the earlier, 1991, paper appeared as a, now well-known, comment in the *Journal of Business and Economic Statistics* (Chamberlain 1992). The "Feedback in Panel Data" paper, however, remained unpublished for many years.

 Prior to his passing in 2020, Gary submitted a version of the feedback paper to the *Journal of Econometrics* in connection with a conference that was held in his honor at Harvard University (Chamberlain 2022) (Manuel Arellano was among the participants at this conference). Before its eventual publication in 2022, the "Feedback in Panel

---

---

[1]   Harvard University, Cambridge, USA

Data" paper was not widely circulated and, to our knowledge, unavailable in any form online.

One version of the 1993 paper included a fifth and final section titled "Binary Response." This material was not published as part of Chamberlain (2022). The "Binary Response" section contained two sets of results. The first, dealing with the identification of static binary choice models, can be found in Chamberlain (2010). The second part, which is reproduced below, deals with the identification of dynamic binary choice models.

Gary establishes that with two consecutive observations of a binary dependent variable, identification is not possible (in a neighborhood of zero), even in the logistic case. This result is not widely known, although it has been cited by, for example, Honoré and Kyriazidou (2000). Publishing this material makes it easily available to researchers worldwide for the first time.

Gary was a great admirer of Manuel Arellano's research, and they engaged in many intellectual exchanges over the years. Gary would no doubt have wanted to celebrate and honor Manuel's considerable achievements and scholarship. Including this short note in this special issue provides a way to do so.

Stéphane Bonhomme, Bryan Graham and Laura Hospido.

2 Identification in dynamic binary response

The outcome variable is binary. There are two periods of observation on each unit ($T = 2$). The random vector $(y_{i1}, y_{i2}, x'_{i1}, x'_{i2}, c_i)$ is independently and identically distributed for $i = 1, ..., n$. We observe $z'_i = (y_{i1}, y_{i2}, x'_{i1}, x'_{i2})$; the (scalar) latent variable $c_i \in \mathcal{R}$ is not observed. The binary variable $y_{it} = 0$ or 1, and $x'_i = (x'_{i1}, x'_{i2})$ has support $X \subset \mathcal{R}^J \times \mathcal{R}^J$.

We assume that

$$
\begin{aligned}
\Pr(y_{i1} = 1 | x_i, c_i) &= F(\beta'_0 x_{i1} + c_i) \\
\Pr(y_{i2} = 1 | x_i, y_{i1}, c_i) &= F(\alpha_0 + \beta'_0 x_{i2} + \gamma_0 y_{i1} + c_i).
\end{aligned}
\tag{1}
$$

The distribution function F is given as part of the prior specification; it is strictly increasing on the whole real line with a bounded, continuous derivative, and with $\lim_{s \to \infty} F(s) = 1$ and $\lim_{s \to -\infty} F(s) = 0$. The parameter space is $\Theta = \Theta_1 \times \Theta_2 \times \Theta_3$, where Θ_1 is an open subset of $\mathcal{R}$, Θ_2 is an open subset of $\mathcal{R}^J$, and Θ_3 is an open subset of $\mathcal{R}$, and $\theta'_0 \equiv (\alpha_0, \beta'_0, \gamma_0) \in \Theta$. We assume that Θ contains a neighborhood of 0. Define

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

Theorem 1 *If X is bounded, then there is a point $(\alpha, \gamma) \in \Theta_1 \times \Theta_3$ such that identification fails in (1) for all θ_0 in a neighborhood of $(\alpha, 0, \gamma)$.*

Proof Define

$$H(x, c_1, ..., c_4, \theta) = [p(x, c_1, \theta), ..., p(x, c_4, \theta)].$$

Suppose that for some $(\alpha, \gamma) \in \Theta_1 \times \Theta_3$ and points $c_1, ..., c_4 \in \mathcal{R}$, we have $H(x, c_1, ..., c_4, \theta^*)$ nonsingular, where $\theta^* = (\alpha, 0, \gamma)$. Consider a convex combination

$$a = \sum_{j=1}^{4} p(x, c_j, \theta^*)\pi_j^* = H(x, c_1, ..., c_4, \theta^*)\pi^*,$$

where $\pi^{*\prime} = (\pi_1^*, ..., \pi_4^*)$, $\pi_j^* > 0$, and $\sum_{j=1}^{4} \pi_j^* = 1$. Let

$$\pi_0(x) = H(x, c_1, ..., c_4, \theta_0)^{-1}a,$$

where $\theta_0 \neq \theta^*$ is any point in Θ sufficiently close to θ^* that the inverse exists and $\pi_{0j}(x) > 0$ for all $x \in X$; there is such a neighborhood of θ^* because X is bounded. Note that $l'H = l'$, where l is a 4×1 vector of ones; hence $l'H^{-1} = l'$ and $l'\pi_0(x) = 1$. Hence

$$\sum_{j=1}^{4} p(x, c_j, \theta_0)\pi_{0j}(x) = \sum_{j=1}^{4} p(x, c_j, \theta^*)\pi_j^*,$$

and we cannot distinguish θ_0 from θ^*.

We conclude that for every $(\alpha, \gamma) \in \Theta_1 \times \Theta_3$ we must have $H(x, c_1, ..., c_4, (\alpha, 0, \gamma))$ singular for every choice of $c_1, ..., c_4 \in \mathcal{R}$. Otherwise identification fails for all θ_0 in some neighborhood of $(\alpha, 0, \gamma)$. So there must exist scalars $\psi_1, ..., \psi_4$ (not all zero) such that

$$\begin{aligned} \psi' p(x, c, (\alpha, 0, \gamma)) = {} &\psi_1 \left[1 - F(c)\right]\left[1 - F(\alpha + c)\right] + \psi_2 \left[1 - F(c)\right] F(\alpha + c) \\ &+ \psi_3 F(c)\left[1 - F(\alpha + \gamma + c)\right] + \psi_4 F(c) F(\alpha + \gamma + c) = 0 \end{aligned}$$

for all $c \in \mathcal{R}$. Letting $c \to \infty$ gives $\psi_4 = 0$. Letting $c \to -\infty$ gives $\psi_1 = 0$. Hence we have

$$\psi_2 \left[1 - F(c)\right] F(\alpha + c) + \psi_3 F(c)\left[1 - F(\alpha + \gamma + c)\right] = 0$$

for all $(\alpha, \gamma) \in \Theta_1 \times \Theta_3$ and all $c \in \mathcal{R}$.

Set $\gamma = 0$. Letting $Q \equiv F/(1 - F)$, we must have

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

setting $c = 0$ gives $\psi_3/\psi_2 = -Q(\alpha)/Q(0)$, and so

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

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

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

then, from $F = Q/(1 + Q)$ it follows

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

Now set $\alpha = 0$. Then, with $G = 1 - F$, we have

$$\psi_2 G(c) + \psi_3 G(\gamma + c) = 0;$$

setting $c = 0$ gives $\psi_2/\psi_3 = -G(\gamma)/G(0)$, and so

$$G(\gamma + c) = G(\gamma)G(c)/G(0)$$

for all $\gamma \in \Theta_3$ and all $c \in \mathcal{R}$. The only positive, continuously differentiable solution to this form of Cauchy's equation is

$$G(s) = \exp(\tau_1 + \tau_2 s),$$

which implies that

$$F(s) = 1 - \exp(\tau_1 + \tau_2 s).$$

So we have

$$\exp(\tau_1 + \tau_2 s) = 1 - F(s) = \left[1 + \exp(\phi_1 + \phi_2 s)\right]^{-1},$$
$$\tau_1 + \tau_2 s = -\log\left[1 + \exp(\phi_1 + \phi_2 s)\right],$$

and taking the derivative with respect to s gives

$$\tau_2 = -\phi_2 F(s).$$

But this is a contradiction, since F is strictly increasing. So it is not possible to maintain the singularity condition on H, and identification fails for all θ_0 in some neighborhood of $(\alpha, 0, \gamma)$.   Q.E.D.   □

Data availability There are no data or codes in this paper.

Declarations

Conflict of interest Not applicable.

References

Chamberlain G (1992) Comment: sequential moment restrictions in panel data. J Bus Econ Stat 10(2):20–26

Chamberlain G (2010) Binary response models for panel data: identification and information. Econometrica 78(1):159–168

Chamberlain G (2022) Feedback in panel data models. J Econom 226(1):4–20

Honoré BE, Kyriazidou E (2000) Panel data discrete choice models with lagged dependent variables. Econometrica 68(4):839–874

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.