NBER WORKING PAPER SERIES

IDENTIFICATION IN A BINARY CHOICE PANEL DATA MODEL WITH A PREDETERMINED COVARIATE

Stéphane Bonhomme
Kevin Dano
Bryan S. Graham

Working Paper 31027
http://www.nber.org/papers/w31027

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2023

We are grateful to Isaiah Andrews, Manuel Arellano, Jesus Carro, Bo Honoré, and the audiences at the conference in honor of Manuel Arellano at the Bank of Spain (July 2022) and other seminars for valuable comments that have improved the paper. All the usual disclaimers apply. Portions of the research reported here were undertaken while Bonhomme and Graham were visiting CEMFI with support from the Spanish State Research Agency under the María de Maeztu Unit of Excellence Programme (Project No: CEX2020-001104-M). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Identification in a Binary Choice Panel Data Model with a Predetermined Covariate
Stéphane Bonhomme, Kevin Dano, and Bryan S. Graham
NBER Working Paper No. 31027
March 2023
JEL No. C23

ABSTRACT

We study identification in a binary choice panel data model with a single predetermined binary covariate (i.e., a covariate sequentially exogenous conditional on lagged outcomes and covariates). The choice model is indexed by a scalar parameter θ, whereas the distribution of unit-specific heterogeneity, as well as the feedback process that maps lagged outcomes into future covariate realizations, are left unrestricted. We provide a simple condition under which θ is never point-identified, no matter the number of time periods available. This condition is satisfied in most models, including the logit one. We also characterize the identified set of θ and show how to compute it using linear programming techniques. While θ is not generally point-identified, its identified set is informative in the examples we analyze numerically, suggesting that meaningful learning about θ is possible even in short panels with feedback.

Stéphane Bonhomme
University of Chicago
1126 East 59th Street
Chicago, IL 60637
sbonhomme@uchicago.edu

Kevin Dano
University of California, Berkeley
530 Evans Hall
Berkeley, CA 94720-3880
kdano@berkeley.edu

Bryan S. Graham
University of California - Berkeley
530 Evans Hall #3880
Berkeley, CA 94720-3880
and NBER
bgraham@econ.berkeley.edu

1 Introduction

Empirical researchers utilizing panel data generally maintain the assumption that covariates are strictly exogenous: realized values of past, current, and future explanatory variables are independent of the time-varying structural disturbances or "shocks".[1] In many settings this assumption is unrealistic. If the covariate is a policy, choice or dynamic state variable, then agents may adjust its level in response to past shocks (as when, for example, a firm adjusts its current capital expenditures in response to past productivity shocks).

When strict exogeneity is untenable, *sequential exogeneity* – sometimes called *predeterminedness* – may be palatable. A predetermined covariate varies independently of current and future time-varying shocks, but general *feedback*, or dependence on past shocks, is allowed. Assumptions of this type play an important role in, for example, production function estimation (Olley and Pakes, 1996; Blundell and Bond, 2000).

In two seminar papers, Arellano and Bond (1991) and Arellano and Bover (1995), Manuel Arellano and his collaborators presented foundational analyses of questions of identification, estimation, efficiency and specification testing in linear panel data models with feedback. Today such models are both well-understood and widely-used (see Arellano (2003) for a textbook review).

In contrast, the properties of nonlinear models with feedback are much less well-understood. In this paper we study binary choice. Most existing work in this area focuses on the case where the covariate is either strictly exogenous or a lagged outcome. Under strict exogeneity, Rasch (1960) and Andersen (1970) show that the coefficient on the covariate is point-identified using two periods of data when shocks are logistic. Chamberlain (2010) provides conditions under which the logit case is the only one admitting point-identification with two periods (Davezies et al. (2020) provide extensions of this result to the case of $T > 2$). In the dynamic case, where the covariate is a lagged outcome, Cox (1958), Chamberlain (1985) and Honoré and Kyriazidou (2000) derive conditions for point-identification of the coefficient on the lagged outcome in the logit case, while Honoré and Tamer (2006) show how to compute bounds on coefficients.

Results for binary choice panel models with predetermined covariates are scarce. Chamberlain (2022) studies identification and semiparametric efficiency bounds in a class of nonlinear panel data models with feedback; he provides both positive and negative results. In an hitherto unpublished section of an early draft of that paper (Chamberlain, 1993), he proves that the coefficient on a lagged outcome is not point-identified in a dynamic logit model when only three periods of outcome data are available. Arellano and Carrasco (2003) and Honoré

[1] Dependence between the covariates and the time-invariant heterogeneity – the so-called "fixed effects" – is, of course, allowed.

and Lewbel (2002) study binary choice models with predetermined covariates. Arellano and Carrasco (2003) assume that the dependence between the time-invariant heterogeneity and the covariates is fully characterized by its conditional mean given current and lagged covariates. Honoré and Lewbel (2002) assume that one of the covariates is independent of the individual effects conditional on the other covariates.

In what follows we pose two questions. First, under what conditions is the coefficient on a predetermined covariate in a binary choice panel data model point-identified? Second, when the coefficient is only set-identified, how extreme is the failure of point-identification, i.e., what is the width of the identified set?

Our analyses leave the dependence between the (time-invariant) unit-specific heterogeneity and the covariates unrestricted. We focus on the special case of a single binary predetermined covariate, leaving the feedback process from lagged outcomes, covariates and the unit-specific heterogeneity onto future covariate realizations fully unrestricted. This is a substantial relaxation of the strict exogeneity assumption.

Regarding point-identification, we provide a simple condition on the model which guarantees that point-identification fails when T periods of data are available (and T is fixed). The condition is satisfied in most familiar models of binary choice, including the logit one. This finding contrasts with the prior work on logit models cited above, where point-identification typically holds for a sufficiently long panel. The exponential binary choice model introduced by Al-Sadoon et al. (2017) does not satisfy our condition. In fact, point-identification holds in that case.

Regarding identified sets, we first show that sharp bounds on the coefficient can be computed using linear programming techniques. Our method builds on Honoré and Tamer (2006), however in contrast to their work, we allow for heterogeneous feedback. Second, we numerically compute examples of identified sets. We find that, relative to the strictly exogenous case, allowing for a predetermined covariate tends to increase the width of the identified set. However, our calculations also suggest that the identified set can remain informative under predeterminedness, even in panels with as few as two periods. Finally, as is true under strict exogeneity, the width of the identified set decreases quickly as the number of periods increases (in the examples we consider).

The balance of this paper proceeds as follows. In Section 2 we present the model. In Section 3 we provide a condition that implies that the common parameter in this model is not point-identified when $T = 2$. In Section 4 we show that our condition implies failure of point-identification for all (finite) T. In Section 5 we show how to compute identified sets, and we report the results of a small set of numerical illustrations. In Section 6 we describe potential restrictions one could impose on the feedback process. These restrictions may restore point identification or shrink the identified set. Section 7 contains the customary conclusion which signals the imminent end of our paper. Proofs are contained in the Appendix. Lastly, replication codes are available as supplementary material.

2 The model

Available to the econometrician is a random sample of n units, each of which is followed for $T \geq 2$ time periods. In the identification analysis we focus on short panels, and keep T fixed while considering a large-n population of individual units.

For any sequence of random variables Z_t and any non-stochastic sequence z_t, we use the shorthand notation $Z^{t:t+s} = (Z_t', ..., Z_{t+s}')'$ and $z^{t:t+s} = (z_t', ..., z_{t+s}')'$. In addition, we simply denote $Z^t = Z^{1:t}$ and $z^t = z^{1:t}$ when the subsequence starts in the first period.

Let $Y_{it} \in \{0, 1\}$ and $X_{it} \in \{0, 1\}$ denote a binary outcome and a binary covariate, respectively. We assume that

$$\Pr(Y_{it} = 1 \,|\, Y_i^{t-1}, X_i^t, \alpha_i; \theta) = F(\theta X_{it} + \alpha_i), \quad t = 1, \ldots, T,$$

where $\alpha_i \in \mathcal{S}_\alpha$ is a scalar individual effect, $F(\cdot)$ is a known differentiable cumulative distribution function, and $\theta \in \Theta$ is a scalar parameter.

We leave the distribution of $\alpha_i \,|\, X_{i1}$ unrestricted. In addition, for each $t \geq 2$ we leave the distribution of $X_{it} \,|\, Y_i^{t-1}, X_i^{t-1}, \alpha_i$ unrestricted. Let

$$\Pr\left(X_{it} = 1 \,|\, Y_i^{t-1} = y^{t-1}, X_i^{t-1} = x^{t-1}, \alpha_i = \alpha\right) = G^t_{y^{t-1},x^{t-1}}(\alpha), \quad t = 2, \ldots, T,$$

denote the *feedback process* through which lagged outcomes, past covariates and heterogeneity affect the current covariate. We leave this process unrestricted, and only assume that $G^t_{y^{t-1},x^{t-1}}(\alpha) \in (0, 1)$. We denote as $G \in \mathcal{G}_T$ the collection of all $G^t_{y^{t-1},x^{t-1}}(\alpha)$, for all $t \in \{2, ..., T\}$, $y^{t-1} \in \{0, 1\}^{t-1}$, $x^{t-1} \in \{0, 1\}^{t-1}$, and all α values.

To keep the formal analysis simple, throughout we assume that α_i takes one of K values, with known support points $\mathcal{S}_\alpha = \{\underline{\alpha}_1, ..., \underline{\alpha}_K\}$. This makes the model fully parametric. However this is not a limitation as our aim is to derive conditions under which point identification *fails*. The conditions we provide will require sufficiently many support points. Define

$$\Pr\left(\alpha_i = \alpha \,|\, X_{i1} = x_1\right) = \pi_{x_1}(\alpha),$$

the *distribution of heterogeneity* given the initial condition x_1. We leave this object unrestricted, only assuming that it belongs to the unit simplex on $\mathcal{S}_\alpha$. We rely on the parameterization given by the $2(K-1)\times 1$ vector $\pi=(\pi_1',\pi_0')'$, where, for all $x_1\in\{0,1\}$, $\pi_{x_1}=(\pi_{x_1}(\underline{\alpha}_1),\ldots,\pi_{x_1}(\underline{\alpha}_{K-1}))'$ and $\pi_{x_1}(\underline{\alpha}_K)=1-\sum_{k=1}^{K-1}\pi_{x_1}(\underline{\alpha}_k)$. The vector $\pi\in\Pi$ is unrestricted, except for the fact that $\pi_{x_1}(\alpha)$, for $\alpha\in\mathcal{S}_\alpha$, belongs to the unit simplex. This parameterization handles the fact that probability mass functions sum to one.

The (integrated) likelihood function conditional on the first period's covariate is

$$\begin{aligned}\Pr\left(Y_i^T=y^T,X_i^{2:T}=x^{2:T}\,|\,X_{i1}=x_1\right)=\sum_{\alpha\in\mathcal{S}_\alpha}&\underbrace{\prod_{t=1}^{T}F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t}}_{\text{outcomes}}\\ &\times\underbrace{\prod_{t=2}^{T}G^t_{y^{t-1},x^{t-1}}(\alpha)^{x_t}[1-G^t_{y^{t-1},x^{t-1}}(\alpha)]^{1-x_t}}_{\text{feedback}}\\ &\times\underbrace{\pi_{x_1}(\alpha)}_{\text{heterogeneity}}\ .\end{aligned}\tag{1}$$

A key feature of a model with predetermined covariates is the dependence of the feedback process on lagged outcomes, as reflected in the dependence of G^t on y^{t-1} in (1). When this dependence is ruled out, the covariate is strictly exogenous, and the likelihood function simplifies.[2] Dynamic responses of covariates to lagged outcome realizations are central to many economic models, including those where X_{it} is a choice variable, policy, or a dynamic state variable.

For any $(\theta,\pi,G)\in\Theta\times\Pi\times\mathcal{G}_T$, and any $(y^T,x^{2:T})\in\{0,1\}^{2T-1}$, let $Q_{x_1}(y^T,x^{2:T};\theta,\pi,G)$ denote the right-hand side of (1). Moreover, let $Q_{x_1}(\theta,\pi,G)$ denote the $2^{2T-1}\times 1$ vector collecting all those elements, for all $(y^T,x^{2:T})\in\{0,1\}^{2T-1}$. Finally, let $Q(\theta,\pi,G)$ denote the $2^{2T}\times 1$ vector stacking $Q_1(\theta,\pi,G)$ and $Q_0(\theta,\pi,G)$. For a given $(\theta,\pi,G)\in\Theta\times\Pi\times\mathcal{G}_T$, we define the *identified set* of θ as

$$\Theta^I=\left\{\widetilde{\theta}\in\Theta\,:\,\exists(\widetilde{\pi},\widetilde{G})\in\Pi\times\mathcal{G}_T\,:\,Q(\widetilde{\theta},\widetilde{\pi},\widetilde{G})=Q(\theta,\pi,G)\right\}.\tag{2}$$

The set in (2) includes all $\widetilde{\theta}\in\Theta$ where, conditional on $\widetilde{\theta}$, it is possible to find a heterogeneity distribution $\widetilde{\pi}\in\Pi$, *and* a feedback process $\widetilde{G}\in\mathcal{G}_T$, such that the resulting conditional

[2]Under strict exogeneity, the likelihood function factors as

$$\begin{aligned}\Pr\left(Y_i^T=y^T,X_i^{2:T}=x^{2:T}\,|\,X_{i1}=x_1\right)=&\left[\sum_{\alpha\in\mathcal{S}_\alpha}\prod_{t=1}^{T}F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t}\pi_{x^T}(\alpha)\right]\\ &\times\Pr\left(X_i^{2:T}=x^{2:T}\,|\,X_{i1}=x_1\right),\end{aligned}$$

where $\pi_{x^T}(\alpha)=\Pr(\alpha_i=\alpha\,|\,X_i^T=x^t)$.

likelihood assigns the same probability to each of the 2^{2T-1} possible data outcomes as the true one (given both $X_{i1} = 0$ and $X_{i1} = 1$).

In the first part of the paper, we provide conditions on the model under which Θ^I is not a singleton. This corresponds to cases where θ is not point-identified. In the second part of the paper, we report numerical calculations of Θ^I for particular parameter values.

Our focus on θ is motivated by the extensive literature on the identification of coefficients in binary choice models. However, in applications average marginal effects may also be of interest. We leave the identification analysis of such quantities in models with predetermined covariates to future work.

3 Failure of point-identification in two-period panels

We first present an analysis of point-identification in the two-period case, since this leads to simple and transparent calculations. In the next section we generalize this result to accommodate $T \geq 2$ periods.

3.1 Assumptions and result

We start by imposing the following assumption on the population parameters.

Assumption 1. *$\theta \in \Theta$, $\pi \in \Pi$, and $G \in \mathcal{G}_T$ are all interior, and $F(\theta x + \alpha) \in (0, 1)$ for all $x \in \{0, 1\}$ and $\alpha \in \mathcal{S}_\alpha$.*

Assumption 1 places restrictions on the underlying parametric binary choice model and heterogeneity distribution. It rules out heterogeneity distributions which induce a point mass of "stayers" (i.e., units with such extreme values of α that they either always take the binary action or they never do).

Next, we assume that the parameter point is regular in the sense of Rothenberg (1971).

Assumption 2. *(θ, π, G) is a regular point of the Jacobian matrix $\nabla Q(\theta, \pi, G)$.*

The assumption of regularity is standard in the literature on the identification of parametric models (Rothenberg, 1971). In Assumption 2, it means that the rank of $\nabla Q(\widetilde{\theta}, \widetilde{\pi}, \widetilde{G})$ is constant for all $(\widetilde{\theta}, \widetilde{\pi}, \widetilde{G})$ in an open neighborhood of (θ, π, G). If $F(\cdot)$ is analytic, the irregular points of $\nabla Q(\theta, \pi, G)$ form a set of measure zero (Bekker and Wansbeek, 2001). Thus, Assumption 2 is satisfied almost everywhere in the parameter space in many binary choice models, including the probit and logit ones.

We aim to provide a simple condition under which point-identification of θ fails when $T = 2$. We start by observing that, when $T = 2$, the $2^{2T-1} = 8$ model outcome probabilities given $X_{i1} = x_1$ are

$$Q_{x_1}(\theta, \pi, G) = \begin{pmatrix} \Pr(Y_{i2} = 1, X_{i2} = 1, Y_{i1} = 1 \mid X_{i1} = x_1; \theta, \pi, G) \\ \Pr(Y_{i2} = 1, X_{i2} = 1, Y_{i1} = 0 \mid X_{i1} = x_1; \theta, \pi, G) \\ \Pr(Y_{i2} = 1, X_{i2} = 0, Y_{i1} = 1 \mid X_{i1} = x_1; \theta, \pi, G) \\ \Pr(Y_{i2} = 1, X_{i2} = 0, Y_{i1} = 0 \mid X_{i1} = x_1; \theta, \pi, G) \\ \Pr(Y_{i2} = 0, X_{i2} = 1, Y_{i1} = 1 \mid X_{i1} = x_1; \theta, \pi, G) \\ \Pr(Y_{i2} = 0, X_{i2} = 1, Y_{i1} = 0 \mid X_{i1} = x_1; \theta, \pi, G) \\ \Pr(Y_{i2} = 0, X_{i2} = 0, Y_{i1} = 1 \mid X_{i1} = x_1; \theta, \pi, G) \\ \Pr(Y_{i2} = 0, X_{i2} = 0, Y_{i1} = 0 \mid X_{i1} = x_1; \theta, \pi, G) \end{pmatrix},$$

which, given the structure of the model, coincide with

$$Q_{x_1}(\theta, \pi, G) = \begin{pmatrix} \sum_{\alpha \in \mathcal{S}_\alpha} F(\theta + \alpha) G^2_{1,x_1}(\alpha) F(\theta x_1 + \alpha) \pi_{x_1}(\alpha) \\ \sum_{\alpha \in \mathcal{S}_\alpha} F(\theta + \alpha) G^2_{0,x_1}(\alpha) [1 - F(\theta x_1 + \alpha)] \pi_{x_1}(\alpha) \\ \sum_{\alpha \in \mathcal{S}_\alpha} F(\alpha) [1 - G^2_{1,x_1}(\alpha)] F(\theta x_1 + \alpha) \pi_{x_1}(\alpha) \\ \sum_{\alpha \in \mathcal{S}_\alpha} F(\alpha) [1 - G^2_{0,x_1}(\alpha)] [1 - F(\theta x_1 + \alpha)] \pi_{x_1}(\alpha) \\ \sum_{\alpha \in \mathcal{S}_\alpha} [1 - F(\theta + \alpha)]\, G^2_{1,x_1}(\alpha) F(\theta x_1 + \alpha) \pi_{x_1}(\alpha) \\ \sum_{\alpha \in \mathcal{S}_\alpha} [1 - F(\theta + \alpha)]\, G^2_{0,x_1}(\alpha) [1 - F(\theta x_1 + \alpha)] \pi_{x_1}(\alpha) \\ \sum_{\alpha \in \mathcal{S}_\alpha} [1 - F(\alpha)]\, [1 - G^2_{1,x_1}(\alpha)] F(\theta x_1 + \alpha) \pi_{x_1}(\alpha) \\ \sum_{\alpha \in \mathcal{S}_\alpha} [1 - F(\alpha)]\, [1 - G^2_{0,x_1}(\alpha)] [1 - F(\theta x_1 + \alpha)] \pi_{x_1}(\alpha) \end{pmatrix}. \tag{3}$$

With this notation in hand we present the following lemma.

Lemma 1. *Let $T = 2$. Suppose that Assumptions 1 and 2 hold, and that θ is point-identified. Then, there exists $x_1 \in \{0, 1\}$ and a non-zero function $\phi_{x_1} : \{0, 1\}^3 \to \mathbb{R}$ such that:*

(i) for all $\alpha \in \mathcal{S}_\alpha$ and $y_1 \in \{0, 1\}$,

$$\sum_{y_2=0}^{1} \phi_{x_1}(y_1, y_2, 1) F(\theta + \alpha)^{y_2} [1 - F(\theta + \alpha)]^{1-y_2} = \sum_{y_2=0}^{1} \phi_{x_1}(y_1, y_2, 0) F(\alpha)^{y_2} [1 - F(\alpha)]^{1-y_2}; \tag{4}$$

(ii) for all $\alpha \in \mathcal{S}_\alpha$ and $x_2 \in \{0,1\}$,

$$\sum_{y_2=0}^{1}\sum_{y_1=0}^{1} \phi_{x_1}(y_1,y_2,x_2)F(\theta x_2+\alpha)^{y_2}[1-F(\theta x_2+\alpha)]^{1-y_2}F(\theta x_1+\alpha)^{y_1}[1-F(\theta x_1+\alpha)]^{1-y_1}=0. \tag{5}$$

The proof of Lemma 1 exploits the fact that, if θ is point-identified, then it is also locally point-identified. Together with the assumption that the parameter is regular, this allows us to apply a result of Bekker and Wansbeek (2001) regarding the identification of subvectors, which guarantees the existence of some $x_1 \in \{0,1\}$ such that $\nabla_{\theta'} Q_{x_1}$ does not belong to the range of the matrix $\begin{bmatrix}\nabla_{\pi'_{x_1}} Q_{x_1} & \nabla_{G'_{x_1}} Q_{x_1}\end{bmatrix}$. We then show, using (3), that this implies the existence of $\phi_{x_1} \neq 0$ such that (4) and (5) hold.

It is instructive to observe that condition (4) in Lemma 1 corresponds to the conditional moment restriction

$$\mathbb{E}\left[\phi_{X_{i1}}\left(Y_{i1},Y_{i2},X_{i2}\right)\middle|\, X_{i1},X_{i2},Y_{i1},\alpha_i\right]=\mathbb{E}\left[\phi_{X_{i1}}\left(Y_{i1},Y_{i2},X_{i2}\right)\middle|\, X_{i1},Y_{i1},\alpha_i\right],$$

while (5) implies the additional requirement that

$$\mathbb{E}\left[\phi_{X_{i1}}\left(Y_{i1},Y_{i2},X_{i2}\right)\middle|\, X_{i1},\alpha_i\right]=0.$$

This formulation clarifies that a necessary condition for point identification of θ is the existence of a non-zero moment function, $\phi_{X_{i1}}\left(Y_{i1},Y_{i2},X_{i2}\right)$, with a mean that is invariant to X_{i2} given α_i and the past (i.e., the first period's covariate and outcome). Such a moment function is "feedback robust", in the sense that it remains valid across all possible feedback processes. This is the content of condition (4) in Lemma 1, while (5) imposes a similar invariance to the distribution of unobserved heterogeneity.

To show that point identification fails, our focus here, we need to show that no such non-zero moment function exists. It turns out that there is a very simple condition for this in our model. Specifically, from Lemma 1 we obtain the following corollary.

Corollary 1. *Let $T=2$. Suppose that Assumptions 1 and 2 hold, and that 1, $F(\alpha)$, and $F(\theta+\alpha)$, for $\alpha \in \mathcal{S}_\alpha$, are linearly independent, then θ is not point-identified.*

Corollary 1 shows that a necessary condition for identification of θ is that 1, $F(\alpha)$, and $F(\theta+\alpha)$, for $\alpha \in \mathcal{S}_\alpha$, are linearly dependent. This is a restrictive condition, as we show in the next subsection.[3]

[3] Note that, when $\theta \neq 0$, this condition cannot hold on the entire real line whenever F is strictly increasing

Remark 1. *Despite the negative result of Corollary 1, the sign of θ is identified provided that $F(\cdot)$ is strictly increasing. Specifically, we show in Appendix C that*

$$\textit{sign}(\theta) = \textit{sign}\left(\mathbb{E}\left[Y_{i2} - Y_{i1} \mid X_{i1} = 0\right]\right) = \textit{sign}\left(\mathbb{E}\left[Y_{i1} - Y_{i2} \mid X_{i1} = 1\right]\right).$$

3.2 The logit model

Consider the logit model with a binary predetermined covariate, which corresponds to $F(u) = \frac{e^u}{1+e^u}$. In this case, the linear dependence condition of Corollary 1 requires that, for some non-zero triplet (A, B, C),

$$A\frac{e^{\theta+\alpha}}{1+e^{\theta+\alpha}} + B\frac{e^{\alpha}}{1+e^{\alpha}} + C = 0, \quad \text{for all } \alpha \in \mathcal{S}_\alpha.$$

However, this implies

$$Ae^\theta e^\alpha(1+e^\alpha) + Be^\alpha(1+e^\theta e^\alpha) + C(1+e^\alpha)(1+e^\theta e^\alpha) = 0, \quad \text{for all } \alpha \in \mathcal{S}_\alpha,$$

which is a quadratic polynomial equation in e^α. Therefore, provided that there are $K \geq 3$ values in $\mathcal{S}_\alpha$, this implies

$$Ae^\theta + Be^\theta + Ce^\theta = 0, \quad Ae^\theta + B + (1+e^\theta)C = 0, \quad C = 0,$$

which, provided that $\theta \neq 0$, entails

$$A = B = C = 0,$$

contradicting the assumption that (A, B, C) is non-zero.

We have thus proved the following corollary.

Corollary 2. *Consider the logit model with $T = 2$. Suppose that Assumptions 1 and 2 hold, that $\theta \neq 0$, and that $\mathcal{S}_\alpha$ contains at least three points, then θ is not point-identified.*

A precedent to Corollary 2 is given in the unpublished working paper by Chamberlain (1993) mentioned in the introduction. In the model he considers, $X_{it} = Y_{i,t-1}$ is a lagged

on $\mathbb{R}$. Indeed, if 1, $F(\alpha)$, and $F(\theta+\alpha)$, for $\alpha \in \mathbb{R}$, are linearly dependent, then for some non-zero triplet (A, B, C) we have

$$AF(\theta+\alpha) + BF(\alpha) + C = 0, \quad \text{for all } \alpha \in \mathbb{R}.$$

This implies, by taking $\alpha \to \pm\infty$ that $C = 0$, and $A + B = 0$, which cannot hold if $\theta \neq 0$ unless $A = B = C = 0$, contradicting the assumption that (A, B, C) is non-zero.

outcome, and $T = 2$ (hence, outcomes are observed for three periods). His model also includes an additional regressor: an indicator for period $t = 2$.

3.3 The exponential model

Suppose now that, for $u \geq 0$, $F(u) = 1 - e^{-u}$. This corresponds to the exponential binary choice model of Al-Sadoon et al. (2017). Note that here the support of $F(\cdot)$ is a strict subset of the real line. In this case, letting

$$A = e^{\theta}, B = -1, C = 1 - e^{\theta},$$

we have

$$A[1 - e^{-(\theta+\alpha)}] + B[1 - e^{-\alpha}] + C = 0.$$

Hence the non point-identification condition of Corollary 1 is not satisfied in the exponential binary choice model.

In fact, in this case (4) and (5) are satisfied for

$$\phi_{x_1}(y_1, y_2, x_2) = (1 - y_2)e^{\theta x_2} - (1 - y_1)e^{\theta x_1},$$

and θ is point-identified based on the conditional moment restriction

$$\mathbb{E}[\phi_{X_{i1}}(Y_{i1}, Y_{i2}, X_{i2}) \mid X_{i1}] = 0,$$

that is,

$$\mathbb{E}[(1 - Y_{i2})e^{\theta X_{i2}} - (1 - Y_{i1})e^{\theta X_{i1}} \mid X_{i1}] = 0.$$

See Wooldridge (1997) for several related results.

4 Failure of point-identification in T-period panels for $T > 2$

In this section we generalize our analysis to an arbitrary number of periods and state our main result.

4.1 Main result

The arguments laid out in the previous section extend to an arbitrary number of time periods, $T \geq 2$. Indeed, using a similar strategy to the proof of Lemma 1 and proceeding by induction, we obtain the following lemma.

Lemma 2. *Let $T \geq 2$. Suppose that Assumptions 1 and 2 hold, and that θ is point-identified. Then, there exists $x_1 \in \{0,1\}$ and a non-zero function $\phi_{x_1} : \{0,1\}^{2T-1} \to \mathbb{R}$ such that:*

(i) for all $\alpha \in \mathcal{S}_\alpha$, $s \in \{0,...,T-2\}$, $y^{T-(s+1)} \in \{0,1\}^{T-(s+1)}$, $x^{T-(s+1)} \in \{0,1\}^{T-(s+1)}$,

$$\sum_{y^{T-s:T} \in \{0,1\}^{s+1}} \phi_{x_1}(y^T, x^{2:T}) \prod_{t=T-s}^{T} F(\theta x_t + \alpha)^{y_t} [1 - F(\theta x_t + \alpha)]^{1-y_t} \tag{6}$$

does not depend on $x^{T-s:T}$;

(ii) for all $\alpha \in \mathcal{S}_\alpha$ and $x^{2:T} \in \{0,1\}^{T-1}$,

$$\sum_{y^T \in \{0,1\}^T} \phi_{x_1}(y^T, x^{2:T}) \prod_{t=1}^{T} F(\theta x_t + \alpha)^{y_t} [1 - F(\theta x_t + \alpha)]^{1-y_t} = 0. \tag{7}$$

Similarly to Lemma 1, Lemma 2 implies the existence of a moment function that is "feedback robust", in the sense that, for all $s \in \{0,...,T-2\}$,

$$\mathbb{E}\left[\phi_{X_{i1}}(Y_i^T, X_i^{2:T}) \,|\, X_i^T, Y_i^{T-(s+1)}, \alpha_i\right] = \mathbb{E}\left[\phi_{X_{i1}}(Y_i^T, X_i^{2:T}) \,|\, X_i^{T-(s+1)}, Y_i^{T-(s+1)}, \alpha_i\right],$$

while also requiring that

$$\mathbb{E}\left[\phi_{X_{i1}}(Y_i^T, X_i^{2:T}) \,|\, X_{i1}, \alpha_i\right] = 0.$$

From Lemma 2 we obtain the following corollary, which we also prove by induction. This is our main result.

Corollary 3. *Let $T \geq 2$. Suppose that Assumptions 1 and 2 hold, and that 1, $F(\alpha)$, and $F(\theta + \alpha)$, for $\alpha \in \mathcal{S}_\alpha$, are linearly independent, then θ is not point-identified.*

4.2 Logit model

Using that, when $\theta \neq 0$, 1, $F(\alpha)$, and $F(\theta + \alpha)$, for $\alpha \in \mathcal{S}_\alpha$, are linearly independent in the logit model, Corollary 3 implies that in the logit model with a binary predetermined covariate, θ is not point-identified irrespective of the number of time periods available. We state this formally in a corollary.

Corollary 4. *Consider the logit model with $T \geq 2$. Suppose that Assumptions 1 and 2 hold, that $\theta \neq 0$, and that $\mathcal{S}_\alpha$ contains at least three points, then θ is not point-identified.*

This non point-identification result contrasts with prior work on logit panel data models. Under strict exogeneity, Rasch (1960) and Andersen (1970) have established that θ is point-identified under mild conditions on X_{it} whenever $T \geq 2$. In the dynamic logit model when $X_{it} = Y_{i,t-1}$, Chamberlain (1993) shows that θ is not point-identified when $T = 2$ (a result also obtained as an implication of Corollary 1). However, Chamberlain (1985), and Honoré and Kyriazidou (2000) in a model with covariates, show that θ is point-identified under suitable conditions whenever $T \geq 3$.[4] By contrast, Corollary 4 shows that, when the feedback process through which current covariates are influenced by lagged outcomes is unrestricted, the failure of point-identification is pervasive irrespective of T, despite the logit structure.

5 Characterizing the identified set

The previous sections show that point-identification often fails in binary choice models with a predetermined covariate. In this section, we explore the extent of non point-identification by presenting numerical calculations of the identified set Θ^I for specific parameter values.

5.1 Linear programming representation

We show that the identified set Θ^I, defined by set (2) above, can be represented as a set of θ values for which a certain linear program has a solution. This characterization facilitates numerical computation of the identified set.

To do so, let us first focus on the $T = 2$ case. Let

$$\psi_{x_1}(x_2, y_1, \alpha) = \Pr(X_{i2} = x_2, Y_{i1} = y_1, \alpha_i = \alpha \mid X_{i1} = x_1; \widetilde{\theta}, \widetilde{\pi}, \widetilde{G}),$$

for some hypothetical values $(\widetilde{\theta}, \widetilde{\pi}, \widetilde{G}) \in \Theta \times \Pi \times \mathcal{G}_2$.

This probability vector satisfies the following restrictions:

$$\psi_{x_1}(x_2, y_1, \alpha) \geq 0, \quad \sum_{x_2=0}^{1} \sum_{y_1=0}^{1} \sum_{\alpha \in \mathcal{S}_\alpha} \psi_{x_1}(x_2, y_1, \alpha) = 1, \tag{8}$$

and

$$\sum_{x_2=0}^{1} \psi_{x_1}(x_2, y_1, \alpha) = F(\widetilde{\theta} x_1 + \alpha)^{y_1} [1 - F(\widetilde{\theta} x_1 + \alpha)]^{1-y_1} \sum_{x_2=0}^{1} \sum_{y_1=0}^{1} \psi_{x_1}(x_2, y_1, \alpha). \tag{9}$$

Here (8) indicates that ψ_{x_1} is a probability vector, and (9) indicates that it needs to be consistent with the outcome distribution in period 1.

[4] Since in the dynamic logit model $X_{it} = Y_{i,t-1}$ is a lagged outcome, $T \geq 2$ (respectively, $T \geq 3$) requires that individual outcomes be available for at least three (resp., four) periods.

Then, $\widetilde{\theta} \in \Theta^I$ if and only if

$$Q_{x_1}(y_2, y_1, x_2; \theta, \pi, G) = \sum_{\alpha \in \mathcal{S}_\alpha} F(\widetilde{\theta} x_2 + \alpha)^{y_2} [1 - F(\widetilde{\theta} x_2 + \alpha)]^{1-y_2} \psi_{x_1}(x_2, y_1, \alpha), \tag{10}$$

for some vectors ψ_{x_1} satisfying (8) and (9) for $x_1 \in \{0, 1\}$. Since all of the equalities and inequalities in (8), (9) and (10) are linear in ψ_{x_1}, it follows that one can verify whether $\widetilde{\theta} \in \Theta^I$ by checking the existence of a solution to a linear program.[5]

The linear programming representation of Θ^I extends to any number $T \geq 2$ of periods. To see this, let, for some $(\widetilde{\theta}, \widetilde{\pi}, \widetilde{G}) \in \Theta \times \Pi \times \mathcal{G}_T$,

$$\psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) = \Pr(X_i^{2:T} = x^{2:T}, Y_i^{T-1} = y^{T-1}, \alpha_i = \alpha \,|\, X_{i1} = x_1; \widetilde{\theta}, \widetilde{\pi}, \widetilde{G}).$$

In the Appendix we derive the following characterization of the (sharp) identified set Θ^I.

Proposition 1. Θ^I *is the set of parameter values* $\widetilde{\theta} \in \Theta$ *such that*

$$Q_{x_1}(y^T, x^{2:T}; \theta, \pi, G) = \sum_{\alpha \in \mathcal{S}_\alpha} F(\widetilde{\theta} x_T + \alpha)^{y_T} [1 - F(\widetilde{\theta} x_T + \alpha)]^{1-y_T} \psi_{x_1}(x^{2:T}, y^{T-1}, \alpha), \tag{11}$$

for some functions $\psi_{x_1} : \{0, 1\}^{2T-2} \times \mathcal{S}_\alpha \to \mathbb{R}$, $x_1 \in \{0, 1\}$, *satisfying*

$$\psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) \geq 0, \qquad \sum_{x^{2:T} \in \{0,1\}^{T-1}} \sum_{y^{T-1} \in \{0,1\}^{T-1}} \sum_{\alpha \in \mathcal{S}_\alpha} \psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) = 1, \tag{12}$$

and, for all $s \in \{2, ..., T\}$,

$$\begin{aligned} &\sum_{x^{s:T} \in \{0,1\}^{T-s+1}} \sum_{y^{s:T-1} \in \{0,1\}^{T-s}} \psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) \\ &= F(\widetilde{\theta} x_{s-1} + \alpha)^{y_{s-1}} [1 - F(\widetilde{\theta} x_{s-1} + \alpha)]^{1-y_{s-1}} \sum_{x^{s:T} \in \{0,1\}^{T-s+1}} \sum_{y^{s-1:T-1} \in \{0,1\}^{T-s+1}} \psi_{x_1}(x^{2:T}, y^{T-1}, \alpha). \end{aligned} \tag{13}$$

Proposition 1 shows that one can verify whether $\widetilde{\theta} \in \Theta^I$ by checking the feasibility of a linear program. In a setting with lagged outcomes and strictly exogenous covariates, Honoré

[5] Note that, to compute the identified set under the assumption of strict exogeneity, one can simply modify this approach by adding to (8), (9) and (10) the additional restriction

$$\frac{\psi_{x_1}(x_2, 1, \alpha)}{F(\widetilde{\theta} x_1 + \alpha)} = \frac{\psi_{x_1}(x_2, 0, \alpha)}{1 - F(\widetilde{\theta} x_1 + \alpha)} \quad \text{for all } (x_2, x_1, \alpha),$$

which is also linear in ψ_{x_1}. The fact that, under strict exogeneity, Θ^I can be computed using linear programming was first established by Honoré and Tamer (2006).

and Tamer (2006) provided an analogous linear programming representation of the identified set. By contrast, in Proposition 1 we characterize the identified set of θ in the general predetermined case where the Granger condition fails; i.e., when $G_{y^{t-1},x^{t-1}}(\alpha)$ may depend on y^{t-1}, a situation that Honoré and Tamer (2006) did not consider but anticipated in their conclusion.

5.2 Numerical illustration

In this section we compute identified sets in logit and probit models for a set of example data generating processes (DGPs). In the DGPs, X_{it} follows a Bernoulli distribution on $\{0,1\}$ with probabilities $(\frac{1}{2},\frac{1}{2})$, independent over time, and α_i takes $K=31$ values with probabilities closely resembling those of a standard normal (a specification we borrow from Honoré and Tamer, 2006), and is drawn independently of $(X_{i1},...,X_{iT})$. In the logit case, $F(u)=\frac{e^u}{1+e^u}$, and in the probit case, $F(u)=\Phi(u)$ for Φ the standard normal cdf. Lastly, we vary θ between -1 and 1. Note that X_{it} is strictly exogenous in this data generating process. We characterize identified sets in two scenarios: assuming that X_{it} are strictly exogenous, and only assuming that X_{it} are predetermined.

In Figure 1 we report our numerical calculations of the identified set Θ^I for the logit model (in the left column panels) and for the probit model (in the right column panels). The three vertical panels correspond to the $T=2,3,4$ cases, respectively. In each graph, we report two sets of upper and lower bounds: those computed while maintaining the strict exogeneity assumption (in dashed lines) and those computed maintaining just predeterminedness (in solid lines). We report the true parameter θ on the x-axis.

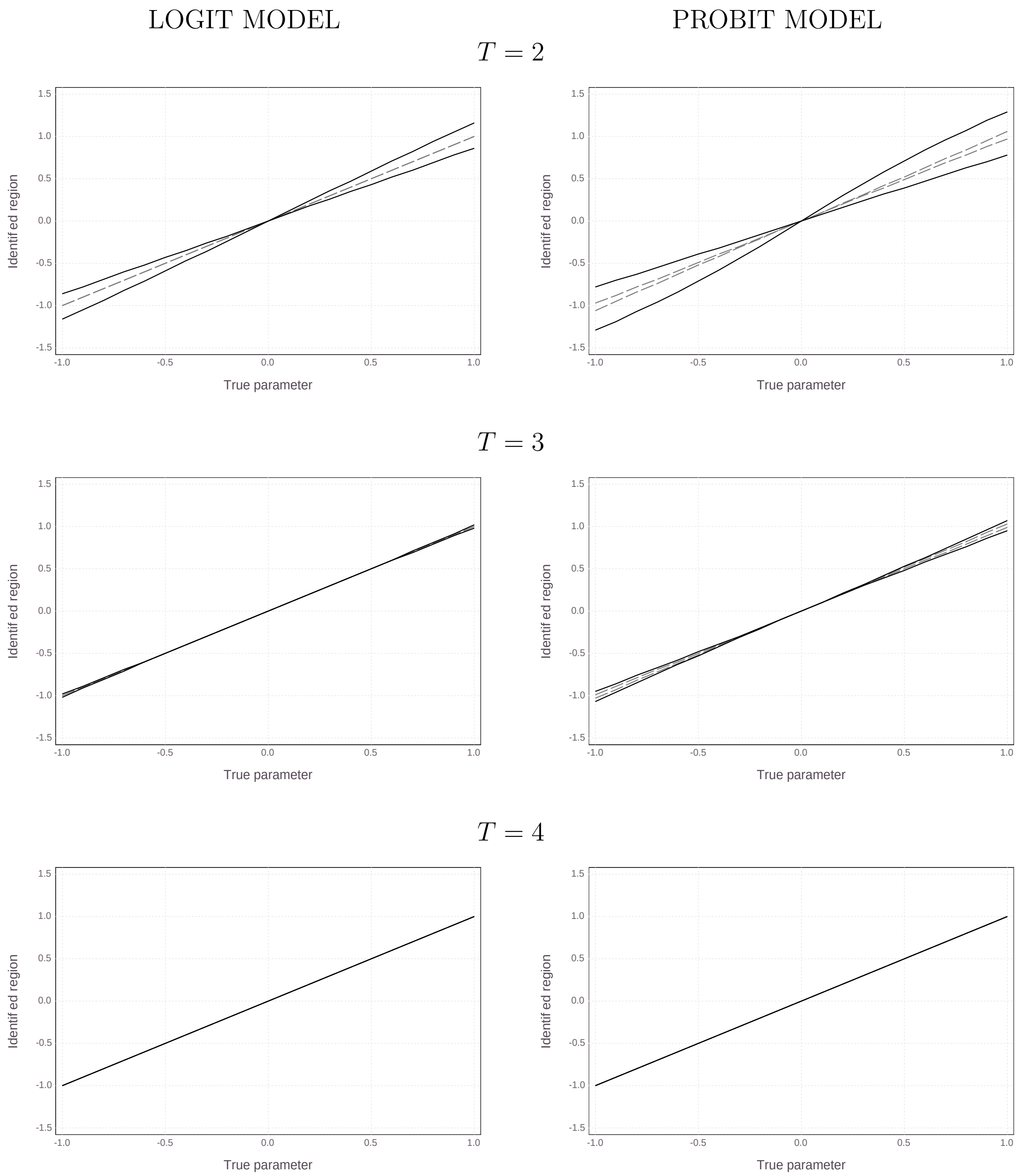
Focusing first on the logit case, shown in the left column of Figure 1, we see that the identified set Θ^I under strict exogeneity is a singleton for any value of θ and irrespective of T. This is not surprising since θ is point-identified in the static logit model. In contrast, the upper and lower bounds of the identified set do not coincide in the predetermined case, consistent with our non point-identification result. At the same time, the identified sets appear rather narrow, even when $T=2$, and the width of the set tends to decrease rapidly when T increases to three and four periods. This is qualitatively similar to the observation of Honoré and Tamer (2006), who focused on dynamic probit models and found that the width of the identified set tends to decrease rapidly with T.[6]

Focusing next on the probit case, shown in the right column of Figure 1, we see that the

[6]It is intuitive that Θ^I gets tighter as T increases. Indeed, under large T, α_i and θ are point-identified based on the moment condition

$$\mathbb{E}\left(\left(\begin{array}{c} X_{it} \\ 1 \end{array}\right)\left(Y_{it}-\frac{\exp(\theta X_{it}+\alpha_i)}{1+\exp(\theta X_{it}+\alpha_i)}\right)\right)=0,$$

Figure 1: Identified sets in logit and probit models



Notes: Upper and lower bounds of the identified set Θ^I in a logit model (left column) and a probit model (right column), for $T = 2, 3, 4$. The identified sets under strict exogeneity are indicated in dashed lines, the sets under predeterminedness are indicated in solid lines. The population value of θ is given on the x-axis.

identified set Θ^I under strict exogeneity is not a singleton in this case. Moreover, allowing the covariate to be predetermined increases the width of the identified set. However, as in the logit case, the sets appear rather narrow, even when $T = 2$, and their widths decrease quickly as T increases.

These calculations suggest that, while relaxing strict exogeneity tends to increase the widths of the bounds, the identified sets under predeterminedness can be informative even when the number of periods is very small. Of course, these conclusions are based on a particular set of specific examples.

6 Restrictions on the feedback process

Our analysis suggests that failures of point-identification are widespread in binary choice models with a predetermined covariate. In this section we describe possible restrictions on the model that can strengthen its identification content. We focus on restrictions on the feedback process, since restrictions on individual heterogeneity are rarely motivated by the economic context.

6.1 Homogeneous feedback

In some applications one may want to restrict the feedback process to not depend on time-invariant heterogeneity; that is, to impose that

$$\Pr\left(X_{it} = 1 \,|\, Y_i^{t-1} = y^{t-1}, X_i^{t-1} = x^{t-1}, \alpha_i = \alpha\right) = G^t_{y^{t-1},x^{t-1}} \tag{14}$$

is independent of α. For example, in structural dynamic discrete choice models, researchers may be willing to model the law of motion of state variables such as dynamic production inputs as homogeneous across units. Kasahara and Shimotsu (2009) show how this assumption can help identification in these models. Here we study how a homogeneity assumption can lead to tighter identified sets in our setting.

To proceed, we focus on the case where $T = 2$. Given (14), the likelihood function takes

provided $\mathbb{E}\left(\frac{\exp(\theta X_{it}+\alpha_i)}{[1+\exp(\theta X_{it}+\alpha_i)]^2}\begin{pmatrix} X_{it} \\ 1 \end{pmatrix}\begin{pmatrix} X_{it} \\ 1 \end{pmatrix}'\right)$ is non-singular.

the form

$$
\begin{aligned}
&\Pr\left(Y_{i2}=y_2, X_{i2}=x_2, Y_{i1}=y_1 \,|\, X_{i1}=x_1\right) \\
&= \left\{ \sum_{\alpha\in\mathcal{S}_\alpha} F(\theta x_2+\alpha)^{y_2}[1-F(\theta x_2+\alpha)]^{1-y_2} F(\theta x_1+\alpha)^{y_1}[1-F(\theta x_1+\alpha)]^{1-y_1}\pi_{x_1}(\alpha) \right\} \\
&\qquad \times [G^2_{y_1,x_1}]^{x_2}[1-G^2_{y_1,x_1}]^{1-x_2},
\end{aligned}
$$

where the likelihood factors due to the fact that the feedback process does not depend on α. Hence, under Assumption 1 we have

$$
\begin{aligned}
&\frac{\Pr\left(Y_{i2}=y_2, X_{i2}=x_2, Y_{i1}=y_1 \,|\, X_{i1}=x_1\right)}{[G^2_{y_1,x_1}]^{x_2}[1-G^2_{y_1,x_1}]^{1-x_2}} \\
&= \sum_{\alpha\in\mathcal{S}_\alpha} F(\theta x_2+\alpha)^{y_2}[1-F(\theta x_2+\alpha)]^{1-y_2} F(\theta x_1+\alpha)^{y_1}[1-F(\theta x_1+\alpha)]^{1-y_1}\pi_{x_1}(\alpha). \qquad (15)
\end{aligned}
$$

A key observation to make about (15) is its right-hand-side coincides with the likelihood function of a binary choice model with a strictly exogenous covariate (where in addition α_i is independent of X_{i2} given X_{i1}). In turn, the left-hand side is weighted by the inverse of the feedback process. This is similar to the inverse-probability-of-treatment-weighting approach to dynamic treatment effect analysis in Jamie Robins' work (e.g., Robins, 2000), with the difference that here we focus on panel data models with fixed effects.

The similarity between (15) and the strictly exogenous case directly delivers point-identification results and consistent estimators. For example, suppose that F is logistic. Given that the left-hand side of (15) is point-identified, it follows from standard arguments (Rasch, 1960, Andersen, 1970) that θ is point-identified. Moreover, a consistent estimator of θ is obtained by maximizing the weighted conditional logit log-likelihood

$$
\sum_{i=1}^n \widehat{\omega}_i \mathbf{1}\{Y_{i1}+Y_{i2}=1\} \left\{ Y_{i1}\ln\left(\frac{\exp(\widetilde{\theta}X_{i1})}{\exp(\widetilde{\theta}X_{i1})+\exp(\widetilde{\theta}X_{i2})}\right) + Y_{i2}\ln\left(\frac{\exp(\widetilde{\theta}X_{i2})}{\exp(\widetilde{\theta}X_{i1})+\exp(\widetilde{\theta}X_{i2})}\right)\right\},
$$

with weights

$$
\widehat{\omega}_i = \left\{ [\widehat{G}^2_{Y_{i1},X_{i1}}]^{X_{i2}}[1-\widehat{G}^2_{Y_{i1},X_{i1}}]^{1-X_{i2}} \right\}^{-1},
$$

for $\widehat{G}^2_{y_1,x_1}$ a consistent estimate of the homogeneous feedback probabilities.[7]

[7]When X_{it} are continuous, demonstrating $\sqrt{n}$ consistency of $\widehat{\theta}$ would generally require imposing rate-of-convergence and other requirements on the first-step estimation of the $\widehat{\omega}_i$ weights.

6.2 Markovian feedback

Another possible restriction on the feedback process is a Markovian condition, such as

$$\Pr\left(X_{it}=1\,|\,Y_i^{t-1}=y^{t-1},X_i^{t-1}=x^{t-1},\alpha_i=\alpha\right)=G^t_{y_{t-1},x_{t-1}}(\alpha) \tag{16}$$

is independent of (y^{t-2},x^{t-2}). Such a condition may be natural in models where X_{it} is the state variable in the agent's economic problem (as in Rust, 1987 and Kasahara and Shimotsu, 2009, for example).

In order to characterize the identified set Θ^I with the Markovian condition (16) added, we augment the restrictions (11), (12) and (13) with the fact that, for all $s\in\{2,...,T-1\}$,

$$\frac{\sum_{x^{s+1:T}\in\{0,1\}^{T-s+1}}\sum_{y^{s:T-1}\in\{0,1\}^{T-s}}\psi_{x_1}(x^{2:T},y^{T-1},\alpha)}{\sum_{x^{s:T}\in\{0,1\}^{T-s+1}}\sum_{y^{s:T-1}\in\{0,1\}^{T-s}}\psi_{x_1}(x^{2:T},y^{T-1},\alpha)}$$

does not depend on (y^{s-2},x^{s-2}).

A difficulty arises in this case since this additional set of restrictions is not linear in ψ_{x_1}. As a result, one would need to use different techniques to characterize the identified set in the spirit of Proposition 1, and to establish conditions for (the failure of) point-identification in the spirit of Corollary 3. Given this, we leave the analysis of identification in models with Markovian feedback processes to future work.

7 Conclusion

In this paper we study a binary choice model with a binary predetermined covariate. The analysis can easily be extended to general discrete covariates. We find that failures of point-identification are widespread in this setting. Point identification fails in many binary choice models, with apparently only a few exceptions (such as the exponential model). At the same time, our numerical calculations of identified sets suggest that the bounds on the parameter may be narrow, even in very short panels. This suggests that, while the strict exogeneity assumption has identifying content, models with predetermined covariates and feedback may still lead to informative empirical conclusions.

Although we have analyzed a binary choice model, our techniques can be used to study other models with stronger identification content, such as models for count data (e.g., Poisson regression, Wooldridge, 1997, Blundell et al., 2002) and models with continuous outcomes (e.g., censored regression, Honore and Hu, 2004, and duration models, Chamberlain, 1985). Deriving sequential moment restrictions in such nonlinear models was considered by Chamberlain (2022) and is further explored in our companion paper (Bonhomme et al., 2022).

References

Al-Sadoon, M. M., Li, T., and Pesaran, H. (2017). Exponential class of dynamic binary choice panel data models with fixed effects. *Econometric Reviews*, 36(6-9).

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(2):283–301.

Arellano, M. (2003). *Panel data econometrics*. OUP Oxford.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics*, 68(1):29 – 51.

Arellano, M. and Carrasco, R. (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics*, 115(1):125 – 157.

Bekker, P. and Wansbeek, T. (2001). Identification in parametric models. *A companion to theoretical econometrics*, pages 144–161.

Blundell, R. and Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3):321 – 340.

Blundell, R., Griffith, R., and Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of econometrics*, 108(1):113–131.

Bonhomme, S., Dano, K., and Graham, B. (2022). Sequential moment restrictions in nonlinear panel data models. *Working Paper*.

Chamberlain, G. (1985). Heterogeneity, duration dependence and omitted variable bias. *Longitudinal Analysis of Labor Market Data. Cambridge University Press New York.*

Chamberlain, G. (1993). Feedback in panel data models. *Working Paper*.

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. *Journal of Econometrics*, 226(1):4 – 20.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society B*, 20(2):215 – 242.

Davezies, L., D'Haultfoeuille, X., and Mugnier, M. (2020). Fixed effects binary choice models with three or more periods. *arXiv preprint arXiv:2009.08108*.

Honore, B. E. and Hu, L. (2004). Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics*, 122(2):293–316.

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

Honoré, B. E. and Lewbel, A. (2002). Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica*, 70(5):2053–2063.

Honoré, B. E. and Tamer, E. (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74(3):611–629.

Kasahara, H. and Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, 77(1):135–175.

Olley, S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Rasch, G. (1960). Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 95–133. Springer.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica: Journal of the Econometric Society*, pages 577–591.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033.

Wooldridge, J. M. (1997). Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory*, 13(5):667–678.

APPENDIX

A Proof of Lemma 1

For any $m \times n$ matrix A, we will denote as

$$\mathcal{R}(A) = \{Au \,:\, u \in \mathbb{R}^n\}$$

the range of A, as

$$\mathcal{N}(A) = \{u \in \mathbb{R}^n \,:\, Au = 0\}$$

the null space of A, and as $A^\dagger$ the Moore-Penrose generalized inverse of A.

We now proceed to prove Lemma 1. Since θ is point-identified, it is locally point-identified. Since (θ, π, G) is a regular point of $\nabla Q(\theta, \pi, G)$ by Assumption 2, it follows from Theorem 8 in Bekker and Wansbeek (2001) that

$$\nabla_{\theta'} Q \notin \mathcal{R}\left(\begin{bmatrix} \nabla_{\pi_1'} Q_1 & \nabla_{G_1'} Q_1 & 0 & 0 \\ 0 & 0 & \nabla_{\pi_0'} Q_0 & \nabla_{G_0'} Q_0 \end{bmatrix}\right). \tag{A1}$$

Therefore, there must exist $x_1 \in \{0, 1\}$ such that

$$\nabla_{\theta'} Q_{x_1} \notin \mathcal{R}\left(\begin{bmatrix} \nabla_{\pi_{x_1}'} Q_{x_1} & \nabla_{G_{x_1}'} Q_{x_1} \end{bmatrix}\right), \tag{A2}$$

and in the rest of the proof we will fix this x_1 value.
Let $\widetilde{\phi}_{x_1}$ denote the projection of $\nabla_{\theta'} Q_{x_1}$ onto the orthogonal complement of the vector space spanned by the columns of $\begin{bmatrix} \nabla_{\pi_{x_1}'} Q_{x_1} & \nabla_{G_{x_1}'} Q_{x_1} \end{bmatrix}$; that is,

$$\widetilde{\phi}_{x_1} = \nabla_{\theta'} Q_{x_1} - \begin{bmatrix} \nabla_{\pi_{x_1}'} Q_{x_1} & \nabla_{G_{x_1}'} Q_{x_1} \end{bmatrix} \begin{bmatrix} \nabla_{\pi_{x_1}'} Q_{x_1} & \nabla_{G_{x_1}'} Q_{x_1} \end{bmatrix}^\dagger \nabla_{\theta'} Q_{x_1}.$$

It follows from (A2) that $\widetilde{\phi}_{x_1} \neq 0$. Moreover, since $\iota' Q_{x_1}(\theta, \pi, G) = 1$, where ι denotes a conformable vector of ones, we have

$$\iota' \nabla_{\theta'} Q_{x_1} = 0, \;\; \iota' \nabla_{\pi_{x_1}'} Q_{x_1} = 0, \;\; \iota' \nabla_{G_{x_1}'} Q_{x_1} = 0. \tag{A3}$$

It follows that $\iota' \widetilde{\phi}_{x_1} = 0$, implying that $\widetilde{\phi}_{x_1}$ cannot be constant.

Now, since by construction

$$\widetilde{\phi}_{x_1} \perp \mathcal{R}\left(\left[\nabla_{\pi'_{x_1}} Q_{x_1} \quad \nabla_{G'_{x_1}} Q_{x_1}\right]\right),$$

we have

$$\widetilde{\phi}_{x_1} \in \mathcal{N}(\nabla_{\pi_{x_1}} Q'_{x_1}) \cap \mathcal{N}(\nabla_{G_{x_1}} Q'_{x_1}).$$

Next, let $P_\theta(x_1, \alpha)$ be the 8×1 vector with elements

$$\Pr(Y_{i2} = y_2, X_{i2} = x_2, Y_{i1} = y_1 \mid X_{i1} = x_1, \alpha_i = \alpha),$$

for $(y_2, x_2, y_1) \in \{0,1\}^3$. Since $\widetilde{\phi}_{x_1} \in \mathcal{N}(\nabla_{\pi_{x_1}} Q'_{x_1})$, we have, for all $\alpha \in \mathcal{S}_\alpha$,

$$\widetilde{\phi}'_{x_1} P_\theta(x_1, \alpha) = \widetilde{\phi}'_{x_1} P_\theta(x_1, \underline{\alpha}_K) \equiv C_{x_1},$$

where we have used the fact that $\pi_{x_1}(\underline{\alpha}_K) = 1 - \sum_{k=1}^{K-1} \pi_{x_1}(\underline{\alpha}_k)$.
Let us define the following demeaned version of $\widetilde{\phi}_{x_1}$:

$$\phi_{x_1} = \widetilde{\phi}_{x_1} - C_{x_1}\iota.$$

The 8×1 vector ϕ_{x_1} represents a function $\phi_{x_1} : \{0,1\}^3 \mapsto \mathbb{R}$. With some abuse of terminology we will sometimes refer to ϕ_{x_1} as a vector and sometimes as a function.
Note that, since $\widetilde{\phi}_{x_1}$ is not constant, it follows that $\phi_{x_1} \neq 0$. Moreover, using (A2) and (A3) we have

$$\phi_{x_1} \in \mathcal{N}(\nabla_{\pi_{x_1}} Q'_{x_1}) \cap \mathcal{N}(\nabla_{G_{x_1}} Q'_{x_1}),$$

from which it follows that

$$\text{(i)} \quad \nabla_{\pi_{x_1}} Q'_{x_1} \phi_{x_1} = 0, \quad \text{(ii)} \quad \nabla_{G_{x_1}} Q'_{x_1} \phi_{x_1} = 0.$$

We are now going to use (i) and (ii) to show (4)-(5). From (ii) we get, for all $\alpha \in \mathcal{S}_\alpha$,

$$\pi_{x_1}(\alpha)\Big(\phi_{x_1}(1,1,1)F(\theta+\alpha)F(\theta x_1+\alpha) - \phi_{x_1}(1,1,0)F(\alpha)F(\theta x_1+\alpha)$$
$$+\phi_{x_1}(1,0,1)[1-F(\theta+\alpha)]F(\theta x_1+\alpha) - \phi_{x_1}(1,0,0)[1-F(\alpha)]F(\theta x_1+\alpha)\Big) = 0,$$
$$\pi_{x_1}(\alpha)\Big(\phi_{x_1}(0,1,1)F(\theta+\alpha)[1-F(\theta x_1+\alpha)] - \phi_{x_1}(0,1,0)F(\alpha)[1-F(\theta x_1+\alpha)]$$
$$+\phi_{x_1}(0,0,1)[1-F(\theta+\alpha)][1-F(\theta x_1+\alpha)] - \phi_{x_1}(0,0,0)[1-F(\alpha)][1-F(\theta x_1+\alpha)]\Big) = 0.$$

This implies, using Assumption 1,

$$\phi_{x_1}(1,1,1)F(\theta+\alpha) - \phi_{x_1}(1,1,0)F(\alpha) + \phi_{x_1}(1,0,1)[1-F(\theta+\alpha)] - \phi_{x_1}(1,0,0)[1-F(\alpha)] = 0,$$
$$\phi_{x_1}(0,1,1)F(\theta+\alpha) - \phi_{x_1}(0,1,0)F(\alpha) + \phi_{x_1}(0,0,1)[1-F(\theta+\alpha)] - \phi_{x_1}(0,0,0)[1-F(\alpha)] = 0,$$

which coincides with (4).
Lastly, from (i) we get, for all $\alpha \in \mathcal{S}_\alpha$,

$$\begin{aligned}
\phi'_{x_1}P_\theta(x_1,\alpha) &= \phi'_{x_1}P_\theta(x_1,\underline{\alpha}_K) \\
&= \widetilde{\phi}'_{x_1}P_\theta(x_1,\underline{\alpha}_K) - C_{x_1}\underbrace{\iota' P_\theta(x_1,\underline{\alpha}_K)}_{=1} \\
&= \widetilde{\phi}'_{x_1}P_\theta(x_1,\underline{\alpha}_K) - \widetilde{\phi}'_{x_1}P_\theta(x_1,\underline{\alpha}_K) \\
&= 0,
\end{aligned}$$

which can be equivalently written as

$$\sum_{y_2=0}^{1}\sum_{x_2=0}^{1}\sum_{y_1=0}^{1}\phi_{x_1}(y_1,y_2,x_2)\Pr(Y_{i2}=y_2, X_{i2}=x_2, Y_{i1}=y_1 \mid X_{i1}=x_1, \alpha_i=\alpha;\theta) = 0.$$

Now, using (4), this implies that, for all $x_2 \in \{0,1\}$,

$$\sum_{y_2=0}^{1}\sum_{y_1=0}^{1}\phi_{x_1}(y_1,y_2,x_2)\Pr(Y_{i2}=y_2 \mid X_{i2}=x_2, \alpha_i=\alpha;\theta)\Pr(Y_{i1}=y_1 \mid X_{i1}=x_1, \alpha_i=\alpha;\theta) = 0,$$

which coincides with (5).

B Proof of Corollary 1

The proof is by contradiction. Suppose that θ is point-identified. Then by (4) we have, for some $x_1 \in \{0,1\}$, and for all $y_1 \in \{0,1\}$ and $\alpha \in \mathcal{S}_\alpha$,

$$\phi_{x_1}(y_1,0,1)[1-F(\theta+\alpha)] + \phi_{x_1}(y_1,1,1)F(\theta+\alpha) = \phi_{x_1}(y_1,0,0)[1-F(\alpha)] + \phi_{x_1}(y_1,1,0)F(\alpha).$$

Since 1, $F(\alpha)$, and $F(\theta+\alpha)$, for $\alpha \in \mathcal{S}_\alpha$, are linearly independent, we thus have, for all $y_1 \in \{0,1\}$,

$$\phi_{x_1}(y_1,0,1) = \phi_{x_1}(y_1,1,1) = \phi_{x_1}(y_1,0,0) = \phi_{x_1}(y_1,1,0). \tag{A4}$$

Next, using (5) at $x_2 = 1$ we have

$$\begin{aligned}&\phi_{x_1}(1,1,1)F(\theta+\alpha)F(\theta x_1+\alpha) + \phi_{x_1}(0,1,1)F(\theta+\alpha)[1-F(\theta x_1+\alpha)]\\ &+ \phi_{x_1}(1,0,1)[1-F(\theta+\alpha)]F(\theta x_1+\alpha) + \phi_{x_1}(0,0,1)[1-F(\theta+\alpha)][1-F(\theta x_1+\alpha)] = 0.\end{aligned}$$

Using (A4) then gives

$$\phi_{x_1}(1,1,1)F(\theta x_1+\alpha) + \phi_{x_1}(0,1,1)[1-F(\theta x_1+\alpha)] = 0.$$

Now, since 1 and $F(\theta x_1+\alpha)$, for $\alpha \in \mathcal{S}_\alpha$, are linearly independent, it follows that

$$\phi_{x_1}(1,1,1) = \phi_{x_1}(0,1,1) = 0.$$

Using (A4) then gives

$$\phi_{x_1}(1,0,1) = \phi_{x_1}(0,0,1) = 0.$$

Lastly, repeating the same argument starting with (5) at $x_2 = 0$ gives

$$\phi_{x_1}(1,1,0) = \phi_{x_1}(0,1,0) = \phi_{x_1}(1,0,0) = \phi_{x_1}(0,0,0) = 0.$$

It follows that $\phi_{x_1} = 0$, which leads to a contradiction.

C Proof of remark 1 (sign identification of θ)

Note that

$$\begin{aligned}\mathbb{E}\left[Y_{i2}-Y_{i1}\,|\,X_{i1}=0\right] &= \mathbb{E}\left[\mathbb{E}\left[Y_{i2}\,|\,X_{i2},Y_{i1},X_{i1}=0,\alpha_i\right]-\mathbb{E}\left[Y_{i1}\,|\,X_{i1}=0,\alpha_i\right]\,|\,X_{i1}=0\right] \\ &= \mathbb{E}\left[F(\theta X_{i2}+\alpha_i)-F(\alpha_i)\,|\,X_{i1}=0\right] \\ &= \mathbb{E}\left[(F(\theta+\alpha_i)-F(\alpha_i))X_{i2}Y_{i1}+(F(\theta+\alpha_i)-F(\alpha_i))X_{i2}(1-Y_{i1})\,|\,X_{i1}=0\right] \\ &= \sum_{\alpha\in\mathcal{S}_\alpha}\sum_{y_1=0}^{1}(F(\theta+\alpha)-F(\alpha))\underbrace{G^2_{y_1,0}(\alpha)F(\alpha)^{y_1}(1-F(\alpha))^{1-y_1}\pi_0(\alpha).}_{>0\text{ by Assumption 1}}\end{aligned} \tag{A5}$$

If $\theta=0$, (A5) implies that $\mathbb{E}\left[Y_{i2}-Y_{i1}\,|\,X_{i1}=0\right]=0$. Moreover, since $F(\cdot)$ is strictly increasing, it follows that $\theta>0$ (respectively, <0) and $\mathbb{E}\left[Y_{i2}-Y_{i1}\,|\,X_{i1}=0\right]>0$ (resp., <0) are equivalent. This implies that $\text{sign}(\theta)=\text{sign}\left(\mathbb{E}\left[Y_{i2}-Y_{i1}\,|\,X_{i1}=0\right]\right)$. A similar argument applied to $X_{i1}=1$ implies that $\text{sign}(\theta)=\text{sign}\left(\mathbb{E}\left[Y_{i1}-Y_{i2}\,|\,X_{i1}=1\right]\right)$.

D Proof of Lemma 2

In what follows we assume $T\geq 3$ having already proved the validity of the claim for $T=2$ in Lemma 1.

Since θ is point-identified it is locally point-identified. Additionally, since (θ,π,G) is a regular point of $\nabla Q(\theta,\pi,G)$ by Assumption 2, we can appeal to Theorem 8 in Bekker and Wansbeek (2001) and follow the same line of arguments as in the proof of Lemma 1 to conclude that there exists $x_1\in\{0,1\}$ and a $2^{2T-1}\times 1$ vector $\phi_{x_1}\neq 0$ such that

$$\text{(i)}\quad \nabla_{\pi_{x_1}}Q'_{x_1}\phi_{x_1}=0,\quad \text{(ii)}\quad \nabla_{G_{x_1}}Q'_{x_1}\phi_{x_1}=0.$$

We will now prove (6) and (7) using finite induction.
Let us start with (6). Given $s\in\{0,...,T-2\}$, let $\mathcal{P}(s)$ denote the statement that, for all $y^{T-(s+1)}\in\{0,1\}^{T-(s+1)}$ and $x^{T-(s+1)}\in\{0,1\}^{T-(s+1)}$,

$$\sum_{y^{T-s:T}\in\{0,1\}^{s+1}}\phi_{x_1}(y^T,x^{2:T})\prod_{t=T-s}^{T}F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t}$$

does not depend on $x^{T-s:T}$.

Base case:
Condition (ii) implies that

$$\left(\frac{\partial Q_{x_1}}{\partial G^T_{y^{T-1},x^{T-1}}(\alpha)}\right)' \phi_{x_1} = 0,$$

or equivalently that

$$\begin{aligned}&\sum_{y_T=0}^{1}\sum_{x_T=0}^{1} \phi_{x_1}(y^T, x^{2:T}) F(\theta x_T + \alpha)^{y_T}[1 - F(\theta x_T + \alpha)]^{1-y_T}(-1)^{1-x_T} \\ &\times \prod_{t=2}^{T-1} F(\theta x_t + \alpha)^{y_t}[1 - F(\theta x_t + \alpha)]^{1-y_t} G^t_{y^{t-1},x^{t-1}}(\alpha)^{x_t}[1 - G^t_{y^{t-1},x^{t-1}}(\alpha)]^{1-x_t} \\ &\times F(\theta x_1 + \alpha)^{y_1}[1 - F(\theta x_1 + \alpha)]^{1-y_1} = 0.\end{aligned}$$

Using Assumption 1, this simplifies to

$$\sum_{y_T=0}^{1}\sum_{x_T=0}^{1} \phi_{x_1}(y^T, x^{2:T}) F(\theta x_T + \alpha)^{y_T}[1 - F(\theta x_T + \alpha)]^{1-y_T}(-1)^{1-x_T} = 0,$$

which implies that

$$\sum_{y_T=0}^{1} \phi_{x_1}(y^T, x^{2:T}) F(\theta x_T + \alpha)^{y_T}[1 - F(\theta x_T + \alpha)]^{1-y_T}$$

does not depend on x_T.
Thus, $\mathcal{P}(0)$ is true.

Induction step:
Suppose that $\mathcal{P}(0), \ldots, \mathcal{P}(s)$ are true for $s \in \{0, \ldots, T-3\}$. We are going to show that $\mathcal{P}(s+1)$ is true.
Condition (ii) implies that

$$\left(\frac{\partial Q_{x_1}}{\partial G^{T-(s+1)}_{y^{T-(s+2)},x^{T-(s+2)}}(\alpha)}\right)' \phi_{x_1} = 0.$$

If $s < (T-3)$, this corresponds to

$$\sum_{y^{T-(s+1):T}\in\{0,1\}^{s+2}} \sum_{x^{T-(s+1):T}\in\{0,1\}^{s+2}} \phi_{x_1}(y^T, x^{2:T})$$
$$\times \prod_{t=T-s}^{T} F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t} G^t_{y^t,x^t}(\alpha)^{x_t}[1-G^t_{y^t,x^t}(\alpha)]^{1-x_t}$$
$$\times F(\theta x_{T-(s+1)}+\alpha)^{y_{T-(s+1)}}[1-F(\theta x_{T-(s+1)}+\alpha)]^{1-y_{T-(s+1)}}(-1)^{1-x_{T-(s+1)}}$$
$$\times \prod_{t=2}^{T-(s+2)} F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t} G^t_{y^{t-1},x^{t-1}}(\alpha)^{x_t}[1-G^t_{y^{t-1},x^{t-1}}(\alpha)]^{1-x_t}$$
$$\times F(\theta x_1+\alpha)^{y_1}[1-F(\theta x_1+\alpha)]^{1-y_1} = 0.$$

While if $s = (T-3)$, this corresponds to

$$\sum_{y^{2:T}\in\{0,1\}^{T-1}} \sum_{x^{2:T}\in\{0,1\}^{T-1}} \phi_{x_1}(y^T, x^{2:T})$$
$$\times \prod_{t=3}^{T} F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t} G^t_{y^t,x^t}(\alpha)^{x_t}[1-G^t_{y^t,x^t}(\alpha)]^{1-x_t}$$
$$\times F(\theta x_2+\alpha)^{y_2}[1-F(\theta x_2+\alpha)]^{1-y_2}(-1)^{1-x_2}$$
$$\times F(\theta x_1+\alpha)^{y_1}[1-F(\theta x_1+\alpha)]^{1-y_1} = 0.$$

Using Assumption 1 this gives, for all $s \in \{0, \dots, T-3\}$,

$$\sum_{y^{T-(s+1):T}\in\{0,1\}^{s+2}} \sum_{x^{T-(s+1):T}\in\{0,1\}^{s+2}} \phi_{x_1}(y^T, x^{2:T})$$
$$\times \prod_{t=T-s}^{T} F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t} G^t_{y^t,x^t}(\alpha)^{x_t}[1-G^t_{y^t,x^t}(\alpha)]^{1-x_t}$$
$$\times F(\theta x_{T-(s+1)}+\alpha)^{y_{T-(s+1)}}[1-F(\theta x_{T-(s+1)}+\alpha)]^{1-y_{T-(s+1)}}(-1)^{1-x_{T-(s+1)}} = 0. \tag{A6}$$

Let L_{s+1} denote the left-hand side of (A6). Exploiting successively the fact that $\mathcal{P}(0), \dots, \mathcal{P}(s)$ are true, alongside the property that, for all $t \in \{T-s, ..., T\}$,

$$\sum_{x_t=0}^{1} G^t_{y^t,x^t}(\alpha)^{x_t}[1-G^t_{y^t,x^t}(\alpha)]^{1-x_t} = 1, \tag{A7}$$

it is easy to see that

$$L_{s+1} = \sum_{y^{T-(s+1):T} \in \{0,1\}^{s+2}} \sum_{x_{T-(s+1)}=0}^{1} \phi_{x_1}(y^T, x^{2:T}) \prod_{t=T-s}^{T} F(\theta x_t + \alpha)^{y_t}[1 - F(\theta x_t + \alpha)]^{1-y_t}$$
$$\times F(\theta x_{T-(s+1)} + \alpha)^{y_{T-(s+1)}}[1 - F(\theta x_{T-(s+1)} + \alpha)]^{1-y_{T-(s+1)}}(-1)^{1-x_{T-(s+1)}} = 0.$$

Recalling that $\mathcal{P}(s)$ is true, this implies that

$$\sum_{y^{T-(s+1):T} \in \{0,1\}^{s+1}} \phi_{x_1}(y^T, x^{2:T}) \prod_{t=T-(s+1)}^{T} F(\theta x_t + \alpha)^{y_t}[1 - F(\theta x_t + \alpha)]^{1-y_t}$$

does not depend on $x^{T-(s+1):T}$. Hence, $\mathcal{P}(s+1)$ is true. This concludes the proof of (6).

Finally, we show (7). As in the proof of Lemma 1, Condition (i) implies that

$$\sum_{y^T \in \{0,1\}^T} \sum_{x^{2:T} \in \{0,1\}^{T-1}} \phi_{x_1}(y^T, x^{2:T})$$
$$\times \prod_{t=2}^{T} F(\theta x_t + \alpha)^{y_t}[1 - F(\theta x_t + \alpha)]^{1-y_t} G^t_{y^t,x^t}(\alpha)^{x_t}[1 - G^t_{y^t,x^t}(\alpha)]^{1-x_t}$$
$$\times F(\theta x_1 + \alpha)^{y_1}[1 - F(\theta x_1 + \alpha)]^{1-y_1} = 0.$$

Using (6) and (A7), it follows that

$$\sum_{y^T \in \{0,1\}^T} \phi_{x_1}(y^T, x^{2:T}) \prod_{t=1}^{T} F(\theta x_t + \alpha)^{y_t}[1 - F(\theta x_t + \alpha)]^{1-y_t} = 0,$$

which coincides with (7).

E Proof of Corollary 3

In what follows we assume $T \geq 3$ having already proved the validity of the claim for $T = 2$ in Corollary 1.

The proof is by contradiction. Suppose that θ is point-identified. We will show that this necessarily leads to $\phi_{x_1} = 0$, which will contradict Lemma 2. To that end, we will first prove via finite induction that ϕ_{x_1} must be a constant function.

For $s \in \{1, ..., T-2\}$, let $\mathcal{P}(s)$ denote the statement that there exists a function $\phi_{x_1}^{T-s} : \{0,1\}^{2T-2s-1} \to \mathbb{R}$ such that, for all $y^T \in \{0,1\}^T$ and $x^{2:T} \in \{0,1\}^{T-1}$, we have

$$\phi_{x_1}(y^T, x^{2:T}) = \phi_{x_1}^{T-s}(y^{T-s}, x^{2:T-s}).$$

Base case:
By (6), the quantity

$$\sum_{y_T=0}^{1} \phi_{x_1}(y^T, x^{2:T}) F(\theta x_T + \alpha)^{y_T} [1 - F(\theta x_T + \alpha)]^{1-y_T} \tag{A8}$$

does not depend on x_T. Hence

$$\begin{aligned} &\phi_{x_1}(y^{T-1}, 1, x^{2:T-1}, 1) F(\theta + \alpha) + \phi_{x_1}(y^{T-1}, 0, x^{2:T-1}, 1)[1 - F(\theta + \alpha)] \\ &= \phi_{x_1}(y^{T-1}, 1, x^{2:T-1}, 0) F(\alpha) + \phi_{x_1}(y^{T-1}, 0, x^{2:T-1}, 0)[1 - F(\alpha)]. \end{aligned}$$

By linear independence of 1, $F(\alpha)$, and $F(\theta + \alpha)$, this implies that $\phi_{x_1}(y^T, x^{2:T})$ does not depend on (y_T, x_T). Hence $\mathcal{P}(1)$ is true.

Induction step

Suppose that $\mathcal{P}(s)$ is true for $s \in \{1, ..., T-3\}$. Let us show that $\mathcal{P}(s+1)$ is true.

Since $\mathcal{P}(s)$ is true, we know that there exists a function $\phi_{x_1}^{T-s} : \{0,1\}^{2T-2s-1} \to \mathbb{R}$ such that

$$\phi_{x_1}(y^T, x^{2:T}) = \phi_{x_1}^{T-s}(y^{T-s}, x^{2:T-s}).$$

Thus, by (6), the quantity:

$$\begin{aligned}
&\sum_{y^{T-s:T}\in\{0,1\}^{s+1}} \phi_{x_1}(y^T, x^{2:T}) \prod_{t=T-s}^{T} F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t} \\
&= \sum_{y_{T-s}=0}^{1} \phi_{x_1}^{T-s}(y^{T-s}, x^{2:T-s}) \sum_{y^{T-(s-1):T}\in\{0,1\}^{s}} \prod_{t=T-(s-1)}^{T} F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t} \\
&\times F(\theta x_{T-s}+\alpha)^{y_{T-s}}[1-F(\theta x_{T-s}+\alpha)]^{1-y_{T-s}} \\
&= \sum_{y_{T-s}=0}^{1} \phi_{x_1}^{T-s}(y^{T-s}, x^{2:T-s}) F(\theta x_{T-s}+\alpha)^{y_{T-s}}[1-F(\theta x_{T-s}+\alpha)]^{1-y_{T-s}}
\end{aligned}$$

does not depend on $x^{T-s:T}$. Therefore,

$$\begin{aligned}
&\phi_{x_1}^{T-s}(y^{T-s-1}, 1, x^{2:T-s-1}, 1)F(\theta+\alpha) + \phi_{x_1}^{T-s}(y^{T-s-1}, 0, x^{2:T-s-1}, 1)[1-F(\theta+\alpha)] \\
&= \phi_{x_1}^{T-s}(y^{T-s-1}, 1, x^{2:T-s-1}, 0)F(\alpha) + \phi_{x_1}^{T-s}(y^{T-s-1}, 0, x^{2:T-s-1}, 0)[1-F(\alpha)].
\end{aligned}$$

Since 1, $F(\alpha)$, and $F(\theta+\alpha)$ are linearly independent, this implies $\mathcal{P}(s+1)$.
It follows from the previous induction argument that there exists a function $\phi_{x_1}^{2} : \{0,1\}^3 \to \mathbb{R}$ such that, for all $(y^T, x^{2:T})$,

$$\phi_{x_1}(y^T, x^{2:T}) = \phi_{x_1}^{2}(y^2, x_2).$$

Using (6), the quantity

$$\begin{aligned}
&\sum_{y^{2:T}\in\{0,1\}^{T-1}} \phi_{x_1}(y^T, x^{2:T}) \prod_{t=2}^{T} F(\theta x_t+\alpha)^{y_t}[1-F(\theta x_t+\alpha)]^{1-y_t} \\
&= \sum_{y_2=0}^{1} \phi_{x_1}^{2}(y^2, x_2) F(\theta x_2+\alpha)^{y_2}[1-F(\theta x_2+\alpha)]^{1-y_2}
\end{aligned}$$

does not depend on $x^{2:T}$. Therefore,

$$\begin{aligned}
&\phi_{x_1}^{2}(y_1, 1, 1)F(\theta+\alpha) + \phi_{x_1}^{2}(y_1, 0, 1)[1-F(\theta+\alpha)] \\
&= \phi_{x_1}^{2}(y_1, 1, 0)F(\alpha) + \phi_{x_1}^{2}(y_1, 0, 0)[1-F(\alpha)].
\end{aligned}$$

Since 1, $F(\alpha)$, and $F(\theta+\alpha)$ are linearly independent, this implies that there exists a function

$\phi^1_{x_1} : \{0,1\} \to \mathbb{R}$ such that, for all $(y^T, x^{2:T})$,

$$\phi_{x_1}(y^T, x^{2:T}) = \phi^1_{x_1}(y_1).$$

Lastly, (7) implies

$$\begin{aligned}
&\sum_{y^T \in \{0,1\}^T} \phi_{x_1}(y^T, x^{2:T}) \prod_{t=1}^{T} F(\theta x_t + \alpha)^{y_t}[1 - F(\theta x_t + \alpha)]^{1-y_t} \\
&= \sum_{y^T \in \{0,1\}^T} \phi^1_{x_1}(y_1) \prod_{t=1}^{T} F(\theta x_t + \alpha)^{y_t}[1 - F(\theta x_t + \alpha)]^{1-y_t} \\
&= \sum_{y_1=0}^{1} \phi^1_{x_1}(y_1) \sum_{y^{2:T} \in \{0,1\}^T} \prod_{t=1}^{T} F(\theta x_t + \alpha)^{y_t}[1 - F(\theta x_t + \alpha)]^{1-y_t} \\
&= \sum_{y_1=0}^{1} \phi^1_{x_1}(y_1) F(\theta x_1 + \alpha)^{y_1}[1 - F(\theta x_1 + \alpha)]^{1-y_1} \\
&= 0.
\end{aligned}$$

Linear independence of 1, $F(\alpha)$, and $F(\theta + \alpha)$ thus implies

$$\phi^1_{x_1}(0) = \phi^1_{x_1}(1) = 0.$$

Therefore, ϕ_{x_1} must be the null function, a contradiction.

F Proof of Proposition 1

It is immediate to verify that, if $\widetilde{\theta} \in \Theta^I$, then (11), (12) and (13) are satisfied. Conversely, suppose that (11), (12) and (13) are satisfied. Let

$$\mu_{x_1}(y^T, x^{2:T}, \alpha) = F(\widetilde{\theta} x_T + \alpha)^{y_T}[1 - F(\widetilde{\theta} x_T + \alpha)]^{1-y_T} \psi_{x_1}(x^{2:T}, y^{T-1}, \alpha). \tag{A9}$$

Using (12) we have

$$\mu_{x_1}(y^T, x^{2:T}, \alpha) \geq 0, \qquad \sum_{y^T \in \{0,1\}^T} \sum_{x^{2:T} \in \{0,1\}^{T-1}} \sum_{\alpha \in \mathcal{S}_\alpha} \mu_{x_1}(y^T, x^{2:T}, \alpha) = 1,$$

so μ_{x_1} is a valid distribution function.

Next, using (11) we have

$$\begin{aligned}
&\sum_{\alpha\in\mathcal{S}_\alpha} \mu_{x_1}(y^T, x^{2:T}, \alpha) \\
&= \sum_{\alpha\in\mathcal{S}_\alpha} F(\widetilde{\theta}x_T+\alpha)^{y_T}[1-F(\widetilde{\theta}x_T+\alpha)]^{1-y_T}\psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) \\
&= Q_{x_1}(y^T, x^{2:T}; \theta, \pi, G),
\end{aligned}$$

so μ_{x_1} is consistent with the conditional distribution $Q_{x_1}(y^T, x^{2:T}; \theta, \pi, G)$ of $(Y_i^T, X_i^{2:T})$ given X_{i1}.
Next, using (13) we have, for all $s \in \{2, ..., T\}$,

$$\begin{aligned}
&\sum_{x^{s:T}\in\{0,1\}^{T-s+1}} \sum_{y^{s:T}\in\{0,1\}^{T-s+1}} \mu_{x_1}(y^T, x^{2:T}, \alpha) \\
&= \sum_{x^{s:T}\in\{0,1\}^{T-s+1}} \sum_{y^{s:T-1}\in\{0,1\}^{T-s}} \left\{\sum_{y_T=0}^{1} F(\widetilde{\theta}x_T+\alpha)^{y_T}[1-F(\widetilde{\theta}x_T+\alpha)]^{1-y_T}\right\}\psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) \\
&= \sum_{x^{s:T}\in\{0,1\}^{T-s+1}} \sum_{y^{s:T-1}\in\{0,1\}^{T-s}} \psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) \\
&= F(\widetilde{\theta}x_{s-1}+\alpha)^{y_{s-1}}[1-F(\widetilde{\theta}x_{s-1}+\alpha)]^{1-y_{s-1}} \sum_{x^{s:T}\in\{0,1\}^{T-s+1}} \sum_{y^{s-1:T-1}\in\{0,1\}^{T-s+1}} \psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) \\
&= F(\widetilde{\theta}x_{s-1}+\alpha)^{y_{s-1}}[1-F(\widetilde{\theta}x_{s-1}+\alpha)]^{1-y_{s-1}} \sum_{x^{s:T}\in\{0,1\}^{T-s+1}} \sum_{y^{s-1:T}\in\{0,1\}^{T-s+2}} \mu_{x_1}(x^{2:T}, y^{T}, \alpha),
\end{aligned}$$

so, for all $t \in \{1, ..., T-1\}$, the conditional distributions of Y_{it} given $(Y_i^{t-1}, X_i^{t-1}, \alpha_i)$ induced by μ_{x_1} coincide with the ones under the model, i.e., with $F(\widetilde{\theta}x_t+\alpha)^{y_t}[1-F(\widetilde{\theta}x_t+\alpha)]^{1-y_t}$.
Lastly, using (A9) we have

$$\begin{aligned}
&\mu_{x_1}(y^T, x^{2:T}, \alpha) = F(\widetilde{\theta}x_T+\alpha)^{y_T}[1-F(\widetilde{\theta}x_T+\alpha)]^{1-y_T}\psi_{x_1}(x^{2:T}, y^{T-1}, \alpha) \\
&= F(\widetilde{\theta}x_T+\alpha)^{y_T}[1-F(\widetilde{\theta}x_T+\alpha)]^{1-y_T} \sum_{y_T=0}^{1} \mu_{x_1}(y^T, x^{2:T}, \alpha),
\end{aligned}$$

so the conditional distribution of Y_{iT} given $(Y_i^{T-1}, X_i^{T-1}, \alpha_i)$ induced by μ_{x_1} also coincides with the one under the model.
This implies that $\widetilde{\theta} \in \Theta^I$.