

Identification of Average Marginal Effects in Fixed Effects Dynamic Discrete Choice Models*

Victor Aguirregabiria[†]
University of Toronto, CEPR

Jesús M. Carro[‡]
Universidad Carlos III de Madrid

September 12, 2024

Abstract

In nonlinear panel data models, fixed effects methods are often criticized because they cannot identify average marginal effects (AMEs) in short panels. The common argument is that identifying AMEs requires knowledge of the distribution of unobserved heterogeneity, but this distribution is not identified in a fixed effects model with a short panel. In this paper, we derive identification results that contradict this argument. In a panel data dynamic logit model, and for T as small as three, we prove the point identification of different AMEs, including causal effects of changes in the lagged dependent variable or the last choice's duration. Our proofs are constructive and provide simple closed-form expressions for the AMEs in terms of probabilities of choice histories. We illustrate our results using Monte Carlo experiments and with an empirical application of a dynamic structural model of consumer brand choice with state dependence.

Keywords: Identification; Average marginal effects; Fixed effects models; Panel data; Dynamic discrete choice; State dependence; Dynamic demand of differentiated products.

JEL codes: C23, C25, C51.

*We thank the editor, three anonymous referees, Manuel Arellano, Dmitry Arkhangelsky, Stephane Bonhomme, Irene Botosaru, Juanjo Dolado, Juan Carlos Escanciano, Christophe Gaillac, Jiaying Gu, Francis Guiton, Bo Honore, Chris Muris, and Enrique Sentana for helpful comments and discussions. We have also benefited from the comments of seminar and conference participants at CEMFI, Toronto, Toulouse, 2020 Econometric Society World Congress, 2021 IAAE Congress, the Panel Data Workshop at Nuffield College, and the Conference in Honor of Manuel Arellano at the Bank of Spain. We thank Susumu Imai for generously sharing the dataset that we use in our empirical application. This research is supported in part by funding from the Social Sciences and Humanities Research Council. The second author gratefully acknowledges support from the *Agencia Estatal de Investigación del Gobierno de España*, grants CEX2021-001181-M (MICIU/AEI/10.13039/501100011033), RTI2018-095231-B-I00 and PID2022-138960OB-I00, and Comunidad de Madrid (Spain) grant EPUC3M11 (VPRICIT).

[†]Department of Economics, University of Toronto. 150 St. George Street, Toronto, ON, M5S 3G7, Canada, victor.aguirregabiria@utoronto.ca.

[‡]Department of Economics, Universidad Carlos III de Madrid. C./ Madrid, 126, 28903 Getafe (Madrid), Spain, jcarro@eco.uc3m.es.

1 Introduction

Ignoring the correlation between unobserved heterogeneity and pre-determined explanatory variables in dynamic panel data models can generate significant biases in estimating dynamic causal effects. The literature distinguishes two approaches to deal with this issue. The *random effects (RE)* approach integrates over the unobserved heterogeneity using a parametric assumption on the distribution of this heterogeneity conditional on the initial values of the predetermined explanatory variables. Typically, this distribution is not identified nonparametrically in short panels, and random effects approaches are not robust to the misspecification of parametric restrictions. This issue is the so-called *initial conditions problem* (Heckman, 1981). In contrast, *fixed effects (FE)* approaches do not restrict this distribution such that identifying parameters of interest is robust to the misspecification of this primitive.

A limitation of FE methods in dynamic discrete choice models with short panels is that they cannot identify the distribution of the time-invariant unobserved heterogeneity. This limitation arises because the data comprises a finite number of probabilities—equivalent to the number of possible choice histories—while the distribution of unobserved heterogeneity has an infinite dimension. This identification problem has generated a more substantial criticism of FE approaches. The applied researcher is often interested in estimating the average marginal effects (AME) of changes in explanatory variables or structural parameters. Since these AMEs are expectations over the distribution of the unobserved heterogeneity, and this distribution is not identified, the common wisdom is that FE approaches cannot (point) identify AMEs.¹

This paper presents new results on the point identification of AMEs in FE dynamic

¹Examples of recent papers describing this common wisdom are [Abrevaya and Hsu \(2021\)](#) (on page 5: "For 'pure' fixed effects models, where the conditional distribution is left unspecified, identification of the partial effects described above would generally require $T \rightarrow \infty$.") and [Honoré and De Paula \(2021\)](#) (on page 2: "It is important to recognize that knowing β [slope parameters] is typically not sufficient for calculating counterfactual distributions or marginal effects. Those will depend on the distribution of α_i [incidental parameters] as well as on β and they are typically not point-identified even if β is.")

logit models. We prove the identification of the AME for a change in the lagged dependent variable, a crucial parameter in dynamic models that measures the causal effect of an agent's past decision on their current decision. Our constructive proofs provide simple closed-form expressions for AMEs based on the probabilities of choice histories in panels where the time dimension can be as small as $T = 3$.

In Section 2, we focus exclusively on binary choice models. Proposition 1 presents a straightforward identification result for dynamic binary logit models. However, it is limited to AMEs where exogenous covariates remain constant over time or are absent from the model. Propositions 2 and 3 present the main contributions of this paper. Proposition 2 provides necessary and sufficient conditions for identifying a general type of AME in a broad class of discrete choice models. Proposition 3 applies these conditions to the dynamic binary logit model, showing the identification of AMEs without requiring covariates to remain constant over time, unlike the more straightforward result in Proposition 1. In Section 2.6, we expand our analysis to include identification results for average transition probabilities (Proposition 4), n-periods forward AMEs (Proposition 5), and AMEs in models with duration dependence (Proposition 6). Section 3 presents our identification and non-identification results for dynamic multinomial logit models (Propositions 7 and 8) and dynamic ordered logit models (Proposition 9).

This paper is related to a large literature on FE estimation of panel data discrete choice models pioneered by Rasch (1961), Andersen (1970), and Chamberlain (1980) for static models, and by Chamberlain (1985) and Honoré and Kyriazidou (2000) for dynamic models. Most papers in this literature focus on identifying and estimating slope parameters and do not present identification results on AMEs. Some important exceptions are Bonhomme (2011), Hoderlein and White (2012), and Chernozhukov, Fernandez-Val, Hahn, and Newey (2013), and, more recently, Dobronyi, Gu, and Kim (2021), Davezies, D'Haultfoeuille, and Laage (2022), Pakel and Weidner (2023), and Botosaru and Muris (2024).²

²Chamberlain (1984), Hahn (2001), and more recently Arellano and Bonhomme (2017), show the identification of a few AMEs in FE nonlinear panel data models. However, these are AMEs for a particular

In an unpublished manuscript, [Bonhomme \(2011\)](#) examines the identification of AMEs within a broad class of non-linear static panel data models, encompassing those with discrete or continuous dependent variables. His work establishes a sufficient condition for identifying AMEs in such models: an injective operator linking the distribution of unobserved heterogeneity to the distribution of the observed dependent variable. However, this condition does not hold in discrete choice models with short panels and continuous distribution of the unobserved heterogeneity. On the other hand, our Proposition 2 establishes both a necessary and sufficient condition for identifying AMEs in a general class of dynamic panel data discrete choice models. Notably, our condition does not necessitate an injective relationship between the distributions of unobserved heterogeneity and the dependent variable. Moreover, we show that our condition holds in dynamic models where the injectivity condition is not satisfied, such as the autoregressive binary logit (Proposition 3), binary logit with duration dependence (Proposition 6), multinomial logit (Proposition 7), and ordered logit (Proposition 9).

[Chernozhukov, Fernandez-Val, Hahn, and Newey \(2013\)](#) – henceforth, CFHN – study the identification of AMEs within nonparametric and semiparametric binary choice models. Their semiparametric model assumes that the transitory shock has a known distribution – e.g., logit model – and corresponds to the model we consider in this paper. They propose a computational method to estimate the bounds in the identified set of the AME. Through numerical examples, they reveal that while the bounds for AME can be notably wide within fully nonparametric models, they rapidly narrow as T increases within the semiparametric model. In contrast to CFHN, our paper adopts a sequential identification strategy, the first step of which is the identification of slope parameters.³ In the second stage, we take the slope

subpopulation of individuals defined by the data. In contrast, we focus on identifying marginal effects defined as averages over the whole population of individuals.

³The sequential approach that we consider in this paper has been used, among others, by [Honoré and De Paula \(2021\)](#) [on page 2 of their paper]: *"It seems that point- or set-identifying and estimating β is a natural first step if one is interested in bounding, say, average marginal effects"*. Previous results on the identification of slope parameters in dynamic logit models include [Chamberlain \(1985\)](#), [Honoré and Kyriazidou \(2000\)](#), [Magnac \(2000, 2004\)](#), [Aguirregabiria, Gu, and Luo \(2021\)](#), [Honoré and Weidner \(2020\)](#), [Dobronyi, Gu, and Kim \(2021\)](#), and [Honoré, Muris, and Weidner \(2021\)](#).

parameters as known to the researcher and consider the identification of AMEs. Our Proposition 2 establishes necessary and sufficient conditions for AME point identification. These conditions consist of infinite equations with a finite number of unknowns. Nonetheless, under the logistic structure, we demonstrate that the infinite system of equations can be effectively encapsulated within a finite framework. Specifically, the unobserved heterogeneity enters through a finite-order polynomial, and the infinite system of equations can be represented as a finite system in terms of the coefficients of each monomial term. We illustrate that this system of equations admits a solution and, through straightforward manipulations, yields closed-form expressions for the targeted AMEs. While CFHN's approach is computationally intensive, owing to the vast dimensionality of unobserved heterogeneity distribution, our method stands out for its computational simplicity as it provides closed-form expressions for AMEs.

Some identification results in our paper are linked to those outlined in [Dobronyi, Gu, and Kim \(2021\)](#) (referred to as DGK hereafter). Within the framework of an autoregressive fixed effects binary logit model, DGK elucidate the sharp identified set of slope parameters, the probability distribution of unobserved heterogeneity, and various functionals derived from these parameters, such as average treatment effects. Their findings stem from a comprehensive examination of the full likelihood function of the model, revealing a polynomial structure in terms of the fixed effects. This polynomial structure also serves as the foundation for our identification results for average treatment effects.⁴ However, it is essential to note that our derivations diverge from those presented in DGK. In particular, our Proposition 2 establishes a necessary and sufficient condition for the point identification of average treatment effects in a general class of panel data discrete choice models beyond logit models. This condition consists of the existence of a solution to a system of equations with analytical expressions. Unlike DGK, we focus on point identification results and a sequential identification approach where slope parameters are initially identified. In contrast, DGK explore both point and

⁴As DGK acknowledge in section 4.1.1 of their paper, our finding of this polynomial structure predates their work.

set identification and delineate the identified set of average marginal effects, even in cases where slope parameters are not point identified. Moreover, while DGK only investigate autoregressive binary logit models, our scope extends to identifying average marginal effects in multinomial and ordered models, where the unobserved heterogeneity is non-scalar. Additionally, we examine models with duration dependence, broadening the applicability of our identification framework across diverse model specifications.

[Davezies, D’Haultfoeuille, and Laage \(2022\)](#), [Pakel and Weidner\(2023\)](#), and [Botosaru and Muris \(2024\)](#) investigate the estimation of AMEs within binary choice logit models, mainly when these parameters are only partially identified. Their research introduces computationally efficient inference methods regarding the identified set. Recognizing the complexities inherent in estimating the sharp identified set, these authors propose diverse approaches for inference on outer bounds for this set.

The remainder of the paper is organized as follows. First, in [Section 2](#), we present our identification results within binary choice logit models. Next, in [Section 3](#), we investigate the identification of AMEs in multinomial and ordered discrete choice models. To illustrate our results, we conduct Monte Carlo experiments, detailed in [Section 4](#), and apply our methodology to a dynamic demand model using consumer scanner data, discussed in [Section 5](#). We summarize and conclude in [section 6](#).

2 Binary choice models

2.1 Model

Consider a panel dataset $\{y_{it}, \mathbf{x}_{it} : i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$ where t represents time, i denotes individuals, and y_{it} can assume two values, $y_{it} \in \mathcal{Y} = \{0, 1\}$. Our focus lies in the investigation of dynamic logit models using panel data. The following equation describes

the binary choice model:

$$y_{it} = \mathbb{1} \{ \alpha_i + \beta y_{i,t-1} + \mathbf{x}'_{it} \boldsymbol{\gamma} + \varepsilon_{it} \geq 0 \}. \quad (1)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function, β and $\boldsymbol{\gamma}$ are slope parameters, and $\boldsymbol{\alpha} \equiv (\alpha_i : i = 1, 2, \dots, N)$ are incidental parameters. The unobservable ε_{it} is i.i.d. over time and across individuals with a Logistic distribution. The explanatory variables in the $K \times 1$ vector \mathbf{x}_{it} are strictly exogenous with respect to the transitory shock ε_{it} . In other words, for any pair of periods (t, s) , the variables \mathbf{x}_{it} and ε_{is} are independently distributed.

The parameter α_i represents permanent unobserved heterogeneity in primitives affecting individuals' decisions, such as preferences or productivity. Its marginal distribution is denoted as $f_\alpha(\alpha_i)$, and $f_{\alpha|\mathbf{x}^{\{1,T\}}}(\alpha_i|\mathbf{x}_i^{\{1,T\}})$ represents the distribution of α_i conditioned on the history of \mathbf{x} variables denoted as $\mathbf{x}_i^{\{1,T\}} = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})$. These distributions are not subject to any constraints. Likewise, the probability of the initial choice y_{i1} conditioned on α_i and $\mathbf{x}_i^{\{1,T\}}$ — represented as $p^*(y_{i1}|\alpha_i, \mathbf{x}_i^{\{1,T\}})$ — is unrestricted. Following the standard setting in fixed effect (FE) approaches, our identification results do not rely on any restrictions on the initial conditions. Assumption 1-BC summarizes the conditions in this binary choice model.

ASSUMPTION 1-BC. (A) (Logit) ε_{it} is i.i.d. over (i, t) with Logistic distribution; (B) (Strict exogeneity of \mathbf{x}_{it}) variable ε_{it} is independent of $(\alpha_i, \mathbf{x}_i^{\{1,T\}})$; and (C) (Fixed effects) the density functions $f_\alpha(\alpha_i)$, $f_{\alpha|\mathbf{x}^{\{1,T\}}}(\alpha_i|\mathbf{x}_i^{\{1,T\}})$, and $p^*(y_{i1}|\alpha_i, \mathbf{x}_i^{\{1,T\}})$ are unrestricted. ■

2.2 Average marginal effects

For the definition of average marginal effects and other parameters of interest, it is convenient to define transition probabilities and their average versions. For $j, k \in \mathcal{Y} = \{0, 1\}$, define the

individual-specific transition probabilities:

$$\pi_{kj}(\alpha_i, \mathbf{x}) \equiv \mathbb{P}(y_{it} = j \mid \alpha_i, y_{i,t-1} = k, \mathbf{x}_{it} = \mathbf{x}), \quad (2)$$

where \mathbf{x} is an arbitrary value chosen by the econometrician. In the binary choice model defined by equation (1) and Assumption 1-BC, we have that:

$$\pi_{11}(\alpha_i, \mathbf{x}) = \Lambda(\alpha_i + \beta + \mathbf{x}'\boldsymbol{\gamma}) \quad \text{and} \quad \pi_{01}(\alpha_i, \mathbf{x}) = \Lambda(\alpha_i + \mathbf{x}'\boldsymbol{\gamma}) \quad (3)$$

where $\Lambda(u)$ is the Logistic function $e^u/[1+e^u]$. Let $\Delta(\alpha_i, \mathbf{x})$ be the *individual-specific* causal effect on y_{it} of a change in $y_{i,t-1}$ from 0 to 1.

$$\begin{aligned} \Delta(\alpha_i, \mathbf{x}) &\equiv \mathbb{E}(y_{it} \mid \alpha_i, y_{i,t-1} = 1, \mathbf{x}_{it} = \mathbf{x}) - \mathbb{E}(y_{it} \mid \alpha_i, y_{i,t-1} = 0, \mathbf{x}_{it} = \mathbf{x}) \\ &= \pi_{11}(\alpha_i, \mathbf{x}) - \pi_{01}(\alpha_i, \mathbf{x}) = \Lambda(\alpha_i + \mathbf{x}'\boldsymbol{\gamma} + \beta) - \Lambda(\alpha_i + \mathbf{x}'\boldsymbol{\gamma}). \end{aligned} \quad (4)$$

This individual-specific causal effect represents the impact of $y_{i,t-1}$ on y_{it} . It measures the persistence of individual i in state 1 generated by *true dynamics* or state dependence.

It is well-established in the literature that parameters β and $\boldsymbol{\gamma}$ are identified in short panels (see section 2.3 below). However, the individual effects α_i are not identified because of the incidental parameters problem (Neyman and Scott, 1948; Heckman, 1981; Lancaster, 2000). Consequently, the individual-specific treatment effects are not identified. Instead, we study the identification of *Average Marginal Effects* (AMEs) defined by integrating the individual effects $\Delta(\alpha_i, \mathbf{x})$ over the distribution of α_i .

In the model without \mathbf{x} covariates (i.e., $\boldsymbol{\gamma} = \mathbf{0}$), the individual-specific effect is $\Delta(\alpha_i) = \Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)$, and the Average Marginal Effect is:

$$AME = \int [\Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)] f_\alpha(\alpha_i) d\alpha_i \quad (5)$$

The sign of the parameter β informs us about the sign of AME . However, the absolute magnitude of β provides minimal insight into the magnitude of AME . Specifically, for any positive value β , we have that AME can take virtually any value within the interval $(0, 1)$ depending on the location of the distribution of α_i . This limitation underscores the crucial importance of identifying Average Marginal Effects.

EXAMPLE 1. Suppose that y_{it} is the indicator for firm i being active in the market during period t . Consider the following thought experiment. At period $t-1$, firms are randomly split into two groups, say groups 0 and 1. Firms in group 0 are designated to be inactive, while those in group 1 are designated to be active. After one period, we examine the proportion of active firms in each group. The Average Marginal Effect defined in equation (5) is the proportion of active firms in group 1 minus the proportion of active firms in group 0. ■

We can extend this type of AME to the model with covariates (when $\gamma \neq \mathbf{0}$). In this scenario, the distribution of α_i depends on the individual's covariate history. Consequently, the AME is not solely determined by the value of \mathbf{x} at which we assess the individual effect $\Delta(\alpha_i, \mathbf{x})$, but also on the values of the covariates that we consider when integrating over the distribution of the individual effect. We provide identification results for two versions of these AMEs. The first AME is defined as an average conditional on a value of the whole history of the covariates between periods 1 and T . For any chosen values of \mathbf{x} and $\mathbf{x}^{\{1,T\}}$, determined by the researcher, we define the following AME:

$$AME(\mathbf{x}, \mathbf{x}^{\{1,T\}}) \equiv \int \Delta(\alpha_i, \mathbf{x}) f_{\alpha|\mathbf{x}^{\{1,T\}}}(\alpha_i | \mathbf{x}_i^{\{1,T\}} = \mathbf{x}^{\{1,T\}}) d\alpha_i \quad (6)$$

This parameter denotes the average value of the marginal effect $\Delta(\alpha_i, \mathbf{x})$ across the population of individuals with a covariate history $\mathbf{x}^{\{1,T\}}$.

In Proposition 1, we provide a straightforward proof of the identification of this AME under two specific conditions: either the model excludes covariates ($\gamma = 0$), or the covariate history $\mathbf{x}^{\{1,T\}}$ is such that the values remain constant $\mathbf{x}_2 = \dots = \mathbf{x}_T = \mathbf{x}$. Proposition 2

establishes necessary and sufficient conditions for identifying the parameter $AME(\mathbf{x}, \mathbf{x}^{\{1,T\}})$ within a broad class of models. By applying these conditions to the binary choice logit model, in Proposition 3, we substantially extend the result of Proposition 1 to a broader context, demonstrating the identification of $AME(\mathbf{x}, \mathbf{x}^{\{1,T\}})$ for any arbitrary selection of \mathbf{x} and $\mathbf{x}^{\{1,T\}}$ as determined by the researcher, provided that $\mathbf{x}_T = \mathbf{x}$.

Often, we are not particularly interested in the causal effect of y_{t-1} on y_t for a subpopulation of individuals with a specific history of covariates, but rather in the average of these effects across the entire distribution of covariate histories. Specifically, the following parameter represents the average value of the marginal effect $\Delta(\alpha_i, \mathbf{x})$ across the population of individuals with a value of the covariates equal to \mathbf{x} at period t :⁵

$$AME_t(\mathbf{x}_{it} = \mathbf{x}) \equiv \int \Delta(\alpha_i, \mathbf{x}_{it}) f_{\alpha|\mathbf{x}_t}(\alpha_i | \mathbf{x}_{it} = \mathbf{x}) d\alpha_i \quad (7)$$

Similarly, the following causal parameter is the average value of $\Delta(\alpha_i, \mathbf{x}_{it})$ across the joint distribution of α_i and \mathbf{x}_{it} at period t :

$$\widetilde{AME}_t \equiv \int \Delta(\alpha_i, \mathbf{x}_{it}) f_{\alpha, \mathbf{x}_t}(\alpha_i, \mathbf{x}_{it}) d\alpha_i d\mathbf{x}_{it} \quad (8)$$

Importantly, the AMEs defined in equations (7) and (8) can be derived by integrating the AME presented in equation (6) over the empirical distribution of the covariate history. Specifically:

$$AME_t(\mathbf{x}_{it} = \mathbf{x}) = \mathbb{E}_{\mathbf{x}_i^{\{1,T\}} | \mathbf{x}_{it} = \mathbf{x}} \left[AME(\mathbf{x}_{it}, \mathbf{x}_i^{\{1,T\}}) \right] \quad (9)$$

$$\widetilde{AME}_t = \mathbb{E}_{\mathbf{x}_i^{\{1,T\}}} \left[AME(\mathbf{x}_{it}, \mathbf{x}_i^{\{1,T\}}) \right]$$

where $\mathbb{E}_{\mathbf{x}_i^{\{1,T\}}} [\cdot]$ represents the expectation operator over the distribution of $\mathbf{x}_i^{\{1,T\}}$. Similarly, $\mathbb{E}_{\mathbf{x}_i^{\{1,T\}} | \mathbf{x}_{it} = \mathbf{x}} [\cdot]$ is the expectation over the distribution of $\mathbf{x}_i^{\{1,T\}}$ conditional on $\mathbf{x}_{it} = \mathbf{x}$.

⁵Since the joint distribution of α_i and \mathbf{x}_{it} can change over time, this AME can vary with t . We can use the identified time-specific AMEs to test the null hypothesis of stationarity of the distribution of $(\alpha_i, \mathbf{x}_{it})$.

2.3 Identification: Slope parameters

We consider a sequential approach to the identification of AMEs. Slope parameters $\theta \equiv (\beta, \gamma)$ are identified in a first step. In the second step, we treat θ as known and focus on establishing the identification of AMEs. The identification of slope parameters in fixed effects dynamic logit models has been well-established in prior literature.⁶ A crucial prerequisite for identification is that the panel must include at least three periods in addition to the initial condition. In models with covariates \mathbf{x}_{it} , the identification of β necessitates the constancy of these covariates for a minimum of two consecutive periods, that is, $Pr(\mathbf{x}_{i2} - \mathbf{x}_{i3} = 0) > 0$. For the identification of γ , an additional condition is necessary: the variance-covariance matrix $Var(\mathbf{x}_{i1} - \mathbf{x}_{i2} | \mathbf{x}_{i2} - \mathbf{x}_{i3} = 0)$ must be full-column rank.

When the support of \mathbf{x}_{it} is discrete, the conditions for consistent estimation of the slope parameters through a Conditional Maximum Likelihood (CML) method remain the same as those for identification. When the support of \mathbf{x}_{it} is continuous, the conditions become somewhat more stringent, as elucidated in Theorem 1 by [Honoré and Kyriazidou \(2000\)](#). For consistent estimation of β , we need the density function of $\mathbf{x}_{i2} - \mathbf{x}_{i3}$ to be strictly positive and continuous in a neighborhood of zero (condition C3 in Theorem 1, [Honoré and Kyriazidou, 2000](#)). Achieving consistent estimation of γ in continuous scenarios necessitates the additional condition that the variance-covariance matrix $Var(\mathbf{x}_{i1} - \mathbf{x}_{i2} | \mathbf{x}_{i2} - \mathbf{x}_{i3} = u)$ is full-column rank for every value u in a neighborhood of zero (condition C6 in Theorem 1, [Honoré and Kyriazidou, 2000](#)).

⁶Specifically, these identification results on slope parameters include: [Chamberlain \(1985\)](#) for the binary choice AR(1) model without exogenous regressors; [Magnac \(2000\)](#) for multinomial AR(1) models; [Honoré and Kyriazidou \(2000\)](#) for binary and multinomial models with exogenous regressors; [Aguirregabiria, Gu, and Luo \(2021\)](#) for models with duration dependence; [Honoré and Weidner \(2020\)](#) and [Dobronyi, Gu, and Kim \(2021\)](#) for binary $AR(p)$ models with $p \geq 2$; and [Honoré, Muris, and Weidner \(2021\)](#) for the dynamic ordered logit.

2.4 Identification of AME: A simple result

The distribution of the fixed effects is not non-parametrically identified in discrete choice models with short panels. Consequently, conventional wisdom dictates that functions dependent on this distribution, such as AMEs, remain non-identifiable. Contrary to this prevailing belief, our research challenges this notion, showing specific instances where AMEs are point-identified without necessitating full knowledge or imposing constraints on the fixed effects distribution. In this section, we present a straightforward yet significant and novel result. In section 2.5, we introduce a method based on new necessary and sufficient conditions, which leads to more general identification results.

Consider the binary choice model in equation (1). Our proof of identification exploits a relationship between the individual effect $\Delta(\alpha_i, \mathbf{x})$, the transition probabilities $\pi_{01}(\alpha_i, \mathbf{x})$ and $\pi_{11}(\alpha_i, \mathbf{x})$, and the parameter β in the Logit model. The following Lemma 1 establishes this relationship.

Lemma 1. *In the binary choice model defined by equation (1) and Assumption 1-BC, the following conditions hold:*

$$\Delta(\alpha_i, \mathbf{x}) = [e^\beta - 1] \pi_{01}(\alpha_i, \mathbf{x}) \pi_{10}(\alpha_i, \mathbf{x}), \quad (10)$$

and

$$e^\beta = \frac{\pi_{11}(\alpha_i, \mathbf{x}) \pi_{00}(\alpha_i, \mathbf{x})}{\pi_{10}(\alpha_i, \mathbf{x}) \pi_{01}(\alpha_i, \mathbf{x})}. \quad \blacksquare \quad (11)$$

Proof of Lemma 1. See Appendix A.1.

Proposition 1 establishes the identification of *AME* in the dynamic binary choice model without covariates, and of $AME(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$ in the model with covariates.

Proposition 1. *Consider the binary choice model defined by equation (1), under Assumption 1-BC, with a fixed $T \geq 3$ and known β . In the model without covariates (i.e., $\boldsymbol{\gamma} = 0$), *AME**

is identified by the following expression:

$$AME = [e^\beta - 1] [\mathbb{P}_{0,1,0} + \mathbb{P}_{1,0,1}] \quad (12)$$

where $\mathbb{P}_{y_1, y_2, y_3}$ is the probability of the choice history (y_1, y_2, y_3) . In the model with covariates (i.e., $\gamma \neq 0$), for any arbitrary values \mathbf{x} and \mathbf{x}_1 chosen by the econometrician, the causal effect parameter $AME(\mathbf{x}, \mathbf{x}^{\{1,3\}} = [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$ is point identified by the following expression

$$AME(\mathbf{x}, \mathbf{x}^{\{1,3\}} = [\mathbf{x}_1, \mathbf{x}, \mathbf{x}]) = [e^\beta - 1] [\mathbb{P}_{0,1,0|\mathbf{x}_1, \mathbf{x}, \mathbf{x}} + \mathbb{P}_{1,0,1|\mathbf{x}_1, \mathbf{x}, \mathbf{x}}] \quad (13)$$

where $\mathbb{P}_{y_1, y_2, y_3|\mathbf{x}^{\{1,3\}}}$ is the probability of the choice history (y_1, y_2, y_3) conditional on the covariate history $\mathbf{x}^{\{1,3\}}$. In equation (13), \mathbf{x}_1 is free to take any value, while $\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}$.

■

Proof of Proposition 1. See Appendix A.2.

Remark 1.1. The proof of identification of $AME(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$ in Proposition 1 relies on the restriction $\mathbf{x}_{i2} = \mathbf{x}_{i3} = \mathbf{x}$. This constraint arises from the application of Lemma 1, which establishes $\Delta(\alpha_i, \mathbf{x}) = [e^\beta - 1] \pi_{01}(\alpha_i, \mathbf{x}) \pi_{10}(\alpha_i, \mathbf{x})$. To relate this Lemma to the probability of a choice history (y_{i1}, y_{i2}, y_{i3}) , it is necessary that $\mathbf{x}_{i2} = \mathbf{x}_{i3} = \mathbf{x}$.

Remark 1.2. Since Proposition 1 applies to arbitrary values of \mathbf{x}_1 and \mathbf{x} , it enables the construction of the following integrated AME:

$$\widetilde{AME}_{\mathbf{x}_2=\mathbf{x}_3} = \int AME(\mathbf{x}, \mathbf{x}_i^{\{1,3\}}) p(\mathbf{x}_i^{\{1,3\}}|\mathbf{x}_{i2} = \mathbf{x}_{i3}) d\mathbf{x}_i^{\{1,3\}} \quad (14)$$

where $p(\mathbf{x}_i^{\{1,3\}}|\mathbf{x}_{i2} = \mathbf{x}_{i3})$ represents the empirical distribution of the covariate history conditional on $\mathbf{x}_{i2} = \mathbf{x}_{i3}$. While this is a relevant average causal parameter, its interest is limited by the constraint of covariate constancy between two consecutive periods, making it applicable only to that specific sub-population. In cases where \mathbf{x} includes continuous covariates,

this integrated AME applies to a subpopulation of mass zero. To address this limitation and obtain an AME without the restrictive condition $\mathbf{x}_{i2} = \mathbf{x}_{i3}$, we introduced a causal parameter denoted as $AME_t(\mathbf{x}_{it} = \mathbf{x})$, as specified in equation (7), and its integrated counterpart, \widetilde{AME}_t , as defined in equation (8). Proposition 3 below establishes the identification of $AME_t(\mathbf{x}_{it} = \mathbf{x})$ and of \widetilde{AME}_t , providing an average causal effect applicable to the entire population.

Remark 1.3 (*Over-identification*). For panels where $T \geq 4$, there are over-identifying restrictions on $AME(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$ because we can construct the empirical distribution of 3-period histories using various groups of periods. As an illustration, when $T = 4$, we can consider two groups: $(y_1, \mathbf{x}_1, y_2, \mathbf{x}_2, y_3, \mathbf{x}_3)$ and $(y_2, \mathbf{x}_2, y_3, \mathbf{x}_3, y_4, \mathbf{x}_4)$. For each of these groups, a distinct estimator of $AME(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$ can be derived, such that this AME is over-identified. These different estimates of the same AME can be used to test the over-identifying restrictions. If the hypothesis is not rejected, the researcher can combine these estimates to obtain a more precise estimate of the AME.

Remark 1.4 (*Estimation*). Equation (13) suggests a simple analog or plug-in estimator for the AMEs. In a first step, we estimate $\boldsymbol{\theta} \equiv (\beta, \boldsymbol{\gamma})$ using Conditional Maximum Likelihood (CML) and the probabilities $\mathbb{P}_{y_1, y_2, y_3 | \mathbf{x}^{\{1,3\}}}$ using a frequency estimator. Then, we plug these estimates in equation (13) to obtain estimates of $AME(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$. When the support of \mathbf{x} is discrete, this estimator is root- N consistent and asymptotically normal. This estimator is a function of slope parameter estimates and frequency estimates of choice histories. When T is greater than three, the set of three-period histories used to derive this estimator may contain overlapping observations. For these reasons, the bootstrap method emerges as a straightforward and convenient approach for inference on $AME(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$. Indeed, this is the methodology employed in our empirical application in Section 5. For the estimation of the integrated $\widetilde{AME}_{\mathbf{x}_2 = \mathbf{x}_3}$ defined in equation (14), we can use a kernel-weighted estimator in line with [Honoré and Kyriazidou \(2000\)](#).

2.5 Identification of AME: Necessary and sufficient conditions

The identification findings in the preceding section indicate that, despite the inherent challenges arising from the non-identification of fixed effects distribution, there exist instances where average marginal effects are identified with fixed T . The proof establishes a specific linear combination (weighted sum) of probability choices that equates to the AME. The question then arises: How can one determine these weights? Is there a systematic procedure to assess the existence of a linear combination that identifies the AME in other models?

In this section, we derive necessary and sufficient conditions, developing a constructive approach to obtain AMEs within a broad class of dynamic logit models. Employing this method, we establish the identification of various AMEs or causal effects and unveil instances of non-identification for specific AMEs.

Let $\mathbf{y}_i^{\{1,T\}} \in \mathcal{Y}^T$ be the vector with individual i 's observed choice history, and let $\mathbf{x}_i^{\{1,T\}} \in \mathcal{X}^T$ represent the history of the exogenous variables. We use $\mathbf{y}_i^{\{2,T\}} \in \mathcal{Y}^{T-1}$ to denote the sub-history from period 2 to T . For arbitrary histories $\mathbf{y}^{\{1,T\}}$ and $\mathbf{x}^{\{1,T\}}$, let $\mathbb{P}_{\mathbf{y}^{\{1,T\}}|\mathbf{x}^{\{1,T\}}}$ represent the probability of a choice history conditional on a covariate history. This probability is identified from the data. Let $\mathbf{P}_{\mathcal{Y}^T|\mathcal{X}^T}$ be the vector with the probabilities $\mathbb{P}_{\mathbf{y}^{\{1,T\}}|\mathbf{x}^{\{1,T\}}}$ for every possible value of $\mathbf{y}^{\{1,T\}}$ and $\mathbf{x}^{\{1,T\}}$ in $\mathcal{Y}^T \times \mathcal{X}^T$. Vector $\mathbf{P}_{\mathcal{Y}^T|\mathcal{X}^T}$ contains all the information in the data that is relevant to identify slope parameters, the distribution of α , and any AME of interest.

According to the model, the probability $\mathbb{P}_{\mathbf{y}^{\{1,T\}}|\mathbf{x}^{\{1,T\}}}$ has the following structure:

$$\mathbb{P}_{\mathbf{y}^{\{1,T\}}|\mathbf{x}^{\{1,T\}}} = \int G(\mathbf{y}^{\{2,T\}} | y_1, \mathbf{x}^{\{1,T\}}, \alpha_i, \boldsymbol{\theta}) p^*(y_1 | \alpha_i, \mathbf{x}^{\{1,T\}}) f_{\alpha|\mathbf{x}^{\{1,T\}}}(\alpha_i | \mathbf{x}^{\{1,T\}}) d\alpha_i, \quad (15)$$

where $G(\mathbf{y}^{\{2,T\}} | y_1, \mathbf{x}^{\{1,T\}}, \alpha_i, \boldsymbol{\theta})$ is the probability of sub-history $\mathbf{y}^{\{2,T\}}$ predicted by the model. For instance, in the binary logit model:

$$G(\mathbf{y}^{\{2,T\}} | y_1, \mathbf{x}^{\{1,T\}}, \alpha_i, \boldsymbol{\theta}) \equiv \prod_{t=2}^T \Lambda([2y_t - 1] [\alpha_i + \beta y_{t-1} + \mathbf{x}'_t \boldsymbol{\gamma}]) \quad (16)$$

We can say that $AME(\mathbf{x}, \mathbf{x}^{\{1,T\}})$ is point identified if there is a function $h(\mathbf{P}_{\mathcal{Y}^T|\mathcal{X}^T}, \boldsymbol{\theta})$ such that $AME(\mathbf{x}, \mathbf{x}^{\{1,T\}}) = h(\mathbf{P}_{\mathcal{Y}^T|\mathcal{X}^T}, \boldsymbol{\theta})$. Proposition 2 states a necessary and sufficient condition for the point identification of this AME.

Proposition 2. *Consider a fixed effects model represented by the probability function G . Let $AME(\mathbf{x}, \mathbf{x}^{\{1,T\}})$ denote the causal parameter defined in equation (6) for arbitrary values of \mathbf{x} and $\mathbf{x}^{\{1,T\}}$ chosen by the researcher. This AME is point-identified if and only if there exists a weighting function $w_{\mathbf{y}^{\{1,T\}}, \mathbf{x}^{\{1,T\}}, \boldsymbol{\theta}}$, mapping from $\mathcal{Y}^T \times \mathcal{X}^T \times \Theta$ to \mathbb{R} , which satisfies the following equation:*

$$\sum_{\mathbf{y}^{\{2,T\}} \in \mathcal{Y}^{T-1}} w_{y_1, \mathbf{y}^{\{2,T\}}, \mathbf{x}^{\{1,T\}}, \boldsymbol{\theta}} G(\mathbf{y}^{\{2,T\}} | y_1, \mathbf{x}^{\{1,T\}}, \alpha_i, \boldsymbol{\theta}) = \Delta(\alpha_i, \mathbf{x}), \quad (17)$$

for every $y_1 \in \mathcal{Y}$ and $\alpha_i \in \mathbb{R}$. Furthermore, this condition implies the following form for the function that identifies $AME(\mathbf{x}, \mathbf{x}^{\{1,T\}})$:

$$AME(\mathbf{x}, \mathbf{x}^{\{1,T\}}) = \sum_{\mathbf{y}^{\{1,T\}} \in \mathcal{Y}^T} w_{\mathbf{y}^{\{1,T\}}, \mathbf{x}^{\{1,T\}}, \boldsymbol{\theta}} \mathbb{P}_{\mathbf{y}^{\{1,T\}} | \mathbf{x}^{\{1,T\}}} \quad \blacksquare \quad (18)$$

Proof of Proposition 2. See Appendix A.3.

Remark 2.1. Proposition 2 places no constraints on the form of function $G(\cdot)$. Specifically, it does not require the structure outlined in equation (16). Consequently, Proposition 2 applies to a broad class of fixed-effects dynamic discrete choice models, extending beyond the logit class.

Remark 2.2. Bonhomme (2011) delves into the identification of AMEs within a broad range of non-linear static panel data models, including those with discrete or continuous dependent variables. His work establishes a sufficient condition, albeit not necessary, for identifying AMEs in such models: the existence of an injective operator connecting the distribution of unobserved heterogeneity to the distribution of the observed dependent variable. However,

this condition fails in discrete choice models with short panels and unobserved heterogeneity with continuous support. In contrast, our Proposition 2 establishes both a necessary and sufficient condition for identifying AMEs in a general class of panel data discrete choice models. Notably, our condition does not require an injective relationship between the distributions of unobserved heterogeneity and the dependent variable.

Remark 2.3. Looking at the structure of the system of equations in (17), one may be inclined to perceive that meeting the necessary and sufficient conditions for identification established in this Proposition seems quite unlikely. Specifically, note that (17) establishes an infinite system of equations – as many equations as values of α_i . The researcher knows the closed-form expressions for functions $G(\cdot)$ and $\Delta(\cdot)$. The unknown variables in this system of equations consist of the weights $w_{\mathbf{y}^{\{1,T\}}, \mathbf{x}^{\{1,T\}}, \boldsymbol{\theta}}$ for every $\mathbf{y}^{\{1,T\}} \in \mathcal{Y}^T$. Given that the set \mathcal{Y}^T is finite, we are in a scenario where the system presents an infinite number of restrictions with only a finite set of unknowns. In the absence of a specific structure, this system lacks a solution. This feature underpins the common belief that AMEs are not identifiable within this class of models.

Proposition 3 below establishes that, in the binary choice logit model defined by equation (1) and Assumption 1-BC, the structure of functions $G(\cdot)$ and $\Delta(\cdot)$ is such that equation (17) can be represented as a finite order polynomial in the variable e^{α_i} . This result implies that there is a solution to the system if and only if the coefficients multiplying every monomial term in this polynomial are all equal to zero. This property transforms the infinite system of equations into a finite linear system with finite unknowns. Furthermore, if a solution exists, this solution implies a closed-form expression for the weights $w_{\mathbf{y}^{\{1,T\}}, \mathbf{x}^{\{1,T\}}, \boldsymbol{\theta}}$, and therefore, for the expression that identifies the *AME*.

Proposition 3. *In the binary choice model defined by equation (1) and Assumption 1-BC:*

(A) *In Proposition 2, we can express equation (17) as a finite-order polynomial in the*

variable e^{α_i} . This formulation leads to a system characterized by a finite number of linear equations, where the unknowns are the weights $w_{\mathbf{y}^{\{1,T\}}, \mathbf{x}^{\{1,T\}}, \theta}$ for every $\mathbf{y}^{\{1,T\}} \in \mathcal{Y}^T$.

(B) For $T \geq 3$, the causal parameter $AME(\mathbf{x}, \mathbf{x}^{\{1,T\}})$, defined in equation (6), is identified for any value \mathbf{x} and $\mathbf{x}^{\{1,T\}}$ chosen by the researcher with $\mathbf{x}_T = \mathbf{x}$. For instance, for $T = 3$:

$$\begin{aligned} AME(\mathbf{x}, \mathbf{x}^{\{1,3\}}) &= w_{0,0,1, \mathbf{x}^{\{1,3\}}} \mathbb{P}_{0,0,1 | \mathbf{x}^{\{1,3\}}} + w_{0,1,0, \mathbf{x}^{\{1,3\}}} \mathbb{P}_{0,1,0 | \mathbf{x}^{\{1,3\}}} \\ &+ w_{1,0,1, \mathbf{x}^{\{1,3\}}} \mathbb{P}_{1,0,1 | \mathbf{x}^{\{1,3\}}} + w_{1,1,0, \mathbf{x}^{\{1,3\}}} \mathbb{P}_{1,1,0 | \mathbf{x}^{\{1,3\}}} \end{aligned} \quad (19)$$

where $\mathbb{P}_{y_1, y_2, y_3 | \mathbf{x}^{\{1,3\}}}$ is the probability of (y_1, y_2, y_3) conditional on $\mathbf{x}^{\{1,3\}}$. The weights are:

$$\begin{aligned} w_{0,0,1, \mathbf{x}^{\{1,3\}}} &= -1 + e^{[\mathbf{x}_2 - \mathbf{x}_3]' \boldsymbol{\gamma}} ; \quad w_{0,1,0, \mathbf{x}^{\{1,3\}}} = -1 + e^{\beta + [\mathbf{x}_3 - \mathbf{x}_2]' \boldsymbol{\gamma}} ; \\ w_{1,0,1, \mathbf{x}^{\{1,3\}}} &= -1 + e^{\beta + [\mathbf{x}_2 - \mathbf{x}_3]' \boldsymbol{\gamma}} ; \quad w_{1,1,0, \mathbf{x}^{\{1,3\}}} = -1 + e^{[\mathbf{x}_3 - \mathbf{x}_2]' \boldsymbol{\gamma}} . \quad \blacksquare \end{aligned} \quad (20)$$

Proof of Proposition 3. See Appendix A.4.

Remark 3.1. Two significant corollaries of Proposition 3 are the identification of the AMEs $AME_t(\mathbf{x}_{it} = \mathbf{x})$ and \widetilde{AME}_t , defined in equations (7) and (8), respectively. As established in equation (9), both AMEs can be derived by integrating the AME identified in Proposition 3 over a distribution of observed covariate histories. Specifically, for $T = 3$ and $t = 3$:

$$\begin{aligned} AME_3(\mathbf{x}_{i3} = \mathbf{x}) &= \mathbb{E}_{\mathbf{x}_i^{\{1,3\}} | \mathbf{x}_{i3} = \mathbf{x}} \left[AME(\mathbf{x}_{i3}, \mathbf{x}_i^{\{1,3\}}) \right] \\ \widetilde{AME}_3 &= \mathbb{E}_{\mathbf{x}_i^{\{1,3\}}} \left[AME(\mathbf{x}_{i3}, \mathbf{x}_i^{\{1,3\}}) \right] \end{aligned} \quad (21)$$

Remark 3.2. It is noteworthy to highlight the broader scope of the identification result presented in Proposition 3 compared to that in Proposition 1. While Proposition 1 enables the identification of $AME(\mathbf{x}, \mathbf{x}^{\{1,3\}} = [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$ for any given value of \mathbf{x}_1 and $\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}$ and allows us to integrate these AMEs across all values of \mathbf{x}_1 and \mathbf{x} , this "integrated AME"

still carries the restriction of covariates at two consecutive periods, rendering this causal effect applicable only to a specific sub-population. The identified AME in Proposition 3, and the implied identification of $AME_t(\mathbf{x}_{it} = \mathbf{x})$ and \widetilde{AME}_t , are not subject to that constraint.

Remark 3.3. When $T = 3$ and the history of the exogenous covariates is such that $\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}$, the identification result in Proposition 3(B) is the one in Proposition 1. That is, the weights that solve the system of equations (17) lead to the expression $AME(\mathbf{x}, \mathbf{x}_i^{\{1,3\}} = [\mathbf{x}_1, \mathbf{x}, \mathbf{x}]) = [e^\beta - 1] [\mathbb{P}_{0,1,0|[\mathbf{x}_1, \mathbf{x}, \mathbf{x}]} + \mathbb{P}_{1,0,1|[\mathbf{x}_1, \mathbf{x}, \mathbf{x}]}]$.

Remark 3.4. (*Over-identification*). For panels with $T \geq 4$, there are over-identifying restrictions on the parameter $AME(\mathbf{x}, \mathbf{x}^{\{1,3\}})$. In Appendix A.4.3, we employ the same procedure to derive the closed-form expression for the weights when T exceeds three. Notably, for panels with $T \geq 4$, there emerge over-identifying restrictions on $AME(\mathbf{x}, \mathbf{x}^{\{1,T\}})$.

Remark 3.5. Proposition 3 places no constraints on the stochastic process of \mathbf{x}_{it} , except that it is strictly exogenous concerning the transitory shock ε_{it} . Furthermore, this identification result applies to scenarios where \mathbf{x}_{it} includes continuous variables.

Remark 3.6. In the dynamic binary logit model, Proposition 3(A) can be extended to AMEs where $\Delta(\mathbf{x}, \alpha_i)$ represents the product, ratio, or a polynomial function of logit transition probabilities, provided these probabilities are evaluated at the same covariate values \mathbf{x} . Proposition 3(B) addresses the existence of a solution to the finite system of equations, which needs to be determined individually for each case.

2.6 Extensions

2.6.1 Other average treatment effects

In a binary choice model without covariates, for $k, j \in \{0, 1\}$, we define Π_{kj} as the average transition probability from k to j that results from integrating the individual-specific

transition probability over the distribution of α_i .

$$\Pi_{kj} \equiv \int \pi_{kj}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \quad (22)$$

For instance, $\Pi_{01} = \int \Lambda(\alpha_i) f_\alpha(\alpha_i) d\alpha_i$, and $\Pi_{11} = \int \Lambda(\alpha_i + \beta) f_\alpha(\alpha_i) d\alpha_i$. In the model with covariates, we define $\Pi_{kj}(\mathbf{x}, \mathbf{x}^{\{1,T\}})$ as the average transition probability from k to j with the current value of the covariates is \mathbf{x} and integrated over the distribution of α_i conditional on the covariate history $\mathbf{x}^{\{1,T\}}$:

$$\Pi_{kj}(\mathbf{x}, \mathbf{x}^{\{1,T\}}) \equiv \int \pi_{kj}(\alpha_i, \mathbf{x}) f_{\alpha|\mathbf{x}^{\{1,T\}}}(\alpha_i | \mathbf{x}_i^{\{1,T\}} = \mathbf{x}^{\{1,T\}}) d\alpha_i \quad (23)$$

Proposition 4 establishes the identification of these average transition probabilities.

Proposition 4. *Consider the binary choice model defined by equation (1), under Assumption 1-BC, with a fixed $T \geq 3$ and known θ . In the model without covariates, all the average transition probabilities Π_{00} , Π_{01} , Π_{10} , and Π_{11} by the following expressions:*

$$\begin{cases} \Pi_{11} &= \mathbb{P}_{1,1} + \mathbb{P}_{0,1,1} + e^\beta \mathbb{P}_{0,1,0} \\ \Pi_{00} &= \mathbb{P}_{0,0} + \mathbb{P}_{1,0,0} + e^\beta \mathbb{P}_{1,0,1} \end{cases} \quad (24)$$

and by definition, $\Pi_{10} = 1 - \Pi_{11}$ and $\Pi_{01} = 1 - \Pi_{00}$. In the model with covariates, for any arbitrary values \mathbf{x} and \mathbf{x}_1 chosen by the econometrician, the average transition probabilities $\Pi_{00}(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$, $\Pi_{01}(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$, $\Pi_{10}(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$, and $\Pi_{11}(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}])$ are point identified by the following expressions:

$$\begin{cases} \Pi_{11}(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}]) &= \mathbb{P}_{1,1|\mathbf{x}_1, \mathbf{x}} + \mathbb{P}_{0,1,1|\mathbf{x}_1, \mathbf{x}, \mathbf{x}} + e^\beta \mathbb{P}_{0,1,0|\mathbf{x}_1, \mathbf{x}, \mathbf{x}} \\ \Pi_{00}(\mathbf{x}, [\mathbf{x}_1, \mathbf{x}, \mathbf{x}]) &= \mathbb{P}_{0,0|\mathbf{x}_1, \mathbf{x}} + \mathbb{P}_{1,0,0|\mathbf{x}_1, \mathbf{x}, \mathbf{x}} + e^\beta \mathbb{P}_{1,0,1|\mathbf{x}_1, \mathbf{x}, \mathbf{x}} \end{cases} \quad (25)$$

where \mathbf{x}_1 is free to take any value, while $\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}$. ■

Proof of Proposition 4. See Appendix A.5.

Remark 4.1. Using the definition of AME , it is straightforward to show that $AME = \Pi_{11} - \Pi_{01}$. Identifying the average transition probabilities is an alternative way of identifying the AME . This identification not only elucidates the AME but also extends to other relevant causal effects, encompassing the ratio Π_{11}/Π_{01} , the percentage change $[(\Pi_{11} - \Pi_{01})/\Pi_{01}]$, the additive effect $\Pi_{01} + \Pi_{11}$, a weighted sum of Π_{01} and Π_{11} , or, more broadly, any well-defined function involving these parameters.

Remark 4.2. Proposition 4 implies also the identification of two interesting average treatment effects ($ATEs$). For concreteness, we describe these $ATEs$ using the application in *Example 1* above. Consider a policy experiment where firms in the experimental group are assigned to the active status at period $t - 1$. For instance, they receive a large temporary subsidy to operate in the market. Firms in the control group are left in their observed status at period $t - 1$. Then, at period t , the researcher observes the proportion of firms that remain active in the experimental and control groups. The difference between these two proportions is the average effect of this policy treatment, which we can denote as $ATE_{11,t}$. According to the model, this average treatment effect is $ATE_{11,t} \equiv \Pi_{11} - \mathbb{E}(y_{it}|t)$, where $\mathbb{E}(y_{it}|t)$ is the mean value of y in the actual distribution of this variable at period t . Since this distribution may change over time, this ATE may also vary with t . We can consider a similar experiment where firms in the experimental group are assigned to be inactive at period $t - 1$ – they receive a large temporary subsidy for being inactive. We use $ATE_{01,t}$ to denote the average effect of this other policy treatment. By definition, $ATE_{01,t} \equiv \Pi_{01} - \mathbb{E}(y_{it}|t)$.

2.6.2 n-periods forward AME

Researchers can be interested in the response to a treatment over multiple periods. Let $\Delta^{(n)}(\alpha_i)$ denote the individual-specific causal effect on $y_{i,t+n}$ resulting from a change in y_{it}

from 0 to 1.

$$\Delta^{(n)}(\alpha_i) \equiv \mathbb{E}(y_{i,t+n} \mid \alpha_i, y_{it} = 1) - \mathbb{E}(y_{i,t+n} \mid \alpha_i, y_{it} = 0) = \pi_{11}^{(n)}(\alpha_i) - \pi_{01}^{(n)}(\alpha_i). \quad (26)$$

As previously mentioned in the context of $\Delta(\alpha_i)$, the individual effect projected forward over n periods remains unidentifiable in a short panel. Our focus, however, lies in the average of this effect:

$$AME^{(n)} \equiv \int \Delta^{(n)}(\alpha_i) f_{\alpha}(\alpha_i) d\alpha_i = \int [\pi_{11}^{(n)}(\alpha_i) - \pi_{01}^{(n)}(\alpha_i)] f_{\alpha}(\alpha_i) d\alpha_i. \quad (27)$$

This n -periods forward average causal effect differs from raising the 1-period AME to the power of n : specifically, $AME^{(n)} \neq [AME]^n$. Consequently, identifying $AME^{(n)}$ is not a straightforward implication from identifying the singular AME .

In the binary choice model with \mathbf{x} covariates, we can define n -periods forward AMEs conditional on a constant value of \mathbf{x} :

$$AME^{(n)}(\mathbf{x}^{\{1,T\}} = [\mathbf{x}, \dots, \mathbf{x}]) = \int [\pi_{11}^{(n)}(\alpha_i, \mathbf{x}) - \pi_{01}^{(n)}(\alpha_i, \mathbf{x})] f_{\alpha|\mathbf{x}}(\alpha_i | \mathbf{x}_i^{\{1,T\}} = [\mathbf{x}, \dots, \mathbf{x}]) d\alpha_i \quad (28)$$

Our proof of the identification of $AME^{(n)}$ builds on Lemma 1 and the following Lemma 2.

Lemma 2. *In the binary choice model defined by equation (1) and Assumption 1-BC, the n -periods forward individual causal effect with constant-over-time \mathbf{x} satisfies the following equation:*

$$\Delta^{(n)}(\alpha_i, \mathbf{x}) = [e^{\beta} - 1]^n [\pi_{10}(\alpha_i, \mathbf{x})]^n [\pi_{01}(\alpha_i, \mathbf{x})]^n. \quad \blacksquare \quad (29)$$

Proof of Lemma 2. See Appendix A.6.

Proposition 5. *Consider the binary choice model defined by equation (1) and Assumption 1-BC. Let n be any positive integer, and let $\widetilde{\mathbf{10}}^n$ be the choice history that consists of the*

n -times repetition of the sequence $(1,0)$, e.g., for $n = 2$, we have that $\widetilde{\mathbf{10}}^2 = (1,0,1,0)$. If $T \geq 2n + 1$, then parameter $AME^{(n)}(\mathbf{x}^{\{1,T\}} = [\mathbf{x}, \dots, \mathbf{x}])$ is identified as:

$$AME^{(n)}(\mathbf{x}^{\{1,T\}} = [\mathbf{x}, \dots, \mathbf{x}]) = [e^\beta - 1]^n \left[\mathbb{P}_{0, \widetilde{\mathbf{10}}^n | \mathbf{x}, \dots, \mathbf{x}} + \mathbb{P}_{\widetilde{\mathbf{10}}^n, 1 | \mathbf{x}, \dots, \mathbf{x}} \right] \quad (30)$$

where $\mathbb{P}_{0, \widetilde{\mathbf{10}}^n | \mathbf{x}, \dots, \mathbf{x}}$ and $\mathbb{P}_{\widetilde{\mathbf{10}}^n, 1 | \mathbf{x}, \dots, \mathbf{x}}$ are the probabilities of choice histories $(0, \widetilde{\mathbf{10}}^n)$ and $(\widetilde{\mathbf{10}}^n, 1)$, respectively, conditional on $\mathbf{x}_1 = \dots = \mathbf{x}_T = \mathbf{x}$. ■

Proof of Proposition 5. See Appendix A.7.

2.6.3 Duration dependence

There are many applications of dynamic models where the dependent variable has duration dependence. For instance, in a model of individual employment, where $y = 1$ represents employment and $y = 0$ unemployment, a worker's productivity often rises with job experience, suggesting that the likelihood of employment increases over the duration in that state. Similarly, in a market entry and exit model, where $y = 1$ denotes a firm's activity in the market, and $y = 0$ denotes inactivity, a firm's profit may rise with its market experience, increasing the probability that the firm stays active.

Let $d_{it} \in \{0, 1, 2, \dots\}$ represent the duration in the choice of alternative $y = 1$. This duration variable can be defined recursively as $d_{it} = y_{i,t-1} (d_{i,t-1} + 1)$. We consider the following dynamic binary choice logit model with duration dependence:

$$y_{it} = \mathbb{1} \{ \alpha_i + \beta y_{i,t-1} + \delta d_{it} + \varepsilon_{it} \geq 0 \}, \quad (31)$$

where β and δ are slope parameters.⁷

⁷We can extend our identification results of AMEs for models with duration dependence to allow for a more flexible specification with a different parameter for each possible value of duration greater than 1. That is, we can replace the term δd_{it} with $\delta_2 \mathbb{1}\{d_{it} = 2\} + \delta_3 \mathbb{1}\{d_{it} = 3\} + \dots$ where $\mathbb{1}\{\cdot\}$ is the indicator function, and $\delta_2, \delta_3, \dots$ are parameters.

We are interested in the causal effect on y_{it} of a change in the duration variable d_{it} . For instance, in a market entry/exit model, we can be interested in the causal effect of one more year of experience on the probability of being active in the market. Let $\Delta_d(\alpha_i, d)$ be the *individual-specific* causal effect on y_{it} of a unit change in duration from d to $d + 1$.

$$\begin{aligned}\Delta_d(\alpha_i, d) &\equiv \mathbb{E}(y_{it} \mid \alpha_i, d_{it} = d + 1) - \mathbb{E}(y_{it} \mid \alpha_i, d_{it} = d) \\ &= \pi_{11}(\alpha_i, d + 1) - \pi_{11}(\alpha_i, d) = \Lambda(\alpha_i + \beta + \delta(d + 1)) - \Lambda(\alpha_i + \beta + \delta d).\end{aligned}\tag{32}$$

where $\pi_{11}(\alpha_i, d) \equiv \mathbb{E}(y_{it} \mid \alpha_i, y_{i,t-1} = 1, d_{it} = d)$. We are interested in the identification of the following *AME*:

$$AME_d(d) \equiv \int \Delta_d(\alpha_i, d) f_\alpha(\alpha_i) d\alpha_i = \int [\pi_{11}(\alpha_i, d + 1) - \pi_{11}(\alpha_i, d)] f_\alpha(\alpha_i) d\alpha_i \tag{33}$$

In this binary choice model with duration dependence, we need to slightly modify Assumption 1-BC to consider that the initial condition of a choice history includes not only y_{i1} but also the initial duration d_{i1} . Therefore, the density function of the initial condition is $p^*(y_{i1}, d_{i1} \mid \alpha_i)$.

Proposition 6 establishes the identification of $AME_d(d)$.

Proposition 6. *Consider the binary choice model with duration dependence defined by equation (31) and Assumption 1-BC. For $T \geq 3 + d$, the causal effect $AME_d(d)$ is identified for*

any $d \geq 1$. For instance, with $T \geq 4$, the causal effect $AME_d(1)$ is identified as:

$$\begin{aligned}
AME_d(1) &= \frac{e^{\beta+2\delta} - e^{\beta+\delta}}{2} [\mathbb{P}_{0,0,1,0} + \mathbb{P}_{0,1,0,0}] + \frac{e^{\beta+2\delta} - e^{\beta+\delta}}{e^{\beta+\delta}} \mathbb{P}_{0,0,1,1} \\
&+ \left(\frac{e^{\beta+2\delta} (1 - e^{\beta+2\delta})}{e^{\beta+\delta}} + e^{\beta+2\delta} - 1 \right) \mathbb{P}_{0,1,1,0} \\
&+ \left(1 - \frac{e^{\beta+\delta}}{e^{\beta+2\delta}} \right) [\mathbb{P}_{1,0,1,0} + \mathbb{P}_{1,0,1,1}] + \left(\frac{e^{\beta+2\delta} - 1}{e^{\beta+\delta}} - 1 + \frac{1}{e^{\beta+2\delta}} \right) \mathbb{P}_{1,1,0,0}. \quad \blacksquare
\end{aligned} \tag{34}$$

Proof of Proposition 6. See Appendix A.9.

3 Multinomial and ordered choice models

EXAMPLE 2. Consider a dynamic panel data model of migration decisions. The set of possible choices, denoted by $\mathcal{Y} = \{0, 1, \dots, J\}$, represents all the cities within a country where an individual can choose to live and work. Individuals are indexed by i and locations by j . Each individual selects the location that maximizes their intertemporal utility. For simplicity, we do not distinguish between the components of utility derived from labor earnings and those from non-pecuniary factors. Some characteristics of both the individual and the location influence the utility of working in a given location j , and these are observable by the researcher, denoted by \mathbf{x}_{it} . Other factors affecting utility are unobservable to the researcher, with some being time-invariant, $\alpha_i(j)$, and others varying over time, $\varepsilon_{it}(j)$. The decision-making is dynamic because relocating implies migration costs, which may include search costs, transportation, and irreversible investments. In this model, the impact of migration costs on the utility of choosing location j is captured by the term $\beta_j \mathbb{1}\{y_{i,t-1} = j\}$, where β_j is a parameter, and $\mathbb{1}\{y_{i,t-1} = j\}$ is an indicator variable that equals one if j is the location where the individual worked previous period, and zero otherwise.

A researcher aims to estimate how an individual's current location influences their choice of location in the next period. Ideally, this would be done through a randomized experiment, where individuals are randomly assigned to one of two groups. In one group, individuals are assigned to work in location j , while in the other group, they are assigned to location k . After one period, we compare the proportion of individuals in each group who work in location j . The difference between these proportions provides an example of AME in this multinomial model. This AME reflects the causal effect of exogenously assigning individuals to a location. In the absence of migration costs, this causal effect would be zero. ■

3.1 Multinomial choice model

Consider a panel dataset $\{y_{it}, \mathbf{x}_{it} : i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$ where y_{it} can take $J + 1$ values: $y_{it} \in \mathcal{Y} = \{0, 1, \dots, J\}$. We can interpret the dependent variable as the choice alternative that maximizes a utility or payoff function. More specifically:

$$y_{it} = \arg \max_{j \in \mathcal{Y}} \{ \alpha_i(j) + \beta_j \mathbb{1}\{y_{i,t-1} = j\} + \mathbf{x}'_{it} \boldsymbol{\gamma}_j + \varepsilon_{it}(j) \}, \quad (35)$$

where $\{\beta_j : j \in \mathcal{Y}\}$ and $\{\boldsymbol{\gamma}_j : j \in \mathcal{Y}\}$ are slope parameters, and $\boldsymbol{\alpha}_i \equiv \{\alpha_i(j) : j \in \mathcal{Y}\}$ are incidental parameters. The unobservables $\{\varepsilon_{it}(j) : j \in \mathcal{Y}\}$ are i.i.d. type 1 extreme value. The explanatory variables in the $K \times 1$ vector \mathbf{x}_{it} are strictly exogenous concerning the transitory shocks $\varepsilon_{it}(j)$.

Parameter β_j represents the impact on the utility of staying in the same choice as in the previous period. It can be interpreted as the effect of habits or as a negative switching cost. The β_j and $\boldsymbol{\gamma}_j$ parameters can be identified only relative to the value of these parameters for a baseline choice alternative. We adopt the conventional normalization of $\beta_0 = 0$ and $\boldsymbol{\gamma}_0 = \mathbf{0}$, such that the identified parameters are actually $\beta_j - \beta_0$ and $\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_0$. The vector $\boldsymbol{\alpha}_i \equiv [\alpha_i(0), \alpha_i(1), \dots, \alpha_i(J)]$ represents unobserved individual heterogeneity in payoffs. Assumption 1-MNL summarizes the conditions in this model.

Assumption 1-MNL. (A) (Logit) $\varepsilon_{it}(j)$ is i.i.d. over (i, t, j) with type 1 extreme value distribution; (B) (Strict exogeneity of \mathbf{x}_{it}) the variables $\varepsilon_{it}(j)$ are independent of $(\boldsymbol{\alpha}_i, \mathbf{x}_i^{\{1, T\}})$; and (C) (Fixed effects) the density functions $f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_i)$, $f_{\boldsymbol{\alpha}|\mathbf{x}^{\{1, T\}}}(\boldsymbol{\alpha}_i|\mathbf{x}_i^{\{1, T\}})$, and $p^*(y_{i1}|\boldsymbol{\alpha}_i, \mathbf{x}_i^{\{1, T\}})$ are unrestricted. ■

In this multinomial model, we can define individual-specific transition probabilities in the same way as we did above in equation (2) for the binary choice model. For any $j, k \in \mathcal{Y}$, and any value $\mathbf{x} \in \mathcal{X}$, $\pi_{kj}(\boldsymbol{\alpha}_i, \mathbf{x}) \equiv \mathbb{P}(y_{i,t+1} = j \mid \boldsymbol{\alpha}_i, y_{it} = k, \mathbf{x}_{i,t+1} = \mathbf{x})$. For this multinomial logit, the transition probability has the following form:

$$\pi_{kj}(\boldsymbol{\alpha}_i, \mathbf{x}) = \frac{\exp\{\alpha_i(j) + \beta_j \mathbb{1}\{j = k\} + \mathbf{x}'\boldsymbol{\gamma}_j\}}{\sum_{\ell=0}^J \exp\{\alpha_i(\ell) + \beta_\ell \mathbb{1}\{\ell = k\} + \mathbf{x}'\boldsymbol{\gamma}_\ell\}} \quad (36)$$

We also define $\Pi_{kj}(\mathbf{x}, \mathbf{x}^{\{1, T\}})$ as the *Average Transition Probability* (ATP) from k to j in the same way as we did for the binary choice model in equation (23). For any $j, k \in \mathcal{Y}$, any value $\mathbf{x} \in \mathcal{X}$, and any history of covariates $\mathbf{x}^{\{1, T\}}$, we have that $\Pi_{kj}(\mathbf{x}, \mathbf{x}^{\{1, T\}}) \equiv \int \pi_{kj}(\boldsymbol{\alpha}_i, \mathbf{x}) f_{\boldsymbol{\alpha}|\mathbf{x}^{\{1, T\}}}(\boldsymbol{\alpha}_i \mid \mathbf{x}_i^{\{1, T\}} = \mathbf{x}^{\{1, T\}}) d\boldsymbol{\alpha}_i$.

For any values $j, k, \ell \in \mathcal{Y}$ with $k \neq \ell$, let $\Delta_{j,k \rightarrow \ell}(\boldsymbol{\alpha}_i, \mathbf{x})$ be the *individual-specific* causal effect on the probability of $y_{it} = j$ of a change in $y_{i,t-1}$ from k to ℓ given that $\mathbf{x}_{it} = \mathbf{x}$.

$$\begin{aligned} \Delta_{j,k \rightarrow \ell}(\boldsymbol{\alpha}_i, \mathbf{x}) &\equiv \mathbb{E}(\mathbb{1}\{y_{it} = j\} \mid \boldsymbol{\alpha}_i, \mathbf{x}, y_{i,t-1} = \ell) - \mathbb{E}(\mathbb{1}\{y_{it} = j\} \mid \boldsymbol{\alpha}_i, \mathbf{x}, y_{i,t-1} = k) \\ &= \pi_{\ell j}(\boldsymbol{\alpha}_i, \mathbf{x}) - \pi_{kj}(\boldsymbol{\alpha}_i, \mathbf{x}). \end{aligned} \quad (37)$$

We are interested in the identification of the following AMEs:

$$\begin{aligned} AME_{j,k \rightarrow \ell}(\mathbf{x}, \mathbf{x}^{\{1, T\}}) &\equiv \int \Delta_{j,k \rightarrow \ell}(\boldsymbol{\alpha}_i, \mathbf{x}) f_{\boldsymbol{\alpha}|\mathbf{x}^{\{1, T\}}}(\boldsymbol{\alpha}_i|\mathbf{x}_i^{\{1, T\}} = \mathbf{x}^{\{1, T\}}) d\boldsymbol{\alpha}_i \\ &= \Pi_{\ell j}(\mathbf{x}, \mathbf{x}^{\{1, T\}}) - \Pi_{kj}(\mathbf{x}, \mathbf{x}^{\{1, T\}}) \end{aligned} \quad (38)$$

3.2 Identification in the multinomial model

To obtain our identification result for the multinomial model in Proposition 7, we apply the following Lemma 3, which is an extension of Lemma 1 to the multinomial case.

Lemma 3. *In the model defined by equation (35) and Assumption 1-MNL, for any triple of choice alternatives j, k, ℓ , with $k \neq \ell$, the following condition holds:*

$$e^{\beta_j [1 - \mathbb{1}\{k=j\} - \mathbb{1}\{\ell=j\}]} = \frac{\pi_{k\ell}(\boldsymbol{\alpha}_i, \mathbf{x}) \pi_{jj}(\boldsymbol{\alpha}_i, \mathbf{x})}{\pi_{kj}(\boldsymbol{\alpha}_i, \mathbf{x}) \pi_{j\ell}(\boldsymbol{\alpha}_i, \mathbf{x})}. \quad \blacksquare \quad (39)$$

Proof of Lemma 3. See Appendix A.8.

Proposition 7 establishes the identification of the average transition probabilities $\Pi_{jj}(\mathbf{x}, \mathbf{x}^{\{1,T\}})$ for covariate histories with $\mathbf{x}_2 = \dots = \mathbf{x}_T$.

Proposition 7. *In the model defined by equation (35) and Assumption 1-MNL, for any choice alternative j , any $\mathbf{x} \in \mathcal{X}$, and covariate histories with $\mathbf{x}_2 = \dots = \mathbf{x}_T = \mathbf{x}$, if $T \geq 3$, the average transition probability $\Pi_{jj}(\mathbf{x}, \mathbf{x}^{\{1,T\}})$ is identified. For instance, for $T = 3$:*

$$\Pi_{jj}(\mathbf{x}, \mathbf{x}^{\{1,3\}}) = \mathbb{P}_{jj|\mathbf{x}^{\{1,2\}}} + \sum_{k \neq j} \left[\mathbb{P}_{kjj|\mathbf{x}^{\{1,3\}}} + \sum_{\ell \neq j} e^{\beta_j [1 - \mathbb{1}\{k=j\} - \mathbb{1}\{\ell=j\}]} \mathbb{P}_{kj\ell|\mathbf{x}^{\{1,3\}}} \right], \quad (40)$$

where $\mathbf{x}^{\{1,3\}}$ is such that $\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}$. \blacksquare

Proof of Proposition 7. See Appendix A.10.

Unfortunately, the procedure described in the proof of Proposition 7 does not provide an identification result for the parameters Π_{jk} with $j \neq k$ when the number of choice alternatives is greater than two. Based on the necessary and sufficient conditions in Proposition 2, the following Proposition 8 establishes that this identification is not possible when $J + 1 \geq 3$.

Proposition 8. *In the multinomial model defined by equation (35) and Assumption 1-MNL, with $J + 1 \geq 3$ and $\boldsymbol{\theta}$ known to the researcher, there is no function $h(\mathbf{P}_{y|\mathbf{x}}, \boldsymbol{\theta})$ that equal the average transition probability Π_{kj} for $k \neq j$. Consequently, no average causal effect $AME_{j,k \rightarrow \ell}$ is identified. ■*

Proof of Proposition 8. See Appendix A.11.

3.3 Ordered logit model

Consider the following ordered logit model with $J + 1$ choice alternatives.⁸

$$\begin{aligned} y_{it} &= j \text{ if } y_{it}^* \in (\lambda_{j-1}, \lambda_j], \text{ with } j = 0, 1, \dots, J \\ y_{it}^* &= \alpha_i + \sum_{k=0}^{J-1} \beta_k \mathbb{1}\{y_{i,t-1} = k\} - \varepsilon_{it}. \end{aligned} \quad (41)$$

where λ 's are parameters with $\lambda_{j-1} < \lambda_j$, $\lambda_{-1} = -\infty$ and $\lambda_{J+1} = +\infty$, and ε_{it} is i.i.d. with Logistic distribution. This model implies the following transition probability function:

$$\pi_{kj}(\alpha_i) = \Lambda(\alpha_i + \beta_k - \lambda_{j-1}) - \Lambda(\alpha_i + \beta_k - \lambda_j) \quad (42)$$

Honoré, Muris, and Weidner (2021) establish the identification of the β and λ parameters. The following proposition establishes the identification of the Average Transition Probabilities Π_{kj} and the causal effects $AME_{j,k \rightarrow \ell} = \Pi_{\ell j} - \Pi_{kj}$ for any values of k , ℓ , and j .

Proposition 9. *Consider the ordered logit model as defined in equation (41) with $J + 1 = 3$ and given the values of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ parameters. If $T \geq 3$, the average transition probabilities*

⁸For simplicity, we do not include covariates \mathbf{x} . It is straightforward to extend the identification result in Proposition 9 below to an AME conditional on $\mathbf{x}_2 = \mathbf{x}_3$, similarly as our results in Propositions and 1 and 7.

$\{\Pi_{kj} : k, j \in \{0, 1, 2\}\}$ are identified. For instance, for $j = 0$:

$$\begin{aligned}
\Pi_{00} &= \sum_{k=0}^J \mathbb{P}_{k,0,0} + \sum_{k=0}^J \sum_{\ell=1}^J e^{\beta_k - \beta_0} \mathbb{P}_{k,0,\ell} \\
\Pi_{10} &= \sum_{k=0}^J \sum_{\ell=0}^1 \mathbb{P}_{k,0,\ell} + \sum_{k=0}^1 e^{\beta_k - \beta_1} \mathbb{P}_{k,0,2} + \mathbb{P}_{2,0,2} + (e^{\beta_2 - \beta_1} - 1) \mathbb{P}_{2,2,0} \\
&\quad + \sum_{k=0}^J \frac{(e^{\beta_1 - \lambda_0} - e^{\beta_k - \lambda_1}) (e^{\beta_k} - e^{\beta_1})}{(e^{\beta_k - \lambda_0} - e^{\beta_k - \lambda_1}) e^{\beta_1}} \mathbb{P}_{k,1,0} \\
\Pi_{20} &= \mathbb{P}_{0,0,0} + \mathbb{P}_{0,0,1} + e^{\lambda_1 - \lambda_0} \mathbb{P}_{0,0,2} + \left(1 - \frac{e^{\beta_2 - \lambda_0}}{e^{\beta_1 - \lambda_1}}\right) \mathbb{P}_{0,2,0} \\
&\quad + \sum_{k=0}^l \sum_{\ell=0}^J \mathbb{P}_{1,k,\ell} + \left(1 - \frac{e^{\beta_2 - \lambda_0}}{e^{\beta_1 - \lambda_1}}\right) \mathbb{P}_{1,2,0} + \sum_{k=0}^{J-1} \mathbb{P}_{2,0,k}
\end{aligned} \tag{43}$$

This implies the identification of $AME_{j,k \rightarrow \ell} = \Pi_{\ell j} - \Pi_{kj}$. ■

Proof of Proposition 9. See Appendix A.12.

4 Monte Carlo experiments

These Monte Carlo experiments serve two main purposes. First, we compare the bias and variance of the FE estimator of AME to those from RE estimators imposing restrictions that we typically find in applications of RE models. Second, we evaluate the power of two testing procedures designed to reject a misspecified RE model: the standard Hausman test, which relies on RE and FE estimators of slope parameters, and an alternative Hausman test based on the RE and FE estimators of AMEs.

The DGP is the binary choice AR1 model described in equation (1) without exogenous covariates. The model for the initial condition is $y_{i1} = \mathbb{1}\{\alpha_i + u_i \geq 0\}$ where u_i is i.i.d. Logistic and independent of α_i and ε_{it} . The number of periods is $T = 4$. We implement experiments for two sample sizes N , 1,000 and 2,000. We consider six DGPs based on

Table 1: **DGPs and true value of AME**

Value of β	Distribution of α_i		
	No α_i	Finite mixture	Mixture of normals
$\beta = -1$	DGP <i>NoUH</i> ($\beta = -1$) $AME = -0.2311$	DGP <i>FinMix</i> ($\beta = -1$) $AME = -0.2164$	DGP <i>MixNor</i> ($\beta = -1$) $AME = -0.113$
$\beta = 1$	DGP <i>NoUH</i> ($\beta = 1$) $AME = 0.2311$	DGP <i>FinMix</i> ($\beta = 1$) $AME = 0.2059$	DGP <i>MixNor</i> ($\beta = 1$) $AME = 0.1108$

two values of parameter β ($\beta = -1$ and $\beta = 1$) and three distributions of the unobserved heterogeneity α_i : no heterogeneity, with $\alpha_i = 0$ for every i ; finite mixture with two points of support, $\alpha_i = -1$ with probability 0.3, and $\alpha_i = 0.5$ with probability 0.7; and a mixture of two normal random variables: $\alpha_i \sim N(-1, 3)$ with probability 0.3, and $\alpha_i \sim N(0.5, 3)$ with probability 0.7.

Table 1 summarizes the six DGPs, the labels we use to represent them, and the corresponding value of AME in the population. When we change the distribution of the unobserved heterogeneity, keeping the value of parameter β constant, the value of AME can vary substantially. For instance, when $\beta = 1$, AME equals 0.23 in the DGP without unobserved heterogeneity, 0.20 for the finite mixture, and 0.11 for a mixture of normal distributions.

For each DGP, we simulate 1,000 random samples with N individuals (with $N = 1,000$ or $N = 2,000$) and $T = 4$. For each sample, we calculate three estimators of β and AME : (1) a FE estimator, that we denote *FE-CMLE*⁹; (2) a maximum likelihood estimator that assumes that the distribution of α_i is discrete with two mass points, that we denote *RE-MLE*; and (3) a maximum likelihood estimator that assumes there is no unobserved heterogeneity, that

⁹The FE estimator of β is the CMLE proposed by Chamberlain (1985). For the estimation of parameter AME , we use a plug-in estimator based on the formula for the identified AME when $T = 4$, which we present in Table 7 in the Appendix. In this formula, we replace parameter β with its CML estimate and the probabilities of choice histories with frequency estimates.

we denote *NoUH-MLE*.¹⁰ Table 2 presents results from experiments with $N = 1000$.¹¹

(i) *Bias of FE estimators relative to MLE*. The mean bias of the FE estimator is very small: between 0.1% and 0.7% of the true value for β , and between 0.2% and 1.4% for *AME*. In this FE approach, the estimation of *AME* does not involve a substantially larger bias than that of β . This bias is of a similar magnitude as the ones of *NoUH-MLE* and *RE-MLE* estimators when these estimators are consistent (i.e., when the DGPs are *NoUH* and *FinMix*, respectively).

(ii) *Variance of FE estimators relative to RE-MLE*. As a percentage of the true value, the standard deviation of the FE estimator is between 10% and 20% for the estimator of β , and between 7% and 30% for the estimator of *AME*. These ratios are substantially smaller for the RE-MLE estimator: between 9% and 13% for the estimator of β , and between 8% and 23% for the estimator of *AME*. As expected, the FE estimators have a larger variance than the RE-MLE estimators. The loss of precision associated with FE estimation is of a similar magnitude when estimating *AME* than when estimating β .

The variance of the FE estimator is substantially larger when β is positive than when it is negative, but this is not the case for the RE-MLE estimators. This result has a clear explanation. The histories that contribute to the identification of the parameters β and *AME* involve some alternation of the two choices over time, e.g., $\{0, 1, 0, 1\}$ or $\{0, 0, 1, 1\}$. These histories occur more frequently when β is negative than when it is positive. Identifying negative state dependence is simpler than identifying positive state dependence because the

¹⁰For the DGPs without unobserved heterogeneity (i.e., *NoUH(-1)* and *NoUH(+1)*), we do not report results for the RE MLE. The reason is that, for these DGPs, the finite mixture (two types) RE model is not identified, and the estimates of β are extremely poor. As expected, the estimate of the mixing probability in the mixture is close to zero, but the points in the support set of α_i are poorly identified, and they take extreme values. This problem also affects the estimation of β , which presents a huge bias and variance. For this reason, we have preferred not to present results for this combination of estimator and DGP. However, it is important to note that avoiding these numerical/identification problems in estimating the distribution of α is a key advantage of FE estimation.

¹¹The results for sample size $N = 2000$ are qualitatively very similar except that, as one would expect, all the estimators have a lower variance when the sample size increases. Therefore, we present only results from experiments with $N = 1000$ here.

Table 2: Monte Carlo Experiments with sample size N=1,000

		<i>Statistics</i>						
		<i>True</i> β	<i>Mean</i> $\hat{\beta}$	<i>Std</i> $\hat{\beta}$	<i>True</i> <i>AME</i>	<i>Mean</i> \widehat{AME}	<i>Std</i> \widehat{AME}	<i>RMSE</i> \widehat{AME}
DGP NoUH(-1)	<i>FE-CMLE</i>	-1.0	-1.0074	0.1310	-0.2311	-0.2314	0.0235	0.0235
	<i>RE-MLE</i>	-1.0	NA	NA	-0.2311	NA	NA	NA
	<i>NoUH-MLE</i>	-1.0	-0.9998	0.0798	-0.2311	-0.2309	0.0175	0.0175
DGP FinMix(-1)	<i>FE-CMLE</i>	-1.0	-1.0012	0.1338	-0.2164	-0.2160	0.0222	0.0221
	<i>RE-MLE</i>	-1.0	-1.0036	0.1214	-0.2164	-0.2165	0.0215	0.0214
	<i>NoUH-MLE</i>	-1.0	-0.5979	0.0781	-0.2164	-0.1430	0.0183	0.0757
DGP MixNor(-1)	<i>FE-CMLE</i>	-1.0	-1.0136	0.2160	-0.1113	-0.1110	0.0178	0.0176
	<i>RE-MLE</i>	-1.0	-0.3604	0.1825	-0.1113	-0.0470	0.0218	0.0679
	<i>NoUH-MLE</i>	-1.0	1.7190	0.1028	-0.1113	0.4022	0.0214	0.5139
DGP NoUH(+1)	<i>FE-CMLE</i>	1.0	1.0013	0.1654	0.2311	0.2344	0.0526	0.0527
	<i>RE-MLE</i>	1.0	NA	NA	0.2311	NA	NA	NA
	<i>NoUH-MLE</i>	1.0	0.9980	0.0778	0.2311	0.2305	0.0176	0.0176
DGP FinMix(+1)	<i>FE-CMLE</i>	1.0	0.9982	0.1841	0.2059	0.2089	0.0539	0.0539
	<i>RE-MLE</i>	1.0	0.9864	0.1296	0.2059	0.2034	0.0315	0.0316
	<i>NoUH-MLE</i>	1.0	1.4100	0.0843	0.2059	0.3212	0.0183	0.1168
DGP MixNor(+1)	<i>FE-CMLE</i>	1.0	1.0055	0.2873	0.1108	0.1169	0.0511	0.0515
	<i>RE-MLE</i>	1.0	1.4863	0.1828	0.1108	0.2120	0.0367	0.1078
	<i>NoUH-MLE</i>	1.0	3.2453	0.1194	0.1108	0.6645	0.0166	0.5541

latter has implications similar to unobserved heterogeneity, whereas negative state dependence does not.

(iii) *Bias of RE-MLE estimators due to misspecification.* The biases due to the misspecification of the RE model are substantial. The bias in the estimation of β from ignoring unobserved heterogeneity, when present, is between 41% of the true value (with the finite mixture DGP) and 270% (with the mixture of normals DGP). The bias is even more significant in estimating *AME*: 60% of the true value in the finite mixture DGP and more than 500% in the mixture of normals DGP. The bias is also substantial for the RE-MLE that accounts for heterogeneity but misspecifies its distribution: between 50% and 65% in the estimation of β ; and between 58% and 93% for *AME*. As a result, the FE estimator dominates the RE-MLE in terms of Root Mean Square Error (RMSE) in the cases where the RE model is misspecified.

(iv) *Testing for misspecification of RE models.* A common approach to test the validity of a RE model consists of using a Hausman test that compares the FE estimator of β (consistent under the null and the alternative) and the RE-MLE of β (efficient under the null but inconsistent under the alternative). See Hausman (1978) and Hausman and Taylor (1981). Given our identification results, we can define a similar Hausman test using the FE and RE estimators of AME instead of β . Therefore, we have two different Hausman statistics to test for the validity of a RE model. The statistic based on the estimators of β is:

$$HS_{\beta} = \frac{\left(\widehat{\beta}_{FE} - \widehat{\beta}_{RE}\right)^2}{\widehat{Var}\left(\widehat{\beta}_{FE}\right) - \widehat{Var}\left(\widehat{\beta}_{RE}\right)} \quad \text{under } H_0 \sim \chi_1^2 \quad (44)$$

And the statistic based on the estimators of AME is:

$$HS_{AME} = \frac{\left(\widehat{AME}_{FE} - \widehat{AME}_{RE}\right)^2}{\widehat{Var}\left(\widehat{AME}_{FE}\right) - \widehat{Var}\left(\widehat{AME}_{RE}\right)} \quad \text{under } H_0 \sim \chi_1^2 \quad (45)$$

The Hausman test based on AME has several advantages over the test based on β . First, the researcher can be particularly interested in the causal effect implied by the model and not in the slope parameter itself. Second, and more substantially, the parameter β test may suffer a scaling problem that does not affect the test on the AME . That is, the parameter β depends on the variance of the transitory shock ε_{it} , which depends on the RE model's specification. For instance, a reason why the estimators $\widehat{\beta}_{FE}$ and $\widehat{\beta}_{NoUH-MLE}$ are different is that, in the model ignoring unobserved heterogeneity, the actual error term is $\alpha_i + \varepsilon_{it}$, and the variance of this variable is larger than the variance of ε_{it} . The estimation of AME – using either FE or RE approaches – is not affected by this scaling problem.

Figures 1 to 6: Empirical distribution of p-values of Hausman tests

Figure 1.

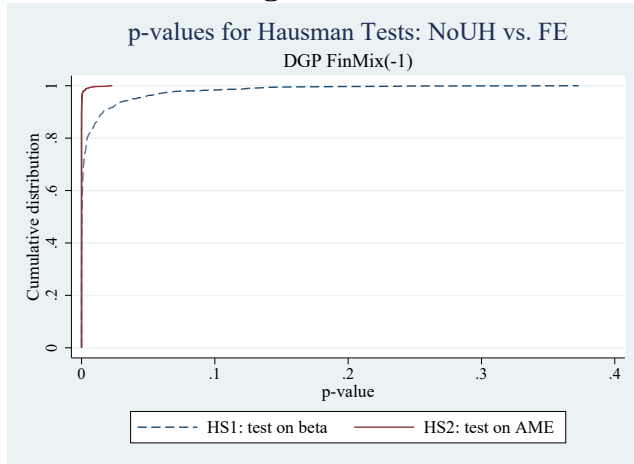


Figure 2

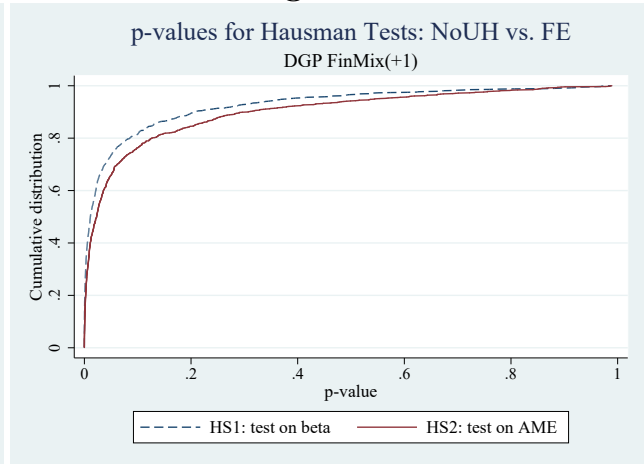


Figure 3

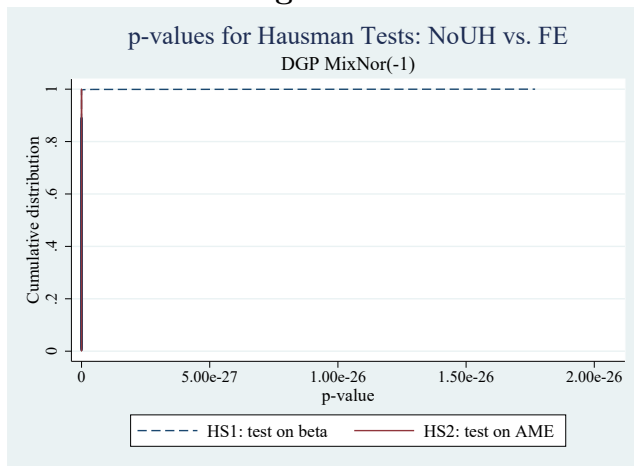


Figure 4

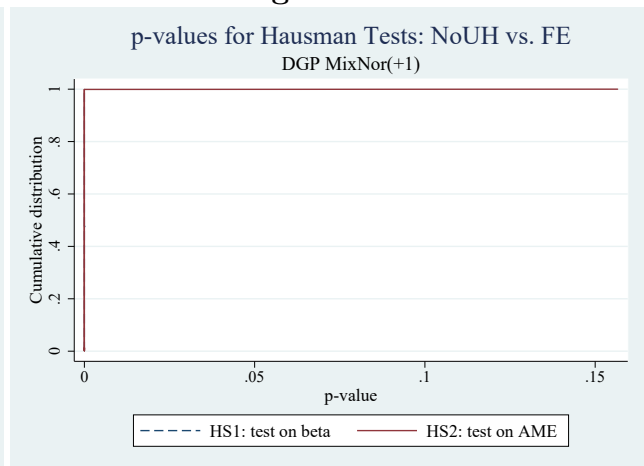


Figure 5

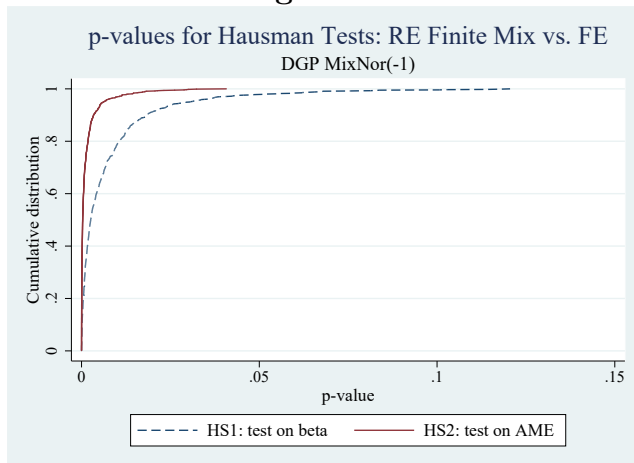
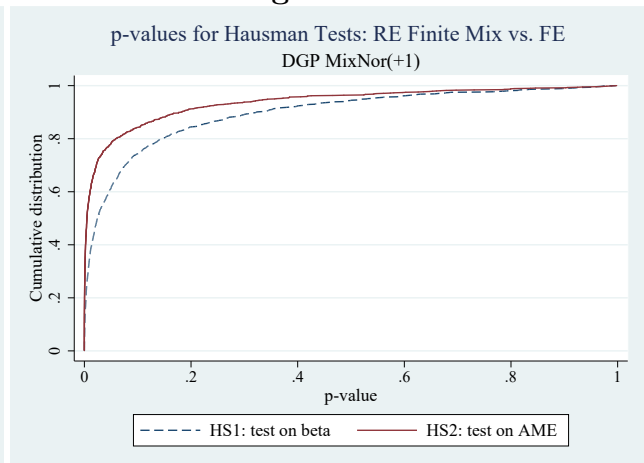


Figure 6



We compare the power of these two tests using our Monte Carlo experiments. Figures 1 to 6 summarize our results. Each figure corresponds to one DGP and presents the cumulative distribution function of the p-value for the null hypothesis that the RE model is valid. More specifically: *Figure 1*: DGP is *FinMix(-1)*, and the null hypothesis is no unobserved heterogeneity; *Figure 2*: DGP is *FinMix(+1)* and the null hypothesis is no unobserved heterogeneity; *Figure 3*: DGP is *MixNor(-1)* and the null hypothesis is no unobserved heterogeneity; *Figure 4*: DGP is *MixNor(+1)* and the null hypothesis is no unobserved heterogeneity; *Figure 5*: DGP is *MixNor(-1)* and the null hypothesis is the *finite mixture* model; *Figure 6*: DGP is *MixNor(+1)* and the null hypothesis is the *finite mixture* model.

Figures 3 and 4 show that both tests have strong statistical power to reject the null of no unobserved heterogeneity when the DGP is a mixture of normals. In Figures 1 and 5, the two tests also have strong power when the true value of β is negative. The relevant comparison appears in Figures 2 and 6. In the DGP with a mixture of normals (Figure 6), the HS_{AME} test has substantially larger power than the test HS_{β} . In particular, HS_{β} has a low-power problem. For this test, with a 5% significance level, we do not reject the null for more than half of the samples. In contrast, the HS_{AME} test has reasonable power. With a 5% significance level for this test, we can reject the null for 80% of the samples. In Figure 2, the HS_{β} test has more power than the HS_{AME} test. However, the differences in power are much smaller than in Figure 6, and neither of the two tests has a low power problem. Overall, the HS_{AME} test has more power than the HS_{β} . This test seems a valuable byproduct of identifying AMEs in FE models.

5 State Dependence in Consumer Brand Choice

We apply our identification results to measure state dependence on consumer brand choices. There is substantial literature on testing and measuring state dependence on consumer brand choices, with seminal papers by [Erdem \(1996\)](#), [Keane \(1997\)](#), and [Roy, Chintagunta, and](#)

Haldar (1996).¹² These applications use consumer scanner panel data and estimate dynamic discrete choice models with persistent unobserved heterogeneity in consumer brand preferences and state dependence generated by purchasing/consumption habits or brand switching costs. The main goal is to determine the relative contribution of unobserved heterogeneity and state dependence to explain the observed time persistence of consumer brand choices. Disentangling the contribution of these two factors has important implications for demand elasticities, competition, consumer welfare, and the evaluation of mergers.¹³

All these previous studies estimate Random Effects (RE) models. In this application, we consider a FE model, estimate average transition probabilities Π_{jj} and average treatment effects $ATE_{jj} \equiv \Pi_{jj} - \mathbb{E}(\mathbb{1}\{y_{it} = j\})$, and use them to measure the contribution of state dependence to brand-choice persistence.

5.1 Data

The dataset comes from A.C. Nielsen scanner panel data for the ketchup product category in the geographic market of Sioux Falls, South Dakota.¹⁴ It contains 996 households and covers a 123-week period from mid-1986 to mid-1988.¹⁵ For our analysis, a period is a *household purchase occasion*. Periods $t = 1, 2, \dots$ represent a household's first, second, and so on ketchup purchases during the sample period. This timing is common in this literature (e.g., Erdem, 1996; Keane, 1997). T_i is the number of purchase occasions for household i . The total number of observations or purchase occasions in this sample is $\sum_{i=1}^N T_i = 9,562$. Table 3 presents the distribution of T_i .

¹²Other contributions to this literature are Seetharaman, Ainslie, and Chintagunta (1999), Erdem, Imai, and Keane (2003), Seetharaman (2004), Dubé, Hitsch, and Rossi (2010), and Osborne (2011), among others. There is also a related literature on the implications of brand-choice state dependence on market competition (see Viard, 2007, and Pakes, Porter, Shepard, and Calder-Wang 2021).

¹³See Erdem, Imai, and Keane (2003) for a detailed discussion of the important economic implications of distinguishing between unobserved heterogeneity and state dependence on consumer demand.

¹⁴Our sample comes from Erdem, Imai, and Keane (2003). We thank the authors for sharing the data with us.

¹⁵The raw data contains 2797 households. Here we use the same working sample of 996 households as in Erdem, Imai, and Keane (2003). This sample focuses on households who are regular ketchup users. See page 30 of that paper for a description of the selection of this working sample.

Table 3: **Distribution of the number of purchase occasions (T_i)**

<i>Minimum</i>	<i>5%</i>	<i>25%</i>	<i>Median</i>	<i>75%</i>	<i>95%</i>	<i>Maximum</i>
3	4	5	8	12	21	52

Table 4: **Matrix of Transition Probabilities of Brand Choices (%)**

<i>Brand choice at t</i>	<i>Brand choice at t + 1</i>				<i>Total</i>
	<i>Heinz (j = 0)</i>	<i>Hunt's (j = 1)</i>	<i>Del Monte (j = 2)</i>	<i>Store (j = 3)</i>	
<i>Heinz (j = 0)</i>	78.95	10.67	6.98	3.40	100.00
<i>Hunt's (j = 1)</i>	45.16	32.30	15.76	6.78	100.00
<i>Del Monte (j = 2)</i>	41.11	18.98	34.07	5.83	100.00
<i>Store (j = 3)</i>	42.32	17.11	13.38	27.19	100.00
<i>Market share (\mathbb{P}_j)</i>	66.65	15.63	12.19	5.53	100.00
<i>Choice persistence ($\mathbb{P}_{jj} - \mathbb{P}_j$)</i>	12.30	16.67	21.88	21.66	

There are four brands in this market: three national brands, Heinz, Hunt's, and Del Monte; and a store brand. We ignore the quantity purchased and focus on brand choice. Table 4 presents brands' market shares constructed using the number of purchasing events, and the matrix of transition probabilities between the four brands. Heinz is the leading brand, with 66% share of purchases, followed by Hunt's at 16%, Del Monte at 12%, and Store brands at 5%.

A measure of *choice persistence for brand j* is the difference between the transition probability $Pr(y_{i,t+1} = j | y_{it} = j)$ and the unconditional probability or market share $Pr(y_{it} = j)$. This measure shows choice persistence for all the brands, with the largest for *Del Monte* and *Store brands* with 21.88% and 21.66%, respectively, followed by *Hunt's* with 16.67%, and *Heinz* with 12.30%. The persistence observed may stem from consumer taste heterogeneity and state dependence. Our main objective in this application is to distinguish the impact of these two factors and examine how they differ across brands.

5.2 Model

Let $y_{it} \in \{0, 1, 2, 3\}$ be the brand choice of household i at purchase occasion t . We consider the following brand choice model with *habit formation*:

$$y_{it} = \arg \max_{j \in \{0,1,2,3\}} \{ \alpha_i(j) + \beta_{jj} \mathbb{1}\{y_{i,t-1} = j\} + \varepsilon_{it}(j) \}. \quad (46)$$

Parameter β_{jj} represents habits in purchasing brand j : the additional utility from buying the same brand as in the previous purchase. Parameter β_{00} (for Heinz) is normalized to zero. Variable $\alpha_i(j)$ represents the household's time-invariant taste for brand j . For simplicity, we ignore duration dependence. We also omit prices.¹⁶

Following [Aguirregabiria, Gu, and Luo \(2021\)](#), we can interpret equation (46) as a model where households are forward-looking. The fixed effects $\alpha_i(j)$ can be interpreted as the sum of two components: a fixed effect in the current utility of choosing brand j and the continuation value (expected and discounted future utility) of choosing brand j today. In this model, these continuation values depend on the current choice j but not on the state variable $y_{i,t-1}$ or the current ε_{it} .

5.3 Estimation

To illustrate our method using a short panel, we split the purchasing histories in the original sample into sub-histories of length T , for $T = 6$ and $T = 8$. Table 5 presents our Fixed Effect estimates of the *brand habit* parameters β_{jj} . We use the Conditional Maximum Likelihood estimator. We obtain standard errors using a bootstrap method that resamples the 996 purchasing histories in the original dataset.¹⁷ Parameter estimates with $T = 6$ and $T = 8$

¹⁶In this dataset, supermarkets follow High-Low pricing, and prices can stay at the high (regular) level for relatively long periods. Omitting prices in our model can be interpreted using choice histories where prices remain constant.

¹⁷Using the original sample of 996 purchasing histories, we resample independently and with replacement 996 histories. Then, we generate all the possible sub-histories of length T from these histories. We also obtained asymptotic standard errors. Bootstrap standard errors are a bit larger (at the second or third significant digit) than the asymptotic ones.

Table 5: **Conditional Maximum Likelihood Estimates**

of Brand Habit (β_{jj}) Parameters				
<i>Parameter</i>	<i>T = 6 sub-histories</i>		<i>T = 8 sub-histories</i>	
β_{jj}	<i>Estimate</i>	<i>(s.e.)</i>	<i>Estimate</i>	<i>(s.e.)</i>
<i>Heinz</i>	0.00	(.)	0.00	(.)
<i>Hunt's</i>	0.2312	(0.0590)	0.2566	(0.0570)
<i>Del Monte</i>	0.1155	(0.0718)	0.1191	(0.0722)
<i>Store</i>	0.3245	(0.1166)	0.4675	(0.1106)
<i># histories of length T</i>	4,764		3,396	

(1) Standard errors (*s.e.*) are obtained using a bootstrap method. We generate 1,000 resamples (independent, with replacement, and with $N = 996$) from the 996 purchasing histories in the original dataset. Then, we split each history of the bootstrap sample into all the possible sub-histories of length T .

are very similar. They are significantly greater than zero at a 5% significance level, showing evidence of state dependence on brand choice. The magnitude of the parameter estimate is not monotonically related to the brand's market share or the degree of brand choice persistence shown in Table 4. However, we need to consider that a larger value of β_{jj} does not imply a larger degree of state dependence as measured by the Average Transition Probabilities or by ATE_{jj} .

Table 6 presents Fixed Effect estimates of ATPs and a decomposition of brand choice persistence into the contributions of state dependence and unobserved heterogeneity. The estimation of the ATPs Π_{jj} is based on equation (40) in Proposition 7. We plug the CML estimates of β_{jj} parameters and frequency estimates of probabilities of choice histories in this equation and obtain standard errors using a bootstrap method.

In Table 6, the column labeled *Pers* provides brand choice persistence as measured by the difference between the transition probability $\mathbb{P}_{j|j}$ and the unconditional probability \mathbb{P}_j . The estimates of ATPs (in the columns labeled *ATP*) are very precise and similar for $T = 6$ and $T = 8$. The column labelled *AME* presents $ATE_{jj} = \Pi_{jj} - \mathbb{E}(\mathbb{1}\{y_{it} = j\})$. This *ATE* is a measure of the contribution of state dependence to brand choice persistence. This contribution is quite small for all the brands: between 1 and 2 percentage points.

Table 6: **Fixed Effects Estimates of ATPs and AMEs**

	<i>T</i> = 6 sub-histories				<i>T</i> = 8 sub-histories			
	<i>Pers</i> (<i>s.e.</i>)	<i>ATP</i> (<i>s.e.</i>)	<i>ATE</i> (<i>s.e.</i>)	<i>UHet</i> (<i>s.e.</i>)	<i>Pers</i> (<i>s.e.</i>)	<i>ATP</i> (<i>s.e.</i>)	<i>ATE</i> (<i>s.e.</i>)	<i>UHet</i> (<i>s.e.</i>)
<i>Heinz</i>	0.1230 (0.0033)	0.6744 (0.0057)	0.0079 (0.0066)	0.1151 (0.0068)	0.1230 (0.0033)	0.6708 (0.0062)	0.0043 (0.0067)	0.1187 (0.0069)
<i>Hunt's</i>	0.1667 (0.0077)	0.1752 (0.0075)	0.0189 (0.0107)	0.1478 (0.0109)	0.1667 (0.0077)	0.1788 (0.0072)	0.0225 (0.0106)	0.1442 (0.0109)
<i>Del Monte</i>	0.2188 (0.0090)	0.1324 (0.0067)	0.0105 (0.0112)	0.2183 (0.0115)	0.2188 (0.0090)	0.1345 (0.0062)	0.0126 (0.0110)	0.2062 (0.0113)
<i>Store</i>	0.2166 (0.0062)	0.0736 (0.0071)	0.0183 (0.0094)	0.1983 (0.0099)	0.2166 (0.0062)	0.0805 (0.0072)	0.0252 (0.0094)	0.1914 (0.0099)

- (1) *Pers* is brand choice persistence, $\mathbb{P}_{j|j} - \mathbb{P}_j$, as measured at the bottom line of Table 4.
- (2) *ATP* is the brand's Average Transition Probability, Π_{jj} .
- (3) $ATE = ATE_{jj} = \Pi_{jj} - \mathbb{E}(\mathbb{1}\{y_{it} = j\})$.
- (4) *UHet* is defined as $\mathbb{P}_{j|j} - \Pi_{jj}$. By construction, $Pers = AME + UHet$.
- (5) Standard errors (*s.e.*) are obtained using the same bootstrap method as for the estimates in Table 6.

In fact, for *Heinz* and *Del Monte*, we cannot reject the null hypothesis that this *AME* is zero at a 5% significance level. The *Store brand* is the one with the largest contribution of state dependence. The column labeled *UHet* presents the contribution of consumer taste heterogeneity to brand choice persistence, as measured by the difference between brand choice persistence and ATE_{jj} . This heterogeneity accounts for most of the persistence of brand choice. This finding contrasts with results found in studies using similar models and data but with a Random Effects specification of consumer unobserved taste heterogeneity (e.g., Keane, 1997).

6 Conclusion

Average marginal effects are informative parameters that represent causal effects. They depend on the model's structural parameters and the unobserved heterogeneity distribution. In fixed effects nonlinear panel data models with short panels, the distribution of the unobserved heterogeneity is not identified, and this problem has been associated with the common belief that AMEs are not identified.

In the context of dynamic logit models, we prove the identification of AMEs associated with changes in lagged dependent variables and duration variables. Our proofs of the identification results are constructive and provide simple closed-form expressions for the AMEs in terms of frequencies of choice histories that can be obtained from the data. We illustrate our identification results using simulated and real-world consumer scanner data in a dynamic demand model with state dependence.

In this paper, we have derived identification results only for logit models, but the procedure that arises from necessary and sufficient conditions may work beyond the logistic. In particular, it may work for any function that shares with the logistic the property of having terms in which the fixed effect appears multiplicatively separated from other parameters of the model so that it is feasible to form polynomials of functions of the fixed effect. We leave this for future research.

References

- Abrevaya, J., and Y.-C. Hsu (2021): "Partial effects in non-linear panel data models with correlated random effects," *The econometrics journal*, 24(3), 519–535.
- Aguirregabiria, V., J. Gu, and Y. Luo (2021): "Sufficient statistics for unobserved heterogeneity in dynamic structural logit models," *Journal of Econometrics*, 223(2), 280–311.
- Andersen, E. (1970): "Asymptotic properties of conditional maximum likelihood estimators," *Journal of the Royal Statistical Society, Series B*, 32, 283–301.
- Arellano, M., and S. Bonhomme (2017): "Nonlinear panel data methods for dynamic heterogeneous agent models," *Annual Review of Economics*, 9, 471–496.
- Bonhomme, S. (2011): "Panel Data, Inverse Problems, and the Estimation of Policy Parameters," *Unpublished Manuscript*.
- Botosaru, I., and C. Muris (2024): "Identification of time-varying counterfactual parameters in nonlinear panel models," *Journal of Econometrics Journal*, In Press.
- Chamberlain, G. (1980): "Analysis of covariance with qualitative data," *Review of Economic Studies*, 47(1), 225–238.
- (1984): "Panel data," in *Handbook of Econometrics, Vol. 2*. North-Holland Press.
- (1985): "Heterogeneity, omitted variable bias, and duration dependence," in *Longitudinal Analysis of Labor Market Data*. Cambridge University Press.
- Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81(2), 535–580.

- Davezies, L., X. D’Haultfoeuille, and L. Laage (2022): “Identification and Estimation of Average Marginal Effects in Fixed Effects Logit Models,” *arXiv preprint arXiv:2105.00879*.
- Dobronyi, C., J. Gu, and K. Kim (2021): “Identification of Dynamic Panel Logit Models with Fixed Effects,” *arXiv preprint arXiv:2104.04590*.
- Dubé, J., G. Hitsch, and P. Rossi (2010): “State dependence and alternative explanations for consumer inertia,” *The RAND Journal of Economics*, 41(3), 417–445.
- Erdem, T. (1996): “A dynamic analysis of market structure based on panel data,” *Marketing Science*, 15(4), 359–378.
- Erdem, T., S. Imai, and M. Keane (2003): “Brand and quantity choice dynamics under price uncertainty,” *Quantitative Marketing and Economics*, 1(1), 5–64.
- Hahn, J. (2001): “Comment: Binary regressors in nonlinear panel-data models with fixed effects,” *Journal of Business & Economic Statistics*, 19(1), 16–17.
- Hausman, J. (1978): “Specification tests in econometrics,” *Econometrica*, pp. 1251–1271.
- Hausman, J., and W. Taylor (1981): “Panel data and unobservable individual effects,” *Econometrica*, pp. 1377–1398.
- Heckman, J. (1981): “The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process,” in *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press.
- Hoderlein, S., and H. White (2012): “Nonparametric identification in nonseparable panel data models with generalized fixed effects,” *Journal of Econometrics*, 168(2), 300–314.
- Honoré, B., and E. Kyriazidou (2000): “Panel data discrete choice models with lagged dependent variables,” *Econometrica*, 68(4), 839–874.

- Honoré, B., and M. Weidner (2020): "Dynamic Panel Logit Models with Fixed Effects," *arXiv preprint arXiv:2005.05942*.
- Honoré, B. E., and Á. De Paula (2021): "Identification in simple binary outcome panel data models," *The Econometrics Journal*, 24(2), C78–C93.
- Honoré, B. E., C. Muris, and M. Weidner (2021): "Dynamic Ordered Panel Logit Models," *arXiv preprint arXiv:2107.03253*.
- Keane, M. (1997): "Modeling heterogeneity and state dependence in consumer choice behavior," *Journal of Business & Economic Statistics*, 15(3), 310–327.
- Lancaster, T. (2000): "The incidental parameter problem since 1948," *Journal of Econometrics*, 95(2), 391–413.
- Magnac, T. (2000): "Subsidised training and youth employment: distinguishing unobserved heterogeneity from state dependence in labour market histories," *The Economic Journal*, 110(466), 805–837.
- (2004): "Panel binary variables and sufficiency: generalizing conditional logit," *Econometrica*, 72(6), 1859–1876.
- Neyman, J., and E. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16(1), 1–32.
- Osborne, M. (2011): "Consumer learning, switching costs, and heterogeneity: A structural examination," *Quantitative Marketing and Economics*, 9(1), 25–70.
- Pakel, C., and M. Weidner (2023): "Bounds on average effects in discrete choice panel data models," *arXiv preprint arXiv:2309.09299*.
- Pakes, A., J. Porter, M. Shepard, and S. Calder-Wang (2021): "Unobserved Heterogeneity, State Dependence, and Health Plan Choices," *National Bureau of Economic Research*.

- Rasch, G. (1961): "On General Laws and the Meaning of Measurement in Psychology," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 4, 321–333.
- Roy, R., P. Chintagunta, and S. Haldar (1996): "A framework for investigating habits," "The Hand of the Past," and heterogeneity in dynamic brand choice," *Marketing Science*, 15(3), 280–299.
- Seetharaman, P. (2004): "Modeling multiple sources of state dependence in random utility models: A distributed lag approach," *Marketing Science*, 23(2), 263–271.
- Seetharaman, P., A. Ainslie, and P. Chintagunta (1999): "Investigating household state dependence effects across categories," *Journal of Marketing Research*, 36(4), 488–500.
- Viard, B. (2007): "Do switching costs make markets more or less competitive? The case of 800-number portability," *The RAND Journal of Economics*, 38(1), 146–163.

A Appendix

A.1 Proof of Lemma 1

For notational simplicity, define $\alpha_{i,\mathbf{x}} \equiv \alpha_i + \mathbf{x}'\boldsymbol{\gamma}$. Using the definition of $\Delta(\alpha_i, \mathbf{x})$, we have:

$$\begin{aligned}\Delta(\alpha_i, \mathbf{x}) &= \frac{e^{\alpha_{i,\mathbf{x}}+\beta}}{1 + e^{\alpha_{i,\mathbf{x}}+\beta}} - \frac{e^{\alpha_{i,\mathbf{x}}}}{1 + e^{\alpha_{i,\mathbf{x}}}} = \frac{e^{\alpha_{i,\mathbf{x}}} [e^\beta - 1]}{[1 + e^{\alpha_{i,\mathbf{x}}}] [1 + e^{\alpha_{i,\mathbf{x}}+\beta}]} \\ &= [e^\beta - 1] \pi_{01}(\alpha_i, \mathbf{x}) \pi_{10}(\alpha_i, \mathbf{x}).\end{aligned}\tag{47}$$

that give us equation (10) in Lemma 1. We also have that:

$$\begin{aligned}\frac{\pi_{11}(\alpha_i, \mathbf{x}) \pi_{00}(\alpha_i, \mathbf{x})}{\pi_{10}(\alpha_i, \mathbf{x}) \pi_{01}(\alpha_i, \mathbf{x})} &= \frac{e^{\alpha_{i,\mathbf{x}}+\beta}/[1 + e^{\alpha_{i,\mathbf{x}}+\beta}]}{1/[1 + e^{\alpha_{i,\mathbf{x}}+\beta}]} \frac{1/[1 + e^{\alpha_{i,\mathbf{x}}}}{e^{\alpha_{i,\mathbf{x}}}/[1 + e^{\alpha_{i,\mathbf{x}}}} \\ &= \frac{e^{\alpha_{i,\mathbf{x}}+\beta}}{e^{\alpha_{i,\mathbf{x}}}} = e^\beta.\end{aligned}\tag{48}$$

that corresponds to equation (11) in Lemma 1. ■

A.2 Proof of Proposition 1

W.l.o.g., we consider $T = 3$.¹⁸ For any choice sequence (y_1, y_2, y_3) and any sequence of covariates $\mathbf{x}^{\{1,3\}}$ with $\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}$, we have:

$$\mathbb{P}_{y_1 y_2 y_3 \mid \mathbf{x}^{\{1,3\}}} = \int p^*(y_1 \mid \alpha_i, \mathbf{x}^{\{1,3\}}) \pi_{y_1 y_2}(\alpha_i, \mathbf{x}) \pi_{y_2 y_3}(\alpha_i, \mathbf{x}) f_{\alpha \mid \mathbf{x}^{\{1,3\}}}(\alpha_i \mid \mathbf{x}^{\{1,3\}}) d\alpha_i \tag{49}$$

Applying equation (10) in Lemma 1 to equation (49) for $\mathbb{P}_{010 \mid \mathbf{x}^{\{1,3\}}}$ and $\mathbb{P}_{101 \mid \mathbf{x}^{\{1,3\}}}$:

$$\begin{cases} \mathbb{P}_{010 \mid \mathbf{x}^{\{1,3\}}} = \frac{1}{e^\beta - 1} \int p^*(0 \mid \alpha_i, \mathbf{x}^{\{1,3\}}) \Delta(\alpha_i, \mathbf{x}) f_{\alpha \mid \mathbf{x}^{\{1,3\}}}(\alpha_i \mid \mathbf{x}^{\{1,3\}}) d\alpha_i \\ \mathbb{P}_{101 \mid \mathbf{x}^{\{1,3\}}} = \frac{1}{e^\beta - 1} \int p^*(1 \mid \alpha_i, \mathbf{x}^{\{1,3\}}) \Delta(\alpha_i, \mathbf{x}) f_{\alpha \mid \mathbf{x}^{\{1,3\}}}(\alpha_i \mid \mathbf{x}^{\{1,3\}}) d\alpha_i \end{cases} \tag{50}$$

Adding up these two equations, multiplying the resulting equation times $e^\beta - 1$, and taking into account that $p^*(0 \mid \alpha_i, \mathbf{x}^{\{1,3\}}) + p^*(1 \mid \alpha_i, \mathbf{x}^{\{1,3\}}) = 1$, we have that $AME(\mathbf{x}) = [e^\beta - 1] [\mathbb{P}_{010 \mid (\mathbf{x}_1, \mathbf{x}, \mathbf{x})} + \mathbb{P}_{101 \mid (\mathbf{x}_1, \mathbf{x}, \mathbf{x})}]$ such that $AME(\mathbf{x}, \mathbf{x}^{\{1,3\}})$ is identified. ■

¹⁸Given identification with $T = 3$, it is obvious that there is also identification for any value of T greater than 3, as we can take sub-histories with three periods.

A.3 Proof of Proposition 2

In this proof, for notational simplicity but w.l.o.g., we omit $\mathbf{x}^{\{1,T\}}$ and $\boldsymbol{\theta}$ as arguments in all the functions. Remember equation (17) in Proposition 2:

$$\sum_{\mathbf{y}^{\{2,T\}} \in \mathcal{Y}^{T-1}} w_{y_1, \mathbf{y}^{\{2,T\}}} G(\mathbf{y}^{\{2,T\}} | y_1, \alpha_i) = \Delta(\alpha_i). \quad (51)$$

(A) Sufficient condition. Multiplying (51) times $p^*(y_1 | \alpha) f_\alpha(\alpha)$, integrating over α , and taking into account that, as defined in (15), $\int G(\mathbf{y}^{\{2,T\}} | y_1, \alpha) p^*(y_1 | \alpha) f_\alpha(\alpha) d\alpha$ is equal to $P_{\mathbf{y}^{\{1,T\}}}$, we obtain:

$$\sum_{\mathbf{y}^{\{2,T\}} \in \mathcal{Y}^{T-1}} w_{y_1, \mathbf{y}^{\{2,T\}}} P_{\mathbf{y}^{\{1,T\}}} = \int \Delta(\alpha) p^*(y_1 | \alpha) f_\alpha(\alpha) d\alpha. \quad (52)$$

We can sum equation (52) over all the possible values of y_1 in \mathcal{Y} . Given that the sum of $p^*(y_1 | \alpha)$ over all values of y_1 in \mathcal{Y} is equal to 1, the right-hand-side becomes $\int \Delta(\alpha) f_\alpha(\alpha) d\alpha$, which is the definition of AME . Furthermore, the sum of the equation (52) over all the possible values of y_1 implies the following equation:

$$\sum_{\mathbf{y} \in \mathcal{D} \times \mathcal{Y}^T} w_{\mathbf{y}} P_{\mathbf{y}^{\{1,T\}}} = AME, \quad (53)$$

which is equation (18) in Proposition 2.

(B) Necessary condition. The proof has two parts. In part (i), we prove that function $h(P_{\mathcal{Y}|\mathcal{X}})$ should be linear in $P_{\mathcal{Y}|\mathcal{X}}$. Then, in part (ii), we show that the linearity of the function $h(P_{\mathcal{Y}|\mathcal{X}})$ implies that equation (51) should hold.

Necessary – Part (i). Equality $h(P_{\mathcal{Y}|\mathcal{X}}) = AME$ should hold for every distribution f_α . Consequently, it should hold for the following three specific cases: (Case 1) a degenerate distribution where $\alpha_i = c$ with probability one, where c is constant; (Case 2) a degenerate distribution where $\alpha_i = c'$ with probability one, where c' is a constant different to c ; and (Case 3) a distribution with c and c' as the only two points of support, with $q \equiv f_\alpha(c)$. For each of these three cases, AME has the following form: $AME = \Delta(c)$ in Case 1; $AME = \Delta(c')$ in Case 2; and $AME = q \Delta(c) + (1 - q) \Delta(c')$ in Case 3. Therefore, function

$h(P_{y|x})$ should satisfy the following three restrictions:

$$\begin{cases} \text{Case 1} : h(\mathbf{P}_{y|x}) = \Delta(\mathbf{c}) \\ \text{Case 2} : h(\mathbf{P}_{y|x}^{(2)}) = \Delta(\mathbf{c}') \\ \text{Case 3} : h(\mathbf{P}_{y|x}^{(3)}) = q \Delta(\mathbf{c}) + (1 - q) \text{text}\Delta(\mathbf{c}') \end{cases} \quad (54)$$

Where $P_{y|x}$, $P_{y|x}^{(2)}$, and $P_{y|x}^{(3)}$ represent the distributions of $\mathbf{y}^{\{1,T\}}$ conditional on $\mathbf{x}^{\{1,T\}}$ under the DGPs of cases 1, 2, and 3, respectively. Note that, by construction:

$$P_{y|x}^{(3)} = q P_{y|x} + (1 - q) P_{y|x}^{(2)} \quad (55)$$

This condition should hold for any arbitrary values of the constants c , c' , and $q \in [0, 1]$. Multiplying equation (54)(Case 1) times q , multiplying equation (54)(Case 2) times $(1 - q)$, adding up these two results, and then subtracting equation (54)(Case 3), we get that function $h(\mathbf{P}_{y|x})$ should satisfy the following equation:

$$q h(\mathbf{P}_{y|x}) + (1 - q) h(\mathbf{P}_{y|x}^{(2)}) = h(q \mathbf{P}_{y|x} + (1 - q) \mathbf{P}_{y|x}^{(2)}). \quad (56)$$

The only possibility that equation (56) holds for any arbitrary value of c , c' , and $q \in [0, 1]$ is that function $h(\mathbf{P}_{y|x})$ is linear in $P_{y|x}$, such that $h(\mathbf{P}_{y|x}) = \sum_{\mathbf{y}^{\{1,T\}}} w_{\mathbf{y}^{\{1,T\}}} P_{\mathbf{y}^{\{1,T\}}}$.

Necessary – Part (ii). We need to prove that, if equation $\sum_{\mathbf{y}^{\{1,T\}}} w_{\mathbf{y}^{\{1,T\}}} P_{\mathbf{y}^{\{1,T\}}} = AME$ holds, then equation (51) should hold for every value α . The proof is by contradiction. Suppose that: (a) equation $\sum_{\mathbf{y}^{\{1,T\}}} w_{\mathbf{y}^{\{1,T\}}} P_{\mathbf{y}^{\{1,T\}}} = AME$ holds for any distribution f_α in the DGP; and (b) there is a value $\alpha = c$ and a value y_1 of the initial condition such that equation (51) does not hold: $\sum_{\mathbf{y}^{\{2,T\}}} w_{y_1, \mathbf{y}^{\{2,T\}}} G(\mathbf{y}^{\{2,T\}}|y_1, c) \neq \Delta(c)$. We show below that condition (b) implies that there is a density function f_α (in fact, a continuum of density functions) such that condition (a) does not hold.

W.l.o.g., consider distributions of α with only two points support, c and c' with $f_\alpha(c) = q$. Define the following function $d(y_1, \alpha)$ that measures the extent in which equation (51) is not satisfied:

$$d(y_1, \alpha) \equiv \sum_{\mathbf{y}^{\{2,T\}}} w_{y_1, \mathbf{y}^{\{2,T\}}} G(\mathbf{y}^{\{2,T\}}|y_1, \alpha) - \Delta(\alpha) \quad (57)$$

Condition (b) implies that $d(y_1, c) \neq 0$. For notational simplicity but w.l.o.g., consider that the initial condition y_1 has binary support $\{0, 1\}$. Applying the same operations as in the

proof of the sufficient condition, we get:

$$\sum_{\mathbf{y}^{\{1,T\}}} w_{\mathbf{y}^{\{1,T\}}} \mathbb{P}_{\mathbf{y}^{\{1,T\}}} - AME = q [p^*(0|c) d(0, c) + p^*(1|c) d(1, c)] + (1 - q) [p^*(0|c') d(0, c') + p^*(1|c') d(1, c')] \quad (58)$$

By definition, each value $d(y_1, \alpha)$ is for a particular value of α , and therefore, it does not depend on the distribution f_α . More specifically, $d(y_1, \alpha)$ does not depend on the value of q . Therefore, there always exists (a continuum of) values of q such that the right-hand side of (58) is different from zero, and condition (a) does not hold. ■

A.4 Proof of Proposition 3

A.4.1 Part A: Polynomial in e^α

In this proof, for notational simplicity but w.l.o.g., we omit \mathbf{x} and $\boldsymbol{\theta}$ as arguments in all the functions. Using the structure of the function G in the binary choice model, as presented in equation (16), and the expression for $\Delta(\alpha)$ in Lemma 1, we can rewrite equation (17) as follows:

$$\sum_{\mathbf{y}^{\{2,T\}}} w_{y_1, \mathbf{y}^{\{2,T\}}} \prod_{t=2}^T \frac{e^{y_t[\alpha + \beta y_{t-1}]}}{1 + e^{\alpha + \beta y_{t-1}}} = (e^\beta - 1) \frac{e^\alpha}{1 + e^\alpha} \frac{1}{1 + e^{\alpha + \beta}} \quad (59)$$

Multiplying this equation times $[1 + e^\alpha]^{T-1} [1 + e^{\alpha + \beta}]^{T-1}$ to eliminate the denominators, and using the Binomial Theorem to expand the terms $[1 + e^\alpha]^n$ as $\sum_{k=0}^n \binom{n}{k} [e^\alpha]^k$, and the terms $[1 + e^{\alpha + \beta}]^n$ as $\sum_{k=0}^n \binom{n}{k} [e^\beta]^k [e^\alpha]^k$, we can represent this equation as a polynomial in e^α . Therefore, this system of equations holds for every value $\alpha \in \mathbb{R}$ if and only if the coefficients multiplying each monomial term in the polynomial are equal to zero. This result defines a finite system of equations. More specifically, the coefficients multiplying each monomial term are linear functions of the weights $w_{y_1, \mathbf{y}^{\{2,T\}}}$. The finite system of equations that makes the monomial coefficients equal to zero is linear in the weights $w_{y_1, \mathbf{y}^{\{2,T\}}}$. ■

A.4.2 Part B: Identification – Model with covariates

W.l.o.g. we consider $T = 3$ and $t = 3$. In this case, equation (17) takes the following form:

$$\sum_{y_2, y_3} w_{y_1, y_2, y_3 | \mathbf{x}^{\{1,3\}}} \frac{e^{y_2[\alpha + \beta y_1 + \mathbf{x}'_2 \boldsymbol{\gamma}]} e^{y_3[\alpha + \beta y_2 + \mathbf{x}'_3 \boldsymbol{\gamma}]} }{1 + e^{\alpha + \beta y_1 + \mathbf{x}'_2 \boldsymbol{\gamma}}} \frac{1}{1 + e^{\alpha + \beta y_2 + \mathbf{x}'_3 \boldsymbol{\gamma}}} = \frac{(e^\beta - 1) e^{\alpha + \mathbf{x}'_3 \boldsymbol{\gamma}}}{(1 + e^{\alpha + \mathbf{x}'_3 \boldsymbol{\gamma}}) (1 + e^{\alpha + \beta + \mathbf{x}'_3 \boldsymbol{\gamma}})} \quad (60)$$

Multiplying this equation by the factor that eliminates the denominators, we get the following equality of polynomials in e^α :

$$\begin{aligned}
& w_{000} + w_{000} e^{\mathbf{x}'_3 \gamma} e^\beta e^\alpha + w_{001} e^{\mathbf{x}'_3 \gamma} e^\alpha + w_{001} (e^{\mathbf{x}'_3 \gamma})^2 e^\beta (e^\alpha)^2 \\
& + w_{010} e^{\mathbf{x}'_2 \gamma} e^\alpha + w_{010} e^{\mathbf{x}'_2 \gamma} e^{\mathbf{x}'_3 \gamma} (e^\alpha)^2 \\
& + w_{011} e^{\mathbf{x}'_2 \gamma} e^{\mathbf{x}'_3 \gamma} e^\beta (e^\alpha)^2 + w_{011} e^{\mathbf{x}'_2 \gamma} (e^{\mathbf{x}'_3 \gamma})^2 e^\beta (e^\alpha)^3 \\
& = e^{\mathbf{x}'_3 \gamma} (e^\beta - 1) e^\alpha + e^{\mathbf{x}'_2 \gamma} e^{\mathbf{x}'_3 \gamma} (e^\beta - 1) (e^\alpha)^2
\end{aligned} \tag{61}$$

To ensure that a solution to this condition holds for every α , the coefficients of each monomial term must be equal on both sides of the equation. Therefore, the solution implies the following four equations:

$$\begin{aligned}
w_{000} &= 0 \\
w_{000} e^{\mathbf{x}'_3 \gamma} e^\beta + w_{001} e^{\mathbf{x}'_3 \gamma} + w_{010} e^{\mathbf{x}'_2 \gamma} &= e^{\mathbf{x}'_3 \gamma} (e^\beta - 1) \\
w_{001} (e^{\mathbf{x}'_3 \gamma})^2 e^\beta + w_{010} e^{\mathbf{x}'_2 \gamma} e^{\mathbf{x}'_3 \gamma} + w_{011} e^{\mathbf{x}'_2 \gamma} e^{\mathbf{x}'_3 \gamma} e^\beta &= e^{\mathbf{x}'_2 \gamma} e^{\mathbf{x}'_3 \gamma} (e^\beta - 1) \\
w_{011} e^{\mathbf{x}'_2 \gamma} (e^{\mathbf{x}'_3 \gamma})^2 e^\beta &= 0
\end{aligned} \tag{62}$$

The unique solution to this system is $w_{000} = w_{011} = 0$, $w_{001} = e^{[\mathbf{x}_2 - \mathbf{x}_3]' \gamma} - 1$, and $w_{010} = e^{[\mathbf{x}_3 - \mathbf{x}_2]' \gamma} e^\beta - 1$.

We can proceed in the same way for the case of $y_1 = 1$ to obtain the following unique solution for the weights: $w_{100} = w_{111} = 0$, $w_{101} = e^{[\mathbf{x}_2 - \mathbf{x}_3]' \gamma} e^\beta - 1$, and $w_{110} = e^{[\mathbf{x}_3 - \mathbf{x}_2]' \gamma} - 1$.

Putting these pieces together, we have that:

$$\begin{aligned}
AME(\mathbf{x}_3, \mathbf{x}^{\{1,3\}}) &= w_{0,0,1, \mathbf{x}^{\{1,3\}}} \mathbb{P}_{0,0,1 | \mathbf{x}^{\{1,3\}}} + w_{0,1,0, \mathbf{x}^{\{1,3\}}} \mathbb{P}_{0,1,0 | \mathbf{x}^{\{1,3\}}} \\
&+ w_{1,0,1, \mathbf{x}^{\{1,3\}}} \mathbb{P}_{1,0,1 | \mathbf{x}^{\{1,3\}}} + w_{1,1,0, \mathbf{x}^{\{1,3\}}} \mathbb{P}_{1,1,0 | \mathbf{x}^{\{1,3\}}}
\end{aligned} \tag{63}$$

with

$$\begin{aligned}
w_{0,0,1, \mathbf{x}^{\{1,3\}}} &= -1 + e^{[\mathbf{x}_2 - \mathbf{x}_3]' \gamma} ; w_{0,1,0, \mathbf{x}^{\{1,3\}}} = -1 + e^{\beta + [\mathbf{x}_3 - \mathbf{x}_2]' \gamma} ; \\
w_{1,0,1, \mathbf{x}^{\{1,3\}}} &= -1 + e^{\beta + [\mathbf{x}_2 - \mathbf{x}_3]' \gamma} ; w_{1,1,0, \mathbf{x}^{\{1,3\}}} = -1 + e^{[\mathbf{x}_3 - \mathbf{x}_2]' \gamma}.
\end{aligned} \tag{64}$$

In the model without covariates, the vector of parameters γ equals zero. Consequently, the weights w_{001} and w_{110} are also zero, and we get $AME = [e^\beta - 1] [\mathbb{P}_{0,1,0} + \mathbb{P}_{1,0,1}]$. ■

A.4.3 Over-identification when T is greater than three

The identification result using only 3 periods proves identification for any $T \geq 3$ because, with more than 3 periods, we can always take 3 periods. Nonetheless, it is possible to obtain close form expression for higher values of T using the same procedure based on Proposition 2.

This expression will use all T periods without having to combine several 3-periods estimates.

For $T > 3$, as said, there is overidentification, so more than one combination of the probability of histories exists. One way to choose one of them is to focus on the probabilities of the sufficient statistics used to identify β in the CMLE. Specifically, for this model and other logit models, the log-probability of a choice history has the following structure:

$$\ln \mathbb{P}(\mathbf{y}_i | \alpha_i, \beta) = \mathbf{s}(\mathbf{y}_i)' \mathbf{g}(\alpha_i) + \mathbf{c}(\mathbf{y}_i)' \beta \quad (65)$$

where $s(\mathbf{y}_i)$ and $c(\mathbf{y}_i)$ are vectors of statistics (functions of \mathbf{y}_i), and $g(\alpha_i)$ is a vector of functions α_i . $s(\mathbf{y}_i)$ is a sufficient statistic for α_i because $P(\mathbf{y}_i | \alpha_i, \beta, s_i) = P(\mathbf{y}_i | \beta, s_i)$.¹⁹ Let S_T be the set of possible values of $s(\mathbf{y}_i)$, let P_s be the probability of a value s of $s(\mathbf{y}_i)$, and let $\mathbb{P}_s \equiv \{P_s : s \in S_T\}$ be the probability distribution of this statistic. Given θ , the empirical distribution P_s contains all the information in the data about the distribution of α_i , and therefore, about AMEs. Taking into account the structure of the probability of a choice history in the equation (65), the model implies:

$$\mathbb{P}_s = \sum_{\mathbf{y}: \mathbf{s}(\mathbf{y})=s} \left[\int e^{\mathbf{s}' \mathbf{g}(\alpha_i) + \mathbf{c}(\mathbf{y})' \theta} f_\alpha(\alpha_i) d\alpha_i \right] \quad (66)$$

If two sequences, say k and l , have the same $s(\mathbf{y}_j)$, the ratio of the probabilities of these two sequences is equal to $e^{c(\mathbf{y}_k)' \beta - c(\mathbf{y}_l)' \beta}$, which is not a function of α_i . This includes the case in which $P(\mathbf{y}_j | \beta, \alpha_i)$ is the same for both sequences. Therefore, the set of sequences with the same or proportional $P(\mathbf{y}_j | \beta, \alpha_i)$ is the set of sequences with the same value of the sufficient statistic $s(\mathbf{y}_j)$. This result leads to an infinite number of combinations of these sequences, with the only restriction being that all the combinations have to sum up to the same number (overall weight). We choose the combination in which all these sequences have the same weight w , and, therefore, look for combinations of P_s instead of $P_{\mathbf{y}}$.

In the BC-AR1 model the sufficient statistics $s(\mathbf{y}_i)$ is the vector $\left(y_{i1}, y_{iT}, \sum_{t=2}^T y_{it} \right)'$ –see [Aguirregabiria, Gu, and Luo \(2021\)](#)– and it can take $4T - 4$ different values, $2T - 2$ values with $y_{i1} = 0$ and $2T - 2$ values with $y_{i1} = 1$. The conditions from Proposition 2 are:

$$\left. \begin{aligned} \sum_{j=1}^{2T-2} w_j \mathbb{P}(\mathbf{s}_j | y_{j1} = 0, \beta, \alpha_i) &= \Delta(\alpha_i) \\ \sum_{j=2T-2+1}^{4T-4} w_j \mathbb{P}(\mathbf{s}_j | y_{j1} = 1, \beta, \alpha_i) &= \Delta(\alpha_i) \end{aligned} \right\} \text{for every } \alpha_i \in \mathbb{R}, \quad (67)$$

¹⁹See [Aguirregabiria, Gu, and Luo \(2021\)](#) for further details on this decomposition of the probability choice and sufficient statistics for discrete choice logit models.

Table 7: **Weights for histories with $y_1 = 0$**

$(y_1, y_T, \sum_{t=2}^T y_t)$	$T = 4$	$T = 5$	$T = 6$	$T = 7$
(0, 0, 0)	0	0	0	0
(0, 0, 1)	$\frac{e^\beta - 1}{2}$	$\frac{e^\beta - 1}{3}$	$\frac{e^\beta - 1}{4}$	$\frac{e^\beta - 1}{5}$
(0, 1, 1)	0	0	0	0
(0, 0, 2)	0	$\frac{e^\beta - 1}{1 + 2e^\beta}$	$\frac{2(e^\beta - 1)}{3 + 3e^\beta}$	$\frac{3(e^\beta - 1)}{6 + 4e^\beta}$
(0, 1, 2)	$\frac{e^\beta - 1}{1 + e^\beta}$	$\frac{e^\beta - 1}{2 + e^\beta}$	$\frac{e^\beta - 1}{3 + e^\beta}$	$\frac{e^\beta - 1}{4 + e^\beta}$
(0, 0, 3)	Not possible	0	$\frac{e^\beta - 1}{2 + 2e^\beta}$	$\frac{(e^\beta - 1)(1 + 2e^\beta)}{1 + 6e^\beta + 3e^{2\beta}}$
(0, 1, 3)	0	$\frac{e^\beta - 1}{2 + e^\beta}$	$\frac{(e^\beta - 1)(1 + e^\beta)}{1 + 4e^\beta + e^{2\beta}}$	$\frac{(e^\beta - 1)(2 + e^\beta)}{3 + 6e^\beta + e^{2\beta}}$
(0, 0, 4)	Not possible	Not possible	0	$\frac{e^\beta - 1}{3 + 2e^\beta}$
(0, 1, 4)	Not possible	0	$\frac{e^\beta - 1}{3 + e^\beta}$	$\frac{(e^\beta - 1)(2 + e^\beta)}{3 + 6e^\beta + e^{2\beta}}$
(0, 0, 5)	Not possible	Not possible	Not possible	0
(0, 1, 5)	Not possible	Not possible	0	$\frac{e^\beta - 1}{4 + e^\beta}$
(0, 1, 6)	Not possible	Not possible	Not possible	0

where $P(\mathbf{s}_j \mid y_{j1} = 0, \beta, \alpha_i) = \sum_{\mathbf{y}:s(\mathbf{y})=\mathbf{s}_j} P(\mathbf{y}^{\{2,T\}} \mid y_1 = 0, \beta, \alpha_i)$

Proceeding similarly, we obtain the weights for *AME* in the binary choice AR(1) model for different values of T . These are in Tables 7 and 8.

A.5 Proof of Proposition 4

We omit \mathbf{x} as an argument throughout this proof for notational simplicity. However, one should understand that the probability of the initial conditions p^* , the density function of α_i , the empirical probabilities of choice histories and the average transition probabilities are all conditional on $\mathbf{x}_i^{\{1,3\}} = [\mathbf{x}_1, \mathbf{x}, \mathbf{x}]$.

By definition of Π_{00} , and taking into account that $p^*(0|\alpha_i) + p^*(1|\alpha_i) = 1$, we have that:

$$\Pi_{00} = \int [p^*(0|\alpha_i) + p^*(1|\alpha_i)] \pi_{00}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \quad (68)$$

This expression includes the term $\int p^*(0|\alpha_i) \pi_{00}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i$, which is equal to the choice history probability \mathbb{P}_{00} . However, it also includes the "counterfactual" $\int p^*(1|\alpha_i) \pi_{00}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i$. Denote this counterfactual as δ_{100} . Given that $\pi_{10}(\alpha_i) + \pi_{11}(\alpha_i) = 1$, we can represent this counterfactual as:

$$\delta_{100} = \int p^*(1|\alpha_i) [\pi_{10}(\alpha_i) + \pi_{11}(\alpha_i)] \pi_{00}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \quad (69)$$

Table 8: **Weights for histories with $y_1 = 1$**

$(y_1, y_T, \sum_{t=2}^T y_t)$	$T = 4$	$T = 5$	$T = 6$	$T = 7$
(1, 0, 0)	0	0	0	0
(1, 0, 1)	$\frac{e^\beta - 1}{1 + e^\beta}$	$\frac{e^\beta - 1}{2 + e^\beta}$	$\frac{e^\beta - 1}{3 + e^\beta}$	$\frac{e^\beta - 1}{4 + e^\beta}$
(1, 1, 1)	0	0	0	0
(1, 0, 2)	0	$\frac{e^\beta - 1}{2 + e^\beta}$	$\frac{(e^\beta - 1)(1 + e^\beta)}{1 + 4e^\beta + e^{2\beta}}$	$\frac{(e^\beta - 1)(2 + e^\beta)}{3 + 6e^\beta + e^{2\beta}}$
(1, 1, 2)	$\frac{e^\beta - 1}{2}$	$\frac{e^\beta - 1}{1 + 2e^\beta}$	$\frac{e^\beta - 1}{2 + 2e^\beta}$	$\frac{e^\beta - 1}{3 + 2e^\beta}$
(1, 0, 3)	Not possible	0	$\frac{e^\beta - 1}{3 + e^\beta}$	$\frac{(e^\beta - 1)(2 + e^\beta)}{3 + 6e^\beta + e^{2\beta}}$
(1, 1, 3)	0	$\frac{e^\beta - 1}{3}$	$\frac{2(e^\beta - 1)}{3 + e^\beta}$	$\frac{(e^\beta - 1)(1 + 2e^\beta)}{1 + 6e^\beta + 3e^{2\beta}}$
(1, 0, 4)	Not possible	Not possible	0	$\frac{e^\beta - 1}{4 + e^\beta}$
(1, 1, 4)	Not possible	0	$\frac{e^\beta - 1}{4}$	$\frac{3(e^\beta - 1)}{6 + 4e^\beta}$
(1, 0, 5)	Not possible	Not possible	Not possible	0
(1, 1, 5)	Not possible	Not possible	0	$\frac{e^\beta - 1}{5}$
(1, 1, 6)	Not possible	Not possible	Not possible	0

This equation shows that δ_{100} is the sum of two terms. The first term is $\int p^*(1|\alpha_i) \pi_{10}(\alpha_i) \pi_{00}(\alpha_i) f_\alpha(\alpha_i)$, which is equal to the choice history probability \mathbb{P}_{100} . The second term is $\int p^*(1|\alpha_i) \pi_{11}(\alpha_i) \pi_{00}(\alpha_i) f_\alpha(\alpha_i)$, which in principle is a counterfactual. However, by Lemma 1, we have that $\pi_{11}(\alpha_i) \pi_{00}(\alpha_i)$ is equal to $e^\beta \pi_{10}(\alpha_i) \pi_{01}(\alpha_i)$. Therefore, the counterfactual $\int p^*(1|\alpha_i) \pi_{11}(\alpha_i) \pi_{00}(\alpha_i) f_\alpha(\alpha_i)$ is equal to $e^\beta \int p^*(1|\alpha_i) \pi_{10}(\alpha_i) \pi_{01}(\alpha_i) f_\alpha(\alpha_i)$, and in turn this is equal to $e^\beta \mathbb{P}_{101}$.

Putting all the pieces together, we have that:

$$\Pi_{00} = \mathbb{P}_{00} + \mathbb{P}_{100} + e^\beta \mathbb{P}_{101} \quad (70)$$

Using the same procedure, we can show that $\Pi_{11} = \mathbb{P}_{11} + \mathbb{P}_{011} + e^\beta \mathbb{P}_{010}$. ■

A.6 Proof of Lemma 2

For clarity in notation, we refrain from explicitly including \mathbf{x} as an argument throughout this proof. It is important to note, however, that all probabilities and expectations in this proof are conditioned on $\mathbf{x}_i^{\{1, T\}} = [\mathbf{x}_1, \mathbf{x}, \dots, \mathbf{x}]$, where \mathbf{x}_1 is free, and $\mathbf{x}_2 = \dots = \mathbf{x}_T = \mathbf{x}$.

Using the Markov structure of the model and the chain rule, we have that:

$$\begin{aligned}\mathbb{E}(y_{i,t+n} \mid \alpha_i, y_{it}) &= \mathbb{P}(y_{i,t+n-1} = 0 \mid \alpha_i, y_{it}) \pi_{01}(\alpha_i) + \mathbb{P}(y_{i,t+n-1} = 1 \mid \alpha_i, y_{it}) \pi_{11}(\alpha_i) \\ &= \pi_{01}(\alpha_i) + \mathbb{E}(y_{i,t+n-1} \mid \alpha_i, y_{it}) [\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)]\end{aligned}\tag{71}$$

Given the definition of $\Delta^{(n)}(\alpha_i)$ as $\mathbb{E}(y_{i,t+n} \mid \alpha_i, y_{it} = 1) - \mathbb{E}(y_{i,t+n} \mid \alpha_i, y_{it} = 0)$, and applying equation (71), we have that:

$$\begin{aligned}\Delta^{(n)}(\alpha_i) &= [\mathbb{E}(y_{i,t+n-1} \mid \alpha_i, y_{it} = 1) - \mathbb{E}(y_{i,t+n-1} \mid \alpha_i, y_{it} = 0)] [\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)] \\ &= \Delta^{(n-1)}(\alpha_i) [\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)]\end{aligned}\tag{72}$$

Applying this expression recursively, we obtain that $\Delta^{(n)}(\alpha_i) = [\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)]^n = [\Delta(\alpha_i)]^n$. Finally, as established in Lemma 1, is that $\Delta(\alpha_i) = [e^\beta - 1] \pi_{10}(\alpha_i) \pi_{01}(\alpha_i)$. Thus, we have that $\Delta^{(n)}(\alpha_i) = [e^\beta - 1]^n [\pi_{10}(\alpha_i)]^n [\pi_{01}(\alpha_i)]^n$. ■

A.7 Proof of Proposition 5

Similarly as in the proof of Lemma 2 above, we omit \mathbf{x} as an argument, but one should understand that all the probabilities in this proof are conditioned on $\mathbf{x}_i^{\{1,T\}} = [\mathbf{x}, \mathbf{x}, \dots, \mathbf{x}]$. W.l.o.g., we consider that $T = 2n + 1$. Given the definition of histories $(0, \widetilde{\mathbf{10}}^n)$ and $(\widetilde{\mathbf{10}}^n, 1)$, it is straightforward to see that:

$$\begin{cases} \mathbb{P}_{0, \widetilde{\mathbf{10}}^n} = \int p^*(0 \mid \alpha_i) [\pi_{10}(\alpha_i)]^n [\pi_{01}(\alpha_i)]^n f_\alpha(\alpha_i) d\alpha_i \\ \mathbb{P}_{\widetilde{\mathbf{10}}^n, 1} = \int p^*(1 \mid \alpha_i) [\pi_{10}(\alpha_i)]^n [\pi_{01}(\alpha_i)]^n f_\alpha(\alpha_i) d\alpha_i \end{cases}\tag{73}$$

Applying equation (29) from Lemma 2, we have that:

$$\begin{cases} \mathbb{P}_{0, \widetilde{\mathbf{10}}^n} = \frac{1}{[e^\beta - 1]^n} \int p^*(0 \mid \alpha_i) \Delta^{(n)}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \\ \mathbb{P}_{\widetilde{\mathbf{10}}^n, 1} = \frac{1}{[e^\beta - 1]^n} \int p^*(1 \mid \alpha_i) \Delta^{(n)}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \end{cases}\tag{74}$$

Adding up these two equations, multiplying the resulting equation times $[e^\beta - 1]^n$, and taking into account that $p^*(0 \mid \alpha_i) + p^*(1 \mid \alpha_i) = 1$, we have that $AME^{(n)} = [e^\beta - 1]^n [\mathbb{P}_{0, \widetilde{\mathbf{10}}^n} +$

$\mathbb{P}_{\mathbf{10}^n, 1}$] such that $AME^{(n)}$ is identified. ■

A.8 Proof of Lemma 3

Given the expression for the choice probabilities in the logit model, it is simple to verify that:

$$\frac{\pi_{k\ell}(\boldsymbol{\alpha}_i, \mathbf{x})}{\pi_{kj}(\boldsymbol{\alpha}_i, \mathbf{x})} = \exp\{\alpha_i(\ell) - \alpha_i(j) + \mathbf{x}'(\gamma_\ell - \gamma_j) - \beta_j \mathbb{1}\{k = j\}\} \quad (75)$$

and

$$\frac{\pi_{jj}(\boldsymbol{\alpha}_i, \mathbf{x})}{\pi_{j\ell}(\boldsymbol{\alpha}_i, \mathbf{x})} = \exp\{\alpha_i(j) - \alpha_i(\ell) + \mathbf{x}'(\gamma_j - \gamma_\ell) + \beta_j - \beta_j \mathbb{1}\{\ell = j\}\} \quad (76)$$

The product of these two expressions is equation (39). ■

A.9 Proof of Proposition 6 – Model with duration dependence

Here, we prove the identification of $AME_d(1)$. Using a similar argument, we can establish the identification of $AME_d(d)$ for any $d \geq 1$. First, we write the expression of the probabilities of choice histories conditional on α_i , $P_{y_1, y_2, y_3, y_4 | \alpha_i}$, implied by the model. Second, for each of these probabilities, we multiply the equation times the weights w_{y_1, y_2, y_3, y_4} in the enunciate of Proposition 6. For the probabilities with non-zero weights, we have:

$$\begin{aligned} \frac{e^{\beta+\delta} - 1}{2} [\mathbb{P}_{0,0,1,0 | \alpha_i} + \mathbb{P}_{0,1,0,0 | \alpha_i}] &= p^*(0|\alpha_i) \frac{(e^{\beta+\delta} - 1) e^{\alpha_i}}{(1 + e^{\alpha_i+\beta+\delta})(1 + e^{\alpha_i})^2} \\ \frac{e^{\beta+\delta} - 1}{e^{\beta+\delta}} \mathbb{P}_{0,0,1,1 | \alpha_i} &= p^*(0|\alpha_i) \frac{(e^{\beta+\delta} - 1) e^{\alpha_i} e^{\alpha_i}}{(1 + e^{\alpha_i+\beta+\delta})(1 + e^{\alpha_i})^2} \\ (e^{\beta+\delta} - 1) [\mathbb{P}_{1,0,1,0 | \alpha_i} + \mathbb{P}_{1,0,1,1 | \alpha_i}] &= p^*(1|\alpha_i) \frac{(e^{\beta+\delta} - 1) e^{\alpha_i} (e^{\alpha_i+\beta+\delta} + 1)}{(1 + e^{\alpha_i+\beta+\delta})^2 (1 + e^{\alpha_i})} \end{aligned} \quad (77)$$

Third, we sum up these three equations. Simplifying factors and taking into account that $p^*(0|\alpha_i) + p^*(1|\alpha_i) = 1$, we get:

$$\begin{aligned} \frac{e^{\beta+\delta} - 1}{2} [\mathbb{P}_{0,0,1,0 | \alpha_i} + \mathbb{P}_{0,1,0,0 | \alpha_i}] + \frac{e^{\beta+\delta} - 1}{e^{\beta+\delta}} \mathbb{P}_{0,0,1,1 | \alpha_i} + (e^{\beta+\delta} - 1) [\mathbb{P}_{1,0,1,0 | \alpha_i} + \mathbb{P}_{1,0,1,1 | \alpha_i}] \\ = \frac{(e^{\beta+\delta} - 1) e^{\alpha_i}}{(1 + e^{\alpha_i+\beta+\delta})(1 + e^{\alpha_i})} = \frac{e^{\alpha_i+\beta+\delta}}{(1 + e^{\alpha_i+\beta+\delta})} - \frac{e^{\alpha_i}}{(1 + e^{\alpha_i})} = \Delta_d(\alpha_i, 1) \end{aligned} \quad (78)$$

Finally, we integrate the two sides of this equation over the distribution of α_i to obtain

$$\frac{e^{\beta+\delta} - 1}{2} [\mathbb{P}_{0,0,1,0} + \mathbb{P}_{0,1,0,0}] + \frac{e^{\beta+\delta} - 1}{e^{\beta+\delta}} \mathbb{P}_{0,0,1,1} + (e^{\beta+\delta} - 1) [\mathbb{P}_{1,0,1,0} + \mathbb{P}_{1,0,1,1}] = AME_d(1) \quad (79)$$

such that $AME_d(1)$ is identified.

We can proceed similarly to prove the identification of $AME_d(d)$ for any value $d \geq 1$. For instance, we can prove that:

$$\begin{aligned} AME_d(2) &= \frac{e^{\beta+2\delta} - 1}{2} [\mathbb{P}_{0,0,1,0} + \mathbb{P}_{0,1,0,0}] + \frac{e^{\beta+2\delta} - 1}{e^{\beta+\delta}} \mathbb{P}_{0,0,1,1} \\ &+ \left(\frac{e^{\beta+2\delta} (1 - e^{\beta+2\delta})}{e^{\beta+\delta}} + e^{\beta+2\delta} - 1 \right) \mathbb{P}_{0,1,1,0} \\ &+ \left(e^{\beta+\delta} - \frac{e^{\beta+\delta}}{e^{\beta+2\delta}} \right) [\mathbb{P}_{1,0,1,0} + \mathbb{P}_{1,0,1,1}] + \left(\frac{e^{\beta+2\delta} - 1}{e^{\beta+\delta}} - 1 + \frac{1}{e^{\beta+2\delta}} \right) \mathbb{P}_{1,1,0,0} \quad \blacksquare \end{aligned} \quad (80)$$

A.10 Proof of Proposition 7

We omit \mathbf{x} as an argument throughout this proof for notational simplicity. However, one should understand that the probability of the initial conditions p^* , the density function of α_i , the empirical probabilities of choice histories and the average transition probabilities are all conditional on $\mathbf{x}_i^{\{1,3\}} = [\mathbf{x}, \mathbf{x}, \mathbf{x}]$. We can write Π_{jj} as:

$$\Pi_{jj} = \int [p^*(0|\alpha_i) + p^*(1|\alpha_i) \dots + p^*(J|\alpha_i)] \pi_{jj}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \quad (81)$$

This expression includes the term $\int p^*(j|\alpha_i) \pi_{jj}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i$ that is equal to the choice history probability $\mathbb{P}_{j,j}$. However, it also includes the "counterfactuals" $\delta_{k,j,j} \equiv \int p^*(k|\alpha_i) \pi_{jj}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i$ for $k \neq j$. We can represent each of these counterfactuals as:

$$\delta_{k,j,j} = \int p^*(k|\alpha_i) [\pi_{k0}(\alpha_i) + \pi_{k1}(\alpha_i) + \dots + \pi_{kJ}(\alpha_i)] \pi_{jj}(\alpha_i) f_\alpha(\alpha_i) \quad (82)$$

That is, we have that $\delta_{k,j,j} = \sum_{\ell=0}^J \delta_{k,\ell,j,j}^{(2)}$, with $\delta_{k,\ell,j,j}^{(2)} \equiv \int p^*(k|\alpha_i) \pi_{k\ell}(\alpha_i) \pi_{jj}(\alpha_i) f_\alpha(\alpha_i)$. For $\ell = j$, we have that $\delta_{k,j,j}^{(2)}$ corresponds to the choice history probability $\mathbb{P}_{k,j,j}$. For the rest of the terms $\delta_{k,\ell,j,j}^{(2)}$, we apply Lemma 3. According to Lemma 3, we have that $\pi_{k\ell}(\alpha_i) \pi_{jj}(\alpha_i) = \exp\{\beta_{k\ell} - \beta_{kj} - \beta_{j\ell}\} \pi_{kj}(\alpha_i) \pi_{j\ell}(\alpha_i)$. Finally, note that $\int p^*(k|\alpha_i) \pi_{kj}(\alpha_i) \pi_{j\ell}(\alpha_i) f_\alpha(\alpha_i)$ is the choice history probability $\mathbb{P}_{k,j,\ell}$. Putting all the pieces together, we have the

expression in equation (40). ■

A.11 Proof of Proposition 8

We omit \mathbf{x} as an argument throughout this proof for notational simplicity. Equation (17), from Proposition 2, provides the necessary and sufficient condition for identifying an AME. Applying this condition to the model defined by equation (35) and Assumption 1-MNL, with $J + 1 = 3$, $T = 3$, and $y_1 = 0$, we get:

$$\begin{aligned}
& w_{0,0,0} \mathbb{P}_{0,0,0 | (y_1=0, \alpha_i)} + w_{0,0,1} \mathbb{P}_{0,0,1 | (y_1=0, \alpha_i)} + w_{0,0,2} \mathbb{P}_{0,0,2 | (y_1=0, \alpha_i)} \\
+ & w_{0,1,0} \mathbb{P}_{0,1,0 | (y_1=0, \alpha_i)} + w_{0,1,1} \mathbb{P}_{0,1,1 | (y_1=0, \alpha_i)} + w_{0,1,2} \mathbb{P}_{0,1,2 | (y_1=0, \alpha_i)} \\
+ & w_{0,2,0} \mathbb{P}_{0,2,0 | (y_1=0, \alpha_i)} + w_{0,2,1} \mathbb{P}_{0,2,1 | (y_1=0, \alpha_i)} + w_{0,2,2} \mathbb{P}_{0,2,2 | (y_1=0, \alpha_i)} = \pi_{10}(\alpha_i)
\end{aligned} \tag{83}$$

Let's denote $d_j \equiv 1 + e^{\beta_j \mathbb{1}\{j=1\} + \alpha_i(1)} + e^{\beta_j \mathbb{1}\{j=2\} + \alpha_i(2)}$, for $j = 0, 1, 2$. Replacing the probabilities by their expression based on the logistic CDF:

$$\begin{aligned}
& w_{0,0,0} \frac{1}{d_0^2} + w_{0,0,1} \frac{e^{\alpha_i(1)}}{d_0^2} + w_{0,0,2} \frac{e^{\alpha_i(2)}}{d_0^2} \\
+ & w_{0,1,0} \frac{e^{\alpha_i(1)}}{d_0 d_1} + w_{0,1,1} \frac{e^{\alpha_i(1)} e^{\beta_1 + \alpha_i(1)}}{d_0 d_1} + w_{0,1,2} \frac{e^{\alpha_i(1)} e^{\alpha_i(2)}}{d_0 d_1} \\
+ & w_{0,2,0} \frac{e^{\alpha_i(2)}}{d_0 d_2} + w_{0,2,1} \frac{e^{\alpha_i(1)} e^{\alpha_i(2)}}{d_0 d_2} + w_{0,2,2} \frac{e^{\alpha_i(2)} e^{\beta_2 + \alpha_i(2)}}{d_0 d_2} = \frac{1}{d_1}
\end{aligned} \tag{84}$$

After some algebra to undo the fractions on both sides:

$$\begin{aligned}
& w_{0,0,0} d_1 d_2 + w_{0,0,1} e^{\alpha_i(1)} d_1 d_2 + w_{0,0,2} e^{\alpha_i(2)} d_1 d_2 \\
+ & w_{0,1,0} e^{\alpha_i(1)} d_0 d_2 + w_{0,1,1} e^{\alpha_i(1)} e^{\beta_1 + \alpha_i(1)} d_0 d_2 + w_{0,1,2} e^{\alpha_i(1)} e^{\alpha_i(2)} d_0 d_2 \\
+ & w_{0,2,0} e^{\alpha_i(2)} d_0 d_1 + w_{0,2,1} e^{\alpha_i(1)} e^{\alpha_i(2)} d_0 d_1 + w_{0,2,2} e^{\beta_2 + \alpha_i(2)} e^{\alpha_i(2)} d_0 d_1 = d_0^2 d_2
\end{aligned} \tag{85}$$

Expanding this equation by doing the products of d_j and of the exponential, we obtain on both sides of the equality a polynomial in $(e^{\alpha_i(1)})^h (e^{\alpha_i(2)})^\ell$, where the minimum value of h and ℓ is 0, and the maximum value is 4. Following Lemma 3, equating the coefficient of each monomial in both sides of the equality, we get a system of linear equation whose

unknowns are the weights $w_{0,0,0}, \dots, w_{0,2,2}$. This condition on the monomials of $(e^{\alpha_i(1)})^2$ and $(e^{\alpha_i(1)})^3$ imply, respectively:

$$w_{0,0,1} + w_{0,1,0} + w_{0,0,1} e^{\beta_1} + w_{0,1,0} = 1 \quad (86)$$

$$w_{0,0,1} e^{\beta_1} + w_{0,1,0} = 0$$

which leads to

$$w_{0,0,1} + w_{0,1,0} = 1. \quad (87)$$

At the same time, the condition on the monomial of $(e^{\alpha_i(1)})^1$ implies:

$$w_{0,0,1} + w_{0,1,0} = 2, \quad (88)$$

which is incompatible with the condition $w_{0,0,1} + w_{0,1,0} = 1$. Therefore, no weights can solve the system of equations in 83. By Proposition 2, this implies that Π_{10} is not point identified.

■

A.12 Proof of Proposition 9 – Ordered Logit

The weights in the statement of Proposition 9 were obtained using the general procedure established in Proposition 2. Here, we present the proof for Π_{20} , but it proceeds the same for other Π_{kj} . By definition, $\Pi_{20} = \int \pi_{20}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i$, and according to equation (42),

$\pi_{20}(\alpha_i) = \frac{1}{1 + e^{\beta_2 - \lambda_0 + \alpha_i}}$. We start with the probabilities of choice histories conditional on α_i , that is, $P_{(y_1, y_2, y_3) | \alpha_i}$. We write the expression for these model probabilities as functions of parameters β , λ , and α_i . For each of these probabilities, we multiply the equation times the weights w_{y_1, y_2, y_3} that appear in the statement of Proposition 9. For the probabilities

with non-zero weights for Π_{20} , we have:

$$\begin{aligned}
\mathbb{P}_{0,0,0 | \alpha_i} + \mathbb{P}_{0,0,1 | \alpha_i} &= p^*(0|\alpha_i) \frac{1}{(1 + e^{\beta_0 - \lambda_1 + \alpha_i})(1 + e^{\beta_0 - \lambda_0 + \alpha_i})} \\
e^{\lambda_1 - \lambda_0} \mathbb{P}_{0,0,2 | \alpha_i} &= p^*(0|\alpha_i) \frac{e^{\beta_0 - \lambda_0 + \alpha_i}}{(1 + e^{\beta_0 - \lambda_1 + \alpha_i})(1 + e^{\beta_0 - \lambda_0 + \alpha_i})} \\
\left(1 - \frac{e^{\beta_2 - \lambda_0}}{e^{\beta_0 - \lambda_1}}\right) \mathbb{P}_{0,2,0 | \alpha_i} &= p^*(0|\alpha_i) \frac{e^{\beta_0 - \lambda_1 + \alpha_i} - e^{\beta_2 - \lambda_0 + \alpha_i}}{(1 + e^{\beta_0 - \lambda_1 + \alpha_i})(1 + e^{\beta_2 - \lambda_0 + \alpha_i})} \\
\sum_{k=0}^l \sum_{\ell=0}^{J-1} \mathbb{P}_{1,k,\ell | \alpha_i} &= p^*(1|\alpha_i) \left(\frac{1}{1 + e^{\beta_2 - \lambda_0 + \alpha_i}} - \frac{1 - e^{\beta_2 - \lambda_0 + \alpha_i}}{(1 + e^{\beta_1 - \lambda_1 + \alpha_i})(1 + e^{\beta_2 - \lambda_0 + \alpha_i})} \right) \\
\left(1 - \frac{e^{\beta_2 - \lambda_0}}{e^{\beta_1 - \lambda_1}}\right) \mathbb{P}_{1,2,0 | \alpha_i} &= p^*(1|\alpha_i) \frac{1 - e^{\beta_2 - \lambda_0 + \alpha_i}}{(1 + e^{\beta_1 - \lambda_1 + \alpha_i})(1 + e^{\beta_2 - \lambda_0 + \alpha_i})} \\
\sum_{k=0}^{J-1} \mathbb{P}_{2,0,k | \alpha_i} &= p^*(2|\alpha_i) \frac{1}{1 + e^{\beta_2 - \lambda_0 + \alpha_i}}
\end{aligned} \tag{89}$$

Summing up these equations, simplifying factors, and taking into account that $p^*(0|\alpha_i) + p^*(1|\alpha_i) + p^*(2|\alpha_i) = 1$, we get:

$$\begin{aligned}
&\mathbb{P}_{0,0,0 | \alpha_i} + \mathbb{P}_{0,0,1 | \alpha_i} + e^{\lambda_1 - \lambda_0} \mathbb{P}_{0,0,2 | \alpha_i} + \left(1 - \frac{e^{\beta_2 - \lambda_0}}{e^{\beta_0 - \lambda_1}}\right) \mathbb{P}_{0,2,0 | \alpha_i} \\
&+ \sum_{k=0}^l \sum_{\ell=0}^J \mathbb{P}_{1,k,\ell | \alpha_i} + \left(1 - \frac{e^{\beta_2 - \lambda_0}}{e^{\beta_1 - \lambda_1}}\right) \mathbb{P}_{1,2,0 | \alpha_i} + \sum_{k=0}^J \mathbb{P}_{2,0,k | \alpha_i} = \pi_{20}(\alpha_i)
\end{aligned} \tag{90}$$

Finally, we integrate the two sides of this equation over the distribution of α_i to obtain the expression for Π_{20} in equation (43). ■