University of Toronto
Department of Economics

Working Paper 701

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

By Victor Aguirregabiria and Jesus Carro

July 08, 2021

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

July 8, 2021

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 the identification of 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 logic model, and for T as small as four, we prove the point identification of different AMEs, including causal effects of changes in the lagged dependent variable or in the duration in last choice. 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 would like to thank Manuel Arellano, Dmitry Arkhangelsky, Stephane Bonhomme, Irene Botosaru, Jiaying Gu, Francis Guiton, Chris Muris, and Enrique Sentana for helpful comments and discussions. We have also benefited from the comments of seminar participants at CEMFI, Toronto, and at the 2020 Econometric Society World Congress. We want to thank Susumu Imai for generously sharing with us the dataset that we use in our empirical application.

†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

In dynamic panel data models, ignoring the correlation between unobserved heterogeneity and pre-determined explanatory variables can generate important biases in the estimation of 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. In short panels, this distribution cannot be identified nonparametrically, and random effects approaches are not robust to misspecification of parametric restrictions. This is the so called *initial conditions problem* (Heckman, 1981). In contrast, *fixed effects (FE)* approaches impose no restriction on this distribution such that the identification of parameters of interest is robust to misspecification of this primitive.

In discrete choice models with short panels, a limitation of FE methods is that they cannot deliver identification of the distribution of the time-invariant unobserved heterogeneity. This is because the data consist of a finite number of probabilities – as many as the number of possible choice histories – but the distribution of the unobserved heterogeneity has infinite dimension. This identification problem has generated a more substantial criticism of FE approaches. The applied researcher is often interested in estimating average marginal effects (AME) of changes in explanatory variables or in 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]

In this paper, we present new results on the point identification of AMEs in FE dynamic logit models. We prove the identification of the AME of a change in the lagged dependent variable. This is a key parameter in dynamic models as it measures the causal effect of an agent's past decision on her current decision. We show that the identification of this parameter does not require knowledge of the full distribution of the unobserved heterogeneity. Our proof is constructive and it provides a simple closed form expression for this AME in terms of probabilities of choice histories in panels where the time dimension can be as small as $T = 4$.

We extend this identification result to more general models and to other AME parameters. First, we show the identification of the AME n periods after the change in the dependent variable, where n can be between 1 and the number of periods in the data minus two. We

[1] Examples 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 $T \to \infty$."*) and Honoré and DePaula (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."*)

denote this parameter the *nperiods forward AME*. This sequence of AMEs provides the *impulse response function* associated to an exogenous change in the dependent variable. Second, we show this identification also holds in dynamic models that include strictly exogenous explanatory variables. Third, we consider a more general dynamic discrete choice model with duration dependence and prove the identification of AMEs where duration is the causal variable. All these identification results of AMEs apply to dynamic multinomial models. We obtain very simple analytical expressions for the AMEs in terms of probabilities of choice histories.

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 the identification and estimation of slope parameters and do not present identification results on AMEs. Two important exceptions of studies that deal with the identification of AMEs in FE models are Chernozhukov, Fernandez-Val, Hahn, and Newey (2013; hereinafter CFHN), and Honoré and Kyriazidou (2019).

CFHN (2013) study the identification of AMEs in nonparametric and semiparametric binary choice models. In their nonparametric model, the distributions of all the unobservables – the time-invariant and the transitory shock – is nonparametric. Their semiparametric model – that corresponds to the model that we consider in this paper – assumes that the transitory shock has a known distribution – e.g., FE dynamic probit and logit models. They propose a computational method to estimate the bounds in the identified set of the AME. Using numerical examples, they find that the bounds for the AME can be very wide for the fully nonparametric model, but that these bounds shrink fast with T in the semiparametric model.

In contrast to CFHN, we consider a sequential identification approach.[2] Given a sufficient statistic for the unobserved heterogeneity (or incidental parameters, α), the log-likelihood function can be written as the sum of two functions: the log-likelihood of the data conditional on the sufficient statistic, $L_1(\beta)$, that depends on the parameters of interest β but not on the incidental parameters α; and the log-likelihood of the sufficient statistic, $L_2(\alpha,\beta)$, that depends on the two sets of parameters. Previous results on conditional likelihood estimation of dynamic logit models establish the identification of slope parameters based on the maximization of the (conditional) log-likelihood $L_1(\beta)$ (Chamberlain, 1985; Honoré and Kyriazidou, 2000; Aguirregabiria, Gu, and Luo, 2021). This is the first stage of our sequential approach. In the second stage, we take the β parameters as known to the researcher and consider the identification of AMEs. These AMEs are defined as functions of the slope parameters β and the distribution of the unobserved het-

[2] The sequential approach that we consider in this paper has been also recently suggested by Honoré and DePaula (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."*

erogeneity. The log-likelihood of the sufficient statistic, $L_2(\alpha,\beta)$, contains all the information in the data about the distribution of the unobserved heterogeneity α. Therefore, given β, the empirical distribution of the sufficient statistic contains all the information about the AMEs. These statistics take a finite number of values such that the information about the AME can be represented using a system of equations. We show that simple manipulations in this system of equations provide a closed form expression for AMEs of interest. While the approach in CFHN is computationally demanding due to the very large dimensionality of the distribution of the unobserved heterogeneity – in fact, it has infinite dimension in FE models – our approach is computationally very simple as it provides closed form expressions for AMEs.[3]

Honoré and Kyriazidou also propose a numerical approach to construct bounds for structural parameters and marginal effects in FE dynamic binary choice models. Their approach is based on Honoré and Tamer (2006). In their numerical examples, they obtain tight bounds for slope parameters, showing that the identification of these parameters can be more general than the existing results of point identification in the literature. Though similar results may apply to AMEs, their numerical exercises do not provide evidence on these parameters.

Chamberlain (1984), Hahn (2001), and more recently Arellano and Bonhomme (2017), show the identification of some AMEs in FE nonlinear panel data models. These are AMEs for a particular subpopulation of individuals defined by the data. In contrast, we focus on the identification of marginal effects that are averaged over the whole population of individuals. As far as we know, the point identification of this type of AMEs has not been previously established in FE dynamic discrete choice models.

As mentioned above, the first step of our sequential identification approach consists in the conditional maximum likelihood estimation of slope parameters. To this respect, it is worth to mention recent results by Honoré and Weidner (2020) and Dobronyi, Gu, and Kim (2021) showing that, in some FE dynamic binary choice models, the conditional likelihood equations do not contain all the identification restrictions on the slope parameters. There are additional moment equalities and inequalities with information on these parameters.

The rest of the paper is organized as follows. Section 2 describes the models and the AMEs of interest. Section 3 presents our main identification results. We illustrate our results using Monte Carlo experiments (in section 4) and an empirical application to a model of dynamic demand using consumer scanner data (in section 5). We summarize and conclude in section 6.

[3]It is important to note that CFHN approach can be used for any AME, while we have shown point identification of some AMEs. However, given the computational complexity, all their numerical examples and empirical illustrations deal with models with only one binary regressor.

2 Model and Average Marginal Effects

2.1 Model

Consider a panel dataset $\{y_{it}, \mathbf{x}_{it} : i = 1, 2, ..., N; t = 1, 2, ..., T\}$ where y_{it} can take $J + 1$ values: $y_{it} \in \mathcal{Y} = \{0, 1, ..., J\}$. We study panel data dynamic logit models. In these models, the dependent variable can be represented as the choice alternative that maximizes a utility or payoff function. That is,

$$y_{it} = \arg\max_{j \in \mathcal{Y}} \left\{ \alpha_i(j) + \sum_{k=0}^{J} \beta_{kj}(d_{it})\ 1\{y_{i,t-1} = k\} + \mathbf{x}_{it}'\ \gamma_j + \varepsilon_{it}(j) \right\}. \tag{1}$$

where $\{\beta_{kj}(d) : k, j \in \mathcal{Y},\ d = 1, 2, ...T\}$ and $\{\gamma_j : j \in \mathcal{Y}\}$ are parameters of interest, and $\alpha_i \equiv \{\alpha_i(j) : j \in \mathcal{Y}\}$ are incidental parameters. The unobservables $\{\varepsilon_{it}(j) : j \in \mathcal{Y}\}$ are i.i.d. type 1 extreme value. Variable $d_{it} \in \{0, 1, 2, ...\}$ represents the duration in the choice at period $t-1$. More formally, $d_{it} = 1\{y_{i,t-1} = y_{i,t-2}\}\ d_{it,-1} + 1$. The explanatory variables in the $K \times 1$ vector $\mathbf{x}_{it}$ are strictly exogenous with respect to the transitory shocks $\varepsilon_{it}(j)$: that is, for any pair of time periods (t,s), variables $\mathbf{x}_{it}$ and ε_{is} are independently distributed.

Parameters $\beta_{kj}(d)$ represent the change in utility associated to switching from alternative k to alternative j given that the agent has been choosing k during the last d periods. This switching cost may vary with the duration in the last choice, such that the parameters $\beta_j(1)$, $\beta_j(2)$, ... can be different. Identification of the β parameters requires some normalization conditions, for instance, $\beta_{jj}(d) = 0$ and $\beta_{j0}(d) = 0$ for any $j \in \mathcal{Y}$ and any d.

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 may increase with job experience and this implies that the probability of employment increases with the duration in that state. Similarly, in a model of firm market entry/exit (where $y = 1$ means a firm is active in the market and $y = 0$ inactive), a firm's profit may increase with its experience in the market.

The vector α_i represents (permanent) unobserved individual heterogeneity in preferences or payoffs. The marginal distribution of α_i is $f_\alpha(\alpha_i)$, and $f_{\alpha|x}(\alpha_i|\mathbf{x}_i^{\{1,T\}})$ is the distribution of α_i conditional on the history of $\mathbf{x}$ variables $\mathbf{x}_i^{\{1,T\}} = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$. These distributions are unrestricted. Similarly, the probability of the initial values (y_{i1}, d_{i1}) conditional on α_i and $\mathbf{x}_i^{\{1,T\}}$ – that we represent as $p^*(y_{i1}, d_{i1}|\alpha_i, \mathbf{x}_i^{\{1,T\}})$ – is unrestricted. Following the standard setting in fixed effect (FE) approaches, our identification results are not based on any restriction on the initial conditions.[4]

Assumption 1 summarizes the conditions in this model.

ASSUMPTION 1: (A) (Logit) $\varepsilon_{it}(j)$ is i.i.d. over (i,t,j) with type 1 extreme value distribution, and is independent of α_i; (B) (Strict exogeneity of $\mathbf{x}_{it}$) for any two periods, t and s, the variables $\varepsilon_{it}(j)$ and $\mathbf{x}_{is}$ are independently distributed; and (C) (Fixed effects) the probability density functions $f_\alpha(\alpha_i)$ and $f_{\alpha|x}(\alpha_i|\mathbf{x}_i^{\{1,T\}})$, and the probability of the initial condition $p^*(y_{i1}, d_{i1}|\alpha_i, \mathbf{x}_i^{\{1,T\}})$ are unrestricted. ■

The form of our identification results vary across different versions of the general model in equation (1). We focus on four models.

(1) Model MNL-AR1. Multinomial choice AR1 model without duration dependence: that is, $\beta_{kj}(d) = \beta_{kj}$ for every value of d.

$$y_{it} = \arg\max_{j\in\mathcal{Y}} \left\{ \alpha_i(j) + \sum_{k=0}^{J} \beta_{kj} \ 1\{y_{i,t-1} = k\} + \mathbf{x}_{it}' \ \gamma_j + \varepsilon_{it}(j) \right\}. \tag{2}$$

(2) Model BC-Dur. Binary choice model ($J + 1 = 1$) with duration dependence in $y = 1$ but not in $y = 0$ (that is, $\beta_{00}(d) = \beta_{00}$ and $\beta_{01}(d) = \beta_{01}$ for every value of d), and without $\mathbf{x}$ variables.

$$y_{it} \quad = \quad 1\{ \ \alpha_i + \beta(d_{it}) \ y_{i,t-1} + \varepsilon_{it} \geq 0\} \tag{3}$$

It is straightforward to verify the following relationship between the parameters in this model and those in the original model in equation (1): $\alpha_i = \alpha_i(1) - \alpha_i(0) + \beta_{01} - \beta_{00}$; and $\beta(d) = \beta_{11}(d) - \beta_{10}(d) - \beta_{01} + \beta_{00}$. Furthemore, we have that $\varepsilon_{it} = \varepsilon_{it}(1) - \varepsilon_{it}(0)$.

(3) Model BC-AR1-X. Binary choice model without duration dependence but with $\mathbf{x}$ explanatory variables.

$$y_{it} = 1\{ \ \alpha_i + \beta \ y_{i,t-1} + \mathbf{x}_{it}' \ \gamma + \varepsilon_{it} \geq 0\}. \tag{4}$$

The relationship between the parameters in this model and those in equation (1) is $\beta = \beta_{11} - \beta_{10} - \beta_{01} + \beta_{00}$, and $\gamma = \gamma_1 - \gamma_0$.

(4) Model BC-AR1. Binary choice, without duration dependence, and without $\mathbf{x}$ variables.

$$y_{it} = 1\{ \ \alpha_i + \beta \ y_{i,t-1} + \varepsilon_{it} \geq 0\}. \tag{5}$$

[4] For instance, some previous studies using random effects models assume that initial choice y_{i1} is a random random draw from the individual-specific steady-state distribution of the endogenous variable.

2.2 Average Marginal Effects (AME)

2.2.1 Average transition probabilities

For the definition of the AMEs and other parameters of interest, it is convenient to define transition probabilities and their average versions. For $j, k \in \mathcal{Y}$, define the individual-specific transition probabilities:

$$\pi_{kj}(\alpha_i, \mathbf{x}, d) \equiv \mathbb{P}\left(y_{i,t+1} = j \mid \alpha_i,\ y_{it} = k,\ \mathbf{x}_{i,t+1} = \mathbf{x},\ d_{it} = d\right) \tag{6}$$

For instance, in the binary choice version of the model, $\pi_{11}(\alpha_i, \mathbf{x}, d) = \Lambda(\alpha_i + \beta(d) + \mathbf{x}'\gamma)$, where $\Lambda(u)$ is the Logistic function $e^u/[1+e^u]$. Similarly, we use $\pi_{kj}(\alpha_i)$ to represent these transition probabilities in models without $\mathbf{x}$ variables and duration.

We define $\Pi_{kj}(\mathbf{x}, d)$ as the average transition probability from k to j that results from integrating the individual-specific transition probability over the distribution of α_i conditional on $\mathbf{x}_i^{\{1,T\}} = (\mathbf{x}, ..., \mathbf{x})$. That is,

$$\Pi_{kj}(\mathbf{x}, d) \equiv \int \pi_{kj}(\alpha_i, \mathbf{x}, d)\ f_\alpha(\alpha_i \mid \mathbf{x}_i^{\{1,T\}} = [\mathbf{x}, ..., \mathbf{x}])\ d\alpha_i \tag{7}$$

Similarly, for models without $\mathbf{x}$ variables and duration, we use Π_{kj} to represent the average transition probability $\int \pi_{kj}(\alpha_i)\ f_\alpha(\alpha_i)\ d\alpha_i$.

For the model without duration, We can extend these definitions to $n - periods$ forward transition probabilities. That is, for any integer $n \geq 1$, we define $\pi_{kj}^{(n)}(\alpha_i, \mathbf{x}) \equiv \mathbb{P}(y_{i,t+n} = j|\alpha_i, \mathbf{x}_{i,t+n} = \mathbf{x}, y_{it} = k)$, and its average $\Pi_{kj}^{(n)}(\mathbf{x}) \equiv \int \pi_{kj}^{(n)}(\alpha_i, \mathbf{x})\ f_\alpha(\alpha_i|\mathbf{x}_i^{\{1,T\}} = (\mathbf{x}, ..., \mathbf{x}))\ d\alpha_i$.

2.2.2 One-period forward AME - Binary choice, no duration, no x's

We start with a simple AME that is very commonly used in empirical applications. Consider the *BC-AR1* model in equation (5). Let $\Delta^{(1)}(\alpha_i)$ be the *individual specific causal effect* on y_{it} of a change in variable y_{it-1} from 0 to 1. That is:

$$\begin{aligned}\Delta^{(1)}(\alpha_i) &\equiv \mathbb{E}\left(y_{it} \mid \alpha_i,\ y_{it-1} = 1\right) - \mathbb{E}\left(y_{it} \mid \alpha_i,\ y_{it-1} = 0\right) \\ &= \pi_{11}(\alpha_i) - \pi_{01}(\alpha_i) = \Lambda(\alpha_i + \beta) - \Lambda(\alpha_i).\end{aligned} \tag{8}$$

This parameter measures the persistence of individual i in state 1 that is generated by true state dependence. It is also an individual-specific treatment (causal) effect.

Using a short panel, parameter β is identified (Chamberlain, 1985; Honoré and Kyriazidou, 2000), but the individual effects α_i are not identified because the incidental parameters problem (Neyman and Scott, 1948; Heckman, 1981; Lancaster, 2000). That is, we can identify $\Delta^{(1)}(\alpha)$ for an hypothetical value of α, but not for the value of α that actually corresponds to individual i. Therefore, the individual-specific treatment effects $\Delta^{(1)}(\alpha_i)$ are not identified. Instead, we study the identification of the following *Average Marginal Effect (AME):*

$$\begin{aligned} AME^{(1)} &\equiv \int \Delta^{(1)}(\alpha_i)\ f_\alpha(\alpha_i)\ d\alpha_i \\ &= \int \left[\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)\right] f_\alpha(\alpha_i)\ d\alpha_i = \Pi_{11} - \Pi_{01}. \end{aligned} \tag{9}$$

The sign of the parameter β tell us the sign of $AME^{(1)}$. However, the absolute magnitude of β provides basically no information about the magnitude of $AME^{(1)}$. For instance, given any positive value β, we have that $AME^{(1)}$ can take any value within the interval $(0, 1)$ depending on the location of the distribution of α_i. This is why the identification of AMEs is so important.

EXAMPLE 1. Consider a model of market entry, where y_{it} is the indicator that firm i is active in the market at period t. Let $V_{it}(1, y_{i,t-1})$ and $V_{it}(0, y_{i,t-1})$ represent firm i's value if active and inactive, respectively. Firms make choices to maximize their value such that firm i chooses to be active if $V_{it}(1, y_{i,t-1}) - V_{it}(0, y_{i,t-1}) \geq 0$, or equivalently, if $V_{it}(1, 0) - V_{it}(0, 0) +\ y_{i,t-1}\ [V_{it}(1, 1)$ $-V_{it}(1, 0)\ -V_{it}(0, 1) + V_{it}(0, 0)] \geq 0$. Our model imposes the restriction that $V_{it}(1, 0) - V_{it}(0, 0) =$ $\alpha_i + \varepsilon_{it}$ and $V_{it}(1, 1)\ -V_{it}(1, 0)\ -V_{it}(0, 1) -\ V_{it}(0, 0) = \beta$. Therefore, parameter β captures the complementarity (or supermodularity) in the value function between the decisions of being active at periods t and $t-1$. It captures true state dependence in market entry and it can interpreted as a sunk entry cost. However, this parameter by itself does not give us a treatment effect or causal effect. Consider the following thought experiment. Suppose that we could split firms randomly in two groups, say groups 0 and 1. Firms in group 0 are assigned to be inactive in the market, and firms in group 1 are assigned to be active. Then, after one period we look at the proportion of firms who are active in the market in each of the two groups. $AME^{(1)}$ is equal to the proportion of active firms in group 1 minus the proportion of active firms in group 0. ■

The parameter $AME^{(1)}$ is also related to the average treatment effects ($ATEs$) from two policy experiments with economic interest. For concreteness, we describe these policy experiments and their corresponding $ATEs$ using the application in Example 1. Consider a policy experiment where firms in the experimental group are assigned to 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 group and in the control group. The difference between these two proportions is the average effect of this policy treatment, that we can denote as $ATE_{11,t}$. According to the model, this average treatment effect has the following form:

$$ATE_{11,t} \quad \equiv \quad \int \pi_{11}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i - \mathbb{E}(y_{it}|t) = \Pi_{11} - \mathbb{E}(y_{it}|t) \tag{10}$$

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 but where firms in the experimental group are assigned to be inactive at period $t-1$ – e.g., 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,

$$ATE_{01,t} \quad \equiv \quad \int \pi_{01}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i - \mathbb{E}(y_{it}|t) = \Pi_{01} - \mathbb{E}(y_{it}|t) \tag{11}$$

Given the definitions of $AME^{(1)}$, $ATE_{11,t}$, and $ATE_{01,t}$ in equations (9), (10), and (11), respectively, it is clear that $AME^{(1)} = ATE_{11,t} - ATE_{01,t}$.

In section 3.3, we show the identification of the parameters Π_{01} and Π_{11}. This implies the identification $ATE_{01,t}$, $ATE_{11,t}$, and $AME^{(1)}$. Knowledge of Π_{01} and Π_{11} also implies the identification of other relevant causal effects, such as the ratio Π_{11}/Π_{01}, the percentage change $(\Pi_{11} - \Pi_{01})/\Pi_{01}$ (as long as $\Pi_{01} \neq 0$), the additive effect $\Pi_{01} + \Pi_{11}$, a weighted sum of Π_{01} and Π_{11}, or more generally, any known function of these parameters.

2.2.3 n-periods forward AME - Binary choice, no duration, no x's

Researchers can be interested in the response to a treatment after more than one period. Let $\Delta^{(n)}(\alpha_i)$ be the individual-specific causal effect on $y_{i,t+n}$ of a change in y_{it} from 0 to 1.

$$\Delta^{(n)}(\alpha_i) \equiv \mathbb{E}\left(y_{i,t+n} \mid \alpha_i, y_{it} = 1\right) - \mathbb{E}\left(y_{i,t+n} \mid \alpha_i, y_{it} = 0\right) = \pi_{11}^{(n)}(\alpha_i) - \pi_{01}^{(n)}(\alpha_i). \tag{12}$$

Similarly as discussed above for $\Delta^{(1)}(\alpha_i)$, this $n-$periods forward individual effect is not identified using a short-panel. We are interested in the average of this effect:

$$AME^{(n)} \equiv \int \Delta^{(n)}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i = \int \left[\pi_{11}^{(n)}(\alpha_i) - \pi_{01}^{(n)}(\alpha_i)\right] f_\alpha(\alpha_i) \ d\alpha_i = \Pi_{11}^{(n)} - \Pi_{01}^{(n)}. \tag{13}$$

In general, this n-periods forward AME is different to the product of n times the 1-period forward AME: that is, $AME^{(n)} \neq [AME^{(1)}]^n$, such that the identification of $AME^{(n)}$ is not a simple corollary that follows from the identification of $AME^{(1)}$.

2.2.4 One-period forward AME - Binary choice, model with x's

Consider the binary choice model with exogenous explanatory variables (*BC-AR1-X*) as described in equation (4). In this model, the AME of the effect of y_{it-1} on y_{it} has to take into account the presence of $\mathbf{x}$ and its correlation with α_i. Let $\Delta^{(1)}(\alpha_i, \mathbf{x})$ be the individual-specific causal effect on y_{it} of a change in variable y_{it-1} from 0 to 1 when $\mathbf{x}_{it} = \mathbf{x}$.

$$
\begin{aligned}
\Delta^{(1)}(\alpha_i, \mathbf{x}) &\equiv \mathbb{E}\left(y_{it} \mid \alpha_i, y_{it-1} = 1, \mathbf{x}_{it} = \mathbf{x}\right) - \mathbb{E}\left(y_{it} \mid \alpha_i, y_{it-1} = 0, \mathbf{x}_{it} = \mathbf{x}\right) \\
&= \pi_{11}(\alpha_i, \mathbf{x}) - \pi_{01}(\alpha_i, \mathbf{x}).
\end{aligned} \tag{14}
$$

This individual-specific marginal effect is not identified using short panels. In section 3.6, we show the identification results of two different average versions of this effect. A first AME is based on the condition that $\mathbf{x}$ remains constant over the T sample periods:

$$
\begin{aligned}
AME^{(1)}(\mathbf{x}) &\equiv \int \left[\pi_{11}(\alpha_i, \mathbf{x}) - \pi_{01}(\alpha_i, \mathbf{x})\right] \; f_{\alpha|\mathbf{x}}(\alpha_i | \mathbf{x}_i^{\{1,T\}} = (\mathbf{x}, ..., \mathbf{x})) \; d\alpha_i \\
&= \Pi_{11}(\mathbf{x}) - \Pi_{01}(\mathbf{x})
\end{aligned} \tag{15}
$$

A second AME is defined as follows:

$$
AME_{x,t}^{(1)} \equiv \int \left[\pi_{11}(\alpha_i, \mathbf{x}_{it}) - \pi_{01}(\alpha_i, \mathbf{x}_{it})\right] \; f_{(\alpha, \mathbf{x}_t)}(\alpha_i, \mathbf{x}_{it}) \; d(\alpha_i, \mathbf{x}_{it}) \tag{16}
$$

This second AME is not conditional to a value of $\mathbf{x}$ but integrated over the joint distribution of α_i and $\mathbf{x}_{it}$ at period t. Chamberlain (1984) describes this AME as the expected causal effect for an individual *randomly drawn* from the distribution of $(\alpha_i, \mathbf{x}_{it})$ at period t. Since this distribution can change over time, these AMEs can vary over time.[5]

EXAMPLE 2. Consider the model of market entry/exit in Example 1, but now we extend this model to include an exogenous explanatory variable x_{it} that represents the population size of the market where the firm considers entry/exit. Then, $AME_x^{(1)}(x)$ represents the average effect on a firm's entry status at period t of going (exogenously) from inactive to active at $t-1$, and for the subpopulation of markets with size x over the T sample periods. The parameter $AME_{x,t}^{(1)}$ is a similar effect but averaged over all the markets (firms) according to their distribution of

[5] Given that we show the identification of these time-specific AMEs at every sample period, they can be used to test the null hypothesis of stationarity of the distribution of $(\alpha_i, \mathbf{x}_{it})$.

population size at period t. ■

Similarly as for the AR1 model, we are also interested in n-periods forward AMEs for this AR1X model. The AME conditional on a constant value of $\mathbf{x}$ is:

$$AME_x^{(n)}(\mathbf{x}) \quad = \quad \int [\pi_{11}(\alpha_i, \mathbf{x}) - \pi_{01}(\alpha_i, \mathbf{x})]^n \;\; f_{\alpha|\mathbf{x}}(\alpha_i | \mathbf{x}_i^{\{1,T\}} = (\mathbf{x}, ..., \mathbf{x})) \; d\alpha_i \tag{17}$$

2.2.5 AME of a change in duration - Binary choice

Consider the binary choice model with duration dependence (*BC-Dur*) as described in equation (3). We are interested in the causal effect on y_{it} of a change in the duration variable d_{it}. For instance, in a model of firm market entry/exit, we can be interested on the causal effect of one more year of experience on the probability of being active in the market.

Let $\Delta_{d \to d'}(\alpha_i)$ be the *individual-specific* causal effect on y_{it} of a change in d_{it} from d to d'.

$$\begin{aligned} \Delta_{d \to d'}(\alpha_i) \;\; &\equiv \;\; \mathbb{E}\left(y_{it} \mid \alpha_i, d_{it} = d'\right) - \mathbb{E}\left(y_{it} \mid \alpha_i, d_{it} = d\right) \\ &= \;\; \pi_{d',1}(\alpha_i) - \pi_{d,1}(\alpha_i) = \Lambda(\alpha_i + \beta(d')) - \Lambda(\alpha_i + \beta(d)). \end{aligned} \tag{18}$$

where $\pi_{d,1}(\alpha_i) \equiv \mathbb{E}(y_{it} | \alpha_i, d_{it} = d)$. Note that, given the definition of the duration variable d_{it}, we have that $d_{it} = d > 0$ implies $y_{i,t-1} = 1$ and $d_{it} = 0$ implies $y_{i,t-1} = 0$, such that we do not need to include explicitly $y_{i,t-1}$ as a conditioning variable in these expectations. We are interested in the identification of the following *AME:*

$$AME_{d \to d'} \;\; \equiv \;\; \int \Delta_{d \to d'}(\alpha_i) \; f_\alpha(\alpha_i) \; d\alpha_i = \int [\pi_{d',1}(\alpha_i) - \pi_{d,1}(\alpha_i)] \;\; f_\alpha(\alpha_i) \; d\alpha_i \tag{19}$$

2.2.6 AMEs in the multinomial choice model

Consider the multinomial model *MNL-AR1* in equation (2). Let $\Delta_{j,k \to j}(\alpha_i)$ be the *individual-specific* causal effect on the probability of $y_{it} = j$ of a change in $y_{i,t-1}$ from k to j.

$$\Delta_{j,k \to j}(\alpha_i) \equiv \mathbb{E}\left(1\{y_{it} = j\} | \alpha_i, y_{i,t-1} = j\right) - \mathbb{E}\left(1\{y_{it} = j\} | \alpha_i, y_{i,t-1} = k\right) = \pi_{jj}(\alpha_i) - \pi_{kj}(\alpha_i). \tag{20}$$

We are interested in identification of the following AMEs:

$$AME_{j,k \to j} \equiv \int \Delta_{j,k \to j}(\alpha_i) \; f_\alpha(\alpha_i) \; d\alpha_i = \Pi_{jj} - \Pi_{kj} \tag{21}$$

and

$$ATE_{jj,t} = \Pi_{jj} - \mathbb{E}\left(1\{y_{it} = j\} \mid t\right) \tag{22}$$

3 Identification

We start, in section 3.1, reviewing existing results on the identification of the slope parameters β and γ using the FE - conditional maximum likelihood (CML) approach. Then, we take these slope parameters as known and study the identification of AMEs. In section 3.2, we establish the identification of average transition probabilities Π_{jj}. We show in section 3.3 that knowledge of these average transitions implies the identification of one-period AMEs, $AME^{(1)}$, $ATE^{(1)}_{11,t}$, and $ATE^{(1)}_{01,t}$ in binary choice models, and $ATE^{(1)}_{jj,t}$ in multinomial models. In section 3.4, we show that these identification results extend to n-periods forward AMEs. In section 3.5 we present a more general approach to show the identification of AMEs in this class of models. We apply this approach to establish the identification of $AME^{(1)}_{x,t}$ (in section 3.6) and of AMEs of changes in duration (in section 3.7).

3.1 Identification of slope parameters β and γ

Here we review previous identification results of slope parameters using a FE sufficient statistics approach. For the binary choice model without exogenous regressors or duration dependence (model *BC-AR1*), the identification of the parameter β when $T \geq 4$ has been proved in Chamberlain (1985). Honoré and Kyriazidou (2000) show the identification of the parameters β and γ in binary and multinomial choice models with exogenous regressors. Their identification results requires the vector of random variables $\mathbf{x}_{it} - \mathbf{x}_{i,t-1}$ to have support in a neighborhood of zero in $\mathbb{R}^K$.[6] Aguirregabiria, Gu, and Luo (2021) establish the identification of the duration-dependence parameters $\beta_{kj}(d)$ (see Proposition 3 in that paper).

Let $\mathbf{y}_i^{\{1,T\}} \equiv (y_{i1}, y_{i2}, ..., y_{iT})$ be the vector with the choice history of individual i, and let $\mathbf{x}_i^{\{1,T\}}$ represent $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$. The vector $\alpha \equiv (\alpha_1, \alpha_2, ..., \alpha_N)$ contains the *fixed effects* or *incidental parameters* for the N individuals. The vector $\theta \equiv (\beta, \gamma)$ contains the parameters of interest. The log-likelihood function where α's are treated as parameters is:

$$\ell\left(\alpha, \theta\right) = \sum_{i=1}^{N} \ln \mathbb{P}\left(\mathbf{y}_i^{\{1,T\}} | \mathbf{x}_i^{\{1,T\}}, \alpha_i, \theta\right) = \sum_{i=1}^{N} \ln \left[\prod_{t=2}^{T} \pi_{y_{i,t-1},y_{it}}\left(\alpha_i, \mathbf{x}_{it}, \theta\right) \; p^*(y_{i1}, d_{i1}|\alpha_i)\right] \tag{23}$$

where $p^*(y_{i1}, d_{i1}|\alpha_i)$ is the probability of the initial condition given α_i.

For the general model in equation (1), the log-probability of a choice history has the following structure:

$$\ln \mathbb{P}\left(\mathbf{y}_i^{\{1,T\}} | \mathbf{x}_i^{\{1,T\}}, \alpha_i, \theta\right) = \mathbf{s}(\mathbf{y}_i^{\{1,T\}})' \; \mathbf{g}(\alpha_i) + \mathbf{c}(\mathbf{y}_i^{\{1,T\}})' \; \theta \tag{24}$$

[6] This condition rules out, for instance, time dummies as explanatory variables in $\mathbf{x}_{it}$

where $\mathbf{s}(\mathbf{y}_i^{\{1,T\}})$ and $\mathbf{c}(\mathbf{y}_i^{\{1,T\}})$ are vectors of statistics (functions of $\mathbf{y}_i^{\{1,T\}}$), and $\mathbf{g}(\alpha_i)$ is a vector of functions α_i. For notational simplicity, we use $\mathbf{s}_i$ and $\mathbf{c}_i$ to represent $\mathbf{s}(\mathbf{y}_i^{\{1,T\}})$ and $c(\mathbf{y}_i^{\{1,T\}})$, respectively. Given this structure of the log-probability of a choice history, it is simple to show that $\mathbf{s}_i$ is a *minimal sufficient statistic* for α_i. That is, $\mathbf{s}_i$ is a *sufficient statistic/* for α_i such that $\mathbb{P}(\mathbf{y}_i^{\{1,T\}}|\mathbf{x}_i^{\{1,T\}}, \alpha_i, \theta, \mathbf{s_i}) = \mathbb{P}(\mathbf{y}_i^{\{1,T\}}|\mathbf{x}_i^{\{1,T\}}, \theta, \mathbf{s_i})$; and $\mathbf{s}_i$ is *minimal* because its elements are linearly independent. More precisely, we have that:

$$
\begin{aligned}
\mathbb{P}(\mathbf{y}_i^{\{1,T\}}|\mathbf{x}_i^{\{1,T\}}, \alpha_i, \theta, \mathbf{s_i}) &= \frac{\mathbb{P}(\mathbf{y}_i^{\{1,T\}}|\mathbf{x}_i^{\{1,T\}}, \alpha_i, \theta)}{\mathbb{P}(\mathbf{s_i}|\mathbf{x}_i^{\{1,T\}}, \alpha_i, \theta)} = \frac{\exp\left\{\mathbf{s}_i' \ \mathbf{g}(\alpha_i) + \mathbf{c}_i' \ \theta\right\}}{\sum\limits_{\mathbf{y}:\mathbf{s}(\mathbf{y})=\mathbf{s}_i} \exp\left\{\mathbf{s}_i' \ \mathbf{g}(\alpha_i) + \mathbf{c}(\mathbf{y})' \ \theta\right\}} \\
&= \frac{\exp\left\{\mathbf{c}_i'\theta\right\}}{\sum\limits_{\mathbf{y}:\mathbf{s}(\mathbf{y})=\mathbf{s}_i} \exp\left\{\mathbf{c}(\mathbf{y})'\theta\right\}}
\end{aligned}
\tag{25}
$$

where $\sum_{\mathbf{y}:\ \mathbf{s}(\mathbf{y})=\mathbf{s}_i}$ represents the sum over all the possible choice histories $\mathbf{y}$ with $\mathbf{s}(\mathbf{y})$ equal to $\mathbf{s}_i$. Furthermore, when $T \geq 4$, the vectors of statistics $\mathbf{c}_i$ and $\mathbf{s}_i$ are linearly independent.

This result implies that we can (point) identify θ using a Conditional Maximum Likelihood (CML) approach. We can write the log-likelihood function as the sum of two likelihoods:

$$
\ell\left(\alpha, \theta\right) = \ell^{C}\left(\mathbf{C}, \mathbf{S}\ ; \theta\right) + \ell^{S}\left(\mathbf{S}\ ; \alpha, \theta\right) \tag{26}
$$

with $\mathbf{C} = \{\mathbf{c}_i : i = 1, 2, ..., N\}$ and $\mathbf{S} = \{\mathbf{s}_i : i = 1, 2, ..., N\}$, and where

$$
\ell^{C}\left(\mathbf{C}, \mathbf{S}\ ; \theta\right) = \sum_{i=1}^{N} \ln \mathbb{P}(\mathbf{y}_i^{\{1,T\}}|\mathbf{x}_i^{\{1,T\}}, \theta, \mathbf{s_i}) = \sum_{i=1}^{N} \mathbf{c}_i'\theta - \sum_{i=1}^{N} \ln\left[\sum_{\mathbf{y}:\ \mathbf{s}(\mathbf{y})=\mathbf{s}_i} \exp\left\{\mathbf{c}(\mathbf{y})'\theta\right\}\right], \tag{27}
$$

and

$$
\ell^{S}\left(\mathbf{S}\ ; \alpha, \theta\right) = \sum_{i=1}^{N} \ln \mathbb{P}(\mathbf{s_i}|\mathbf{x}_i^{\{1,T\}}, \alpha_i, \theta) = \sum_{i=1}^{N} \ln\left[\sum_{\mathbf{y}:\mathbf{s}(\mathbf{y})=\mathbf{s}_i} \exp\left\{\mathbf{s}_i' \ \mathbf{g}(\alpha_i) + \mathbf{c}(\mathbf{y})' \ \theta\right\}\right]. \tag{28}
$$

Function $\ell^{C}\left(\mathbf{C}, \mathbf{S}\ ; \theta\right)$ is the *conditional log-likelihood function*. It has two important properties: it does not depend on the incidental parameters α; and the maximization of this function with respect to θ uniquely identifies these parameters of interest. In fact, the (sample) conditional log-likelihood function is globally concave in θ. Function $\ell^{S}\left(\mathbf{S}\ ; \alpha, \theta\right)$ is the likelihood for the sufficient statistic $\mathbf{s}_i$. All the information in the sample about the incidental parameters appears in this function. Note that it depends on the data only through the sufficient statistics

$\mathbf{s}_i$. Therefore, given θ, all the information in the data about the incidental parameters appears in the empirical distribution of the sufficient statistic $\mathbf{s}_i$. This result plays an important role in our identification of the AMEs. For the rest of this section 3 we treat the parameters of interest θ as known to the researcher.

3.2 Identification of average transition probabilities

Consider the general model in equation (1). Lemma 1 establishes a relationship between switching cost parameters $\beta_{kj}(d)$ and individual-specific transition probabilities that plays a key role in our identification results.

Lemma 1. *In the model defined by equation (1) and assumption 1, for any triple of choice alternatives j, k, ℓ (not necessarily all different, but with $j \neq k$ and $j \neq \ell$, the following condition holds:*

$$\exp\left\{\beta_{k\ell}(d) - \beta_{kj}(d) + \beta_{jj}(d) - \beta_{j\ell}(d)\right\} = \frac{\pi_{k\ell}(\alpha_i, \mathbf{x}, d)\ \pi_{jj}(\alpha_i, \mathbf{x}, d)}{\pi_{kj}(\alpha_i, \mathbf{x}, d)\ \pi_{j\ell}(\alpha_i, \mathbf{x}, d)}. \quad \blacksquare \tag{29}$$

Proof. Given the expression for the choice probabilities in the logit model, it is simple to verify that $\pi_{k\ell}(\alpha_i, \mathbf{x}, d)/\pi_{kj}(\alpha_i, \mathbf{x}, d) = \exp\{\alpha_i(\ell) - \alpha_i(j) + \beta_{k\ell}(d) - \beta_{kj}(d)\}$, and similarly, $\pi_{jj}(\alpha_i, \mathbf{x}, d)/\pi_{j\ell}(\alpha_i, \mathbf{x}, d) = \exp\{\alpha_i(j) - \alpha_i(\ell) + \beta_{jj}(d) - \beta_{j\ell}(d)\}$. The product of these two expressions is equation (29). $\square$

Proposition 1 establishes the identification of the average transition probabilities $\Pi_{jj}(\mathbf{x})$ in the logit model without duration dependence.

Proposition 1. *Consider the model without duration dependence in equation (2) under Assumption 1. If $T \geq 3$, the average transition probabilities $\{\Pi_{jj}(\mathbf{x}) : j \in \mathcal{Y}\}$ are identified using the following equation,*

$$\Pi_{jj}(\mathbf{x}) \;=\; \mathbb{P}_{j,j}(\mathbf{x}) + \sum_{k \neq j} \left[\mathbb{P}_{k,j,j}(\mathbf{x}) + \sum_{\ell \neq j} \exp\left\{\beta_{k\ell} - \beta_{kj} + \beta_{jj} - \beta_{j\ell}\right\}\ \mathbb{P}_{k,j,\ell}(\mathbf{x}) \right], \tag{30}$$

where $\mathbb{P}_{y_1,y_2,y_3}(\mathbf{x})$ and $\mathbb{P}_{y_1,y_2}(\mathbf{x})$ represent the probability of choice histories $(y_{i1}, y_{i2}, y_{i3}) = (y_1, y_2, y_3)$ and $(y_{i1}, y_{i2}) = (y_1, y_2)$, respectively, conditional on $\mathbf{x}_i^{\{1,3\}} = (\mathbf{x}, \mathbf{x}, \mathbf{x})$. $\blacksquare$

Proof. For notational simplicity, we omit $\mathbf{x}$ as an argument throughout this proof. However, it should be understood 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:

$$\Pi_{jj} = \int [p^*(0|\alpha_i) + p^*(1|\alpha_i) ... + p^*(J|\alpha_i)] \ \pi_{jj}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i \tag{31}$$

This expression includes the term $p^*(j|\alpha_i) \ \pi_{jj}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i$ that is equal to the choice history probability $\mathbb{P}_{j,j}$. However, it also includes the "counterfactuals" $\delta^{(1)}_{k,j,j} \equiv p^*(k|\alpha_i) \ \pi_{jj}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i$ for $k \neq j$. We can represent each of these counterfactuals as:

$$\delta^{(1)}_{k,j,j} = \int p^*(k|\alpha_i) \ [\pi_{k0}(\alpha_i) + \pi_{k1}(\alpha_i) + ... + \pi_{kJ}(\alpha_i)] \ \pi_{jj}(\alpha_i) \ f_\alpha(\alpha_i) \tag{32}$$

That is, we have that $\delta^{(1)}_{k,j,j} = \sum_{\ell=0}^{J} \delta^{(2)}_{k,\ell,j,j}$, with $\delta^{(2)}_{k,\ell,j,j} \equiv p^*(k|\alpha_i) \ \pi_{k\ell}(\alpha_i) \ \pi_{jj}(\alpha_i) \ f_\alpha(\alpha_i)$. For $\ell = j$, we have that $\delta^{(2)}_{k,j,j,j}$ corresponds to the choice history probability $\mathbb{P}_{k,j,j}$. Therefore, we have that:

$$\Pi_{jj} = \mathbb{P}_{j,j} + \sum_{k \neq j} \left[\mathbb{P}_{k,j,j} + \sum_{\ell \neq j} \delta^{(2)}_{k,\ell,j,j} \right]. \tag{33}$$

For the rest of the terms $\delta^{(2)}_{k,\ell,j,j}$, we apply Lemma 1. According to Lemma 1, we have that $\pi_{k\ell}(\alpha_i)$ $\pi_{jj}(\alpha_i) = \exp\{\beta_{k\ell} - \beta_{kj} + \beta_{jj} - \beta_{j\ell}\} \ \pi_{kj}(\alpha_i) \ \pi_{j\ell}(\alpha_i)$. Finally, note that $p^*(k|\alpha_i) \ \pi_{kj}(\alpha_i) \ \pi_{j\ell}(\alpha_i)$ $f_\alpha(\alpha_i)$ is the choice history probability $\mathbb{P}_{k,j,\ell}$. Putting all the pieces together, we have that the expression in equation (30). □

Unfortunately, the procedure described in the proof of Proposition 1 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. Nevertheless, we show in the next section 3.3 how Proposition 1 implies the identification of important AMEs.

3.3 Identification of one-period AMEs

Corollary 1.1. For the binary choice model without duration dependence, as described in equation (4), Proposition 1 implies the identification of $\Pi_{01}(\mathbf{x})$, $\Pi_{11}(\mathbf{x})$, $AME^{(1)}(\mathbf{x})$, $ATE_{01,t}$, and $ATE_{11,t}$. We now develop these results in more detail. First, remember that, in this model, the slope parameter β is equal to $\beta_{11} - \beta_{10} - \beta_{01} + \beta_{00}$. Therefore, by Lemma 1 we have that:

$$\exp\{\beta\} = \frac{\pi_{11}(\alpha_i, \mathbf{x}) \ \pi_{00}(\alpha_i, \mathbf{x})}{\pi_{10}(\alpha_i, \mathbf{x}) \ \pi_{01}(\alpha_i, \mathbf{x})} \tag{34}$$

The application of Proposition 1 to this binary model implies the identification of $\Pi_{11}(\mathbf{x})$ and $\Pi_{00}(\mathbf{x})$ with the following expressions:

$$\left\{ \begin{array}{lcl} \Pi_{11}(\mathbf{x}) & = & \mathbb{P}_{1,1}(\mathbf{x}) + \mathbb{P}_{0,1,1}(\mathbf{x}) + \exp\{\beta\} \ \mathbb{P}_{0,1,0}(\mathbf{x}) \\ \\ \Pi_{00}(\mathbf{x}) & = & \mathbb{P}_{0,0}(\mathbf{x}) + \mathbb{P}_{1,0,0}(\mathbf{x}) + \exp\{\beta\} \ \mathbb{P}_{1,0,1}(\mathbf{x}) \end{array} \right. \tag{35}$$

As shown in equation (15), $AME^{(1)}(\mathbf{x}) = \Pi_{11}(\mathbf{x}) - \Pi_{01}(\mathbf{x})$. In this binary model, we have that $\Pi_{01}(\mathbf{x}) = 1 - \Pi_{00}(\mathbf{x})$. Therefore, it is clear that $AME^{(1)}(\mathbf{x})$ is identified as $\Pi_{11}(\mathbf{x}) + \Pi_{00}(\mathbf{x}) - 1$. After some simple algebra, we can obtain the following expression:

$$AME^{(1)}(\mathbf{x}) = [\exp\{\beta\} - 1] \ \ [\mathbb{P}_{0,1,0}(\mathbf{x}) + \mathbb{P}_{1,0,1}(\mathbf{x})] \tag{36}$$

These results clearly imply the identification of the unconditional parameters Π_{11}, Π_{01}, and $AME^{(1)}$ in the model without $\mathbf{x}$ variables. Finally, given the definitions $ATE_{01,t} = \Pi_{01} - \mathbb{E}(y_{it}|t)$ and $ATE_{11,t} = \Pi_{11} - \mathbb{E}(y_{it}|t)$, these causal effect parameters are also identified.

Corollary 1.2. Proposition 1 implies the identification of any parameter that is a nonlinear function of the average transitions Π_{jj}. For instance, in the binary choice model, for $\Pi_{01} > 0$, the marginal effect in percentage change, $(\Pi_{11} - \Pi_{01})/\Pi_{01}$, is identified. Similarly, in the multinomial case we can identify the log-odds ratio parameter $\ln(\Pi_{jj}/\Pi_{00})$. This parameter measures the degree of state dependence in choice alternative j relative to a baseline alternative 0.

Corollary 1.3. The identification of Π_{jj} implies the identification $ATE_{jj,t}$. Remember that $ATE_{jj,t}$ is the average treatment effect on $1\{y_{it} = j\}$ from a randomized experiment where individuals in the experimental group are assigned to $y_{t-1} = j$, and individuals in the control group receive no treatment. By definition, $ATE_{jj,t} = \Pi_{jj} - \mathbb{E}(y_{it}|t)$, such that $ATE_{jj,t}$ is identified at any period t in the sample.

3.4 Identification of n-periods forward AMEs

We now establish, in Proposition 2, the identification of n-periods forward AMEs in the binary choice model without duration dependence in equation (4). Our proof of this Proposition builds on the following Lemma.

Lemma 2. *Consider the binary choice model defined by equation (4) and Assumption 1. Suppose that* $\mathbf{x}_i^{\{1,n\}} = (\mathbf{x}, ..., \mathbf{x})$. *Then, the n-periods forward individual-specific causal effect* $\Delta^{(n)}(\alpha_i, \mathbf{x})$

satisfies the following equation:

$$\Delta^{(n)}(\alpha_i, \mathbf{x}) = [\exp\{\beta\} - 1]^n \quad [\pi_{10}(\alpha_i, \mathbf{x})]^n \quad [\pi_{01}(\alpha_i, \mathbf{x})]^n . \qquad \blacksquare \tag{37}$$

Proof. For notational simplicity, we omit $\mathbf{x}$ as an argument throughout this proof. However, it should be understood that all the probabilities are conditional on $\mathbf{x}_i^{\{1,n\}} = (\mathbf{x}, \mathbf{x}, \mathbf{x})$. Using the Markov structure of the model and the chain rule, we have that:

$$\begin{aligned} \mathbb{E}(y_{i,t+n} \mid \alpha_i, y_{it}) &= \mathbb{P}\left(y_{i,t+n-1} = 0 \mid \alpha_i, y_{it}\right) \ \pi_{01}(\alpha_i) + \mathbb{P}\left(y_{i,t+n-1} = 1 \mid \alpha_i, y_{it}\right) \ \pi_{11}(\alpha_i) \\ &= \pi_{01}(\alpha_i) + \mathbb{E}\left(y_{i,t+n-1} \mid \alpha_i, y_{it}\right) \ [\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)] \end{aligned} \tag{38}$$

Given the definition of $\Delta^{(n)}(\alpha_i)$ as $\mathbb{E}(y_{i,t+n} \mid \alpha_i, y_{it} = 1) - \mathbb{E}(y_{i,t+n}|\alpha_i, y_{it} = 0)$, and applying equation (38), we have that:

$$\begin{aligned} \Delta^{(n)}(\alpha_i) &= [\mathbb{E}\left(y_{i,t+n-1} \mid \alpha_i, y_{it} = 1\right) - \mathbb{E}\left(y_{i,t+n-1} \mid \alpha_i, y_{it} = 0\right)] \ [\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)] \\ &= \Delta^{(n-1)}(\alpha_i) \ [\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)] \end{aligned} \tag{39}$$

Applying this expression recursively, we obtain that $\Delta^{(n)}(\alpha_i) = [\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)]^n = [\Delta^{(1)}(\alpha_i)]^n$. Finally, an implication of Lemma 1 is that $\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i) = [\exp\{\beta\} - 1] \ \pi_{10}(\alpha_i) \ \pi_{01}(\alpha_i)$. To see this, note that by Lemma 1, $\exp\{\beta\} \ \pi_{10}(\alpha_i) \ \pi_{01}(\alpha_i) = \pi_{11}(\alpha_i) \ \pi_{00}(\alpha_i)$. This implies that $[\exp\{\beta\} - 1] \ \pi_{10} \ \pi_{01} = \pi_{11} \ \pi_{00} - \pi_{10} \ \pi_{01} = \pi_{11}(1 - \pi_{01}) - (1 - \pi_{11})\pi_{01} = \pi_{11} - \pi_{01}$. $\square$

Proposition 2. *Consider the binary choice model defined by equation (4) and Assumption 1. Let n be any positive integer, suppose that $\mathbf{x}_i^{\{1,2n+1\}} = (\mathbf{x}, ..., \mathbf{x})$, 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 $AME^{(n)}$ is identified as:*

$$AME^{(n)}(\mathbf{x}) = [\exp\{\beta\} - 1]^n \quad \left[\mathbb{P}_{0,\widetilde{\mathbf{10}}^n}(\mathbf{x}) + \mathbb{P}_{\widetilde{\mathbf{10}}^n,1}(\mathbf{x})\right] \tag{40}$$

where $\mathbb{P}_{0,\widetilde{\mathbf{10}}^n}(\mathbf{x})$ and $\mathbb{P}_{\widetilde{\mathbf{10}}^n,1}(\mathbf{x})$ are the probabilities of choice histories $(0, \widetilde{\mathbf{10}}^n)$ and $(\widetilde{\mathbf{10}}^n, 1)$ conditional on $\mathbf{x}_i^{\{1,2n+1\}} = (\mathbf{x}, ..., \mathbf{x})$. $\blacksquare$

Proof. For notational simplicity, we omit $\mathbf{x}$ as an argument throughout this proof. 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:

$$\begin{cases} \mathbb{P}_{0,\widetilde{\mathbf{10}}^n} &= \displaystyle\int p^*(0|\alpha_i) \ [\pi_{10}(\alpha_i)]^n \ [\pi_{01}(\alpha_i)]^n \ f_\alpha(\alpha_i) \ d\alpha_i \\ \\ \mathbb{P}_{\widetilde{\mathbf{10}}^n,1} &= \displaystyle\int p^*(1|\alpha_i) \ [\pi_{10}(\alpha_i)]^n \ [\pi_{01}(\alpha_i)]^n \ f_\alpha(\alpha_i) \ d\alpha_i \end{cases} \tag{41}$$

Applying equation (37) from Lemma 2, we have that:

$$\begin{cases} \mathbb{P}_{0,\widetilde{\mathbf{10}}^n} &= \dfrac{1}{[\exp\{\beta\}-1]^n} \displaystyle\int p^*(0|\alpha_i) \ \Delta^{(n)}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i \\ \\ \mathbb{P}_{\widetilde{\mathbf{10}}^n,1} &= \dfrac{1}{[\exp\{\beta\}-1]^n} \displaystyle\int p^*(1|\alpha_i) \ \Delta^{(n)}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i \end{cases} \tag{42}$$

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

3.5 A general procedure for identification of AMEs

In the identification results presented above, we show that AMEs are weighted sums of probabilities of choice histories. In this section, we present a general method to obtain these weights. The derivation of this method provides a better understanding of our approach for the identification of AMEs. In sections 3.6 and 3.7, we use this procedure to obtain the identifications of AMEs with $\mathbf{x}$ variables and duration dependence, respectively.

As in Section 3.1, let $\mathbf{s}(\mathbf{y})$ be the sufficient statistic for α_i. Let $\mathbb{S}_T$ be the set of possible values of $\mathbf{s}$, $\mathbb{P}_\mathbf{s}$ the probability of a value $\mathbf{s}$, and $\mathbf{P_s} \equiv \{\mathbb{P}_\mathbf{s} : s \in \mathbb{S}_T\}$ the probability distribution of this statistic. Given θ, the empirical distribution $\mathbf{P_s}$ contains all the information in the data about the distribution of α_i, and therefore, about AMEs. Let $AME = \int \Delta(\alpha_i, \theta) f_\alpha(\alpha_i) d\alpha_i$ be one of the AMEs described in section 2.2. We are interested in obtaining a function that depends only on $\mathbf{P_s}$ and θ, say function $h(\mathbf{P_s}, \theta)$, such that $AME = h(\mathbf{P_s}, \theta)$. Lemma 3 presents a necessary and sufficient condition to obtain this result.

The model determines how $\mathbb{P}_\mathbf{s}$ depends on the parameters of interest θ, the distribution of the incidental parameters, f_α, and the probability of the initial conditions. Taking into account the structure of the probability of a choice history in equation (24), the model implies:

$$\mathbb{P}_{\mathbf{s}} = \sum_{\mathbf{y}:\ \mathbf{s}(\mathbf{y})=\mathbf{s}} \left[\int \exp\{\mathbf{s}'\ \mathbf{g}(\alpha_i) + \mathbf{c}(\mathbf{y})'\ \theta\}\ f_\alpha(\alpha_i)\ d\alpha_i \right] \tag{43}$$

For the presentation of our results, it is convenient to distinguish two components in the sufficient statistic: the initial condition (y_1, d_1), and the rest of statistics which we represent as $\widetilde{\mathbf{s}}$ such that $\mathbf{s} = (y_1, d_1, \widetilde{\mathbf{s}})$. Similarly, we distinguish two components in the vector $\mathbf{g}(\alpha_i)$ such that we have:

$$\exp\{\mathbf{s}'\ \mathbf{g}(\alpha_i)\} = p^*(y_1, d_1|\alpha_i)\ \exp\{\widetilde{\mathbf{s}}'\ \widetilde{\mathbf{g}}(\alpha_i)\} \tag{44}$$

Lemma 3. *Consider the general model defined by equation (1) and Assumption 1. A necessary and sufficient condition for the existence of a function $h(\mathbf{P}_{\mathbf{s}}, \theta)$ such that $AME = h(\mathbf{P}_{\mathbf{s}}, \theta)$ is that there are weights $\{m_{\widetilde{\mathbf{s}}}(\theta) : \widetilde{\mathbf{s}} \in \widetilde{\mathbb{S}}_T\}$ which are known functions of θ and satisfy the following equation for every $\alpha_i \in \mathbb{R}^J$:*

$$\sum_{\widetilde{\mathbf{s}} \in \widetilde{\mathbb{S}}_T} m_{\widetilde{\mathbf{s}}}(\theta)\ \exp\{\widetilde{\mathbf{s}}'\ \widetilde{\mathbf{g}}(\alpha_i)\} = \Delta(\alpha_i, \theta). \tag{45}$$

Furthermore, this condition implies the following form for the function $h(\mathbf{P}_{\mathbf{s}}, \theta)$:

$$AME = h(\mathbf{P}_{\mathbf{s}}, \theta) = \sum_{\mathbf{s} \in \mathbb{S}_T} w_{\mathbf{s}}\ \mathbb{P}_{\mathbf{s}}, \tag{46}$$

where $w_{\mathbf{s}}(\theta) \equiv m_{\widetilde{\mathbf{s}}}(\theta) / \sum_{\mathbf{y}:\ \mathbf{s}(\mathbf{y})=\mathbf{s}} \exp\{\mathbf{c}(\mathbf{y})'\ \theta\}$. ■

Proof. In the Appendix, section 7.1. □

Based on equation (45) in Lemma 3, we develop a general and simple procedure to obtain the analytical expression of the weights $m_{\widetilde{\mathbf{s}}}(\theta)$ in the binary choice versions of the model. Equation (45) defines an infinite system of equations – as many as values of α_i. The researcher knows the functions $\widetilde{g}(\alpha_i)$ and $\Delta(\alpha_i, \theta)$ for every possible value α_i. The unknowns in this system are the weights $m_{\widetilde{\mathbf{s}}}(\theta)$ for every $\widetilde{\mathbf{s}} \in \widetilde{\mathbb{S}}_T$. The vector of unknowns has finite dimension: the number of points in set $\mathbb{S}_T$.

Proposition 3 presents our main result in this subsection. We show that equation (45) can be represented as a finite order polynomial in the variable $\exp\{\alpha_i\}$, and that this implies that we can obtain close-form expressions for the weights $m_{\widetilde{\mathbf{s}}}(\theta)$ as the analytical solution of a (finite dimensional) system of linear equations.

It is helpful to remember that, for the binary choice model in equation (5), we have that:[7] $\theta = \beta$; $\mathbf{s}(\mathbf{y}) = (y_1, y_T, n_1)$ with $n_1 = \sum_{t=2}^{T} y_t$; $\mathbf{c} = n_{11} \equiv \sum_{t=2}^{T} y_{t-1} y_t$; the individual effect $\Delta(\alpha_i, \theta)$ is equal to $e^{\alpha_i}(e^{\beta} - 1)/(1 + e^{\alpha_i + \beta})(1 + e^{\alpha_i})$; and

$$\widetilde{\mathbf{s}}' \, \widetilde{\mathbf{g}}(\alpha) = (T-1) \; \ln\left(\frac{1}{1+e^{\alpha}}\right) + y_T \; \ln\left(\frac{1+e^{\alpha_i+\beta}}{1+e^{\alpha}}\right) + n_1 \; \ln\left(\frac{e^{\alpha}(1+e^{\alpha})}{1+e^{\alpha+\beta}}\right) \tag{47}$$

Therefore, equation (45) takes the following form:

$$\sum_{y_T=0}^{1} \sum_{n_1=y_T}^{T-2+y_T} m_{y_T,n_1}(\beta) \left(\frac{1}{1+e^{\alpha}}\right)^{T-1} \left(\frac{1+e^{\alpha+\beta}}{1+e^{\alpha}}\right)^{y_T} \left(\frac{e^{\alpha}\,(1+e^{\alpha})}{1+e^{\alpha+\beta}}\right)^{n_1} = \frac{e^{\alpha}(e^{\beta}-1)}{(1+e^{\alpha+\beta})(1+e^{\alpha})}. \tag{48}$$

Proposition 3. *Consider the binary choice model defined by equation (5) and Assumption 1, and suppose that $T \geq 3$. The RHS and LHS of equation (48) can be represented as polynomials of order $2T-3$ in variable e^{α_i}. For this equation to hold for every value of α_i, we need that each of the $2T-2$ monomials in this polynomial have the same coefficient in the RHS and LHS of the equation. This condition implies a system of $2T-2$ linear equations (i.e., the number of monomials) and $2T-2$ unknowns (i.e., the weights $\{m_{y_T,n_1}(\beta) : y_T \in \{0,1\}, n_1 - y_T \in \{0,1,...,T-2\}\}$). The solution to this system exists, is unique,and it provides a closed-form expression for the weights $m_{\widetilde{s}}(\beta)$.* ■

Proof. In the Appendix, section 7.2. □

Given weights $\{m_{\widetilde{s}}(\beta)\}$, it is straightforward to obtain weights $\{w_{y_1,\widetilde{s}}(\beta)\}$ by simply applying its definition in Lemma 3, and then calculate $AME^{(1)}$.

In section 7.3 in the Appendix, we use Proposition 3 to obtain the closed-form expression of the weights when T is equal to 3, 4, 5, 6, and 7, respectively. For a given length T of the choice histories, the weights that solve the system are unique. However, this does not mean that $AME^{(1)}$ is just-identified. In fact, for panels with $T \geq 4$ there are over-identifying restrictions on $AME^{(1)}$. This is because we can use the panel to construct the empirical distribution of 3-period histories, of 4-period histories, and so on up to T-period histories. For each of these groups of histories, we can use the corresponding weights that solve the system in Proposition 3 to obtain a separate estimator of $AME^{(1)}$. Therefore, the model implies $T-3$ over-identifying restrictions on $AME^{(1)}$.

[7] See, for example, Aguirregabiria, Gu, and Luo (2021) for more details on this.

3.6 Identification of $\mathbf{AME_{x,t}}$

The AMEs conditional on $\mathbf{x}$ that we have identified in sections 3.3 and 3.4 apply to the subpopulation of individuals where the exogenous variable is constant over the sample. Since we can identify this AME for any value of $\mathbf{x}$, it is clear that we can obtain an integrated AME over all the values of $\mathbf{x}$. However, that integrated AME is still imposing the restriction that the exogenous variables are constant over time, and therefore, it is an AME for that subpopulation of individuals. We would like to obtain an AME that does not impose this restriction. This type of AME corresponds to $AME_{x,t}^{(1)}$ that we have defined in equation (16). Proposition 4 establishes the identification of $AME_{x,t}^{(1)}$.

Proposition 4. *Consider the binary choice model defined by equation (5) and Assumption 1, and suppose that* $T \geq 3$. *Then,* $AME_{x,t}^{(1)}$, *as defined in equation (16), is identified for any period* $t \geq 3$ *in the sample. For instance, for* $T = 3$ *and* $t = 3$, *we have that:*

$$AME_{x,3}^{(1)} = \sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\in\mathcal{X}^{\{1,3\}}} \mathbb{P}_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \begin{bmatrix} w_{(0,0,1;\mathbf{x})}\ \mathbb{P}_{(0,0,1)\ |\ (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)} \\ +w_{(0,1,0;\mathbf{x})}\ \mathbb{P}_{(0,1,0)\ |\ (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)} \\ +w_{(1,0,1;\mathbf{x})}\ \mathbb{P}_{(1,0,1)\ |\ (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)} \\ +w_{(1,1,0;\mathbf{x})}\ \mathbb{P}_{(1,1,0)\ |\ (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)} \end{bmatrix} \tag{49}$$

where $\mathbb{P}_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3}$ *and* $\mathbb{P}_{(y_1,y_2,y_3)\ |\ (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)}$ *are the density functions of* $(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)$ *and* (y_1,y_2,y_3) *conditional on* $(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)$, *respectively. The weights* $w_{(y_1,y_2,y_3;\mathbf{x})}$ *are:*

$$\begin{aligned} w_{(0,0,1;\mathbf{x})} &= \frac{e^{\mathbf{x}_2'\gamma} - e^{\mathbf{x}_3'\gamma}}{e^{\mathbf{x}_3'\gamma}};\ \ w_{(0,1,0;\mathbf{x})} = \frac{e^{\beta+\mathbf{x}_3'\gamma} - e^{\beta+\mathbf{x}_2'\gamma}}{e^{\mathbf{x}_2'\gamma}}; \\ w_{(1,0,1;\mathbf{x})} &= \frac{e^{\beta+\mathbf{x}_2'\gamma} - e^{\mathbf{x}_3'\gamma}}{e^{\mathbf{x}_3'\gamma}};\ \ w_{(1,1,0;\mathbf{x})} = \frac{e^{\mathbf{x}_3'\gamma} - e^{\mathbf{x}_2'\gamma}}{e^{\mathbf{x}_2'\gamma}}. \qquad \blacksquare \end{aligned} \tag{50}$$

Proof. In section 7.4 in the Appendix. □

Remark 4.1. Proposition 4 does not impose any restriction on the stochastic process of $\mathbf{x}_{it}$ – other than it is strictly exogenous with respect the transitory shock ε_{it}. Furthermore, though the notation in the enunciate and proof of Proposition 4 assumes that the support of $\mathbf{x}_{it}$ is discrete, this identification result trivially extends to the case of continuous $\mathbf{x}$ variables.

Remark 4.2. There is a relationship between the identification of $AME_{x,t}^{(1)}$ in Proposition 4 and the identification of $AME^{(1)}(\mathbf{x})$ in Corollary 1.1 of Proposition 1. These two AMEs are the same if $\mathbf{x}_{it}$ is constant over time – with probability one – for every individual in the sample. Under this condition, the (sub)population of individuals with constant $\mathbf{x}_{it}$ is simply the population of all the individuals, and we can confirm that the weights to obtain $AME_{x,t}^{(1)}$ in equation (50) are equal to the weights to obtain $AME^{(1)}(\mathbf{x})$ in equation (36). That is:

$$w_{(0,0,1;\mathbf{x})} = w_{(1,1,0;\mathbf{x})} = \frac{e^{\mathbf{x}'\gamma} - e^{\mathbf{x}'\gamma}}{e^{\mathbf{x}'\gamma}} = 0 \; ; \; w_{(0,1,0;\mathbf{x})} = w_{(1,0,1;\mathbf{x})} = \frac{e^{\beta+\mathbf{x}'\gamma} - e^{\mathbf{x}'\gamma}}{e^{\mathbf{x}'\gamma}} = e^{\beta} - 1. \tag{51}$$

3.7 Identification of AMEs of changes in duration

Proposition 5. *Consider the binary choice model with duration dependence defined by equation (3) and Assumption 1, and suppose that $T \geq 4$. Under these conditions, $AME_{0\to 1}$, $AME_{1\to 2}$, and $AME_{0\to 2}$ – as defined in equation (19) – are identified.*

$$\begin{aligned} AME_{0\to 1} &= \frac{e^{\beta(1)} - 1}{2} \; [\mathbb{P}_{0,0,1,0} + \mathbb{P}_{0,1,0,0}] + \frac{e^{\beta(1)} - 1}{e^{\beta(1)}} \; \mathbb{P}_{0,0,1,1} \\ &+ \; \left(e^{\beta(1)} - 1\right) \; [\mathbb{P}_{1,0,1,0} + \mathbb{P}_{1,0,1,1}] \end{aligned} \tag{52}$$

In section 7.5 in the Appendix, we provide the expression for the identification of $AME_{0\to 2}$ and $AME_{0\to 2}$. ■

Proof. In section 7.5 in the Appendix. □

4 Monte Carlo experiments

The purpose of these Monte Carlo experiments is twofold. First, we illustrate the precision of the FE estimator of $AME^{(1)}$ using sample sizes that we find in actual applications, and compare the bias and variance of this FE estimator to those from a RE estimator that imposes restrictions that we typically find in applications of RE models. Second, we compare the power of two testing procedures for rejecting a misspecified RE model: the standard Hausman test based on the difference between RE and FE estimates of slope parameters, and a new Hausman test that we propose based on the difference between RE and FE estimates of AMEs.

The DGP is the binary choice AR1 model without exogenous explanatory variables in equation (5). The model for the initial condition is $y_{i1} = 1\{\alpha_i + u_i \geq 0\}$ where u_i is i.i.d. Logistic and independent of α_i and ε_{it}. The number of periods in the sample is $T = 4$. We present results for two different sample sizes N, 1000 and 2000. We consider six different DGPs based on two different values of parameter β (i.e., $\beta = -1$ and $\beta = 1$) and three distributions of the unobserved heterogeneity α_i: no unobserved heterogeneity, such that $\alpha_i = 0$ for any individual

i; finite mixture with two points of support such that $\alpha_i = -1$ with probability 0.3, and $\alpha_i = 0.5$ with probability 0.7; and a mixture of two normal random variables such that $\alpha_i \sim N(-1, 3)$ with probability 0.3, and $\alpha_i \sim N(0.5, 3)$ with probability 0.7. Note that this mixture of normals implies an asymmetric and bimodal distribution.

Table 1 summarizes the six DGPs, the labels we use to represent each of them, and the corresponding true value of $AME^{(1)}$ in the population. Note that, keeping parameter β constant, the AME can vary substantially when we change the distribution of the unobserved heterogeneity. This is why the identification of AME, and not only β, is important to measure causal effects. For instance, when $\beta = 1$, AME is equal to 0.23 in the DGP without unobserved heterogeneity, 0.20 for the finite mixture, and 0.11 when α has a mixture of normals distribution. We find similar variation when $\beta = -1$.

Table 1

DGPs and true values of AME

	Distribution of α_i		
Value of β	$\alpha_i = 0$	*Finite mixture*	*Mixture of normals*
$\beta = -1$	**NoUH(-1)** $\text{AME}^{(1)} = -0.2311$	**FinMix(-1)** $\text{AME}^{(1)} = -0.2164$	**MixNor(-1)** $\text{AME}^{(1)} = -0.113$
$\beta = 1$	**NoUH(+1)** $\text{AME}^{(1)} = 0.2311$	**FinMix(+1)** $\text{AME}^{(1)} = 0.2059$	**MixNor(+1)** $\text{AME}^{(1)} = 0.1108$

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 $AME^{(1)}$: (1) a FE estimator, that we denote *FE-CMLE*;[8] (2) a maximum likelihood estimator of a RE model 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 we denote *NoUH-MLE*.[9]

[8]For parameter β, the fixed effect estimator is the CMLE proposed by Chamberlain (1985). For parameter $AME^{(1)}$, we use a plug-in estimator based on the formula for the identified $AME^{(1)}$ when $T = 4$ that we present in Table 5 (see section 7.3 in the Appendix). In that formula, we replace parameter β with its CML estimate, and the probabilities of choice histories with their frequency estimates.

[9]For the DGPs without unobserved heterogeneity (i.e., *NoUH(-1)* and *NoUH(+1)*), we do not report results for the RE MLE. This is because, 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 of α_i are not identified and they take extreme values. This also affects the estimation of β that presents very large 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 the estimation of the distribution of α is one of the key advantages of FE estimator.

Tables 2 and 3 present results from the experiments with sample sizes 1000 and 2000, respectively. The results are quite similar for the two samples sizes except that, as one would expect, all the estimators have lower bias and variance when the sample size increases. Therefore, we focus our discussion in the results with $N = 1000$ in Table 2.

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

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

Table 3
Monte Carlo Experiments with sample size N=2,000

		Statistics						
		True β	*Mean* $\widehat{\beta}$	*Std* $\widehat{\beta}$	*True* AME	*Mean* $\widehat{AME}$	*Std* $\widehat{AME}$	*RMSE* $\widehat{AME}$
DGP	*FE-CMLE*	-1.0	-1.0045	0.0951	-0.2311	-0.2315	0.0169	0.0169
NoUH(-1)	*RE-MLE*	-1.0	NA	NA	-0.2311	NA	NA	NA
	NoUH-MLE	-1.0	-1.0030	0.0566	-0.2311	-0.2316	0.0123	0.0123
DGP	*FE-CMLE*	-1.0	-1.0006	0.0985	-0.2164	-0.2160	0.0165	0.0164
FinMix(-1)	*RE-MLE*	-1.0	-1.0018	0.0898	-0.2164	-0.2163	0.0159	0.0158
	NoUH-MLE	-1.0	-0.5941	0.0576	-0.2164	-0.1421	0.0135	0.0755
DGP	*FE-CMLE*	-1.0	-1.0096	0.1514	-0.1113	-0.1116	0.0126	0.0124
MixNor(-1)	*RE-MLE*	-1.0	-0.3522	0.1281	-0.1113	-0.0467	0.0156	0.0665
	NoUH-MLE	-1.0	1.7142	0.0757	-0.1113	0.4012	0.0158	0.5128
DGP	*FE-CMLE*	1.0	1.0051	0.1171	0.2311	0.2339	0.0366	0.0367
NoUH(+1)	*RE-MLE*	1.0	NA	NA	0.2311	NA	NA	NA
	NoUH-MLE	1.0	1.0008	0.0527	0.2311	0.2311	0.0120	0.0120
DGP	*FE-CMLE*	1.0	1.0059	0.1306	0.2059	0.2089	0.0370	0.0371
FinMix(+1)	*RE-MLE*	1.0	0.9986	0.0925	0.2059	0.2055	0.0228	0.0228
	NoUH-MLE	1.0	1.4166	0.0598	0.2059	0.3229	0.0130	0.1177
DGP	*FE-CMLE*	1.0	1.0028	0.2005	0.1108	0.1137	0.0327	0.0327
MixNor(+1)	*RE-MLE*	1.0	1.4992	0.1274	0.1108	0.2146	0.0255	0.1069
	NoUH-MLE	1.0	3.2432	0.0844	0.1108	0.6645	0.0118	0.5538

(i) Bias of FE estimators relative to MLE. The mean biases 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 $AME^{(1)}$. FE estimation of $AME^{(1)}$ does not involve a substantially larger bias than the FE estimation of β. This bias is of 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 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 $AME^{(1)}$. 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 $AME^{(1)}$. As expected, the FE estimators have larger variances than the RE-MLE

estimators. However, the loss of precision associated with FE estimation is of similar magnitude when estimating $AME^{(1)}$ 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 has a clear explanation. The histories that contribute to the identification of the parameters β and $AME^{(1)}$ 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. It is easier to identify negative state dependence than positive state dependence because the former has very different implications than unobserved heterogeneity, while the later have similarities with unobserved heterogeneity.

(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 larger in the estimation of $AME^{(1)}$: 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 $AME^{(1)}$. As a result, the FE estimator clearly 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 in 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 but using the FE and RE estimators of $AME^{(1)}$. Therefore, we have two different Hausman statistics to test for the validity of a RE model. The statistic based on the estimators of β:

$$HS_{\beta} = \frac{\left(\widehat{\beta}_{FE} - \widehat{\beta}_{RE}\right)^2}{\widehat{Var}\left(\widehat{\beta}_{FE}\right) - \widehat{Var}\left(\widehat{\beta}_{RE}\right)} \quad \text{under } H_0 \quad \sim \chi^2_1 \tag{53}$$

And the statistic based on the estimators of $AME^{(1)}$:

$$HS_{AME} = \frac{\left(\widehat{AME}_{FE} - \widehat{AME}_{RE}\right)^2}{\widehat{Var}\left(\widehat{AME}_{FE}\right) - \widehat{Var}\left(\widehat{AME}_{RE}\right)} \quad \text{under } H_0 \quad \sim \chi^2_1 \tag{54}$$

The Hausman test based on AME has several advantages with respect the test based on β.

First, the researcher can be particularly interested in the causal effect implied by the model and not on the slope parameter itself. Second, and more substantially, the test on the parameter β may suffer of 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}, and this variance depends on the specification of RE model. For instance, when we compare $\widehat{\beta}_{FE}$ with $\widehat{\beta}_{NoUH-MLE}$ part of the reason why these two estimators are different is because in the model that does not account for 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.

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 each of the two tests– of the null hypothesis of valid RE model. More specifically:

Figure 1: DGP is *FinMix(-1)* and null hypothesis is no unobserved heterogeneity.
Figure 2: DGP is *FinMix(+1)* and null hypothesis is no unobserved heterogeneity.
Figure 3: DGP is *MixNor(-1)* and null hypothesis is no unobserved heterogeneity.
Figure 4: DGP is *MixNor(+1)* and null hypothesis is no unobserved heterogeneity.
Figure 5: DGP is *MixNor(-1)* and null hypothesis is the *finite mixture* model.
Figure 6: DGP is *MixNor(+1)* and null hypothesis is the *finite mixture* model.

Figures 3 and 4 show that both tests have very strong power to reject the null of no unobserved heterogeneity when the DGP is a mixture of normals. In Figures 1 and 5, we can see that the two test have also strong power when the true value of β is negative. The relevant comparison appears in Figures 2 and 6. The results are mixed. 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 serious problem of low power. For this test, the p-value is greater than 5% for more than half of the samples, such that with a 5% significance level we could reject the null for less than half of the samples. In contrast, the HS_{AME} test has reasonable power. For this test, the p-value is greater than 5% for one-fifth of the samples, such that with a 5% significance level we do reject the null for 80% of the samples. In Figure 1, 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 serious problem of low power. Overall, the HS_{AME} test has larger power than the test HS_{β}. Therefore, this test seems a useful byproduct of identification of AMEs in Fixed Effects models.

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

Figure 1

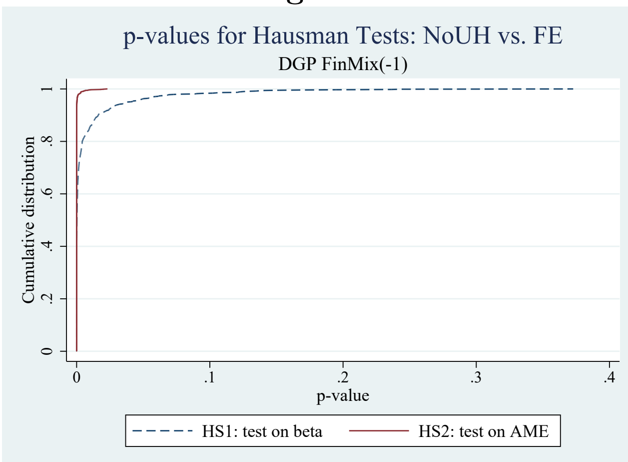


Figure 2

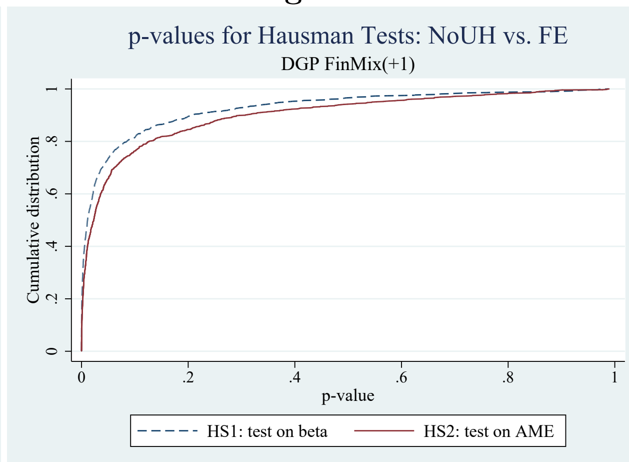


Figure 3

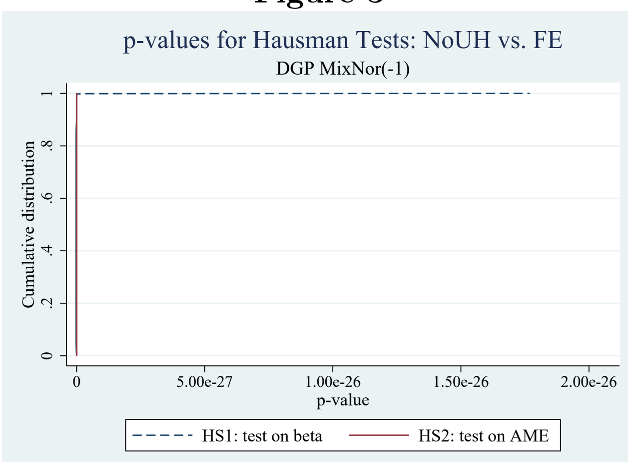


Figure 4

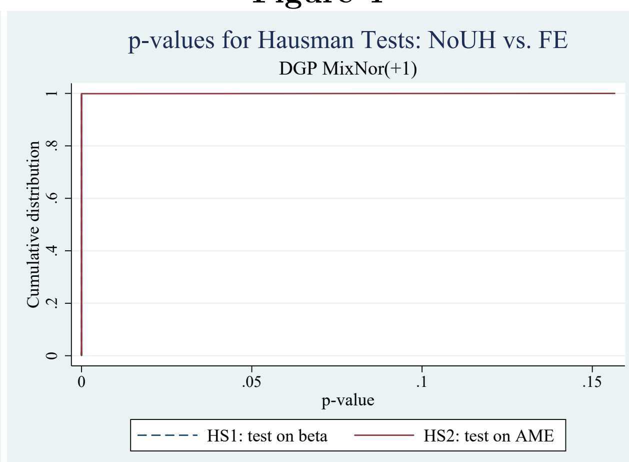


Figure 5

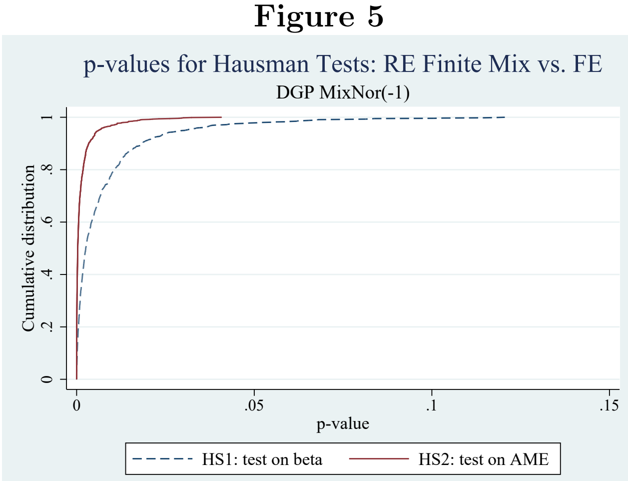
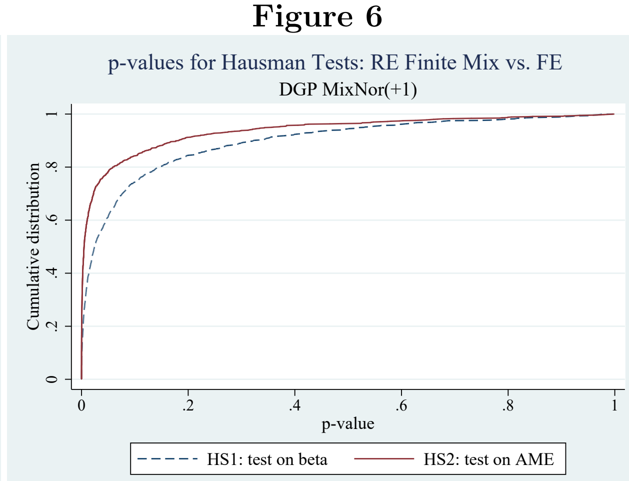


Figure 6



5 State Dependence in Consumer Brand Choice

We apply our identification results to measure state dependence in consumer brand choices. There is an important literature on testing and measuring state dependence in consumer brand choices, with seminal papers by Erdem (1996), Keane (1997), and Roy, Chintagunta, and Haldar (1996).[10] 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 habits or/and brand switching costs. The main goal is to determine the relative importance 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 on demand elasticities, competition, consumer welfare, and the evaluation of mergers.[11] All these previous studies estimate Random Effects (RE) models. In this application, we consider a FE model, estimate average transition probabilities Π_{jj}, and use them to measure the contribution of state dependence to brand-choice persistence.

5.1 Data

The dataset is A.C. Nielsen scanner panel data from Sioux Falls, South Dakota, for the ketchup product category.[12] It contains 996 households and covers a 123-week period from mid-1986 to mid-1988.[13] For our analysis, a time period is a *household purchase occasion*. That is, periods $t = 1, 2, ...$ represent a household's first, second, ... purchase of ketchup 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 4 presents the distribution of T_i.

Table 4
Distribution of number of purchase occasions (T_i)

Minimum	*5%*	*25%*	*Median*	*75%*	*95%*	*Maximum*
3	4	5	8	12	21	52

[10] Other contributions in 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 growing literature on the implications of brand-choice state dependence on market competition (see Viard, 2007, and Pakes, Porter, Shepard, and Calder-Wang 2021.

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

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

[13] The 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 in that paper for a description of the selection of this working sample.

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 5 presents brands' market shares (i.e., shares in number of purchases) and the matrix of transition probabilities between the four brands. Heinz is the leading brand, with 66% share of purchases, followed by Hunts 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 *Hunts* with 16.67%, and *Heinz* with 12.30%. This persistence may be due both to consumer taste heterogeneity and state dependence. Our maim goal in this application is to disentangle the contribution of these two factors.

Table 5
Matrix of Transition Probabilities of Brand Choices
(percentage points)

	Brand choice at $t+1$				*Total*
Brand choice at t	*Heinz* ($j = 0$)	*Hunts* ($j = 1$)	*Del Monte* ($j = 2$)	*Store* ($j = 3$)	
Heinz ($j = 0$)	**78.95**	10.67	6.98	3.40	100.00
Hunts ($j = 1$)	45.16	**32.30**	15.76	6.78	100.00
Del Monte ($j = 2$)	41.11	18.98	**34.07**	5.83	100.00
Store ($j = 3$)	42.32	17.11	13.38	**27.19**	100.00
Market share ($\mathbb{P}_j$)	66.65	15.63	12.19	5.53	100.00
Choice persistence ($\mathbb{P}_{j\|j} - \mathbb{P}_j$)	12.30	16.67	21.88	21.66	

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*:

$$y_{it} = \arg \max_{j \in \{0,1,2,3\}} \ \{ \ \alpha_i(j) + \beta_{jj} \ 1\{y_{i,t-1} = j\} + \varepsilon_{it}(j) \ \} \ . \tag{55}$$

Parameter β_{jj} represents habits in the purchase/consumption of brand j: the additional utility from keeping purchasing the same brand as in 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.[14]

Following Aguirregabiria, Gu, and Luo (2021), equation (55) can be interpreted as a model where households are forward-looking. That is, 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 on current ε_{it}.

5.3 Estimation

To illustrate our method using a short panel, we split the purchasing histories in the original sample into subs-histories of length T, where T is small. We present results for $T = 6$ and $T = 8$.

Table 6
Conditional Maximum Likelihood Estimates
of Brand Habit (β_{jj}) Parameters

Parameter	$T = 6$ **sub-histories**		$T = 8$ **sub-histories**	
β_{jj}	*Estimate*	*(s.e.)*$^{(1)}$	*Estimate*	*(s.e.)*$^{(1)}$
Heinz	0.00	(.)	0.00	(.)
Hunts	0.2312	(0.0590)	0.2566	(0.0570)
Del Monte	0.1155	(0.0718)	0.1191	(0.0722)
Store	0.3245	(0.1166)	0.4675	(0.1106)
# histories of length T	4,764		3,396	

(1) Standard errors ($s.e$) are obtained using a boostrap 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.

Table 6 presents our Fixed Effect estimates of the *brand habit* parameters β_{jj}. We use the Conditional Maximum Likelihood estimator described in section 3.1. Standard errors are obtained using a bootstrap method that resamples the 996 purchasing histories in the original

[14] In 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 in terms of estimating the model using choice histories where prices remain constant.

dataset.[15] Parameter estimates with $T = 6$ and $T = 8$ are very similar. They are significantly greater than zero at 5% significance level, showing evidence of state dependence in brand choice. The magnitude of the parameter estimate is not monotonically related to the brand's market share, or to the degree of brand choice persistence shown in Table 5. However, we need to take into account that a larger value of β_{jj} does not imply a larger degree of state dependence as measured by the Average Transition Probabilities or AME_{jj}. The parameters β_{jj} do not provide a measure of the contribution of state dependence to the observed brand choice persistence.

Table 7

Fixed Effects Estimates of Average Transition Probabilities (ATPs) and AMEs

	$T = 6$ **sub-histories**				$T = 8$ **sub-histories**			
	Pers (s.e.)	*ATP* (*s.e.*)	*AME* (s.e.)	*UHet* (*s.e.*)	*Pers* (s.e.)	*ATP* (*s.e.*)	*AME* (s.e.)	*UHet* (*s.e.*)
Heinz	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)
Hunts	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)
Del Monte	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)
Store	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 6.

(2) *ATP* is the brand's Average Transition Probability, Π_{jj}.

(3) *AME* is the one defined in equation (10): $\Pi_{jj} - \mathbb{E}(1\{y_{it} = j\})$.

(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 boostrap method as for the estimates in Table 6.

[15]Using 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 only a bit larger (at the second or third significant digit) than the asymptotic ones.

Table 7 presents Fixed Effect estimates of average transition probabilities (ATPs), and provides a decomposition of brand choice persistence into the contributions of state dependence and unobserved heterogeneity. Column labelled *Pers* provides brand choice persistence as measured by the difference between the transition probability $\mathbb{P}_{j|j}$ and the uncodnitional probability $\mathbb{P}_j$. The estimates of ATPs (in the columns labelled *ATP*) are very precise and similar for $T = 6$ and $T = 8$. The column labelled AME presents the AME defined in equation (22): $ATE_{jj} = \Pi_{jj} - \mathbb{E}(1\{y_{it} = j\})$. This AME is a measure of the contribution of state dependence to brand choice persistence. For all the brands, this contribution is quite small: between 1 and 2 percentage points. In fact, for *Heinz* and *Del Monte*, we cannot reject the null hypothesis that this AME is zero at 5% significance level. The *Store brand* is the one with the largest contribution of state dependence. The column labelled $UHet$ presents the contribution of consumer taste heterogeneity to brand choice persistence, as measured by the difference between brand choice persistence and AME_{jj}. This heterogeneity accounts for most of the brand choice persistence. 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 (AMEs) are useful parameters to represent causal effects in econometric applications. AMEs depend on the structural parameters of the model but also on the distribution of the unobserved heterogeneity. 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 in 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 both simulated data and real-world consumer scanner data. In contrast to previous studies using similar models and data but with a Random Effects specification of consumer unobserved taste heterogeneity, we find that most brand choice persistent can be attributed to consumer taste heterogeneity, and state dependence has a negligible contribution.

7 Appendix

7.1 Proof of Lemma 3

It is convenient to define the weights $w_{(y_1,d_1,\widetilde{\mathbf{s}})}(\theta) \equiv m_{\widetilde{\mathbf{s}}}(\theta) \;/\; \sum_{\mathbf{y}:\, \mathbf{s}(\mathbf{y})=(y_1,d_1,\widetilde{\mathbf{s}})} \exp\{\mathbf{c}(\mathbf{y})'\theta\}$, such that equation (45) can be re-written as:

$$\sum_{\widetilde{\mathbf{s}}\in\widetilde{\mathbb{S}}_T} w_{(y_1,d_1,\widetilde{\mathbf{s}})}(\theta)\ \exp\{\widetilde{\mathbf{s}}'\ \widetilde{\mathbf{g}}(\alpha_i)\} \left[\sum_{\mathbf{y}:\, \mathbf{s}(\mathbf{y})=(y_1,d_1,\widetilde{\mathbf{s}})} \exp\{\mathbf{c}(\mathbf{y})'\ \theta\}\right] = \Delta(\alpha_i,\theta). \tag{56}$$

Given the structure of the probability of a choice history in (24), we have that (56) is equivalent to:

$$\sum_{\widetilde{\mathbf{s}}\in\widetilde{\mathbb{S}}_T} w_{(y_1,d_1,\widetilde{\mathbf{s}})}(\theta)\ \left[\sum_{\mathbf{y}:\, \mathbf{s}(\mathbf{y})=(y_1,d_1,\widetilde{\mathbf{s}})} P\left(\mathbf{y} \mid y_1.d_1,\alpha_i,\theta\right)\right] = \Delta(\alpha_i,\theta). \tag{57}$$

(A) Sufficient condition. Multiplying (57) times $p^*(y_1,d_1|\alpha_i)\ f_\alpha(\alpha_i)$, integrating over α_i, and taking into account that $\int P\left(\mathbf{y} \mid y_1,d_1,\alpha_i,\theta\right)\ p^*(y_1,d_1|\alpha_i)\ f_\alpha(\alpha_i)\ d\alpha_i$ is equal to $\mathbb{P}_{\mathbf{y}}$, we obtain:

$$\sum_{\widetilde{\mathbf{s}}\in\widetilde{\mathbb{S}}_T} w_{(y_1,d_1,\widetilde{\mathbf{s}})}(\theta) \left[\sum_{\mathbf{y}:\, \mathbf{s}(\mathbf{y})=(y_1,d_1,\widetilde{\mathbf{s}})} \mathbb{P}_{\mathbf{y}}\right] = \int \Delta(\alpha_i,\theta)\ p^*(y_1,d_1|\alpha_i)\ f_\alpha(\alpha_i)\ d\alpha_i. \tag{58}$$

We can sum equation (58) over all the possible values of (y_1,d_1). Given that the sum of $p^*(y_1,d_1|\alpha_i)$ over all values of (y_1,d_1) is equal to 1, the right-hand-side becomes $\int \Delta(\alpha_i,\theta)\ f_\alpha(\alpha_i)\ d\alpha_i$ which is the AME. Furthermore, $\sum_{\mathbf{y}:\, \mathbf{s}(\mathbf{y})=\mathbf{s}} \mathbb{P}_{\mathbf{y}} = \mathbb{P}_{\mathbf{s}}$. Therefore, $h\left(\mathbf{P}_{\mathbf{s}},\theta\right)$ has the form in equation (46):

$$\sum_{\mathbf{s}\in\mathbb{S}_T} w_{\mathbf{s}}(\theta)\ \mathbb{P}_{\mathbf{s}} = AME. \tag{59}$$

(B) Necessary condition. The proof has two parts. First, we prove that function $h\left(\mathbf{P}_{\mathbf{s}},\theta\right)$ should be linear in $\mathbf{P}_{\mathbf{s}}$. Second, we show that the system of equations in (45) should hold.

Necessary (i). Equality $h\left(\mathbf{P}_{\mathbf{s}},\theta\right) = AME$ should hold for every distribution f_α. In particular, it should hold for: (Case 1) a degenerate distribution where $\alpha_i = c$ with probability one; (Case 2) a degenerate distribution where $\alpha_i = c' \neq c$ with probability one; and (Case 3) a distribution with two points of support, c and c', with $q \equiv f_\alpha(c)$. AME has the following form: (Case 1) $AME = \Delta(c)$; (Case 2) $AME = \Delta(c')$; and (Case 3) $AME = q\ \Delta(c) + (1-q)\ \Delta(c')$. Function $h\left(\mathbf{P}_{\mathbf{s}},\theta\right)$ should satisfy:

$$\begin{cases} \text{Case 1} & : \quad h\left(\mathbf{P}_{\mathbf{s}}^{(1)},\theta\right) = \Delta(c) \\ \text{Case 2} & : \quad h\left(\mathbf{P}_{\mathbf{s}}^{(2)},\theta\right) = \Delta(c') \\ \text{Case 3} & : \quad h\left(\mathbf{P}_{\mathbf{s}}^{(3)},\theta\right) = q\ \Delta(c) + (1-q)\ \Delta(c') \end{cases} \tag{60}$$

where $\mathbf{P}_{\mathbf{s}}^{(1)}$, $\mathbf{P}_{\mathbf{s}}^{(2)}$, and $\mathbf{P}_{\mathbf{s}}^{(3)}$ represent the distributions of the statistic $\mathbf{s}$ under the DGPs of cases 1, 2, and 3, respectively. Note that $\mathbf{P}_\mathbf{s}^{(3)} = q\,\mathbf{P}_\mathbf{s}^{(1)} + (1-q)\,\mathbf{P}_\mathbf{s}^{(2)}$. These conditions are for arbitrary values for c, $c^{'}$, and $q \in [0,1]$. Multiplying (60 - Case 1) times q, multiplying (60 - Case 2) times $(1-q)$, adding up these two conditions, and then subtracting (60 - Case 3), we get that $h\left(\mathbf{P}_\mathbf{s}, \theta\right)$ should satisfy:

$$q\ h\left(\mathbf{P}_\mathbf{s}^{(1)}, \theta\right) + (1-q)\ h\left(\mathbf{P}_\mathbf{s}^{(2)}, \theta\right) = h\left(q\ \mathbf{P}_\mathbf{s}^{(1)} + (1-q)\ \mathbf{P}_\mathbf{s}^{(2)}, \theta\right). \tag{61}$$

The only possibility that equation (61) holds for every $c, c' \in \mathbb{R}$ and $q \in [0,1]$ is that the function $h\left(\mathbf{P}_\mathbf{s}, \theta\right)$ is linear in $\mathbf{P}_\mathbf{s}$, such that $h\left(\mathbf{P}_\mathbf{s}, \theta\right) = \sum_{\mathbf{s} \in \mathbb{S}_T} w_\mathbf{s}(\theta)\ \mathbb{P}_\mathbf{s}$.

Necessary (ii): Given equation $\sum_{\mathbf{s} \in \mathbb{S}_T} w_\mathbf{s}(\theta)\ \mathbb{P}_\mathbf{s} = AME$, then (57) holds for every value $\alpha_i \in \mathbb{R}^J$. The proof is by contradiction. Suppose that: (a) $\sum_{\mathbf{s} \in \mathbb{S}_T} w_\mathbf{s}(\theta)\ \mathbb{P}_\mathbf{s} = AME$ is satisfied for any distribution f_α; and (b) there is a value of α_i – say $\alpha_i = c$ – and a value of the initial condition $s_1 \equiv (y_1, d_1)$, w.l.o.g. $s_1 = 0$ such that $\sum_{\widetilde{\mathbf{s}} \in \widetilde{\mathbb{S}}_T} w_{0,\widetilde{\mathbf{s}}}(\theta)\ \left[\sum_{\mathbf{y}:\ \mathbf{s}(\mathbf{y}) = (0,\widetilde{\mathbf{s}})} P\left(\mathbf{y} \mid 0, c, \theta\right)\right] \neq \Delta(c)$ 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_i with only two points support, c and c', with $f_\alpha(c) \equiv q$. Also, for notationaly simplicity but W.l.o.g. suppose that $s_1 \in \{0, 1\}$. Define:

$$d(\alpha_i, s_1) \equiv \sum_{\widetilde{\mathbf{s}} \in \widetilde{\mathbb{S}}_T} w_{s_1,\widetilde{\mathbf{s}}}(\theta)\ \left[\sum\nolimits_{\mathbf{y}:\ \mathbf{s}(\mathbf{y}) = \mathbf{s}} P\left(\mathbf{y} \mid s_1, c, \theta\right)\right] - \Delta(\alpha_i) \tag{62}$$

Condition (b) implies that $d(c, 0) \neq 0$. Applying the same operations as in the proof of the sufficient condition, we get:

$$\sum_{\mathbf{s} \in \mathbb{S}_T} w_\mathbf{s}(\theta)\ \mathbb{P}_\mathbf{s} - AME = q\left[p^*(0|c)\ d(c,0) + p^*(1|c)\ d(c,1)\right]\ +\ (1-q)\ \left[p^*(0|c')\ d(c',0) + p^*(1|c')\ d(c',1)\right] \tag{63}$$

By definition, the values of $d(\alpha_i, s_1)$ do not depend on the distribution f_α. Therefore, there always exist (a continuum of) values of q such that the right hand side of (63) is different to zero, and condition (a) does not hold. ■

7.2 Proof of Proposition 3

We use a to represent e^α and b to represent e^β. We multiple equation (48) times $[1+ab]^{T-2}\ [1+a]^{T-1}$ to eliminate the denominator. We get:

$$\sum_{\widetilde{s}} m_{\widetilde{s}}(\beta)\ a^{n_1}\ [1+a]^{n_1 - y_T}\ [1+ab]^{T-2-n_1+y_T} = [b-1]\,a\,[1+a]^{T-2}\,[1+ab]^{T-3}. \tag{64}$$

We apply the *Binomial Theorem* to expand the terms $[1+x]^n$ as $\sum_{k=0}^{n} \binom{n}{k} x^k$.

$$\begin{aligned} & \sum_{\widetilde{s}} m_{\widetilde{s}}(\beta)\ a^{n_1} \left[\sum_{k=0}^{n_1 - y_T} \binom{n_1 - y_T}{k} a^k\right] \left[\sum_{k=0}^{T-2-n_1+y_T} \binom{T-2-n_1+y_T}{k} b^k a^k\right] \\ = \quad & [b-1]\,a \left[\sum_{k=0}^{T-2} \binom{T-2}{k} a^k\right] \left[\sum_{k=0}^{T-3} \binom{T-3}{k} b^k a^k\right]. \end{aligned} \tag{65}$$

To solve (65) with respect to the $2T-2$ weights $\{m_{y_T,n_1}(\beta)\}$, note that each side of this equation is a polynomial in $a \equiv e^{\alpha}$. The order of this polynomial is $2T-3$. For this equation to hold for every value of α, the coefficient of each monomial should be the same in RHS and LHS. This requirement implies a system $2T-2$ linear equations (i.e., the number of monomials in a polynomial of order $2T-3$, including the constant term) and $2T-2$ unknowns (i.e., the weights m_{y_T,n_1}).

7.3 Applying Proposition 3

Tables 8 and 9 present the weights for $AME^{(1)}$ in the binary choice AR(1) model for different values of T. These weights have been derived using the procedure in Proposition 3. For the sake of illustration, we describe here the derivation of these weights for $T=3$.

When $T=3$, the statistics (y_T, n_1) can take four possible values: $(0,0)$, $(0,1)$, $(1,1)$, and $(1,2)$. Therefore, there are four weights m_{y_T,n_1}. Equation (65) takes the following form:

$$\begin{aligned} & m_{0,0}\ a^0 \left[\sum_{k=0}^{0}\binom{0}{k}a^k\right]\left[\sum_{k=0}^{1}\binom{1}{k}b^k a^k\right] + m_{0,1}\ a^1 \left[\sum_{k=0}^{1}\binom{1}{k}a^k\right]\left[\sum_{k=0}^{0}\binom{0}{k}b^k a^k\right] \\ + \quad & m_{1,1}\ a^1 \left[\sum_{k=0}^{0}\binom{0}{k}a^k\right]\left[\sum_{k=0}^{1}\binom{1}{k}b^k a^k\right] + m_{1,2}\ a^2 \left[\sum_{k=0}^{1}\binom{1}{k}a^k\right]\left[\sum_{k=0}^{0}\binom{0}{k}b^k a^k\right] \\ = \quad & [b-1]\, a \left[\sum_{k=0}^{1}\binom{1}{k}a^k\right]\left[\sum_{k=0}^{0}\binom{0}{k}b^k a^k\right]. \end{aligned} \tag{66}$$

Or equivalently,

$$m_{0,0}\ \ [1+b\ a] + m_{0,1}\ \ \left[a+a^2\right] + m_{1,1}\ \ \left[a+b\ a^2\right] + m_{1,2}\ \ \left[a^2+b\ a^3\right] = [b-1]\left[a+a^2\right]. \tag{67}$$

Making equal the coefficients in the RHS and LHS for each monomial, we get the following system of four equations with four unknowns.

$$\left\{ \begin{array}{rcl} m_{0,0} & = & 0 \\ m_{0,0}\ b + m_{0,1} + m_{1,1} & = & b-1 \\ m_{0,1} + m_{1,1}\ b + m_{1,2} & = & b-1 \\ m_{1,2} & = & 0 \end{array} \right. \tag{68}$$

The solution to this system is $m_{0,0} = m_{1,2} = m_{1,1} = 0$ and $m_{0,1} = b-1$. Therefore, with $T=3$, only histories with $(y_T, n_1) = (0,1)$ receive positive weight in the identification of $AME^{(1)}$. There are two choice histories with this condition: $(0,1,0)$ and $(1,1,0)$. Finally, to obtain the weights w_s of these histories we apply the formula: $w_{y_1,y_T,n_1} = m_{y_T,n_1} \ / \sum_{y:s(y)=(y_1,y_T,n_1)} \exp\{\beta n_{11}(y)\}$. That is, $w_{0,1,0} = (b-1)$

Table 8

Weights $w_{\mathbf{s}}$ 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

Table 9

Weights $w_{\mathbf{s}}$ 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

7.4 Proof of Proposition 4

W.l.o.g. we consider $T = 3$ and $t = 3$. We first obtain the expression for the probabilities of (y_1, y_2, y_3) conditional on $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ and on α_i, that we represent as $\mathbb{P}_{(y_1,y_2,y_3)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)}$.

$$\begin{cases} \mathbb{P}_{(0,0,1)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} &=& p^*(0|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\;\dfrac{1}{1+e^{\alpha_i+\mathbf{x}_2'\gamma}}\;\dfrac{e^{\alpha_i+\mathbf{x}_3'\gamma}}{1+e^{\alpha_i+\mathbf{x}_3'\gamma}} \\ \mathbb{P}_{(0,1,0)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} &=& p^*(0|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\;\dfrac{e^{\alpha_i+\mathbf{x}_2'\gamma}}{1+e^{\alpha_i+\mathbf{x}_2'\gamma}}\;\dfrac{1}{1+e^{\alpha_i+\beta+\mathbf{x}_3'\gamma}} \\ \mathbb{P}_{(1,0,1)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} &=& p^*(1|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\;\dfrac{1}{1+e^{\alpha_i+\beta+\mathbf{x}_2'\gamma}}\;\dfrac{e^{\alpha_i+\mathbf{x}_3'\gamma}}{1+e^{\alpha_i+\mathbf{x}_3'\gamma}} \\ \mathbb{P}_{(1,1,0)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} &=& p^*(1|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\;\dfrac{e^{\alpha_i+\beta+\mathbf{x}_2'\gamma}}{1+e^{\alpha_i+\beta+\mathbf{x}_2'\gamma}}\;\dfrac{1}{1+e^{\alpha_i+\beta+\mathbf{x}_3'\gamma}} \end{cases} \tag{69}$$

Now, consider that we multiply each of these four probabilities $\mathbb{P}_{(y_1,y_2,y_3)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)}$ by the corresponding weighting value $w_{(y_1,y_2,y_3;\mathbf{x})}$ as defined in (50). We have that:

$$\begin{cases} w_{(0,0,1;\mathbf{x})}\;\mathbb{P}_{(0,0,1)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} &=& p^*(0|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\;\dfrac{e^{\alpha_i+\mathbf{x}_2'\gamma}\;e^{\alpha_i+\mathbf{x}_3'\gamma}}{\left[1+e^{\alpha_i+\mathbf{x}_2'\gamma}\right]\left[1+e^{\alpha_i+\mathbf{x}_3'\gamma}\right]} \\ w_{(0,1,0;\mathbf{x})}\;\mathbb{P}_{(0,1,0)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} &=& p^*(0|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\;\dfrac{e^{\alpha_i+\beta+\mathbf{x}_3'\gamma}-e^{\alpha_i+\mathbf{x}_2'\gamma}}{\left[1+e^{\alpha_i+\mathbf{x}_2'\gamma}\right]\left[1+e^{\alpha_i+\beta+\mathbf{x}_3'\gamma}\right]} \\ w_{(1,0,1;\mathbf{x})}\;\mathbb{P}_{(1,0,1)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} &=& p^*(1|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\;\dfrac{e^{\alpha_i+\beta+\mathbf{x}_2'\gamma}-e^{\alpha_i+\mathbf{x}_3'\gamma}}{\left[1+e^{\alpha_i+\beta+\mathbf{x}_2'\gamma}\right]\left[1+e^{\alpha_i+\mathbf{x}_3'\gamma}\right]} \\ w_{(1,1,0;\mathbf{x})}\;\mathbb{P}_{(1,1,0)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} &=& p^*(1|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\;\dfrac{e^{\alpha_i+\beta+\mathbf{x}_3'\gamma}-e^{\alpha_i+\beta+\mathbf{x}_2'\gamma}}{\left[1+e^{\alpha_i+\beta+\mathbf{x}_2'\gamma}\right]\left[1+e^{\alpha_i+\beta+\mathbf{x}_3'\gamma}\right]} \end{cases} \tag{70}$$

To obtain the expression in the right hand side of equation (49) we first add these four terms. After some operations and taking into account that $p^*(0|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)+\,p^*(1|\alpha_i,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)=1$, we get:

$$\sum_{y_1,y_2,y_3} w_{(y_1,y_2,y_3;\mathbf{x})}\;\mathbb{P}_{(y_1,y_2,y_3)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)} \;=\; \Lambda(\alpha_i+\beta+\mathbf{x}_3'\gamma)-\Lambda(\alpha_i+\mathbf{x}_3'\gamma)=\Delta^{(1)}(\alpha_i,\mathbf{x}_3) \tag{71}$$

Equation (71) holds for every value of $(\alpha_i,\ \mathbf{x}_{i1},\mathbf{x}_{i2},\mathbf{x}_{i3})$. We can integrate (71) over the distribution of $(\alpha_i,\ \mathbf{x}_{i1},\mathbf{x}_{i2},\mathbf{x}_{i3})$. In the RHS, we obtain $AME^{(1)}_{x,3}$. For the LHS, we take into that the empirical distribution $\mathbb{P}_{(y_1,y_2,y_3)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)}$ is equal to $\int \mathbb{P}_{(y_1,y_2,y_3)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\alpha_i)}\ f_{\alpha_i}(\alpha_i|\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\ d\alpha_i$. We get:

$$\sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\in\mathcal{X}^{\{1,3\}}} \mathbb{P}_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3}\left[\sum_{y_1,y_2,y_3} w_{(y_1,y_2,y_3;\mathbf{x})}\;\mathbb{P}_{(y_1,y_2,y_3)\mid(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)}\right]=AME^{(1)}_{x,3} \tag{72}$$

and $AME^{(1)}_{x,3}$ is identified. ■

7.5 Proof of Proposition 5

We proceed similarly as for the proof of Proposition 4 in section 7.4. For the sake of illustration, we present this proof for $AME_{0\to 1}$, but it proceeds the same for the other AMEs. We start with the probabilities of choice histories conditional on α_i, that is, $\mathbb{P}_{(y_1,y_2,y_3,y_4)\mid\alpha_i}$. First, we write the expression for these probabilities implied by the model as functions of parameters β and α_i. Second, for each of these probabilities, we multiply the equation times the weights $w_{(y_1,y_2,y_3,y_4)}$ that appear in the enunciate of Proposition 4. For the probabilities with non-zero weights for $AME_{0\to 1}$, we have:

$$\begin{aligned}
\frac{e^{\beta(1)}-1}{2}\left[\mathbb{P}_{(0,0,1,0)\mid\alpha_i}+\mathbb{P}_{(0,1,0,0)\mid\alpha_i}\right] &= p^*(0|\alpha_i)\ \frac{\left(e^{\beta(1)}-1\right)e^{\alpha_i}}{\left(1+e^{\alpha_i+\beta(1)}\right)\left(1+e^{\alpha_i}\right)^2}\\
\frac{e^{\beta(1)}-1}{e^{\beta(1)}}\ \mathbb{P}_{(0,0,1,1)\mid\alpha_i} &= p^*(0|\alpha_i)\ \frac{\left(e^{\beta(1)}-1\right)e^{\alpha_i}e^{\alpha_i}}{\left(1+e^{\alpha_i+\beta(1)}\right)\left(1+e^{\alpha_i}\right)^2}\\
\left(e^{\beta(1)}-1\right)\left[\mathbb{P}_{(1,0,1,0)\mid\alpha_i}+\mathbb{P}_{(1,0,1,1)\mid\alpha_i}\right] &= p^*(1|\alpha_i)\ \frac{\left(e^{\beta(1)}-1\right)e^{\alpha_i}\left(e^{\alpha_i+\beta(1)}+1\right)}{\left(1+e^{\alpha_i+\beta(1)}\right)^2\left(1+e^{\alpha_i}\right)}
\end{aligned} \tag{73}$$

Third, we sum these (three) equations for every choice history (y_1, y_2, y_3, y_4) with non-zero weight. Simplifying factors and taking into account that $p^*(0|\alpha_i)+p^*(1|\alpha_i)=1$, we get:

$$\begin{aligned}
&\frac{e^{\beta(1)}-1}{2}\left[\mathbb{P}_{(0,0,1,0)\mid\alpha_i}+\mathbb{P}_{(0,1,0,0)\mid\alpha_i}\right]+\frac{e^{\beta(1)}-1}{e^{\beta(1)}}\mathbb{P}_{(0,0,1,1)\mid\alpha_i}+\left(e^{\beta(1)}-1\right)\left[\mathbb{P}_{(1,0,1,0)\mid\alpha_i}+\mathbb{P}_{(1,0,1,1)\mid\alpha_i}\right]\\
&=\frac{\left(e^{\beta(1)}-1\right)e^{\alpha_i}}{\left(1+e^{\alpha_i+\beta(1)}\right)\left(1+e^{\alpha_i}\right)}=\frac{e^{\alpha_i+\beta(1)}}{\left(1+e^{\alpha_i+\beta(1)}\right)}-\frac{e^{\alpha_i}}{\left(1+e^{\alpha_i}\right)}=\Delta_{0\to1}(\alpha_i)
\end{aligned} \tag{74}$$

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

$$\frac{e^{\beta(1)}-1}{2}\left[\mathbb{P}_{0,0,1,0}+\mathbb{P}_{0,1,0,0}\right]+\frac{e^{\beta(1)}-1}{e^{\beta(1)}}\mathbb{P}_{0,0,1,1}+\left(e^{\beta(1)}-1\right)\left[\mathbb{P}_{1,0,1,0}+\mathbb{P}_{1,0,1,1}\right]=AME_{0\to1} \tag{75}$$

such that $AME_{0\to 1}$ is identified. We can proceed similarly to prove the identification of the others $AME_{d\to d'}$. In particular, we can prove that:

$$\begin{aligned}
AME_{1\to2} &= \frac{e^{\beta(2)}-e^{\beta(1)}}{2}\ \left[\mathbb{P}_{0,0,1,0}+\mathbb{P}_{0,1,0,0}\right]+\frac{e^{\beta(2)}-e^{\beta(1)}}{e^{\beta(1)}}\ \mathbb{P}_{0,0,1,1}\\
&+\left(\frac{e^{\beta(2)}\left(1-e^{\beta(2)}\right)}{e^{\beta(1)}}+e^{\beta(2)}-1\right)\mathbb{P}_{0,1,1,0}\\
&+\left(1-\frac{e^{\beta(1)}}{e^{\beta(2)}}\right)\left[\mathbb{P}_{1,0,1,0}+\mathbb{P}_{1,0,1,1}\right]+\left(\frac{e^{\beta(2)}-1}{e^{\beta(1)}}-1+\frac{1}{e^{\beta(2)}}\right)\mathbb{P}_{1,1,0,0}
\end{aligned} \tag{76}$$

and

$$\begin{aligned}
AME_{0\to2} &= \frac{e^{\beta(2)}-1}{2}\left[\mathbb{P}_{0,0,1,0}+\mathbb{P}_{0,1,0,0}\right]+\frac{e^{\beta(2)}-1}{e^{\beta(1)}}\ \mathbb{P}_{0,0,1,1}\\
&+\left(\frac{e^{\beta(2)}\left(1-e^{\beta(2)}\right)}{e^{\beta(1)}}+e^{\beta(2)}-1\right)\mathbb{P}_{0,1,1,0}\\
&+\left(e^{\beta(1)}-\frac{e^{\beta(1)}}{e^{\beta(2)}}\right)\left[\mathbb{P}_{1,0,1,0}+\mathbb{P}_{1,0,1,1}\right]+\left(\frac{e^{\beta(2)}-1}{e^{\beta(1)}}-1+\frac{1}{e^{\beta(2)}}\right)\mathbb{P}_{1,1,0,0}\ \blacksquare
\end{aligned} \tag{77}$$

References

Abrevaya, J., and Y. Hsu (2021): "Partial effects in non-linear panel data models with correlated random effects," *The Econometrics Journal*.

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.

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.

Dobronyi, C., J. Gu, and K. Kim (2021): "Identification of Dynamic Panel Logit Models with Fixed Effects," .

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.

Honoré, B., and A. DePaula (2021): "Identification in simple binary outcome panel data models," *The Econometrics Journal.*

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

——— (2019): "Identification in Binary Response Panel Data Models: Is Point-Identification More Common Than We Thought?," *Annals of Economics and Statistics*, 134, 207–226.

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

Honoré, B., and M. Weidner (2020): "Dynamic Panel Logit Models with Fixed Effects," *arXiv preprint arXiv:2005.05942.*

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.

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.

Pakes, A., J. Porter, M. Shepard, and S. Calder-Wang (2021): "Unobserved Heterogeneity, State Dependence, and Health Plan Choices," .

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.