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

October 29, 2022

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 three, 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, Juanjo Dolado, Juan Carlos Escanciano, Christophe Gaillac, Jiaying Gu, Francis Guiton, Bo Honore, Chris Muris, and Enrique Sentana for helpful comments and discussions. We have also benefited from the comments of seminar participants at CEMFI, Toronto, at the 2020 Econometric Society World Congress, at the 2021 IAAE Congress, at the Panel Data Workshop in Oxford, and at the Conference in Honor of Manuel Arellano at the Bank of Spain. We thank Susumu Imai for generously sharing with us the dataset that we use in our empirical application. The second author gratefully acknowledges support from the *Agencia Estatal de Investigación del Gobierno de España*, grant RTI2018-095231-B-I00, and Comunidad de Madrid (Spain), grant EPUC3M11 (V PRICIT).

†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 $T = 3$. This result can be used to have a root-N consistent estimator of the AME.

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 denote this parameter the $n-$*periods 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 show identification of average transition probabilities in a multinomial logit model, and identification of the AME in an ordered logit model. Fourth, we consider a more general dynamic discrete choice model with duration dependence and

[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."*)

prove the identification of AMEs where duration is the causal variable. All these identification results provide 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 Bonhomme (2011) and Chernozhukov, Fernandez-Val, Hahn, and Newey (2013; hereinafter CFHN).

Bonhomme (2011) considers the identification of AMEs in non-linear panel data models. It makes clear the difficulties in point identifying the AMEs in fixed effect discrete choice models with fixed T. A sufficient condition for point identifiying any AME is the existance of an injective operator relating the distribution of the unobserved heterogeneity with the observed distribution, which amounts to being able to recover the distribution function of the unobserved heterogeneity. As Bonhomme (2011) shows, such condition is not satisfied in discrete choice models. Nonetheless, Proposition 1 in Bonhomme (2011) gives a condition for point identification for which injectivity suffices but it is not necessary. Bonhomme (2011) does not give any case in discrete choice models where the condition is satisfied and there is point identification of an AME. In this regard, our contribution is to find such cases and provide the expressions that identifies the mentioned AMEs.

CFHN (2013) study the identification of AMEs in nonparametric and semiparametric binary choice models. In the nonparametric model, the distribution 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, whose first step is the identification and estimation of the β parameters.[2] Previous results on dynamic logit models establish the identification of slope parameters (Chamberlain, 1985; Honoré and Kyriazidou, 2000; Magnac, 2000, 2004; Aguirregabiria, Gu, and Luo, 2021; Honoré and Weidner, 2020; Dobronyi, Gu, and Kim,2021; and Honoré, Muris, and Weidner,2021). In the second stage, which is where we start, 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 heterogeneity. The distribution of the Probability Choices contains all the information in the data to identify the AME. They take a finite number of values, whereas the necessary and sufficient conditions for identification of the AMEs (without having any knowledge of the distribution of the fixed effects) impose

[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."*

an infinite number of restrictions. However, as we show, the logistic structure allows to transform that into a finite system of linear equations. We show cases in which the system of equations has a solution and simple manipulations 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.[3,4]

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

The rest of the paper is organized as follows. Section 2 describes the models and the AMEs of interest. Sections 3 and 4 present our main identification results. We illustrate our results using Monte Carlo experiments (in section 5) and an empirical application to a model of dynamic demand using consumer scanner data (in section 6). We summarize and conclude in section 7.

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 $\{\beta_{kj}(d) : k, j \in \mathcal{Y}, \ d = 0, 1, ...\}$ 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 $d_{it} \in \{0, 1, ...\}$ 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

[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 exogenous regressor.

[4] Other recent papers studying identification of AMEs in FE discrete choice models are Davezies, D'Haultfoeuille, and Laage (2021) and Pakel and Weidner (2021). They have only set identification results, and they require more computationally intensive methods than our simple close-forms expressions.

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. 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 varies across different versions of the general model in equation (1). We focus on four models.

(1) Model MNL-AR1. Multinomial 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 $(J + 1 = 2)$ 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.

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

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

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

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

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 $\mathbf{x}$ and duration variables.

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. 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 $\mathbf{x}$ and duration variables, 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 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 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). 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):*

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

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 virtually 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 $V_{it}(1, 0) - V_{it}(0, 0) + y_{i,t-1} \left[V_{it}(1, 1) - V_{it}(1, 0) - V_{it}(0, 1) + V_{it}(0, 0)\right] \geq 0$. 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 state dependence in market entry and it can be 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, 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 1-period AME to the power of n: 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. We show the identification results of three different average versions of this effect. A first AME is based on the condition that $\mathbf{x}$ remains constant over the T sample periods:

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

In the other two AMEs, whose identification is shown in section 4.3, the condition that $\mathbf{x}$ remains constant is not imposed. In particulat a second AME is defined as follows:

$$AME^{(1)}(\mathbf{x}_{i3} = \mathbf{x}) \equiv \int \left[\pi_{11}(\alpha_i, \mathbf{x}) - \pi_{01}(\alpha_i, \mathbf{x})\right]\ f_{\alpha|\mathbf{x}}(\alpha_i | \mathbf{x}_{i3} = \mathbf{x})\ d\alpha_i \tag{16}$$

A third AME is defined as follows:

$$AME^{(1)}_{x,t} \equiv \int [\pi_{11}(\alpha_i, \mathbf{x}_{it}) - \pi_{01}(\alpha_i, \mathbf{x}_{it})] \ f_{(\alpha,\mathbf{x}_t)}(\alpha_i, \mathbf{x}_{it}) \ d(\alpha_i, \mathbf{x}_{it}) \tag{17}$$

This third 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, $AME^{(1)}_x(x)$ 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 $AME^{(1)}_{x,t}$ is a similar effect but averaged over all the markets (firms) according to their distribution of 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:

$$AME^{(n)}_x(\mathbf{x}) = \int [\pi_{11}(\alpha_i, \mathbf{x}) - \pi_{01}(\alpha_i, \mathbf{x})]^n \ f_{\alpha|\mathbf{x}}(\alpha_i | \mathbf{x}_i^{\{1,T\}} = (\mathbf{x}, ..., \mathbf{x})) \ d\alpha_i \tag{18}$$

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'.

$$\begin{aligned} \Delta_{d \to d'}(\alpha_i) &\equiv \mathbb{E}(y_{it} \mid \alpha_i, d_{it} = d') - \mathbb{E}(y_{it} \mid \alpha_i, d_{it} = d) \\ &= \pi_{d',1}(\alpha_i) - \pi_{d,1}(\alpha_i) = \Lambda(\alpha_i + \beta(d')) - \Lambda(\alpha_i + \beta(d)). \end{aligned} \tag{19}$$

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:*

$$AME_{d \to d'} \equiv \int \Delta_{d \to d'}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i = \int [\pi_{d',1}(\alpha_i) - \pi_{d,1}(\alpha_i)] \ f_\alpha(\alpha_i) \ d\alpha_i \tag{20}$$

[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 $(\alpha_i, \mathbf{x}_{it})$.

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.

$$\Delta_{j,k\to j}(\alpha_i) \equiv \mathbb{E}\left(1\{y_{it}=j\}|\alpha_i, y_{i,t-1}=j\right) - \mathbb{E}\left(1\{y_{it}=j\}|\alpha_i, y_{i,t-1}=k\right) = \pi_{jj}(\alpha_i) - \pi_{kj}(\alpha_i). \tag{21}$$

We are interested in identification of the following AMEs:

$$AME_{j,k\to j} \equiv \int \Delta_{j,k\to j}(\alpha_i)\ f_\alpha(\alpha_i)\ d\alpha_i = \Pi_{jj} - \Pi_{kj} \tag{22}$$

and

$$ATE_{jj,t} = \Pi_{jj} - \mathbb{E}\left(1\{y_{it}=j\} \mid t\right) \tag{23}$$

3 Identification of $AME^{(1)}$, $AME^{(n)}$ in BC-AR1, and Π_{jj} in *MNL-AR1*

Since the distribution of the fixed effects cannot be non-parametrically identified in discrete choice models, the default situation is one of no-identification of any function that depends on the distribution of the fixed effects like the AME. To show that there are models and AMEs that can be point identified without any knowledge or restriction of the distribution of the fixed effects, we start by proving a simple but relevant result. It is simple both in term of the identifying expression and its proof. In subsection 3.1 we show the identification of one-period AME, $AME^{(1)}$, and of the n-periods forward AME, $AME^{(n)}$ in the simple BC-AR1 model in which the only explanatory variable is a lag of the dependent variable and there is no duration. Second, we show identification of the Average transition probabilities, Π_{jj}, in a multinomial logit without covariates nor duration dependence. This result is also easy to prove. In Section 4 we will use a procedure based on new necessary and sufficient conditions to obtain more identification results.

We take the vector of slope parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \boldsymbol{\gamma})$ as known. The identification of these parameters in FE dynamic logit models has been established in previous papers: Chamberlain (1985) for the binary choice AR(1) model without exogenous regressors; Magnac (2000) for multinomial AR(1) models; Honoré and Kyriazidou (2000) for binary and multinomial models with exogenous regressors; Aguirregabiria, Gu, and Luo (2021) for models with duration dependence; Honoré and Weidner (2020) and Dobronyi, Gu, and Kim (2021) for binary $AR(p)$ models with $p \geq 2$; and Honoré, Muris, and Weidner (2021) for the dynamic ordered logit.

3.1 Identification of AME$^{(1)}$in BC-AR1 model

Consider the binary choice AR1 model without duration or x regressors, as described by equation (5). Our proof of identification exploits a relationship between the individual effect $\Delta^{(1)}(\alpha_i)$, the transition probabilities $\pi_{01}(\alpha_i)$ and $\pi_{11}(\alpha_i)$ and the parameter β in the Logit model. The following Lemma 1 establishes this relationship.

Lemma 1. *LEMMA 1. In the BC-AR1 model, the following conditions hold:*

$$\Delta^{(1)}(\alpha_i) = [\exp\{\beta\} - 1] \ \ \pi_{01}(\alpha_i) \ \pi_{10}(\alpha_i). \tag{24}$$

$$\exp\{\beta\} = \frac{\pi_{11}(\alpha_i) \ \pi_{00}(\alpha_i)}{\pi_{10}(\alpha_i) \ \pi_{01}(\alpha_i)}. \qquad \blacksquare \tag{25}$$

Proof of Lemma 1. By definition, we have that:

$$\begin{aligned} \Delta^{(1)}(\alpha_i) &= \frac{\exp\{\alpha_i + \beta\}}{1 + \exp\{\alpha_i + \beta\}} - \frac{\exp\{\alpha_i\}}{1 + \exp\{\alpha_i\}} = \frac{\exp\{\alpha_i\} \ \ [\exp\{\beta\} - 1]}{[1 + \exp\{\alpha_i\}] \ \ [1 + \exp\{\alpha_i + \beta\}]} \\ &= [\exp\{\beta\} - 1] \ \ \pi_{01}(\alpha_i) \ \pi_{10}(\alpha_i), \end{aligned} \tag{26}$$

that give us equation (24). We also have that:

$$\begin{aligned} \frac{\pi_{11}(\alpha_i)}{\pi_{10}(\alpha_i)} \frac{\pi_{00}(\alpha_i)}{\pi_{01}(\alpha_i)} &= \frac{\exp\{\alpha_i + \beta\}/[1 + \exp\{\alpha_i + \beta\}]}{1/[1 + \exp\{\alpha_i + \beta\}]} \frac{1/[1 + \exp\{\alpha_i\}}{\exp\{\alpha_i\}/[1 + \exp\{\alpha_i\}]} \\ &= \frac{\exp\{\alpha_i + \beta\}}{\exp\{\alpha_i\}} = \exp\{\beta\}. \qquad \blacksquare \end{aligned} \tag{27}$$

Proposition 1 establishes the identification of $AME^{(1)}$ in the binary choice model *BC-AR1.*

Proposition 1. *Consider the binary choice model defined by equation (5), Assumption 1, and that β is given. For any given T, $T \geq 3$, $AME^{(1)}$ is identified as:*

$$AME^{(1)} \ \ = \ \ [\exp\{\beta\} - 1] \ \ [\mathbb{P}_{0,1,0} + \mathbb{P}_{1,0,1}] \tag{28}$$

where $\mathbb{P}_{y_1,y_2,y_3}$ represents the probability of the choice history $(y_{i1}, y_{i2}, y_{i3}) = (y_1, y_2, y_3)$. $\blacksquare$

Proof of Proposition 1: w.l.o.g. we consider $T = 3$.[6] For any sequence (y_1, y_2, y_3):

$$\mathbb{P}_{y_1,y_2,y_3} = \int p^*(y_1|\alpha_i) \ \pi_{y_1,y_2}(\alpha_i) \ \pi_{y_2,y_3}(\alpha_i) \ f_\alpha(\alpha_i) d\alpha_i \tag{29}$$

Applying equation (24) in Lemma 1 to $\mathbb{P}(0,1,0)$ and $\mathbb{P}(1,0,1)$, we have that:

$$\left\{ \begin{aligned} \mathbb{P}_{0,1,0} &= \frac{1}{\exp\{\beta\} - 1} \int p^*(0|\alpha_i) \ \Delta^{(1)}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i \\ \mathbb{P}_{1,0,1} &= \frac{1}{\exp\{\beta\} - 1} \int p^*(1|\alpha_i) \ \Delta^{(1)}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i \end{aligned} \right. \tag{30}$$

[6]Given identification with $T = 3$, it is obvious that there is also identification for any value of T greater than 3, as we can take subhistories with three periods.

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

Remark 1.1 *(Estimation).* Equation (28) provides simple analog or plug-in estimator for $AME^{(1)}$. In a first step, we estimate β using CML and the probabilities $\mathbb{P}_{y_1,y_2,y_3}$ using a frequency estimator. Then, we plug these estimates in equation (28) to obtain estimates of $AME^{(1)}$. This estimator is root$-N$ consistent.

Remark 1.2 *(Identification with $T>3$).* For $T=3$ the model is just identified. For panels with $T\geq 4$ there are over-identifying restrictions on $AME^{(1)}$. When $T\geq 4$ we can use the panel to construct the empirical distribution of 3-period histories. For each of these groups of histories, we can obtain a separate estimator of $AME^{(1)}$ such that the model implies $T-3$ over-identifying restrictions on $AME^{(1)}$. This is why to prove identification is enough to prove it with 3 periods. Nonetheless, in Appendix we will derive closed form identifying expressions for $T\geq 4$ using the procedure that we will describe in Section 4.

Corollary 1.1. The identification result in Proposition 1 and its proof holds in the model with covariates (4) if we condition on $\mathbf{x}$ that remain constant over time, so that $AME^{(1)}(\mathbf{x})$ *in (15) is identified as*

$$AME^{(1)}(\mathbf{x})=[\exp\{\beta\}-1]\ \ [\mathbb{P}_{0,1,0}(\mathbf{x})+\mathbb{P}_{1,0,1}(\mathbf{x})] \tag{31}$$

Proof of Corollary 1.1 It is straightforward to show that Lemma 1 applies also to this model with exogenous explanatory variables such that, conditional on $\mathbf{x}_{it}=\mathbf{x}$, $\Delta^{(1)}(\alpha_i,\mathbf{x})=[\exp\{\beta\}-1]\ \pi_{10}(\alpha_i,\mathbf{x})\ \pi_{01}(\alpha_i,\mathbf{x})$; and $\exp\{\beta\}=\pi_{11}(\alpha_i,\mathbf{x})\ \pi_{00}(\alpha_i,\mathbf{x})/\ \pi_{10}(\alpha_i,\mathbf{x})\ \pi_{01}(\alpha_i,\mathbf{x})$. W.l.o.g. we consider that $T=3$. Then, we can prove this Corollary using exactly the same procedure we have used for the proof of Proposition 1. $\blacksquare$

Remark 1.3. When $\mathbf{x}$ can take many values, like when having continuos covariates, the identification result in this corollary is on a set of mass zero. However, in those cases, a marginal effect conditional on a specific value of $\mathbf{x}$, i.e. an AME only for a set of individuals with mass zero, is not very interesting, as we will have a different AME for each of the many values of $\mathbf{x}$. In this case, the $AME_{x,t}^{(1)}$ defined in (17) is more interesting because it gives the average effect integrating also over the distribution of the continuos x. It gives the AME for the entire population. More important, $AME_{x,t}^{(1)}$ is not restricted to the case in which x remain constant over time. We will show the identification of $AME_{x,t}^{(1)}$ in section 4.3.

3.2 Identification of AME$^{(n)}$ in BC-AR1 model

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

Lemma 2. *Consider the binary choice model defined by equation (5) and Assumption 1. Then, the n-periods forward individual-specific causal effect $\Delta^{(n)}(\alpha_i)$ satisfies the following equation:*

$$\Delta^{(n)}(\alpha_i)=[\exp\{\beta\}-1]^n\ \ [\pi_{10}(\alpha_i)]^n\ \ [\pi_{01}(\alpha_i)]^n\,. \qquad \blacksquare \tag{32}$$

Proof of Lemma 2 Using the Markov structure of the model and the chain rule, we have that:

$$\begin{aligned}\mathbb{E}(y_{i,t+n} \mid \alpha_i, y_{it}) &= \mathbb{P}\left(y_{i,t+n-1}=0 \mid \alpha_i, y_{it}\right)\ \pi_{01}(\alpha_i) + \mathbb{P}\left(y_{i,t+n-1}=1 \mid \alpha_i, y_{it}\right)\ \pi_{11}(\alpha_i) \\ &= \pi_{01}(\alpha_i) + \mathbb{E}\left(y_{i,t+n-1} \mid \alpha_i, y_{it}\right)\ \left[\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)\right]\end{aligned} \tag{33}$$

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 (33), we have that:

$$\begin{aligned}\Delta^{(n)}(\alpha_i) &= \left[\mathbb{E}\left(y_{i,t+n-1} \mid \alpha_i, y_{it}=1\right) - \mathbb{E}\left(y_{i,t+n-1} \mid \alpha_i, y_{it}=0\right)\right]\ \left[\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)\right] \\ &= \Delta^{(n-1)}(\alpha_i)\ \left[\pi_{11}(\alpha_i) - \pi_{01}(\alpha_i)\right]\end{aligned} \tag{34}$$

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 3 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 3, $\exp\{\beta\}\ \pi_{10}(\alpha_i)\ \pi_{01}(\alpha_i) = \pi_{11}(\alpha_i)\ \pi_{00}(\alpha_i)$. This implies that $[\exp\{\beta\}-1]\ \pi_{10}\ \pi_{01} = \pi_{11}\ \pi_{00} - \pi_{10}\ \pi_{01} = \pi_{11}(1-\pi_{01}) - (1-\pi_{11})\pi_{01} = \pi_{11} - \pi_{01}$. ■

Proposition 2. *Consider the binary choice model defined by equation (5) and Assumption 1. Let n be any positive integer, and let $\widetilde{\mathbf{10}}^n$ be the choice history that consists of the n-times repetition of the sequence (1,0), e.g., for $n=2$, we have that $\widetilde{\mathbf{10}}^2 = (1,0,1,0)$. If $T \geq 2n+1$, then parameter $AME^{(n)}$ is identified as:*

$$AME^{(n)} = [\exp\{\beta\}-1]^n\ \left[\mathbb{P}_{0,\widetilde{\mathbf{10}}^n} + \mathbb{P}_{\widetilde{\mathbf{10}}^n,1}\right] \tag{35}$$

Proposition 3. *where $\mathbb{P}_{0,\widetilde{\mathbf{10}}^n}$ and $\mathbb{P}_{\widetilde{\mathbf{10}}^n,1}$ are the probabilities of choice histories $(0,\widetilde{\mathbf{10}}^n)$ and $(\widetilde{\mathbf{10}}^n,1)$.* ■

Proof of Proposition 2: 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:

$$\left\{\begin{aligned}\mathbb{P}_{0,\widetilde{\mathbf{10}}^n} &= \int p^*(0|\alpha_i)\ [\pi_{10}(\alpha_i)]^n\ \ [\pi_{01}(\alpha_i)]^n\ f_\alpha(\alpha_i)\ d\alpha_i \\ \mathbb{P}_{\widetilde{\mathbf{10}}^n,1} &= \int p^*(1|\alpha_i)\ [\pi_{10}(\alpha_i)]^n\ \ [\pi_{01}(\alpha_i)]^n\ f_\alpha(\alpha_i)\ d\alpha_i\end{aligned}\right. \tag{36}$$

Applying equation (32) from Lemma 2, we have that:

$$\left\{\begin{aligned}\mathbb{P}_{0,\widetilde{\mathbf{10}}^n} &= \frac{1}{[\exp\{\beta\}-1]^n}\int p^*(0|\alpha_i)\ \Delta^{(n)}(\alpha_i)\ f_\alpha(\alpha_i)\ d\alpha_i \\ \mathbb{P}_{\widetilde{\mathbf{10}}^n,1} &= \frac{1}{[\exp\{\beta\}-1]^n}\int p^*(1|\alpha_i)\ \Delta^{(n)}(\alpha_i)\ f_\alpha(\alpha_i)\ d\alpha_i\end{aligned}\right. \tag{37}$$

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. ■

Corollary 2.1. The identification result in Proposition 2 and its proof holds in the model with covariates (4) if we condition on $\mathbf{x}$ that remain constant over time, $\mathbf{x}_i^{\{1,2n+1\}}=(\mathbf{x},...,\mathbf{x})$, so that

$$AME^{(n)}(\mathbf{x})=[\exp\{\beta\}-1]^n \quad \left[\mathbb{P}_{0,\widetilde{\mathbf{10}}^n}(\mathbf{x})+\mathbb{P}_{\widetilde{\mathbf{10}}^n,1}(\mathbf{x})\right] \tag{38}$$

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 $\mathbf{x}_i^{\{1,2n+1\}}=(\mathbf{x},...,\mathbf{x})$. ■

3.3 Identification of Π_{jj} in multinomial model

Consider the multinomial model without duration in equation (2). To obtain our identification results for the multinomial model, we apply the following Lemma, which is an extension of Lemma 1 to the multinomial case.

Lemma 3. *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:*

$$\exp\left\{\beta_{k\ell}(d)-\beta_{kj}(d)+\beta_{jj}(d)-\beta_{j\ell}(d)\right\}=\frac{\pi_{k\ell}(\alpha_i,\mathbf{x},d)\ \pi_{jj}(\alpha_i,\mathbf{x},d)}{\pi_{kj}(\alpha_i,\mathbf{x},d)\ \pi_{j\ell}(\alpha_i,\mathbf{x},d)}. \quad ■ \tag{39}$$

Proof of Lemma 3: 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 (39).

Proposition 4 establishes the identification of the average transition probabilities $\Pi_{jj}(\mathbf{x})$ in the logit model without duration dependence.

Proposition 4. *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,*

$$\Pi_{jj}(\mathbf{x}) \;=\; \mathbb{P}_{j,j}(\mathbf{x})+\sum_{k\neq j}\left[\mathbb{P}_{k,j,j}(\mathbf{x})+\sum_{\ell\neq j}\exp\left\{\beta_{k\ell}-\beta_{kj}+\beta_{jj}-\beta_{j\ell}\right\}\ \mathbb{P}_{k,j,\ell}(\mathbf{x})\right], \tag{40}$$

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 $\mathbf{x}_i^{\{1,3\}}=(\mathbf{x},\mathbf{x},\mathbf{x})$. ■

Proof of Proposition 4: 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:

$$\Pi_{jj} = \int \left[p^*(0|\alpha_i) + p^*(1|\alpha_i) ... + p^*(J|\alpha_i)\right] \ \pi_{jj}(\alpha_i) \ f_\alpha(\alpha_i) \ d\alpha_i \tag{41}$$

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

$$\delta^{(1)}_{k,j,j} = \int p^*(k|\alpha_i) \ \left[\pi_{k0}(\alpha_i) + \pi_{k1}(\alpha_i) + ... + \pi_{kJ}(\alpha_i)\right] \ \pi_{jj}(\alpha_i) \ f_\alpha(\alpha_i) \tag{42}$$

That is, we have that $\delta^{(1)}_{k,j,j} = \sum_{\ell=0}^{J} \delta^{(2)}_{k,\ell,j,j}$, with $\delta^{(2)}_{k,\ell,j,j} \equiv \int p^*(k|\alpha_i) \ \pi_{k\ell}(\alpha_i) \ \pi_{jj}(\alpha_i) \ f_\alpha(\alpha_i)$. For $\ell = j$, we have that $\delta^{(2)}_{k,j,j,j}$ corresponds to the choice history probability $\mathbb{P}_{k,j,j}$. For the rest of the terms $\delta^{(2)}_{k,\ell,j,j}$, we apply Lemma 3. According to Lemma 3, we have that $\pi_{k\ell}(\alpha_i)$ $\pi_{jj}(\alpha_i) = \exp\{\beta_{k\ell} - \beta_{kj} - \beta_{j\ell}\} \ \pi_{kj}(\alpha_i) \ \pi_{j\ell}(\alpha_i)$. Finally, note that $\int p^*(k|\alpha_i) \ \pi_{kj}(\alpha_i) \ \pi_{j\ell}(\alpha_i)$ $f_\alpha(\alpha_i)$ is the choice history probability $\mathbb{P}_{k,j,\ell}$. Putting all the pieces together, we have that the expression in equation (40). ■

Corollary 3.1. Proposition 4 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 3.2. Proposition 4 implies the identification result in Proposition *1* for the binary choice model. In particular, it is straightforward to verify that for the binary choice model, equation (40) implies (note that $\beta_0 = 0$): $\Pi_{11} = \mathbb{P}_{1,1} + \mathbb{P}_{0,1,1} + \exp\{\beta_1\} \ \mathbb{P}_{0,1,0}$, and $\Pi_{00} = \mathbb{P}_{0,0} + \mathbb{P}_{1,0,0} + \exp\{\beta_1\} \ \mathbb{P}_{1,0,1}$. In the binary case, we have that $\Pi_{01} = 1 - \Pi_{00}$ such that Π_{01} is also identified, and so is $AME^{(1)} = \Pi_{11} - \Pi_{01}$.

Corollary 3.3. 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.

Remark 3.1. Unfortunately, the procedure described in the proof of Proposition 4 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. In next section, based on necessary and sufficient conditions, we will prove that this is not possible.

4 Identification results from necessary and sufficient conditions

Identification results in the previous section show that, despite the general difficulty because of the non-identification of the fixed effects distribution in discrete choice models, there are cases in which parameters of interest can be fixed effect identified with fixed-T. However, they raise a number of questions. The proofs start from a given linear combination (weighted sum) of the Probability choices and show they are equal to the AME, but, how can those weights be obtained? Is there a procedure to check whether a linear combination identifying the AME exists for other AMEs and models?

In section 4.1, starting from necessary and sufficient conditions, a constructive approach is provided to obtain AMEs in a general class of dynamic logit models. We apply this approach to obtain different AMEs or causal effects, such as: in the binary AR(1) model, the effect of last period choice when having different T, and the AME when having exogenous explanatory variables (in section 4.3); the average transition probabilities Π_{jj} and AME in the ordered logit (in section 4.5); and the effect of a change in duration (in section 4.6). We also use the approach to show that the AME in the (unordered) multinomial model is not identified (in section 4.4).

4.1 Procedure to study identification of AMEs in discrete choice models based on necessary and sufficient conditions

Let $\mathbf{y}_i \equiv (d_{i1}, y_{i1}, y_{i2}, ..., y_{iT}) \in \mathcal{D} \times \mathcal{Y}^T$ be the vector with individual i's choice history, including the initial duration, and let $\mathbf{x}_i \in \mathcal{X}^T$ represent the history of the exogenous variables, $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$. We use $\mathbf{y}_i^{\{2,T\}} \in \mathcal{Y}^{T-1}$ to denote the sub-history from period 2 to T. For arbitrary histories $\mathbf{y} \in \mathcal{D} \times \mathcal{Y}^T$ and $\mathbf{x} \in \mathcal{X}^T$, let $\mathbb{P}_{\mathbf{y}|\mathbf{x}}$ represent the probability $P(\mathbf{y}_i = \mathbf{y}|\mathbf{x}_i = \mathbf{x})$. This probability is identified from the data. Let $\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}}$ be the vector with the probabilities $\mathbb{P}_{\mathbf{y}|\mathbf{x}}$ for every possible value of $\mathbf{y}$ and $\mathbf{x}$: i.e., $\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}} \equiv \{\mathbb{P}_{\mathbf{y}|\mathbf{x}} : \mathbf{y} \in \mathcal{D} \times \mathcal{Y}^T, \mathbf{x} \in \mathcal{X}^T\}$. Vector $\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}}$ contains all the information in the data that is relevant to identify the parameters of interest $\boldsymbol{\theta}$, the distribution of α, and any AME of interest.

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

$$\mathbb{P}_{\mathbf{y}|\mathbf{x}} \;=\; \int G\left(\mathbf{y}^{\{2,T\}}|y_1, d_1, \mathbf{x}, \boldsymbol{\alpha}; \boldsymbol{\theta}\right)\; p^*(y_1, d_1|\boldsymbol{\alpha}, \mathbf{x})\; f_\alpha(\boldsymbol{\alpha}|\mathbf{x})\; d\alpha, \tag{43}$$

where $p^*(y_1, d_1|\boldsymbol{\alpha})$ is the probability of the initial condition given α and $\mathbf{x}$, and $G(\mathbf{y}^{\{2,T\}}|y_1, d_1, \mathbf{x}, \boldsymbol{\alpha}; \boldsymbol{\theta})$ is the probability of sub-history $\mathbf{y}^{\{2,T\}}$ predicted by the model. More specifically:

$$G\left(\mathbf{y}^{\{2,T\}}|y_1, d_1, \mathbf{x}, \boldsymbol{\alpha}; \boldsymbol{\theta}\right) \;\equiv\; \prod_{t=2}^{T} \Lambda\left(y_t|y_{t-1}, d_t, \mathbf{x}_t, \boldsymbol{\alpha}; \boldsymbol{\theta}\right), \tag{44}$$

and $\Lambda\left(y_t|y_{t-1}, d_t, \mathbf{x}_t, \boldsymbol{\alpha}; \boldsymbol{\theta}\right)$ is the logit transition probability from the model.

Let $\Delta(\boldsymbol{\alpha}_i, \mathbf{x}, \boldsymbol{\theta})$ be an individual marginal effect, and let $AME(\mathbf{x}) \equiv \int \Delta(\boldsymbol{\alpha}_i, \mathbf{x}, \boldsymbol{\theta})\; f_\alpha(\boldsymbol{\alpha}_i|\mathbf{x})\; d\boldsymbol{\alpha}_i$ be the corresponding average marginal effect. In general, we can say that this AME is point identified if there is a function $h(\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}}, \boldsymbol{\theta})$ such that $AME(\mathbf{x}) = h(\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}}, \boldsymbol{\theta})$. Lemma 4 states a necessary and sufficient condition for the point identification of AME in a broad class of FE dynamic discrete choice models that includes our logit models as a particular case.

Lemma 4. *Consider a FE dynamic discrete choice model characterized by the probability function* $G(\mathbf{y}^{\{2,T\}}|y_1, d_1, \mathbf{x}, \boldsymbol{\alpha}; \boldsymbol{\theta})$*. Let* $AME(\mathbf{x}) \equiv \int \Delta(\boldsymbol{\alpha}_i, \mathbf{x}, \boldsymbol{\theta}) f_\alpha(\boldsymbol{\alpha}_i|\mathbf{x})\; d\boldsymbol{\alpha}_i$ *be an average* *marginal effect of interest. Then, this AME is point identified if and only if there is a weighting function* $w(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})$ *from* $\mathcal{D} \times \mathcal{Y}^T \times \mathcal{X}^T \times \Theta \rightarrow \mathbb{R}$ *that satisfies the following equation:*

$$\sum_{\mathbf{y}^{\{2,T\}} \in \mathcal{Y}^{T-1}} w(d_1, y_1, \mathbf{y}^{\{2,T\}}, \mathbf{x}, \boldsymbol{\theta})\; G\left(\mathbf{y}^{\{2,T\}} | y_1, d_1, \mathbf{x}, \boldsymbol{\alpha}; \boldsymbol{\theta}\right) \;=\; \Delta(\boldsymbol{\alpha}, \mathbf{x}, \boldsymbol{\theta}), \tag{45}$$

for every value $(d_1, y_1) \in \mathcal{D} \times \mathcal{Y}$ *and every* $\boldsymbol{\alpha} \in \mathbb{R}^J$*. Furthermore, this condition implies the following form for the function that identifies* $AME(\mathbf{x})$*:*

$$AME(\mathbf{x}) \;=\; h(\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}}, \boldsymbol{\theta}) \;=\; \sum_{\mathbf{y} \in \mathcal{D} \times \mathcal{Y}^T} w(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})\; \mathbb{P}_{\mathbf{y}|\mathbf{x}} \qquad \blacksquare \tag{46}$$

Proof. In the Appendix, section 8.1. □

While the proofs are different, the results in Lemma 4, requiring the AME to be expressed as a linear function of the observables, are not new. They can be found in Bonhomme (2011) for non-linear panel data models, and in Severini and Tripathi (2012) Severini and Tripathi (2012) in the context of nonparametric IV regression.

Lemma 4 does not impose any restriction on the form of function G. For instance, it does not require the structure in equation (44) above. Therefore, Lemma 4 applies to a general class of FE dynamic discrete choice models, and not only to the logit class. Equation (45) defines, for every given value of (d_1, y_1), an infinite system of equations – as many as values of $\boldsymbol{\alpha}_i$. The researcher knows the closed-form expressions for functions $G(.|\boldsymbol{\alpha}, \mathbf{x}, \boldsymbol{\theta})$ and $\Delta(\boldsymbol{\alpha}, \mathbf{x}, \boldsymbol{\theta})$. The unknowns in this infinite system of equations are the weights $w(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})$ for every $\mathbf{y} \in \mathcal{Y}$. Since the set $\mathcal{Y}$ is finite, we have a system with infinite restrictions and a finite number of unknowns. Without some particular structure, this system does not have a solution.

Lemma 5 shows that in the FE dynamic logit model, the structure of functions $G(.|\boldsymbol{\alpha}, \mathbf{x}, \boldsymbol{\theta})$ and $\Delta(\boldsymbol{\alpha}, \mathbf{x}, \boldsymbol{\theta})$ is such that equation (45) can be represented as a finite order polynomial in the variables $\exp\{\alpha_i(j)\}$ for $j = 1, 2, ..., J$. This implies that there is a solution to the system if and only if the coefficients multiplying every monomial term in this polynomial are all equal to zero. This property transforms the infinite system of equations into a finite linear system with finite unknowns. Furthermore, if a solution exists, this solution implies a closed-form expression for the weights $w(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})$, and therefore, for AME.

Lemma 5. *Consider the FE dynamic logit model defined by equation (1) and Assumption 1. Equation* (45) *can be represented as a finite order polynomial in the variables* $\exp\{\alpha_i(j)\}$ *for* $j = 1, 2, ..., J$*. This implies a system with a finite number of linear equations with respect to the unknown finite number of weights* $w(\mathbf{y}, \mathbf{x}, \boldsymbol{\theta})$ *for every* $\mathbf{y} \in \mathcal{Y}$*.* $\blacksquare$

Proof. In the Appendix, section 8.2. □

Once we have the identification problem reduced to a system of linear equations, we can check whether or not the weights that identify $AME(\mathbf{x})$ exist: that is, whether the system of equations is compatible such that at least one solution exists. If it is compatible, then we can obtain the weights by solving the system. Uniqueness or multiplicity of the solution does not affect the identification result. Given the weights, we can calculate $AME(\mathbf{x})$ using equation (46). In sections 4.2 to 4.6, we apply this approach to obtain closed-form expressions for different AMEs and different versions of the FE dynamic logit model. Those applications will help understanding the result in Lemma 5, and how it is used to study identification.

4.2 One-period AMEs in the BC-AR1 model

We start using Lemmas 4 and 5 in the BC-AR1 model without $\mathbf{x}$ covariates as described in equation (5), and the identification of $AME^{(1)}$ defined in equation (9). The identification of this has already being shown in section 3.1. Thus, the purpose of this section is to illustrate to use of the procedure that arises from section 4.1. We do this in the Appendix, section 8.3, showing that the weights that satisfy (45) in the BC-AR1 model and $AME^{(1)}$ lead to the linear combination in (28), $AME^{(1)} = [\exp\{\beta\} - 1] \, [\mathbb{P}_{0,1,0} + \mathbb{P}_{1,0,1}]$. Additionally, this way of obtain the weights and proving identifications shows that the weights are unique and the model does not provide additional restrictions on AME when $T = 3$.

In section 8.4 in the Appendix, we use the same procedure presented in previous subsection to obtain the closed-form expression of the weights for different values of T. As said, for panels with $T \geq 4$ there are over-identifying restrictions on $AME^{(1)}$, because we can use the panel to construct the empirical distribution of 3-period histories. There we obtain one of the possible combinations, using all T periods without having to make combinations of 3 periods.

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

The identification results in section 3.1 without $\mathbf{x}$ have been extended in Corollary 1.1 only to the case where the $\mathbf{x}$ variables are constant over time. 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. Furthermore, when $\mathbf{x}$ can take many values, like when having continuos covariates, the identification result for constant $\mathbf{x}$ is on a set of mass zero. We would like to obtain an AME that does not have these limitations and it is not restricted to $\mathbf{x}$ taking a constant value. This type of AME corresponds to $AME_{x,t}^{(1)}$ that we have defined in equation (17). Proposition 5 establishes the identification of $AME_{x,t}^{(1)}$.

Proposition 5. *Consider the binary choice model defined by equation (4) and Assumption 1, and suppose that $T \geq 3$. Then, $AME_{x,t}^{(1)}$, as defined in equation (17), is identified for any period $t \geq 3$ in the sample. For instance, for $T = 3$ and $t = 3$, we have that:*

$$AME_{x,3}^{(1)} = \sum_{(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)\in\mathcal{X}^{\{1,3\}}} \mathbb{P}_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \begin{bmatrix} w_{(0,0,1;\mathbf{x})} \; \mathbb{P}_{(0,0,1) \mid (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)} \\ +w_{(0,1,0;\mathbf{x})} \; \mathbb{P}_{(0,1,0) \mid (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)} \\ +w_{(1,0,1;\mathbf{x})} \; \mathbb{P}_{(1,0,1) \mid (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)} \\ +w_{(1,1,0;\mathbf{x})} \; \mathbb{P}_{(1,1,0) \mid (\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3)} \end{bmatrix} \quad (47)$$

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:

$$\begin{aligned} w_{(0,0,1;\mathbf{x})} &= \frac{e^{\mathbf{x}_2'\gamma}-e^{\mathbf{x}_3'\gamma}}{e^{\mathbf{x}_3'\gamma}};\ w_{(0,1,0;\mathbf{x})} = \frac{e^{\beta+\mathbf{x}_3'\gamma}-e^{\beta+\mathbf{x}_2'\gamma}}{e^{\mathbf{x}_2'\gamma}}; \\ w_{(1,0,1;\mathbf{x})} &= \frac{e^{\beta+\mathbf{x}_2'\gamma}-e^{\mathbf{x}_3'\gamma}}{e^{\mathbf{x}_3'\gamma}};\ w_{(1,1,0;\mathbf{x})} = \frac{e^{\mathbf{x}_3'\gamma}-e^{\mathbf{x}_2'\gamma}}{e^{\mathbf{x}_2'\gamma}}. \qquad \blacksquare \end{aligned} \tag{48}$$

Proof. In section 8.5 in the Appendix. □

Remark 5.1. Proposition 5 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 5 assumes that the support of $\mathbf{x}_{it}$ is discrete, this identification result trivially extends to the case of continuous $\mathbf{x}$ variables.

Remark 5.2. There is a relationship between the identification of $AME^{(1)}_{x,t}$ in Proposition 5 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 $AME^{(1)}_{x,t}$ in equation (48) are equal to the weights to obtain $AME^{(1)}(\mathbf{x})$ in equation (31). That is:

$$w_{(0,0,1;\mathbf{x})} = w_{(1,1,0;\mathbf{x})} = \frac{e^{\mathbf{x}'\gamma}-e^{\mathbf{x}'\gamma}}{e^{\mathbf{x}'\gamma}} = 0\ ;\ w_{(0,1,0;\mathbf{x})} = w_{(1,0,1;\mathbf{x})} = \frac{e^{\beta+\mathbf{x}'\gamma}-e^{\mathbf{x}'\gamma}}{e^{\mathbf{x}'\gamma}} = e^{\beta}-1. \tag{49}$$

Corollary 5.1. *Everything in Proposition 5 and in its proof holds if we condition everywhere on $x_{i3}=x$, so that we could estimate*

$$AME^{(1)}(x_{i3}=x) = \sum_{(\mathbf{x}_1,\mathbf{x}_2)\in\mathcal{X}^{\{1,2\}}} \mathbb{P}_{(\mathbf{x}_1,\mathbf{x}_2)|\,\mathbf{x}_3=x} \sum_{\mathbf{y}_1^3\in\Gamma_T} w_{(y_1,y_2,y_3;\mathbf{x})}\ \mathbb{P}_{(y_1,y_2,y_3)\,|\,(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3=x)} \tag{50}$$

where the weights $w_{(y_1,y_2,y_3;\mathbf{x})}$ are the same as in Proposition 5. ■

$AME^{(1)}(x_{i3}=x)$ is like $AME^{(1)}(x)$ but without imposing that x remains constant over the sample periods. In the first two periods x can take any value, and the AME is only conditional on $x_{i3}=x$.

4.4 No identification of the AME in the multinomial model

Consider the multinomial model without duration in equation (2). In section 3.3 we have shown identification of the average transition probabilities Π_{jj} and of $ATE_{jj,t}$. However we would also like to study identification of $AME_{j,k\to j}$ defined in equation (22) and that it is equal to $\Pi_{jj}-\Pi_{kj}$. To that end we need to identify Π_{kj} with $j\neq k$. Unfortunately, the identification of Π_{jj} in Proposition 4 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. Furthermore, in the following proposition we prove that the identification of Π_{10} in a model with 3 choices and with $T=3$ is not possible because the necessary conditions specified in Lemma 4 are not satisfied.

Proposition 6. *Consider the model without duration dependence in equation (2) under Assumption 1, given β and γ, and 3 choices (i.e. $J+1=3$). If $T=3$, there is no function $h\left(\mathbf{P}_{\mathbf{y}|\mathbf{x}},\theta\right)$ of the observed probability choices $\mathbf{P}_{\mathbf{y}|\mathbf{x}}$ and θ that equal the average transition probability $\Pi_{10}(\mathbf{x})$.* ■

Proof. In section 8.6 in the Appendix. □

Proving this for Π_{10} is enough to prove it is not possible to identify Π_{jk} for all possible values of j and k, with $j \neq k$. If we cannot identify Π_{jk}, then we cannot obtain $AME_{j,k\to j} \equiv \Pi_{jj} - \Pi_{kj}$ in model (2) with more than two alternatives, because knowing two of the three elements imply knowing the third. Contrary to this negative result, in next subsection we show the identification in an ordered logit, which is a restricted version of our multinomial choice model in which the choices are ordered and the individual specific effects are common to all alternatives, that is $\alpha_i(j) = \alpha_i$.

4.5 Identification of average transition probabilities and AMEs in an ordered logit model

Consider the following ordered logit model without duration dependence, for $J+1$ possible choices,[7]

$$\begin{aligned} y_{it} &= j \quad \text{if} \quad y_{it}^* \in (\lambda_{j-1}, \lambda_j], \text{ with } j = 0, 1, ..., J \\ y_{it}^* &= \alpha_i + \sum_{k=0}^{J-1} \beta_k \, 1\{y_{i,t-1} = k\} \; + \varepsilon_{it}. \end{aligned} \tag{51}$$

with, $\lambda_{-1} := -\infty$ and $\lambda_{J+1} := +\infty$, and ε_{it} follows a logistic distribution. This implies

$$\Pr\left(y_{it} = j | y_{it-1} = k, \alpha_i\right) = \Lambda\left(\beta_k + \alpha_i - \lambda_{j-1}\right) - \Lambda\left(\beta_k + \alpha_i - \lambda_j\right) \tag{52}$$

Honoré, Muris, and Weidner (2021) establish the identification of the β and λ parameters (under some normalization, like, for example, $\beta_0 = \lambda_0 = 0$). The following proposition establishes the identification of all the Average Transition Probabilities from any previous choice in $t-1$ to choice 0 in period t, Π_{k0} and, therefore, the identification of the Average Marginal Effect of a previous choice on the probability of choosing 0 in period t.

Proposition 7. *Consider the ordered logit model without duration dependence in equations (51) with $J+1=3$ and given the values of β and λ parameters. If $T \geq 3$, the average*

[7] For simplicity we write it without $\mathbf{x}$, but the extension of the results conditioning to $\mathbf{x}$ taking constant values, as in some of the previous results, is straight forward.

transition probabilities $\{\Pi_{k0} : k \in \{0, 1, 2\}\}$ *are identified using the following equation,*

$$\Pi_{00} = \sum_{k=0}^{J} \mathbb{P}_{k,0,0} + \sum_{k=0}^{J} \sum_{l=1}^{J} \exp\{\beta_k - \beta_0\} \mathbb{P}_{k,0,l} \tag{53}$$

$$\Pi_{10} = \sum_{k=0}^{J} \sum_{l=0}^{1} \mathbb{P}_{k,0,l} + \sum_{k=0}^{1} \exp\{\beta_k - \beta_1\} \mathbb{P}_{k,0,2} + \mathbb{P}_{2,0,2} + (\exp\{\beta_2 - \beta_1\} - 1) \mathbb{P}_{2,2,0} \tag{54}$$

$$+ \sum_{k=0}^{J} \frac{(\exp\{\beta_1 - \lambda_0\} - \exp\{\beta_k - \lambda_1\}) (\exp\{\beta_k\} - \exp\{\beta_1\})}{(\exp\{\beta_k - \lambda_0\} - \exp\{\beta_k - \lambda_1\}) \exp\{\beta_1\}} \mathbb{P}_{k,1,0} \tag{55}$$

$$\Pi_{20} = \mathbb{P}_{0,0,0} + \mathbb{P}_{0,0,1} + \exp\{\lambda_1 - \lambda_0\} \mathbb{P}_{0,0,2} + \left(1 - \frac{\exp\{\beta_2 - \lambda_0\}}{\exp\{\beta_0 - \lambda_1\}}\right) \mathbb{P}_{0,2,0} \tag{56}$$

$$+ \sum_{k=0}^{l} \sum_{l=0}^{J} \mathbb{P}_{1,k,l} + \left(1 - \frac{\exp\{\beta_2 - \lambda_0\}}{\exp\{\beta_1 - \lambda_1\}}\right) \mathbb{P}_{1,2,0} + \sum_{k=0}^{J-1} \mathbb{P}_{2,0,k} \tag{57}$$

where $\mathbb{P}_{y_1,y_2,y_3}$ *represent the probability of choice histories* $(y_{i1}, y_{i2}, y_{i3}) = (y_1, y_2, y_3)$. . ■

Proof. In section 8.7 in the Appendix. □

Corollary 6.1. From Π_{k0} we can obtain the average marginal effect on the Probability of having $y_{it} = 0$ of moving from $y_{it-1} = k$ to $y_{it-1} = 0$:

$$\begin{aligned} AME_{k0} &\equiv \int (\pi_{00}(\alpha_i) - \pi_{k0}(\alpha_i)) f_\alpha(\alpha_i) d\alpha_i \\ &= \Pi_{00} - \Pi_{k0} \end{aligned}$$

As has been done for Π_{k0}, the general procedure and result in Section 4.1 can be used to obtain the combination of observed probabilities that identify any other Π_{kj}.

4.6 Identification of AMEs of changes in duration

Proposition 8. *Consider the binary choice model with duration dependence defined by equation (3) and Assumption 1, and suppose that* $T \geq 4$. *Under these conditions,* $AME_{0\to1}$, $AME_{1\to2}$, *and* $AME_{0\to2}$ *– as defined in equation (20) – are identified.*

$$\begin{aligned} AME_{0\to1} \quad &= \quad \frac{e^{\beta(1)} - 1}{2} \left[\mathbb{P}_{0,0,1,0} + \mathbb{P}_{0,1,0,0}\right] + \frac{e^{\beta(1)} - 1}{e^{\beta(1)}} \mathbb{P}_{0,0,1,1} \\ &+ \quad \left(e^{\beta(1)} - 1\right) \left[\mathbb{P}_{1,0,1,0} + \mathbb{P}_{1,0,1,1}\right] \end{aligned} \tag{58}$$

In section 8.8 in the Appendix, we provide the expression for the identification of $AME_{1\to2}$ *and* $AME_{0\to2}$. ■

Proof. In section 8.8 in the Appendix. □

5 Monte Carlo experiments

The purpose of these Monte Carlo experiments is twofold. First, we compare the bias and variance of the FE estimator of $AME^{(1)}$ to those from RE estimators imposing restrictions that we 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 RE and FE estimators of slope parameters, and a Hausman test based on the RE and FE estimators of AMEs.

The DGP is a binary choice AR1 model as 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 is $T = 4$. We implement experiments for two sample sizes N, 1000 and 2000. We consider six DGPs based on two values of parameter β (i.e., $\beta = -1$ and $\beta = 1$) and three distributions of the unobserved heterogeneity α_i: no heterogeneity, such that $\alpha_i = 0$ for every i; finite mixture with two points of support, $\alpha_i = -1$ with probability 0.3, and $\alpha_i = 0.5$ with probability 0.7; and a mixture of two normal random variables: $\alpha_i \sim N(-1, 3)$ with probability 0.3, and $\alpha_i \sim N(0.5, 3)$ with probability 0.7.

Table 1 summarizes the six DGPs, the labels we use to represent them, and the corresponding value of $AME^{(1)}$ in the population. Keeping parameter β constant, the AME can vary substantially when we change the distribution of the unobserved heterogeneity. 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 for mixture of normal distributions.

Table 1
DGPs and true value of AME

	Distribution of α_i		
Value of β	*No α_i*	*Finite mixture*	*Mixture of normals*
$\beta = -1$	DGP *NoUH(-1)* $AME^{(1)} = -0.2311$	DGP *FinMix(-1)* $AME^{(1)} = -0.2164$	DGP *MixNor(-1)* $AME^{(1)} = -0.113$
$\beta = 1$	DGP *NoUH(+1)* $AME^{(1)} = 0.2311$	DGP *FinMix(+1)* $AME^{(1)} = 0.2059$	DGP *MixNor(+1)* $AME^{(1)} = 0.1108$

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 that

[8]The FE estimator of β 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 7 in the Appendix. In this formula, we replace parameter β with its CML estimate, and the probabilities of choice histories with frequency estimates.

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] Table 2 present results from experiments with $N = 1000$.[10]

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 NoUH(-1)	*FE-CMLE*	-1.0	-1.0074	0.1310	-0.2311	-0.2314	0.0235	0.0235
	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 FinMix(-1)	*FE-CMLE*	-1.0	-1.0012	0.1338	-0.2164	-0.2160	0.0222	0.0221
	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 MixNor(-1)	*FE-CMLE*	-1.0	-1.0136	0.2160	-0.1113	-0.1110	0.0178	0.0176
	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 NoUH(+1)	*FE-CMLE*	1.0	1.0013	0.1654	0.2311	0.2344	0.0526	0.0527
	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 FinMix(+1)	*FE-CMLE*	1.0	0.9982	0.1841	0.2059	0.2089	0.0539	0.0539
	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 MixNor(+1)	*FE-CMLE*	1.0	1.0055	0.2873	0.1108	0.1169	0.0511	0.0515
	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

(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 $AME^{(1)}$. In this FE approach, the estimation of $AME^{(1)}$ does not involve a substantially larger bias than the 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).

[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 a key advantage of FE estimation.

[10]The results for sample size $N = 2000$ are qualitative very similar except that, as one would expect, all the estimators have lower variance when the sample size increases. For this reason, we present here only results from experiments with $N = 1000$.

(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. 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:

$$HS_{\beta} = \frac{\left(\widehat{\beta}_{FE} - \widehat{\beta}_{RE}\right)^2}{\widehat{Var}\left(\widehat{\beta}_{FE}\right) - \widehat{Var}\left(\widehat{\beta}_{RE}\right)} \quad \text{under } H_0 \quad \sim \chi^2_1 \tag{59}$$

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

$$HS_{AME} = \frac{\left(\widehat{AME}_{FE} - \widehat{AME}_{RE}\right)^2}{\widehat{Var}\left(\widehat{AME}_{FE}\right) - \widehat{Var}\left(\widehat{AME}_{RE}\right)} \quad \text{under } H_0 \quad \sim \chi^2_1 \tag{60}$$

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

Figure 1.

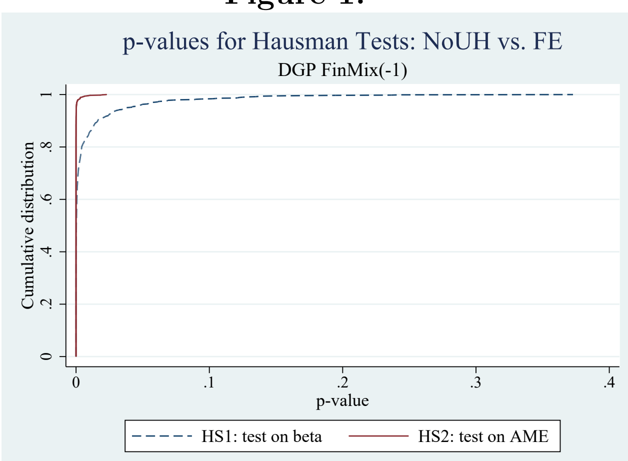


Figure 2

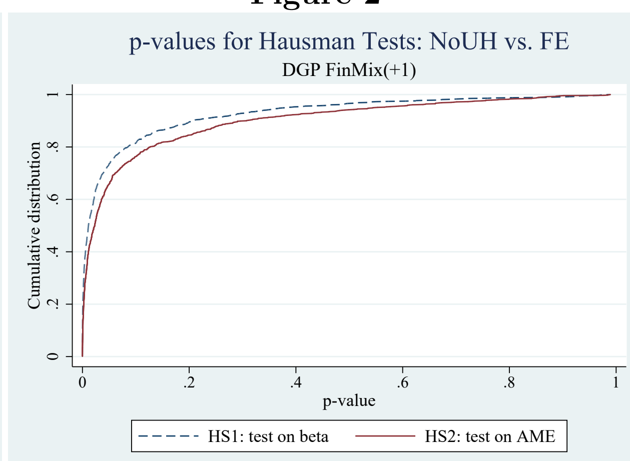


Figure 3

Figure 4

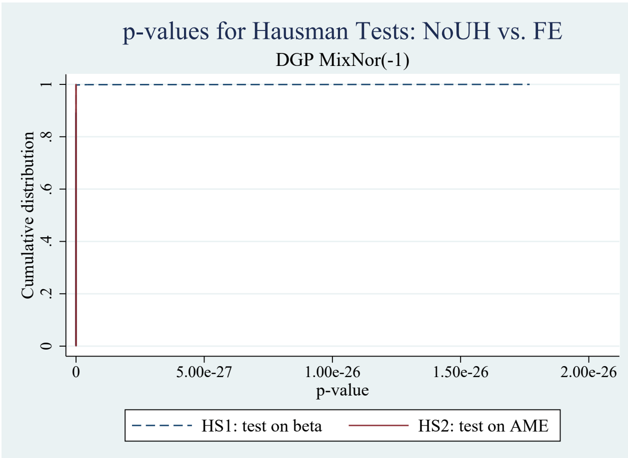


Figure 5

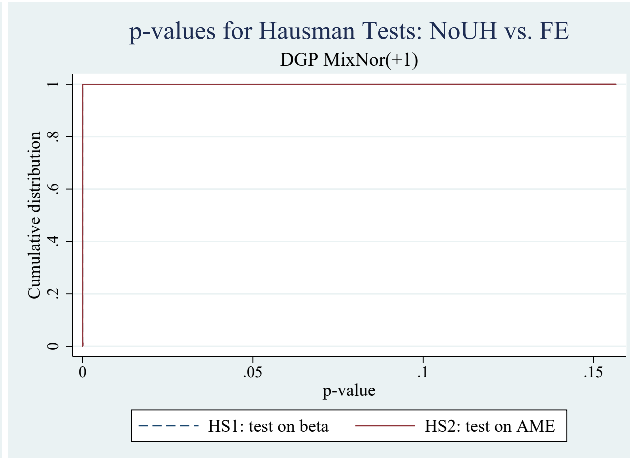
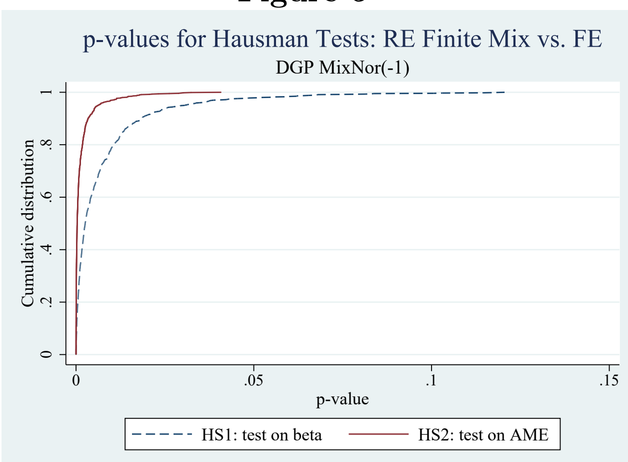
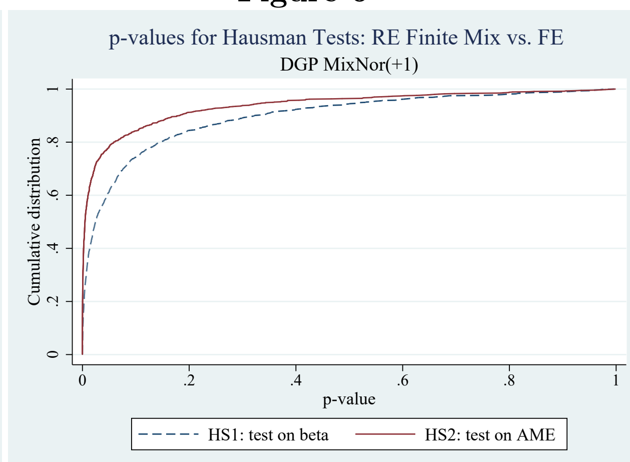


Figure 6





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

6 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).[11] These applications use consumer scanner panel data and estimate dynamic discrete choice models with persistent unobserved heterogeneity in consumer brand preferences and state dependence generated by purchasing/consumption habits or/and brand switching costs. The main goal is to determine the relative contribution of unobserved heterogeneity and state dependence to explain the observed time persistence of consumer brand choices. Disentangling the contribution of these two factors has important implications on demand elasticities, competition, consumer welfare, and the evaluation of mergers.[12]

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

6.1 Data

The dataset is A.C. Nielsen scanner panel data from Sioux Falls, South Dakota, for the ketchup product category.[13] It contains 996 households and covers a 123-week period from mid-1986 to mid-1988.[14] 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 3 presents the distribution of T_i.

Table 3
Distribution of number of purchase occasions (T_i)

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

There are four brands in this market: three national brands, Heinz, Huntâ€Âs and Del Monte; and a store brand. We ignore the quantity purchased and focus on brand choice. Table 4 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

[11]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.

[12]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.

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

[14]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.

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 4
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	

6.2 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*:

$$y_{it} = \arg \max_{j \in \{0,1,2,3\}} \left\{ \alpha_i(j) + \beta_{jj} \; 1\{y_{i,t-1} = j\} + \varepsilon_{it}(j) \right\}. \tag{61}$$

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.[15]

Following Aguirregabiria, Gu, and Luo (2021), equation (61) 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}.

[15]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.

6.3 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 5
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 5 presents our Fixed Effect estimates of the *brand habit* parameters β_{jj}. We use the Conditional Maximum Likelihood estimator. Standard errors are obtained using a bootstrap method that resamples the 996 purchasing histories in the original dataset.[16] 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 4. 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 by ATE_{jj}.

[16] 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 6

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

	$T = 6$ sub-histories				**$T = 8$ sub-histories**			
	Pers (s.e.)	*ATP* (*s.e.*)	*ATE* (s.e.)	*UHet* (*s.e.*)	*Pers* (s.e.)	*ATP* (*s.e.*)	*ATE* (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 4.

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

(3) *ATE* is the one defined in equation (23): $ATE_{jj} = \Pi_{jj} - \mathbb{E}(1\{y_{it} = j\})$.

(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.

Table 6 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. The estimation of the ATPs Π_{jj} is based on equation 40 in Proposition 4. In this equation, we plug-in the CML estimates of β_{jj} parameters and frequency estimates of probabilities of choice histories. Standard errors are obtained using a bootstrap method.

In Table 6, 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 (23): $ATE_{jj} = \Pi_{jj} - \mathbb{E}(1\{y_{it} = j\})$. 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).

7 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 dynamic demand model with state dependence.

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

8 Appendix

8.1 Proof of Lemma 4

For notational simplicity but w.o.l.g., in this section we omit x and θ as arguments in all the functions. Remember that equation (45) is:

$$\sum_{\mathbf{y}^{\{2,T\}} \in \mathcal{Y}^{T-1}} w(d_1, y_1, \mathbf{y}^{\{2,T\}}) \; G\left(\mathbf{y}^{\{2,T\}} | y_1, d_1, \boldsymbol{\alpha}\right) \;=\; \Delta(\boldsymbol{\alpha}). \tag{62}$$

(A) Sufficient condition. Multiplying (45) times $p^*(d_1, y_1|\alpha)\; f_\alpha(\alpha)$, integrating over α, and taking into account that, as defined in (43), $\int G(y^{\{2,T\}}|y_1, d_1, \alpha)\; p^*(d_1, y_1|\alpha)\; f_\alpha(\alpha)\; d\alpha$ is equal to $P_{\mathbf{y}|\mathbf{x}}$, we obtain:

$$\sum_{\mathbf{y}^{\{2,T\}} \in \mathcal{Y}^{T-1}} w(d_1, y_1, \mathbf{y}^{\{2,T\}}) \; \mathbb{P}_{\mathbf{y}|\mathbf{x}} \;=\; \int \Delta(\boldsymbol{\alpha}) \; p^*(d_1, y_1|\boldsymbol{\alpha}) \; f_\alpha(\boldsymbol{\alpha}) \; d\boldsymbol{\alpha}. \tag{63}$$

We can sum equation (63) over all the possible values of (d_1, y_1). Given that the sum of $p^*(d_1, y_1|\alpha)$ over all values of (d_1, y_1) is equal to 1, the right-hand-side becomes $\int \Delta(\alpha)\; f_\alpha(\alpha)\; d\alpha$, which is the definition of AME. Furthermore, the sum of equation (63) over all the possible values of (d_1, y_1) implies equation (46):

$$\sum_{\mathbf{y} \in \mathcal{D} \times \mathcal{Y}^T} w(\mathbf{y}) \; \mathbb{P}_{\mathbf{y}|\mathbf{x}} \;=\; AME. \tag{64}$$

(B) Necessary condition. The proof has two parts. First, we prove that function $h(P_{\mathcal{Y}|\mathcal{X}})$ should be linear in $P_{\mathcal{Y}|\mathcal{X}}$. Second, we show that equation (45) should hold.

Necessary (i). Equality $h(P_{\mathcal{Y}|\mathcal{X}}) = AME$ 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, c is constant; (Case 2) a degenerate distribution where $\alpha_i = c'$ with probability one, where c' is a constant different to c; and (Case 3) a distribution with two points of support, c and c', with $q \equiv f_\alpha(c)$. Then, 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 $h(P_{\mathcal{Y}|\mathcal{X}})$ should satisfy:

$$\begin{cases} \text{Case 1} \;\; : \;\; h\left(\boldsymbol{P}^{(1)}_{\mathcal{Y}|\mathcal{X}}\right) = \Delta(\mathbf{c}) \\ \text{Case 2} \;\; : \;\; h\left(\boldsymbol{P}^{(2)}_{\mathcal{Y}|\mathcal{X}}\right) = \Delta(\mathbf{c}') \\ \text{Case 3} \;\; : \;\; h\left(\boldsymbol{P}^{(3)}_{\mathcal{Y}|\mathcal{X}}\right) = q\; \Delta(\mathbf{c}) + \; (1-q)\; \Delta(\mathbf{c}') \end{cases} \tag{65}$$

where $P^{(1)}_{\mathcal{Y}|\mathcal{X}}$, $P^{(2)}_{\mathcal{Y}|\mathcal{X}}$, and $P^{(3)}_{\mathcal{Y}|\mathcal{X}}$ represent the distributions of y conditional on x under the DGPs of cases 1, 2, and 3, respectively. Note that, by construction, $P^{(3)}_{\mathcal{Y}|\mathcal{X}} = q\; P^{(1)}_{\mathcal{Y}|\mathcal{X}} + (1-q)\; P^{(2)}_{\mathcal{Y}|\mathcal{X}}$. These conditions are for arbitrary values of c, c', and $q \in [0,1]$. Multiplying equation (65)(Case 1) times q, multiplying equation (65)(Case 2) times $(1-q)$, adding up these two results, and then subtracting equation (65)(Case 3), we get that function $h\left(\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}}\right)$ should satisfy the following equation:

$$q\; h\left(\boldsymbol{P}^{(1)}_{\mathcal{Y}|\mathcal{X}}\right) + (1-q)\;\; h\left(\boldsymbol{P}^{(2)}_{\mathcal{Y}|\mathcal{X}}\right) = h\left(q\; \boldsymbol{P}^{(1)}_{\mathcal{Y}|\mathcal{X}} + (1-q)\;\; \boldsymbol{P}^{(1)}_{\mathcal{Y}|\mathcal{X}}\right). \tag{66}$$

The only possibility that equation (66) holds for any arbitrary value of c, c', and $q \in [0,1]$ is that function $h\left(\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}}\right)$ is linear in $P_{\mathcal{Y}|\mathcal{X}}$, such that $h\left(\boldsymbol{P}_{\mathcal{Y}|\mathcal{X}}\right) = \sum_{\mathbf{y}} w(y)\ P_{\mathbf{y}|\mathbf{x}}$.

Necessary (ii). We need to prove that, given equation $\sum_{\mathbf{y}} w(y) = AME$, then equation (45) should hold for every value $\alpha \in R^J$. The proof is by contradiction. Suppose that: (a) equation $\sum_{\mathbf{y}} w(y) = AME$ holds for any distribution f_α in the DGP; and (b) there is a value $\alpha = c$ and a value (d_1, y_1) of the initial condition such that equation (45) does not hold: $\sum_{\mathbf{y}^{\{2,T\}}} w(d_1, y_1, y^{\{2,T\}})\ G\left(\mathbf{y}^{\{2,T\}}|y_1, d_1, \boldsymbol{c}\right) \neq \Delta(c)$. 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 with only two points support, c and c', with $f_\alpha(c) = q$. Define:

$$d(\boldsymbol{\alpha}, d_1, y_1) \equiv \sum_{\mathbf{y}^{\{2,T\}}} w(d_1, y_1, \mathbf{y}^{\{2,T\}})\ G\left(\mathbf{y}^{\{2,T\}}|y_1, d_1, \boldsymbol{\alpha}\right) \ - \ \Delta(\boldsymbol{\alpha}) \tag{67}$$

Condition (b) implies that $d(c, d_1, y_1) \neq 0$. For notational simplicity but w.l.o.g., consider that the initial condition (d_1, y_1) has binary support $\{0, 1\}$. Applying the same operations as in the proof of the sufficient condition, we get:

$$\begin{aligned} &\sum_{\mathbf{y}} w(\mathbf{y})\ \mathbb{P}_{\mathbf{y}|\mathbf{x}} - AME \ = \\ &q\ \left[p^*(0|\mathbf{c})\ d(\mathbf{c}, 0) + p^*(1|\mathbf{c})\ d(\mathbf{c}, 1)\right] \ + \ (1-q)\ \left[p^*(0|\mathbf{c}')\ d(\mathbf{c}', 0) + p^*(1|\mathbf{c}')\ d(\mathbf{c}', 1)\right] \end{aligned} \tag{68}$$

By definition, each value $d(\alpha, d_1, y_1)$ is for a particular value of α, and therefore, it does not depend on distribution f_α. More specifically, $d(\alpha, d_1, y_1)$ does not depend on the value of q. Therefore, there always exist (a continuum of) values of q such that the right hand side of (68) is different to zero, and condition (a) does not hold. ■

8.2 Proof of Lemma 5

For notational simplicity but w.o.l.g., in this section we omit x and θ as arguments in all the functions. Lemma 4 applies under the conditions of Lemma 5 such that equation (45) holds. Using the structure of function G in equation (44), and the definition of the transition probabilities $\pi_{kj}(\alpha)$, we can rewrite equation (45) as follows:

$$\sum_{\mathbf{y}^{\{2,T\}} \in \mathcal{Y}^{T-1}} w(d_1, y_1, \mathbf{y}^{\{2,T\}}) \prod_{t=2}^{T} \pi_{y_{t-1} \cdot y_t}(\boldsymbol{\alpha}) \ = \ \Delta(\boldsymbol{\alpha}). \tag{69}$$

Given the dynamic logit in equation (1), the transition probabilities are:

$$\pi_{kj}(\boldsymbol{\alpha}) \ = \ \frac{\exp\left\{\beta_{kj}(d) + \mathbf{x}'\boldsymbol{\gamma}_j\right\}\ \ \exp\left\{\alpha(j)\right\}}{1 + \sum_{\ell=1}^{J} \exp\left\{\beta_{k\ell}(d) + \mathbf{x}'\boldsymbol{\gamma}_\ell\right\}\ \ \exp\left\{\alpha(\ell)\right\}} \tag{70}$$

And in the right-hand-side of equation (69), the individual effect $\Delta(\alpha)$ is either a difference of transition probabilities – as in $\pi_{11}(\alpha) - \pi_{10}(\alpha)$ – or a transition probability – as in $\pi_{jj}(\alpha)$. For concreteness, suppose that $\Delta(\alpha) = \pi_{jj}(\alpha)$, but it is straightforward to extend the proof to the case where $\Delta(\alpha)$ is the difference of two transition probabilities. Therefore, we have:

$$\sum_{\mathbf{y}^{\{2,T\}}} w(d_1, y_1, \mathbf{y}^{\{2,T\}}) \prod_{t=2}^{T} \frac{\exp\left\{\beta_{y_{t-1},y_t}(d_t) + \mathbf{x}_t'\boldsymbol{\gamma}_{y_t}\right\} \ \exp\left\{\alpha(y_t)\right\}}{1 + \sum_{\ell=1}^{J} \exp\left\{\beta_{y_{t-1},\ell}(d_t) + \mathbf{x}_t'\boldsymbol{\gamma}_{\ell}\right\} \ \exp\left\{\alpha(\ell)\right\}}$$
$$- \quad \frac{\exp\left\{\beta_{jj}(d) + \mathbf{x}'\boldsymbol{\gamma}_j\right\} \ \exp\left\{\alpha(j)\right\}}{1 + \sum_{\ell=1}^{J} \exp\left\{\beta_{j\ell}(d) + \mathbf{x}'\boldsymbol{\gamma}_\ell\right\} \ \exp\left\{\alpha(\ell)\right\}} \ = \ 0 \tag{71}$$

Multiplying this equation times $\prod_{t=2}^{T}\left(1 + \sum_{\ell=1}^{J} \exp\left\{\beta_{y_{t-1},\ell}(d_t) + \mathbf{x}_t'\boldsymbol{\gamma}_\ell\right\} \exp\left\{\alpha(\ell)\right\}\right)$ to eliminate the denominators, and using the Binomial Theorem to expand the terms $[1+z]^n$ as $\sum_{k=0}^{n}\binom{n}{k} z^k$, we obtain a polynomial in $\left\{e^{\alpha(j)}\right\}_{j=1}^{J}$. Therefore, this system of equations holds for every value $\alpha \in R^J$ if and only if the coefficients multiplying each monomial term in the polynomial are all equal to zero. This defines a finite system of equations.

In this polynomial, the coefficients multiplying each monomial term are linear functions of products between the weights $w(d_1, y_1, y^{\{2,T\}})$ and the terms $\exp\left\{\beta_{y_{t-1},y_t}(d_t) + \mathbf{x}_t'\boldsymbol{\gamma}_{y_t}\right\}$. The finite system of equations, as many equations as the order of the polynomial, that makes the monomial coefficients equal to zero is a linear system in the weights $w(d_1, y_1, y^{\{2,T\}})$. ■

8.3 Proof of Proposition 1 using Lemmas 4 and 5

For the binary choice model in equation (5) and $AME^{(1)}$, we have

$$\begin{cases} \Delta(\boldsymbol{\alpha}) & = \dfrac{e^{\alpha}\left(e^{\beta}-1\right)}{\left(1+e^{\alpha+\beta}\right)\left(1+e^{\alpha}\right)} \\ \\ G\left(\mathbf{y}^{\{2,T\}}|y_1, \alpha\right) & = \dfrac{\left(e^{\alpha}\right)^{n_1}\left(e^{\beta}\right)^{n_{11}}}{\left(1+e^{\alpha}\right)^{T-1+y_T-y_1-n_1}\left(1+e^{\alpha+\beta}\right)^{n_1-y_T+y_1}} \end{cases} \tag{72}$$

where $n_1 \equiv \sum_{t=2}^{T} y_t$ and $n_{11} \equiv \sum_{t=2}^{T} y_{t-1}y_t$. With $T=3$ there are (2^{T-1}) 4 possible values of $y^{\{2,T\}}$ with each of the two values of y_1. In this case, condition (45) in Lemma 4 for $y_1 = 0$ is

$$w_1\mathbb{P}\left(\mathbf{y}=(0,0,0)\,|y_{i1}=0,\alpha_i\right) + w_2\mathbb{P}\left(\mathbf{y}=(0,0,1)\,|y_{i1}=0,\alpha_i\right) + w_3\mathbb{P}\left(\mathbf{y}=(0,1,0)\,|y_{i1}=0,\alpha_i\right) +$$
$$+ w_4\mathbb{P}\left(\mathbf{y}=(0,1,1)\,|y_{i1}=0,\alpha_i\right) = \Delta(\alpha_i)$$

Replacing $P\left(\mathbf{y}|y_{i1}=0,\alpha_i\right)$ and $\Delta(\alpha_i)$ by its expressions according to (72), multiplying the result times $\left(1+e^{\alpha}\right)^2\left(1+e^{\alpha+\beta}\right)$ to remove denominators, doing some algebra and putting everything in the left hand side of the equality, we have a polynomial in $\exp\left(\alpha_i\right)$ as Lemma 5 established:

$$w_1 + \left[w_1\exp\left(\beta\right) + w_2 + w_3 - \left(\exp\left(\beta\right)-1\right)\right]\exp\left(\alpha_i\right) +$$
$$+ \left[w_2\exp\left(\beta\right) + w_3 + w_4\exp\left(\beta\right) - \left(\exp\left(\beta\right)-1\right)\right]\exp\left(\alpha_i\right)^2 + w_4\exp\left(\beta\right)\exp\left(\alpha_i\right)^3 = 0 \tag{73}$$

As said in Lemma 5, this reduces the infinite conditions in (45) to a system of linear equations that we have to solve for the weights:

$$\begin{cases} w_1 = 0 \\ w_1 \exp(\beta) + w_2 + w_3 - \exp(\beta) + 1 = 0 \\ w_2 \exp(\beta) + w_3 + w_4 \exp(\beta) - \exp(\beta) + 1 = 0 \\ w_4 \exp(\beta) = 0 \end{cases} \tag{74}$$

Solving it we obtain $\{w_1 = w_2 = w_4 = 0; w_3 = \exp(\beta) - 1\}$ as the only solution. For $y_1 = 1$ we have

$$w_5 \mathbb{P}(\mathbf{y} = (1,0,0)\,|y_{i1} = 1, \alpha_i) + w_6 \mathbb{P}(\mathbf{y} = (1,0,1)\,|y_{i1} = 1, \alpha_i) + w_7 \mathbb{P}(\mathbf{y} = (1,1,0)\,|y_{i1} = 1, \alpha_i) + \\ + w_8 \mathbb{P}(\mathbf{y} = (1,1,1)\,|y_{i1} = 1, \alpha_i) = \Delta^{(1)}(\alpha_i)$$

and it leads to the system of linear equations

$$\begin{cases} w_5 = 0 \\ w_5 \exp(\beta) + w_3 + w_3 \exp(\beta) - \exp(\beta) + 1 = \\ w_6 \exp(\beta) + w_7 \exp(\beta) + w_8 \exp(\beta)^2 - \exp(\beta)(\exp(\beta) - 1) = 0 \\ w_8 \exp(\beta)^2 = 0 \end{cases} \tag{75}$$

whose only solutions is $\{w_5 = w_7 = w_8 = 0; w_6 = \exp(\beta) - 1\}$. Therefore,

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

This is exactly the expression that we have in Proposition 1. This way of obtain the weights and proving identification also shows that the weights are unique and the model does not provide additional restrictions on AME when $T = 3$.

8.4 Applying Lemma 5 to obtain expression for $AME^{(1)}$ with $T > 3$

The identification result using only 3 periods proves identification for any $T \geq 3$, because with more than 3 periods we can always take 3 periods. Nonetheless, it is possible to obtain close form expression for higher values of T using the same procedure based on section 4.1. This expression will use all T periods without having to combine several 3-periods estimates.

For $T > 3$, as said, there is overidentification, so more than one combination of the probability of histories exits. A way of choosing one of them in the BC-AR1 model is to focus on the probabilities of the sufficient statistics that are used to identify β in the CMLE. This is as follows. For this model and other logit models, the log-probability of a choice history has the following structure:

$$\ln \mathbb{P}(\mathbf{y}_i|\alpha_i, \beta) = \mathbf{s}(\mathbf{y}_i)' \,\mathbf{g}(\alpha_i) + \mathbf{c}(\mathbf{y}_i)' \,\beta \tag{76}$$

where $s(y_i)$ and $c(y_i)$ are vectors of statistics (functions of y_i), and $g(\alpha_i)$ is a vector of functions α_i. $s(y)$ is a sufficient statistic for α_i because $P(y_i|\alpha_i, \beta, s_i) = P(y_i|\beta, s_i)$.[17] Let S_T be the set of

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

possible values of $s(y)$, let $P_{\mathbf{s}}$ be the probability of a value s of $s(y)$, and let $P_{\mathbf{s}} \equiv \{P_{\mathbf{s}} : s \in S_T\}$ be the probability distribution of this statistic. Given θ, the empirical distribution $P_{\mathbf{s}}$ contains all the information in the data about the distribution of α_i, and therefore, about AMEs. Taking into account the structure of the probability of a choice history in equation (76), the model implies:

$$\mathbb{P}_{\mathbf{s}} = \sum_{\mathbf{y}:\ \mathbf{s}(\mathbf{y})=\mathbf{s}} \left[\int \exp\{\mathbf{s}'\ \mathbf{g}(\alpha_i) + \mathbf{c}(\mathbf{y})'\ \theta\}\ f_\alpha(\alpha_i)\ d\alpha_i \right] \tag{77}$$

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

In the BC-AR1 model the sufficient statistics $s(y_i)$ is the vector $\left(y_{i1}, y_{iT}, \sum_{t=2}^{T} y_{it}\right)'$ –see Aguirregabiria, Gu, and Luo (2021)– and it can take $4T-4$ different values, $2T-2$ values with $y_{i1}=0$ and $2T-2$ values with $y_{i1}=1$. The conditions in (45) are here

$$\left.\begin{array}{rcl} \sum_{j=1}^{2T-2} w_j\ \mathbb{P}\left(\mathbf{s_j} \mid y_{j1}=0, \beta, \alpha_i\right) & = & \Delta(\alpha_i) \\ \sum_{j=2T-2+1}^{4T-4} w_j\ \mathbb{P}\left(\mathbf{s_j} \mid y_{j1}=1, \beta, \alpha_i\right) & = & \Delta(\alpha_i) \end{array}\right\} \text{for every } \alpha_i \in \mathbb{R}, \tag{78}$$

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

Proceeding as in 8.3, we obtain the weights for $AME^{(1)}$ in the binary choice AR(1) model for different values of T. These are in Tables 7 and 8.

Table 7
Weights w_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 8
Weights w_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

8.5 Proof of Proposition 5

W.l.o.g. we consider $T=3$ and $t=3$. According to Lemma 4 to prove identification of the AME, we have to find weights $w_{(y_1,y_2,y_3;\mathbf{x})}$ that satisfy condition (45) for any value of α. Equation (45) for $AME_{x,3}^{(1)}$ takes the following form for each value of y_1:

$$\sum_{y_2,y_3} w_{(y_1,y_2,y_3;\mathbf{x})}\ \mathbb{P}_{(y_1,y_2,y_3)\ |\ (y_1,\mathbf{x},\alpha_i)} \quad = \quad \Lambda(\alpha_i+\beta+\gamma x_3)-\Lambda(\alpha_i+\gamma x_3)=\Delta^{(1)}(\alpha_i,x_3) \tag{79}$$

For shortness $w_{(0,0,0;\mathbf{x})}$ is written as w_{1i}, $w_{(0,0,1;\mathbf{x})}$ as w_{2i}, ..., and $w_{(1,1,1;\mathbf{x})}$ as w_{8i}. We start by finding w_{1i}, w_{2i}, w_{3i}, and w_{4i} such that

$$\begin{aligned}&w_{1i}*\mathbb{P}_{(0,0,0)\ |\ (y_1=0,\mathbf{x},\alpha_i)}+w_{2i}*\mathbb{P}_{(0,0,1)\ |\ (y_1=0,\mathbf{x},\alpha_i)}+w_{3i}*\mathbb{P}_{(0,1,0)\ |\ (y_1=0,\mathbf{x},\alpha_i)}\\&+w_{4i}*\mathbb{P}_{(0,1,1)\ |\ (y_1=0,\mathbf{x},\alpha_i)}=\Lambda(\alpha_i+\beta+\gamma x_{i3})-\Lambda(\alpha_i+\gamma x_{i3})\end{aligned}$$

Replacing the Probabilities by their expression:

$$\begin{aligned}&w_{1i}\left(1-\Lambda\left(\alpha_i+\gamma x_{i2}\right)\right)\left(1-\Lambda\left(\alpha_i+\gamma x_{i3}\right)\right)+w_{2i}\left(1-\Lambda\left(\alpha_i+\gamma x_{i2}\right)\right)\Lambda\left(\alpha_i+\gamma x_{i3}\right)\\&+w_{3i}\Lambda\left(\alpha_i+\gamma x_{i2}\right)\left(1-\Lambda\left(\alpha_i+\beta+\gamma x_{i3}\right)\right)+w_{4i}\Lambda\left(\alpha_i+\gamma x_{i2}\right)\Lambda\left(\alpha_i+\beta+\gamma x_{i3}\right)\\&=\Lambda(\alpha_i+\beta+\gamma x_{i3})-\Lambda(\alpha_i+\gamma x_{i3})\end{aligned}$$

Replacing the logistic cdf by its expression:

$$\begin{aligned}&w_{1i}\tfrac{1}{(1+\exp(\alpha_i+\gamma x_{i2}))(1+\exp(\alpha_i+\gamma x_{i3}))}+w_{2i}\tfrac{\exp(\alpha_i+\gamma x_{i3})}{(1+\exp(\alpha_i+\gamma x_{i2}))(1+\exp(\alpha_i+\gamma x_{i3}))}+w_{3i}\tfrac{\exp(\alpha_i+\gamma x_{i2})}{(1+\exp(\alpha_i+\gamma x_{i2}))(1+\exp(\alpha_i+\beta+\gamma x_{i3}))}\\&+w_{4i}\tfrac{\exp(\alpha_i+\gamma x_{i2})\exp(\alpha_i+\beta+\gamma x_{i3})}{(1+\exp(\alpha_i+\gamma x_{i2}))(1+\exp(\alpha_i+\beta+\gamma x_{i3}))}=\tfrac{\exp(\alpha_i+\beta+\gamma x_{i3})}{1+\exp(\alpha_i+\beta+\gamma x_{i3})}-\tfrac{\exp(\alpha_i+\gamma x_{i3})}{1+\exp(\alpha_i+\gamma x_{i3})}\end{aligned}$$

Operating to undo the fractions in both sides and simplifying:

$$\begin{aligned}&w_{1i}+w_{1i}\exp\left(\gamma x_{i3}\right)\exp\left(\beta\right)\exp\left(\alpha_i\right)+w_{2i}\exp\left(\gamma x_{i3}\right)\exp\left(\alpha_i\right)+w_{2i}\exp\left(\gamma x_{i3}\right)^2\exp\left(\beta\right)\exp\left(\alpha_i\right)^2\\&+w_{3i}\exp\left(\gamma x_{i2}\right)\exp\left(\alpha_i\right)+w_{3i}\exp\left(\gamma x_{i2}\right)\exp\left(\gamma x_{i3}\right)\exp\left(\alpha_i\right)^2\\&+w_{4i}\exp\left(\gamma x_{i2}\right)\exp\left(\gamma x_{i3}\right)\exp\left(\beta\right)\exp\left(\alpha_i\right)^2+w_{4i}\exp\left(\gamma x_{i2}\right)\exp\left(\gamma x_{i3}\right)^2\exp\left(\beta\right)\exp\left(\alpha_i\right)^3\\&=\exp\left(\gamma x_{i3}\right)\left(\exp\left(\beta\right)-1\right)\exp\left(\alpha_i\right)+\exp\left(\gamma x_{i2}\right)\exp\left(\gamma x_{i3}\right)\left(\exp\left(\beta\right)-1\right)\exp\left(\alpha_i\right)^2\end{aligned}$$

Since w_{1i}, w_{2i}, w_{3i}, and w_{4i} cannot depend on α_i all the terms in the polynomial of $\exp\left(\alpha_i\right)$ in both sides must be the same. Therefore, as in lemma 5, this implies the following (finite) system of linear equations:

$$w_{1i}=0$$

$$w_{1i}\exp\left(\gamma x_{i3}\right)\exp\left(\beta\right)+w_{2i}\exp\left(\gamma x_{i3}\right)+w_{3i}\exp\left(\gamma x_{i2}\right)=\exp\left(\gamma x_{i3}\right)\left(\exp\left(\beta\right)-1\right)$$

$$w_{2i}\exp\left(\gamma x_{i3}\right)^2\exp\left(\beta\right)+w_{3i}\exp\left(\gamma x_{i2}\right)\exp\left(\gamma x_{i3}\right)w_{4i}\exp\left(\gamma x_{i2}\right)\exp\left(\gamma x_{i3}\right)\exp\left(\beta\right)=\exp\left(\gamma x_{i2}\right)\exp\left(\gamma x_{i3}\right)(\mathrm{ex}$$

$$w_{4i}\exp\left(\gamma x_{i2}\right)\exp\left(\gamma x_{i3}\right)^2\exp\left(\beta\right)=0$$

Solving this system of equation we obtain

$$\begin{aligned}w_{1i}&=0\\w_{2i}&=\frac{\exp\left(\gamma x_{i2}\right)-\exp\left(\gamma x_{i3}\right)}{\exp\left(\gamma x_{i3}\right)}\\w_{3i}&=\frac{\exp\left(\gamma x_{i3}\right)\exp\left(\beta\right)-\exp\left(\gamma x_{i2}\right)}{\exp\left(\gamma x_2\right)}\\w_{4i}&=0\end{aligned}$$

Proceeding the same way with

$$w_{5i} * \mathbb{P}_{(1,0,0)\ |\ (y_1=1,\mathbf{x},\alpha_i)} + w_{6i} * \mathbb{P}_{(1,0,1)\ |\ (y_1=1,\mathbf{x},\alpha_i)} + w_{7i} * \mathbb{P}_{(1,1,0)\ |\ (y_1=1,\mathbf{x},\alpha_i)}$$
$$+w_{8i} * \mathbb{P}_{(1,1,1)\ |\ (y_1=1,\mathbf{x},\alpha_i)} = \Lambda(\alpha_i + \beta + \gamma x_{i3}) - \Lambda(\alpha_i + \gamma x_{i3})$$

we obtain

$$\begin{aligned}
w_{5i} &= 0 \\
w_{6i} &= \frac{\exp(\gamma x_{i2})\exp(\beta) - \exp(\gamma x_{i3})}{\exp(\gamma x_{i3})} \\
w_{7i} &= \frac{\exp(\gamma x_{i3}) - \exp(\gamma x_{i2})}{\exp(\gamma x_{i2})} \\
w_{8i} &= 0
\end{aligned}$$

Using the previous results,

$$\sum_{y_1,y_2,y_3} w_{(y_1,y_2,y_3;\mathbf{x})}\ \mathbb{P}_{(y_1,y_2,y_3)\ |\ \mathbf{x}_i} =$$
$$\int \left[w_{2i} * \mathbb{P}_{(0,0,1)\ |\ (y_1=0,\mathbf{x}_i,\alpha_i)} + w_{3i} * \mathbb{P}_{(0,1,0)\ |\ (y_1=0,\mathbf{x}_i,\alpha_i)}\right] p^*(0|\alpha_i,\mathbf{x}_i) f_{\alpha|x}(\alpha_i|\mathbf{x}_i) d\alpha_i$$
$$+\int \left[w_{6i} * \mathbb{P}_{(1,0,1)\ |\ (y_1=1,\mathbf{x}_i,\alpha_i)} + w_{7i} * \mathbb{P}_{(1,1,0)\ |\ (y_1=1,\mathbf{x}_i,\alpha_i)}\right] p^*(1|\alpha_i,\mathbf{x}_i) f_{\alpha|x}(\alpha_i|\mathbf{x}_i) d\alpha_i$$
$$= \int \Delta^{(1)}(\alpha_i,\mathbf{x}_3) p^*(0|\alpha_i,\mathbf{x}_i) f_{\alpha|x}(\alpha_i|\mathbf{x}_i) d\alpha_i + \int \Delta^{(1)}(\alpha_i,\mathbf{x}_3) p^*(1|\alpha_i,\mathbf{x}_i) f_{\alpha|x}(\alpha_i|\mathbf{x}_i) d\alpha_i$$
$$= \int \Delta^{(1)}(\alpha_i,\mathbf{x}_3) \left(p^*(0|\alpha_i,\mathbf{x}_i) + p^*(1|\alpha_i,\mathbf{x}_i)\right) f_{\alpha|x}(\alpha_i|\mathbf{x}_i) d\alpha_i$$
$$\int \Delta^{(1)}(\alpha_i,\mathbf{x}_3) f_{\alpha|x}(\alpha_i|\mathbf{x}_i) d\alpha_i$$

So, we have shown that those weights are such that

$$\sum_{y_1,y_2,y_3} w_{(y_1,y_2,y_3;\mathbf{x})}\ \mathbb{P}_{(y_1,y_2,y_3)\ |\ \mathbf{x}_i} = \int \Delta^{(1)}(\alpha_i,\mathbf{x}_3) f_{\alpha|x}(\alpha_i|\mathbf{x}_i) d\alpha_i$$

This implies

$$\begin{aligned}
&\sum_{x_i}\left(\sum_{y_1,y_2,y_3} w_{(y_1,y_2,y_3;\mathbf{x})}\ \mathbb{P}_{(y_1,y_2,y_3)\ |\ \mathbf{x}_i}\right)\mathbb{P}(\mathbf{x}_i) \\
&= \sum_{x_i}\left(\int \Delta^{(1)}(\alpha_i,\mathbf{x}_3) f_{\alpha|x}(\alpha_i|\mathbf{x}_i) d\alpha_i\right)\mathbb{P}(\mathbf{x}_i) \\
&= \int \Delta^{(1)}(\alpha_i,\mathbf{x}_3) f_{(\alpha,x)}(\alpha_i,\mathbf{x}_i) d(\alpha_i,\mathbf{x}_i) \\
&= AME^{(1)}_{x,3}.
\end{aligned} \tag{80}$$

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

8.6 Proof of Proposition 6

For notational simplicity, we omit x as an argument throughout this proof. In model (2) with $J+1=3$ and $T=3$, the necessary and sufficient condition in equation (45) of Lemma 4 for $y_1=0$ is

$$\begin{aligned} & w_1 * \mathbb{P}_{(0,0,0)\,|\,(y_1=0,\alpha_i)} + w_2 * \mathbb{P}_{(0,0,1)\,|\,(y_1=0,\alpha_i)} + w_3 * \mathbb{P}_{(0,0,2)\,|\,(y_1=0,\alpha_i)} \\ & + w_4 * \mathbb{P}_{(0,1,0)\,|\,(y_1=0,\alpha_i)} + w_5 * \mathbb{P}_{(0,1,1)\,|\,(y_1=0,\alpha_i)} + w_6 * \mathbb{P}_{(0,1,2)\,|\,(y_1=0,\alpha_i)} \\ & + w_7 * \mathbb{P}_{(0,2,0)\,|\,(y_1=0,\alpha_i)} + w_8 * \mathbb{P}_{(0,2,1)\,|\,(y_1=0,\alpha_i)} + w_9 * \mathbb{P}_{(0,2,2)\,|\,(y_1=0,\alpha_i)} = \pi_{10}(\alpha_i) \end{aligned} \tag{81}$$

Let's denote $d_j \equiv 1+\exp\{\beta_{j1}+\alpha_i(1)\}+\exp\{\beta_{j2}+\alpha_i(2)\}$ for $j=0,1,2$. Replacing the Probabilities by their expression based on the logistic cdf,

$$\begin{aligned} & w_1 \tfrac{1}{d_0^2} + w_2 \tfrac{\exp\{\beta_{01}+\alpha_i(1)\}}{d_0^2} + w_3 \tfrac{\exp\{\beta_{02}+\alpha_i(1)\}}{d_0^2} + w_4 \tfrac{\exp\{\beta_{01}+\alpha_i(1)\}}{d_0 d_1} \\ & + w_5 \tfrac{\exp\{\beta_{01}+\alpha_i(1)\}\exp\{\beta_{11}+\alpha_i(1)\}}{d_0 d_1} + w_6 \tfrac{\exp\{\beta_{12}+\alpha_i(2)\}\exp\{\beta_{01}+\alpha_i(1)\}}{d_0 d_1} + w_7 \tfrac{\exp\{\beta_{02}+\alpha_i(2)\}}{d_0 d_2} \\ & + w_8 \tfrac{\exp\{\beta_{21}+\alpha_i(1)\}\exp\{\beta_{02}+\alpha_i(2)\}}{d_0 d_2} + w_9 \tfrac{\exp\{\beta_{22}+\alpha_i(2)\}\exp\{\beta_{02}+\alpha_i(2)\}}{d_0 d_2} = \tfrac{1}{d_1} \end{aligned}$$

After some algebra to undo the fractions in both sides and simplifying,

$$\begin{aligned} & w_1 d_1 d_2 + w_2 \exp\{\beta_{01}+\alpha_i(1)\} d_1 d_2 + w_3 \exp\{\beta_{02}+\alpha_i(2)\} d_1 d_2 + w_4 \exp\{\beta_{01}+\alpha_i(1)\} d_0 d_2 \\ & + w_5 \exp\{\beta_{01}+\alpha_i(1)\}\exp\{\beta_{11}+\alpha_i(1)\} d_0 d_2 + w_6 \exp\{\beta_{12}+\alpha_i(2)\}\exp\{\beta_{01}+\alpha_i(1)\} d_0 d_2 \\ & + w_7 \exp\{\beta_{02}+\alpha_i(2)\} d_0 d_1 + w_8 \exp\{\beta_{21}+\alpha_i(1)\}\exp\{\beta_{02}+\alpha_i(2)\} d_0 d_1 \\ & + w_9 \exp\{\beta_{22}+\alpha_i(2)\}\exp\{\beta_{02}+\alpha_i(2)\} d_0 d_1 = d_0^2 d_2 \end{aligned}$$

Expanding this equation by doing the products of d_j and of the exponential, we obtain in both sides of the equality a polynomial in $\exp\{\alpha_i(1)\}^h * \exp\{\alpha_i(2)\}^l$, where the minimum value of h and l is 0, and the maximum value is 4. Following lemma 5, equating the coefficient of each monomial in both sides of the equality, this results on a system of linear equation whose unknowns are the weights $w_1, ..., w_9$. This condition on the monomials of $\exp\{\alpha_i(1)\}^3$ and $\exp\{\alpha_i(1)\}^2$ imply respectively:

$$w_2 \exp(\beta_{11}) + w_4 \exp(\beta_{01}) = 0 \tag{82}$$

$$w_2 \exp(\beta_{21}) + w_4 \exp(\beta_{21}) + w_2 \exp(\beta_{11}) + w_4 \exp(\beta_{01}) = \exp(\beta_{21}) \tag{83}$$

which leads to

$$w_2 + w_4 = 1. \tag{84}$$

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

$$w_2 \exp(\beta_{01}) + w_4 \exp(\beta_{01}) = \exp(\beta_{01}) + \exp(\beta_{21}). \tag{85}$$

That is,

$$w_2 + w_4 = 1 + \frac{\exp(\beta_{21})}{\exp(\beta_{01})}, \tag{86}$$

which is incompatible with condition (84). Therefore, there do not exist weights that satisfy (45) of Lemma 4 –represented by equation (81) in this particular case–, and, according to Lemma 4, there is no function of the observed probability choices that identifies Π_{10}. ■

8.7 Proof of Proposition 7

The weights in the enunciate of Proposition 7 were obtained using the general procedure and result in Section 4.1. However, given the weights, it is easier to prove directly that the linear combination give the average transition probability. We present this proof for Π_{20}, but it proceeds the same for other Π_{kj}. Notice that $\Pi_{20} = \int \pi_{20}(\alpha_i)\ f_\alpha(\alpha_i)\ d\alpha_i$, where, according to the probability (52) of this model, $\pi_{20}(\alpha_i) = \frac{1}{1+\exp(\beta_2-\lambda_0+\alpha_i)}$. We start with the probabilities of choice histories conditional on α_i, that is, $P_{(y_1,y_2,y_3)\ |\ \alpha_i}$. First, we write the expression for these probabilities implied by the model as functions of parameters β, λ and α_i. Second, for each of these probabilities, we multiply the equation times the weights $w_{(y_1,y_2,y_3)}$ that appear in the enunciate of Proposition 7. For the probabilities with non-zero weights for Π_{20}, we have:

$$\begin{aligned}
\mathbb{P}_{(0,0,0)\ |\ \alpha_i} + \mathbb{P}_{(0,0,1)\ |\ \alpha_i} &= p^*(0|\alpha_i)\ \frac{1}{(1+\exp(\beta_0-\lambda_1+\alpha_i))(1+\exp(\beta_0-\lambda_0+\alpha_i))} \\
\exp\left(\lambda_1-\lambda_0\right)\ \mathbb{P}_{(0,0,2)\ |\ \alpha_i} &= p^*(0|\alpha_i)\ \frac{\exp(\beta_0-\lambda_0+\alpha_i)}{(1+\exp(\beta_0-\lambda_1+\alpha_i))(1+\exp(\beta_0-\lambda_0+\alpha_i))} \\
\left(1-\frac{\exp(\beta_2-\lambda_0)}{\exp(\beta_0-\lambda_1)}\right)\left[\mathbb{P}_{(0,2,0)\ |\ \alpha_i}\right] &= p^*(0|\alpha_i)\ \frac{\exp(\beta_0-\lambda_1+\alpha_i)-\exp(\beta_2-\lambda_0+\alpha_i)}{(1+\exp(\beta_0-\lambda_1+\alpha_i))(1+\exp(\beta_2-\lambda_0+\alpha_i))} \\
\sum_{k=0}^{l}\sum_{l=0}^{J-1}\mathbb{P}_{(1,k,l)\ |\ \alpha_i} &= p^*(1|\alpha_i)\left(\frac{1}{1+\exp(\beta_2-\lambda_0+\alpha_i)} - \frac{1-\exp(\beta_2-\lambda_0+\alpha_i)}{(1+\exp(\beta_1-\lambda_1+\alpha_i))(1+\exp(\beta_2-\lambda_0+\alpha_i))}\right) \\
\left(1-\frac{\exp(\beta_2-\lambda_0)}{\exp(\beta_1-\lambda_1)}\right)\mathbb{P}_{(1,2,0)\ |\ \alpha_i} &= p^*(1|\alpha_i)\frac{1-\exp(\beta_2-\lambda_0+\alpha_i)}{1+\exp(\beta_1-\lambda_1+\alpha_i)1+\exp(\beta_2-\lambda_0+\alpha_i)} \\
\sum_{k=0}^{J-1}\mathbb{P}_{(2,0,k)\ |\ \alpha_i} &= p^*(2|\alpha_i)\frac{1}{1+\exp(\beta_2-\lambda_0+\alpha_i)}
\end{aligned} \tag{87}$$

Notice that $\exp\left(\lambda_1-\lambda_0\right) = \frac{\exp(\beta_0-\lambda_0+\alpha_i)}{\exp(\beta_0-\lambda_1+\alpha_i)}$. Third, we sum these equations. Simplifying factors and taking into account that $p^*(0|\alpha_i)+ p^*(1|\alpha_i) + p^*(2|\alpha_i) = 1$, we get:

$$\begin{aligned}
&\mathbb{P}_{(0,0,0)\ |\ \alpha_i} + \mathbb{P}_{(0,0,1)\ |\ \alpha_i} + \exp\left\{\lambda_1-\lambda_0\right\}\mathbb{P}_{(0,0,2)\ |\ \alpha_i} + \left(1-\frac{\exp\left\{\beta_2-\lambda_0\right\}}{\exp\left\{\beta_0-\lambda_1\right\}}\right)\mathbb{P}_{(0,2,0)\ |\ \alpha_i} \\
&+\sum_{k=0}^{l}\sum_{l=0}^{J}\mathbb{P}_{(1,k,l)\ |\ \alpha_i} + \left(1-\frac{\exp\left\{\beta_2-\lambda_0\right\}}{\exp\left\{\beta_1-\lambda_1\right\}}\right)\mathbb{P}_{(1,2,0)\ |\ \alpha_i} + \sum_{k=0}^{J}\mathbb{P}_{(2,0,k)\ |\ \alpha_i} \\
&= \frac{1}{1+\exp\left(\beta_2-\lambda_0+\alpha_i\right)} = \pi_{20}(\alpha_i)
\end{aligned} \tag{88}$$

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

$$\mathbb{P}_{0,0,0} + \mathbb{P}_{0,0,1} + \exp\left\{\lambda_1-\lambda_0\right\}\mathbb{P}_{0,0,2} + \left(1-\frac{\exp\left\{\beta_2-\lambda_0\right\}}{\exp\left\{\beta_0-\lambda_1\right\}}\right)\mathbb{P}_{0,2,0} \tag{89}$$

$$+\sum_{k=0}^{l}\sum_{l=0}^{J-1}\mathbb{P}_{1,k,l} + \left(1-\frac{\exp\left\{\beta_2-\lambda_0\right\}}{\exp\left\{\beta_1-\lambda_1\right\}}\right)\mathbb{P}_{1,2,0} + \sum_{k=0}^{J-1}\mathbb{P}_{2,0,k} = \Pi_{20} \tag{90}$$

8.8 Proof of Proposition 8

As in 8.7, given the weights, it is easier to prove directly that they give the AME. We present this proof for $AME_{0\to1}$, but it proceeds the same for the other AMEs. We start with the probabilities of choice histories conditional on α_i, that is, $P_{(y_1,y_2,y_3,y_4)\mid\alpha_i}$. 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 Proposition 8. For the probabilities with non-zero weights for $AME_{0\to 1}$, we have:

$$\begin{aligned}\frac{e^{\beta(1)}-1}{2}\left[\mathbb{P}_{(0,0,1,0)\mid\alpha_i}+\mathbb{P}_{(0,1,0,0)\mid\alpha_i}\right] &= p^*(0|\alpha_i)\,\frac{\left(e^{\beta(1)}-1\right)e^{\alpha_i}}{\left(1+e^{\alpha_i+\beta(1)}\right)\left(1+e^{\alpha_i}\right)^2}\\ \frac{e^{\beta(1)}-1}{e^{\beta(1)}}\,\mathbb{P}_{(0,0,1,1)\mid\alpha_i} &= p^*(0|\alpha_i)\,\frac{\left(e^{\beta(1)}-1\right)e^{\alpha_i}e^{\alpha_i}}{\left(1+e^{\alpha_i+\beta(1)}\right)\left(1+e^{\alpha_i}\right)^2}\\ \left(e^{\beta(1)}-1\right)\left[\mathbb{P}_{(1,0,1,0)\mid\alpha_i}+\mathbb{P}_{(1,0,1,1)\mid\alpha_i}\right] &= p^*(1|\alpha_i)\,\frac{\left(e^{\beta(1)}-1\right)e^{\alpha_i}\left(e^{\alpha_i+\beta(1)}+1\right)}{\left(1+e^{\alpha_i+\beta(1)}\right)^2\left(1+e^{\alpha_i}\right)}\end{aligned} \tag{91}$$

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

$$\begin{aligned}&\frac{e^{\beta(1)}-1}{2}\left[\mathbb{P}_{(0,0,1,0)\mid\alpha_i}+\mathbb{P}_{(0,1,0,0)\mid\alpha_i}\right]+\frac{e^{\beta(1)}-1}{e^{\beta(1)}}\mathbb{P}_{(0,0,1,1)\mid\alpha_i}+\left(e^{\beta(1)}-1\right)\left[\mathbb{P}_{(1,0,1,0)\mid\alpha_i}+\mathbb{P}_{(1,0,1,1)\mid\alpha_i}\right]\\ &\quad=\frac{\left(e^{\beta(1)}-1\right)e^{\alpha_i}}{\left(1+e^{\alpha_i+\beta(1)}\right)\left(1+e^{\alpha_i}\right)}=\frac{e^{\alpha_i+\beta(1)}}{\left(1+e^{\alpha_i+\beta(1)}\right)}-\frac{e^{\alpha_i}}{\left(1+e^{\alpha_i}\right)}=\Delta_{0\to 1}(\alpha_i)\end{aligned} \tag{92}$$

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

$$\frac{e^{\beta(1)}-1}{2}\left[\mathbb{P}_{0,0,1,0}+\mathbb{P}_{0,1,0,0}\right]+\frac{e^{\beta(1)}-1}{e^{\beta(1)}}\mathbb{P}_{0,0,1,1}+\left(e^{\beta(1)}-1\right)\left[\mathbb{P}_{1,0,1,0}+\mathbb{P}_{1,0,1,1}\right]=AME_{0\to 1} \tag{93}$$

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:

$$\begin{aligned}AME_{1\to 2} &= \frac{e^{\beta(2)}-e^{\beta(1)}}{2}\left[\mathbb{P}_{0,0,1,0}+\mathbb{P}_{0,1,0,0}\right]+\frac{e^{\beta(2)}-e^{\beta(1)}}{e^{\beta(1)}}\,\mathbb{P}_{0,0,1,1}\\ &+ \left(\frac{e^{\beta(2)}\left(1-e^{\beta(2)}\right)}{e^{\beta(1)}}+e^{\beta(2)}-1\right)\mathbb{P}_{0,1,1,0}\\ &+ \left(1-\frac{e^{\beta(1)}}{e^{\beta(2)}}\right)\left[\mathbb{P}_{1,0,1,0}+\mathbb{P}_{1,0,1,1}\right]+\left(\frac{e^{\beta(2)}-1}{e^{\beta(1)}}-1+\frac{1}{e^{\beta(2)}}\right)\mathbb{P}_{1,1,0,0}\end{aligned} \tag{94}$$

and

$$\begin{aligned}AME_{0\to 2} &= \frac{e^{\beta(2)}-1}{2}\left[\mathbb{P}_{0,0,1,0}+\mathbb{P}_{0,1,0,0}\right]+\frac{e^{\beta(2)}-1}{e^{\beta(1)}}\,\mathbb{P}_{0,0,1,1}\\ &+ \left(\frac{e^{\beta(2)}\left(1-e^{\beta(2)}\right)}{e^{\beta(1)}}+e^{\beta(2)}-1\right)\mathbb{P}_{0,1,1,0}\\ &+ \left(e^{\beta(1)}-\frac{e^{\beta(1)}}{e^{\beta(2)}}\right)\left[\mathbb{P}_{1,0,1,0}+\mathbb{P}_{1,0,1,1}\right]+\left(\frac{e^{\beta(2)}-1}{e^{\beta(1)}}-1+\frac{1}{e^{\beta(2)}}\right)\mathbb{P}_{1,1,0,0}\ \blacksquare\end{aligned} \tag{95}$$

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.

Bonhomme, S. (2011): "Panel Data, Inverse Problems, and the Estimation of Policy Parameters," .

Chamberlain, G. (1980): "Analysis of covariance with qualitative data," *Review of Economic Studies*, 47(1), 225–238.

——— (1984): "Panel data," in *Handbook of Econometrics, Vol. 2*. North-Holland Press.

——— (1985): "Heterogeneity, omitted variable bias, and duration dependence," in *Longitudinal Analysis of Labor Market Data*. Cambridge University Press.

Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81(2), 535–580.

Davezies, L., X. D'Haultfoeuille, and L. Laage (2021): "Identification and Estimation of average marginal effects in fixed effect logit models," .

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.

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

Honoré, B. E., C. Muris, and M. Weidner (2021): "Dynamic Ordered Panel Logit Models," *arXiv preprint arXiv:2107.03253*.

Keane, M. (1997): "Modeling heterogeneity and state dependence in consumer choice behavior," *Journal of Business & Economic Statistics*, 15(3), 310–327.

Lancaster, T. (2000): "The incidental parameter problem since 1948," *Journal of Econometrics*, 95(2), 391–413.

Magnac, T. (2000): "Subsidised training and youth employment: distinguishing unobserved heterogeneity from state dependence in labour market histories," *The Economic Journal*, 110(466), 805–837.

——— (2004): "Panel binary variables and sufficiency: generalizing conditional logit," *Econometrica*, 72(6), 1859–1876.

Neyman, J., and E. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16(1), 1–32.

Osborne, M. (2011): "Consumer learning, switching costs, and heterogeneity: A structural examination," *Quantitative Marketing and Economics*, 9(1), 25–70.

Pakel, C., and M. Weidner (2021): "Bounds on Average Effects in Discrete Choice Panel Data Models," .

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.

Severini, T., and G. Tripathi (2012): "Efficiency bounds for estimating linear functionals of nonparametric regression models with endogenous regressors," *Journal of Econometrics*, 170, 491–498.

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.