

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

February 8, 2020

Abstract

In nonlinear panel data models, fixed effects methods are often criticized based on the argument that they cannot identify average marginal effects (AMEs). The common argument is that: (1) the identification of AMEs requires knowledge of the distribution of unobserved heterogeneity; but (2) this distribution is not fixed- T identified in a fixed effects model because the data consist only of a finite number of probabilities. In this paper, we show that point (1) in this argument is incorrect. In a panel data dynamic logit model, we prove the point identification of the AME of a change in the lagged dependent variable. Despite the data comprise a finite number of probabilities, there is a combination of these probabilities that identifies this AME. We build on this result to show the identification of other AMEs of interest such as n periods forward AME of changes in the lagged dependent variable.

Keywords: Panel Data; Fixed Effects Models; Dynamic Discrete Choice; Identification; Average marginal effects.

JEL classifications: C23, C25, C51

Victor Aguirregabiria. Department of Economics, University of Toronto. 150 St. George Street, Toronto, Ontario, M5S 3G7.

Email: victor.aguirregabiria@utoronto.ca

Jesús M. Carro. Department of Economics, Universidad Carlos III de Madrid. C./ Madrid, 126, 28903 Getafe (Madrid), Spain.

Email: jcarro@eco.uc3m.es

*We would like to thank Jiaying Gu, Stephane Bonhomme, Irene Botosaru, and Chris Muris for helpful comments and discussions.

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 control for time-invariant unobserved heterogeneity. *Random effects (RE)* models use parametric assumptions on the distribution of the unobserved heterogeneity. In contrast, an attractive feature of *fixed effects (FE)* approaches is that they impose no restriction on this distribution such that the identified parameters are robust to misspecification of this primitive. Furthermore, FE approaches do not suffer from the *initial conditions problem* in dynamic models.

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 when the number of time periods is fixed. This is because the data consist only of a finite number of probabilities – the probabilities of every choice history – but the distribution of the unobserved heterogeneity has infinite dimension. This identification problem has generated another criticism of FE approaches that is important in empirical applications. The applied researcher is often interested in estimating average marginal effects (AMEs) of changes in the explanatory variables or in the structural parameters. These parameters are expectations over the distribution of the unobserved heterogeneity. Since the distribution of the unobserved heterogeneity is not identified, a common criticism of FE approaches is that they cannot (point) identify AMEs.

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 $T = 3$.

We extend this result to 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. Again, we obtain a

very simple analytical expression for the this AME in terms of probabilities of choice histories. We also prove that this identification results extend to a version of the model that includes exogenous explanatory variables.

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. These papers focus on the identification and estimation of the slope parameters in the index that defines the logit model but do not present identification results on AMEs. The papers that study identification of AMEs in FE models and that are more closely related to our paper are Chernozhukov, Fernandez-Val, Hahn, and Newey (2013; hereinafter CFHN), and Honoré and Kyriazidou (2019).

CFHN (2013) study the identification of AMEs in dynamic binary choice models. They study both nonparametric and semiparametric models. In the nonparametric model, the distributions of the two unobservables – the time-invariant and the transitory shock – are nonparametric. The semiparametric model assumes that the transitory shock has a known distribution – e.g., FE dynamic probit and logit models. The later corresponds to the model that we consider in this paper. They propose and implement a computational method to estimate the bounds that define the identified set of an AME. This method can be applied when the AME parameter is either point or partially identified. Using numerical examples, they find that the bounds for the AME can be very wide for the fully nonparametric model. For the semiparametric FE models, they find that the bounds for the AME shrink fast with T in both the dynamic and static models. For static probit and logit models with an exogenous binary regressor as the only covariate, they calculate numerically the bounds of the AME and they are quite tight, relative to the nonparametric bounds. Nevertheless, they find partial identification. They do not report any case for which they find point identification of the AME.

In contrast to CFHN, we consider a sequential identification approach. The log-likelihood function can be written as the sum of two functions: a function that contains all the terms that depend on the unobserved heterogeneity (or incidental parameters, α) – that we denote as $L_1(\alpha, \beta)$ – and the remaining part of the likelihood, that does not depend on the incidental parameters and we denote as $L_2(\beta)$. Previous results on conditional likelihood estimation of dynamic logit models (Chamberlain, 1985; Honoré and Kyriazidou, 2000) establish the identification of slope structural parameters based on the maximization of the (conditional) log-likelihood $L_2(\beta)$. This is the first

stage of our sequential approach. In the second stage, we take the β parameters as known to the researcher and consider the identification of AMEs. These AMEs are known functions of the slope parameters β and the distribution of the unobserved heterogeneity. We know that the likelihood function $L_1(\alpha, \beta)$ contains all the information in the data about the distribution of the unobserved heterogeneity. Therefore, given β , the empirical distribution of the sample statistics that enter in $L_1(\alpha, \beta)$ contain all the information about the AMEs. These statistics take a finite number of values such that the information about the AME can be represented by a simple system of equations. We show that simple manipulations in this system of equations provide a closed form expression for the AMEs.

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 model – our approach is computationally very simple as it provides closed form expressions for AME. However, using our approach, we have proved identification for only some AMEs, while CFHN provide a very general procedure that, in principle, can be used for any AME.¹

Honoré and Kyriazidou (2019) propose a numerical approach – in the spirit of Honoré and Tamer (2007) – to construct bounds for structural parameters and for marginal effects in FE dynamic binary choice models. The motivation of that paper is to illustrate that identification of these parameters can be more general than the existing results of point identification in the literature. The authors present numerical exercises to obtain bounds of the slope parameters but not of AMEs.

In a similar spirit, Arellano and Bonhomme (2017) mention that there may be room for point identification of some AME in FE nonlinear panel data models. As an example, they mention the point identification of AME for a particular subpopulation of individuals defined by the data. That identification result was also pointed out by Chamberlain (1982) and Hahn (2001). 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 AME has not been previously established in FE dynamic binary choice models.

Section 2 describes the models under study and the AME of interest. Section 3 presents our identification results. We summarize and conclude in section 4.

¹In all their numerical examples and two empirical illustrations, Chernozhukov, Fernandez-Val, Hahn, and Newey (2013) limit the use of their procedure to compute the bounds to a model with only one binary regressor.

2 Model and Average Marginal Effects

2.1 Model

Consider a panel dataset $\{y_{it}, x_{it} : i = 1, 2, \dots, N; t = 1, 2, \dots, T\}$. We study panel data dynamic logit models with the following form:

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

The transitory shock ε_{it} is *i.i.d.* with Logistic distribution. Variable x_{it} is strictly exogenous with respect to the transitory shock ε_{it} : that is, for any pair of time periods (t, s) , variables x_{it} and ε_{is} are independently distributed. We denote α_i as the (permanent) unobserved heterogeneity. The marginal distribution of α_i – that we represent as $f_\alpha(\alpha_i)$ – and the distribution of α_i conditional on the history of x variables – that we represent as $f_{\alpha|x}(\alpha_i|\mathbf{x}_i)$ with $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iT})$ – are completely unrestricted. Similarly, the probability of the initial value y_{i1} conditional on α_i and \mathbf{x}_i – that we represent as $p^*(y_{i1}|\alpha_i, \mathbf{x}_i)$ – is unrestricted. Following the standard setting in fixed effect (FE) approaches, our identification results are not based on any restriction on the initial choices, for instance, the assumption that initial choices are random draws from the individual-specific ergodic distribution of the endogenous variable.

Assumption 1 summarizes the conditions in this model.

ASSUMPTION 1: (A) ε_{it} is i.i.d. Logistic and independent of α_i ; (B) (strict exogeneity of x_{it}) for any two periods (t, s) , the variables ε_{it} and x_{is} are independently distributed; (C) the probability density functions $f_\alpha(\alpha_i)$ and $f_{\alpha|x}(\alpha_i|\mathbf{x}_i)$, and the probability of the initial condition $p^(y_{i1}|\alpha_i, \mathbf{x}_i)$ are completely unrestricted; and (D) the random variable $x_{it} - x_{it-1}$ has support in a neighborhood of zero. ■*

Assumption 1(D) is used by Honoré and Kyriazidou (2000) in their proof for the identification of the parameters β and γ . This condition basically rules out time dummies. It is straightforward to extend our results to a model with multiple x variables. For notational simplicity, we have preferred to present a model with only one exogenous x variable.

For some of our results, we focus on a restricted version of this model with $\gamma = 0$, i.e., pure AR1 model without exogenous covariates. For the rest of the paper we use the terms *AR1 model* and *AR1X* to indicate the models with $\gamma = 0$ and with unrestricted γ , respectively.

2.2 Average Marginal Effects (AME)

2.2.1 AR1 model: One period forward AME

The AR1 model implies the individual-specific transition probabilities $\mathbb{P}(y_{it} = 1 | y_{it-1} = 0; \alpha_i) = \Lambda(\alpha_i)$ and $\mathbb{P}(y_{it} = 1 | y_{it-1} = 1; \alpha_i) = \Lambda(\alpha_i + \beta)$, where $\Lambda(u)$ is the CDF of the Logistic distribution, that is, $\Lambda(u) = \exp(u) / [1 + \exp(u)]$.

Let $\Delta^{(1)}(\alpha_i)$ be the *individual effect* on y_{it} of an exogenous change in variable y_{it-1} from 0 to 1. That is, $\Delta^{(1)}(\alpha_i) \equiv \mathbb{E}(y_{it} | \alpha_i, y_{it-1} = 1) - \mathbb{E}(y_{it} | \alpha_i, y_{it-1} = 0)$. For the AR1 model, we have that:

$$\Delta^{(1)}(\alpha_i) = \Lambda(\alpha_i + \beta) - \Lambda(\alpha_i) \quad (2)$$

This parameter measures individual i 's persistence in state 1 generated by the true state dependence. It is also an individual specific treatment 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 of the incidental parameters problem (Neyman and Scott, 1948; Heckman, 1981; Lancaster, 2000). Therefore, we can identify $\Delta^{(1)}(\alpha)$ for an hypothetical value of α , but not for the value of α that actually corresponds to individual i . That is, the individual-specific $\Delta^{(1)}(\alpha_i)$ are not identified. Instead, we study the identification of the following *Average Marginal Effect (AME)*:

$$AME^{(1)} \equiv \int \Delta^{(1)}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i = \int [\Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)] f_\alpha(\alpha_i) d\alpha_i \quad (3)$$

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

EXAMPLE 1. Consider a model of employment / unemployment status, where y_{it} is the indicator that individual i is employed at period t . Let $U_{it}(1, y_{i,t-1})$ and $U_{it}(0, y_{i,t-1})$ be the utilities if employed and unemployed, respectively. And individual decides to work if $U_{it}(1, y_{i,t-1}) - U_{it}(0, y_{i,t-1}) \geq 0$, or equivalently, if $U_{it}(1, 0) - U_{it}(0, 0) + y_{i,t-1} [U_{it}(1, 1) - U_{it}(1, 0) - U_{it}(0, 1) - U_{it}(0, 0)] \geq 0$. Our model imposes the restriction that $U_{it}(1, 0) - U_{it}(0, 0) = \alpha_i + \varepsilon_{it}$ and $U_{it}(1, 1) - U_{it}(1, 0) - U_{it}(0, 1) - U_{it}(0, 0) = \beta$. Therefore, parameter β captures the additional utility of being employed (relative to being unemployed) due to being unemployed at previous period. It

captures true state dependence in employment. 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 individuals randomly in two groups, say groups 0 and 1. Individuals in group 0 are assigned to the unemployment status; and individuals in group 1 to employment status. Then, after one period we look at the proportion of individuals who are employed in each of the two groups. $AME^{(1)}$ is equal to the proportion of employed individuals in group 1 minus the proportion of employed individual in group 0.

The parameter $AME^{(1)}$ is also related to the average treatment effects from two policy experiments. Consider a policy experiment where individuals in the experimental group are assigned to employment status at period $t - 1$ – e.g., they receive a large temporary subsidy for being employed – and individuals in the control group are left in their observed status at period $t - 1$. Then, at period t the researcher observes the proportion of individuals that remain employed 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_1 . We can consider a similar experiment but where individuals in the experimental group are assigned to unemployment at period $t - 1$ – e.g., they receive a large temporary subsidy for being unemployed. We use ATE_0 to denote the average effect of this other policy treatment. It is straightforward to show that $AME^{(1)} = ATE_1 - ATE_0$. ■

The average treatment effects ATE_1 and ATE_0 described in Example 1 are of interest to the applied researcher. They can be defined as follows:

$$\begin{aligned}
 ATE_{1,t} &\equiv \int \Lambda(\alpha_i + \beta) f_\alpha(\alpha_i) d\alpha_i - \int \Lambda(\alpha_i + \beta y_{it}) f_{(\alpha,y)t}(\alpha_i, y_{it}) d(\alpha_i, y_{it}) \\
 ATE_{0,t} &\equiv \int \Lambda(\alpha_i) f_\alpha(\alpha_i) d\alpha_i - \int \Lambda(\alpha_i + \beta y_{it}) f_{(\alpha,y)t}(\alpha_i, y_{it}) d(\alpha_i, y_{it})
 \end{aligned}
 \tag{4}$$

where $f_{(\alpha,y)t}(\alpha_i, y_{it})$ represents the joint distribution of α_i and y_{it} at period t . In section 3, we show that these ATEs are also identified.

2.2.2 AR1 model: n periods forward AME

Researchers can be interested in how the response to a treatment evolves over time. Let $\Delta^{(n)}(\alpha_i)$ be the individual effect on y_{it+n-1} of an exogenous change in y_{it-1} from 0 to 1. By definition, $\Delta^{(n)}(\alpha_i) \equiv \mathbb{E}(y_{it+n-1} \mid \alpha_i, y_{it-1} = 1) - \mathbb{E}(y_{it+n-1} \mid \alpha_i, y_{it-1} = 0)$. Similarly as $\Delta^{(1)}(\alpha_i)$, this n -periods forward individual effect is not identified using a short-panel due to the incidental

parameters problem. We are interested in the average of this n periods forward effect:

$$AME^{(n)} \equiv \int \Delta^{(n)}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \quad (5)$$

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

LEMMA 1. The first order Markov structure of the AR1 model implies that – at the individual level – $\Delta^{(n)}(\alpha_i) = [\Delta^{(1)}(\alpha_i)]^n$. Therefore, the n -periods forward AME can be represented as:

$$AME^{(n)} = \int [\Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)]^n f_\alpha(\alpha_i) d\alpha_i. \quad \blacksquare \quad (6)$$

Proof: First, note that $\mathbb{E}(y_{it+n-1} \mid \alpha_i, y_{it-1}) = \mathbb{P}(y_{it+n-1} = 1 \mid \alpha_i, y_{it-1})$. Using the Markov structure of the model and the chain rule, we have that

$$\begin{aligned} \mathbb{P}(y_{it+n-1} = 1 \mid \alpha_i, y_{it-1}) &= \mathbb{P}(y_{it+n-2} = 0 \mid \alpha_i, y_{it-1}) \mathbb{P}(y_{it+n-1} = 1 \mid \alpha_i, y_{it+n-2} = 0) \\ &+ \mathbb{P}(y_{it+n-2} = 1 \mid \alpha_i, y_{it-1}) \mathbb{P}(y_{it+n-1} = 1 \mid \alpha_i, y_{it+n-2} = 1) \\ &= \Lambda(\alpha_i) + \mathbb{P}(y_{it+n-2} = 1 \mid \alpha_i, y_{it-1}) [\Lambda(\beta + \alpha_i) - \Lambda(\alpha_i)] \end{aligned} \quad (7)$$

Using the same argument, we can obtain that $\mathbb{P}(y_{it+n-1} = 1 \mid \alpha_i, y_{it-1}) = \Lambda(\alpha_i) + \mathbb{P}(y_{it+n-2} = 1 \mid \alpha_i, y_{it-1}) [\Lambda(\beta + \alpha_i) - \Lambda(\alpha_i)]$. Taking these expressions into account and the definition $\Delta^{(n)}(\alpha_i) = \mathbb{P}(y_{it+n-1} = 1 \mid \alpha_i, y_{it-1} = 1) - \mathbb{P}(y_{it+n-1} = 1 \mid \alpha_i, y_{it-1} = 0)$, we have that:

$$\Delta^{(n)}(\alpha_i) = \Delta^{(n-1)}(\alpha_i) [\Lambda(\beta + \alpha_i) - \Lambda(\alpha_i)] \quad (8)$$

Applying this expression recursively to $\Delta^{(n-1)}(\alpha_i)$, then to $\Delta^{(n-2)}(\alpha_i)$, and so on, we obtain that $\Delta^{(n)}(\alpha_i) = [\Lambda(\beta + \alpha_i) - \Lambda(\alpha_i)]^n = [\Delta^{(1)}(\alpha_i)]^n$. \blacksquare

2.2.3 AR1X model: AMEs of change in y_{t-1}

In the AR1X model, the AME of the effect of y_{it-1} on y_{it} has to take into account the presence of x and its potential (and unrestricted) relation with α_i . Let $\Delta^{(1)}(\alpha_i, x)$ be the individual effect on y_{it} of an exogenous change in variable y_{it-1} from 0 to 1 when $x_{it} = x$. That is,

$$\begin{aligned} \Delta^{(1)}(\alpha_i, x) &\equiv \mathbb{E}(y_{it} \mid \alpha_i, x_{it} = x, y_{it-1} = 1) - \mathbb{E}(y_{it} \mid \alpha_i, x_{it} = x, y_{it-1} = 0) \\ &= \Lambda(\alpha_i + \beta + \gamma x) - \Lambda(\alpha_i + \gamma x) \end{aligned} \quad (9)$$

This marginal effect is individual specific and, as we have discussed above, it is not identified due to the incidental parameters problem.

We present identification results for two different average effects in the AR1X model. A first AME is based on the condition that variable x remains constant over the T sample periods:

$$\begin{aligned} AME_x^{(1)}(x) &\equiv \int \Delta^{(1)}(\alpha_i, x) f_{\alpha|x}(\alpha_i | \mathbf{x}_i = (x, x, \dots, x)) d\alpha_i \\ &= \int [\Lambda(\alpha_i + \beta + \gamma x) - \Lambda(\alpha_i + \gamma x)] f_{\alpha|\mathbf{x}}(\alpha_i | \mathbf{x}_i = (x, x, \dots, x)) d\alpha_i \end{aligned} \tag{10}$$

Note that value x for the exogenous variable enters in two different ways in the definition of this AME: in the individual effect $\Delta^{(1)}(\alpha_i, x)$; and in the conditional distribution of α_i .

We also show the identification of the following AME.

$$\begin{aligned} AME_{x,t}^{(1)} &\equiv \int \Delta^{(1)}(\alpha_i, x_{it}) f_{(\alpha, x_t)}(\alpha_i, x_{it}) d(\alpha_i, x_{it}) \\ &= \int [\Lambda(\alpha_i + \beta + \gamma x_{it}) - \Lambda(\alpha_i + \gamma x_{it})] f_{(\alpha, x_t)}(\alpha_i, x_{it}) d(\alpha_i, x_{it}) \end{aligned} \tag{11}$$

Note that this AME is not conditional to a value of x but integrated over the distribution of x_{it} at period t . Therefore, there are as many of these AMEs as periods in the sample. Under the stationarity of the joint distribution $f_{(\alpha, x_t)}(\alpha_i, x_{it})$, all these AMEs should be equal. We show that – given the identification of these AMEs at every sample period – this is a testable restriction.

Chamberlain (1984) presents this kind of AME as the expected effect over y_{it} of changing y_{it-1} for a randomly drawn individual. By integrating over the joint distribution of x_t and α_i , this takes into account any potential relation between the observable and unobservable characteristics that might exist, which is one of the main advantages of the fixed effects approach.

EXAMPLE 2. Consider the model of employment / unemployment status in Example 1, but now we extend this model to include the exogenous explanatory variable x_{it} that represents the individual's health status at period t . For simplicity, suppose that x_{it} is binary such that $x_{it} = 1$ and $x_{it} = 0$ represent "good" and "bad" health status, respectively. Then, $AME_x^{(1)}(x = 1)$ represents the AME of employment versus unemployment given that the individual has good health at the period of treatment, and for the subpopulation of individuals that are observed with good health over the three sample periods. The parameter $AME_{x,t}^{(1)}$ is the effect of employment versus unemployment average over all the individuals according to their distribution of health status at period t . ■

Similarly as for the AR1 model, we are also interested in n-periods forward AMEs for this AR1X model. This definition is straightforward for the AME conditional on a constant value of x :

$$AME_x^{(n)}(x) = \int [\Lambda(\alpha_i + \beta + \gamma x) - \Lambda(\alpha_i + \gamma x)]^n f_{\alpha|\mathbf{x}}(\alpha_i|\mathbf{x}_i = (x, x, \dots, x)) d\alpha_i \quad (12)$$

3 Identification

We first review existing results on the identification of the parameters β in the AR1 model using a conditional maximum likelihood (CML) approach. Second, based on the structure of the likelihood function, we show that the sufficient statistic used in the CML approach contains all the information in the sample on the distribution of the unobserved heterogeneity. Third, we show that the *average marginal effect* $AME^{(1)}$ is identified as a linear combination of empirical probabilities of the sufficient statistic, where the weights in this linear combination are known functions of the parameter β . For the AR1X model, we show the identification of β and γ using a CML approach. Given these slope parameters, we show the identification of the average marginal effects $AME_x^{(1)}(x)$, $AME_x^{(n)}(x)$, and $AME_{x,t}^{(1)}$.

3.1 AR1 model: Preliminary results

The identification of the parameter β in the AR1 model has been proved in Chamberlain (1985). Let \mathbf{y}_i be the vector with the choice history of individual i , that is $\mathbf{y}_i \equiv (y_{i1}, y_{i2}, \dots, y_{iT})$. And let α represent the vector of *fixed effects* or *incidental parameters* for the N individuals, that is $\alpha \equiv (\alpha_1, \alpha_2, \dots, \alpha_N)$. The log-likelihood function where α 's are treated as parameters is:

$$\ell(\alpha, \beta) = \sum_{i=1}^N \ln \mathbb{P}(\mathbf{y}_i|\alpha_i, \beta) = \sum_{i=1}^N \ln \left[\prod_{t=2}^T \frac{\exp\{y_{it}[\alpha_i + \beta y_{i,t-1}]\}}{1 + \exp\{\alpha_i + \beta y_{i,t-1}\}} p^*(y_{i1}|\alpha_i) \right] \quad (13)$$

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

In this model, the log-probability of a choice history, $\ln \mathbb{P}(\mathbf{y}_i|\alpha_i, \beta)$, has the following structure:

$$\ln \mathbb{P}(\mathbf{y}_i|\alpha_i, \beta) = \mathbf{s}(\mathbf{y}_i)' g(\alpha_i) + c(\mathbf{y}_i) \beta \quad (14)$$

where: $\mathbf{s}(\mathbf{y}_i)$ is a vector of statistics (functions of \mathbf{y}_i) with elements $(y_{i1}, y_{iT}, \sum_{t=2}^T y_{it})$; $c(\mathbf{y}_i)$ is the statistic $\sum_{t=2}^T y_{it} y_{i,t-1}$; and $g(\alpha_i)$ is a vector of functions α_i with the following elements: $\ln p^*(1|\alpha_i) - \ln p^*(0|\alpha_i) + \ln(1 + \exp\{\alpha_i\}) - \ln(1 + \exp\{\alpha_i + \beta\})$, $\ln(1 + \exp\{\alpha_i + \beta\}) - \ln(1 + \exp\{\alpha_i\})$,

and $\alpha_i + \ln(1 + \exp\{\alpha_i + \beta\}) - \ln(1 + \exp\{\alpha_i\})$.² For notational simplicity, we use \mathbf{s}_i and c_i to represent $\mathbf{s}(\mathbf{y}_i)$ and $c(\mathbf{y}_i)$, respectively.

Given this structure of the log-probability of a choice history, it is simple to show that \mathbf{s}_i is a *minimal sufficient statistic* for α_i . That is, (1) \mathbf{s}_i is a sufficient statistic for α_i such that $\mathbb{P}(\mathbf{y}_i|\alpha_i, \beta, \mathbf{s}_i) = \mathbb{P}(\mathbf{y}_i|\beta, \mathbf{s}_i)$; and (2) \mathbf{s}_i is *minimal* because its elements are linearly independent.³ More precisely, we have that:

$$\mathbb{P}(\mathbf{y}_i|\alpha_i, \beta, \mathbf{s}_i) = \frac{\mathbb{P}(\mathbf{y}_i|\alpha_i, \beta)}{\mathbb{P}(\mathbf{s}_i|\alpha_i, \beta)} = \frac{\exp\{\mathbf{s}_i' g(\alpha_i) + c_i \beta\}}{\sum_{\mathbf{y}: \mathbf{s}(\mathbf{y})=\mathbf{s}_i} \exp\{\mathbf{s}_i' g(\alpha_i) + c(\mathbf{y}) \beta\}} = \frac{\exp\{c_i \beta\}}{\sum_{\mathbf{y}: \mathbf{s}(\mathbf{y})=\mathbf{s}_i} \exp\{c(\mathbf{y}) \beta\}} \quad (15)$$

where $\sum_{\mathbf{y}: \mathbf{s}(\mathbf{y})=\mathbf{s}_i}$ represents the sum over all the possible (hypothetical) choice histories \mathbf{y} with $\mathbf{s}(\mathbf{y})$ equal to \mathbf{s}_i . Furthermore, when $T \geq 4$, the statistic c_i has independent variation with respect to \mathbf{s}_i .

These results imply that we can (point) identify β using a Conditional Maximum Likelihood (CML) approach. We now describe this approach. From the sufficiency of \mathbf{s}_i , we have that $\mathbb{P}(\mathbf{y}_i|\beta, \mathbf{s}_i) = \mathbb{P}(\mathbf{y}_i|\alpha_i, \beta) / \mathbb{P}(\mathbf{s}_i|\alpha_i, \beta)$, and this implies that $\ln \mathbb{P}(\mathbf{y}_i|\alpha_i, \beta) = \ln \mathbb{P}(\mathbf{y}_i|\beta, \mathbf{s}_i) + \ln \mathbb{P}(\mathbf{s}_i|\alpha_i, \beta)$. Based on this property, we can write the log-likelihood function $\ell(\alpha, \beta)$ as the sum of two likelihoods:

$$\ell(\alpha, \beta) = \ell^C(\mathbf{C}, \mathbf{S}; \beta) + \ell^S(\mathbf{S}; \alpha, \beta) \quad (16)$$

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

$$\ell^C(\mathbf{C}, \mathbf{S}; \beta) = \sum_{i=1}^N \ln \mathbb{P}(\mathbf{y}_i|\beta, \mathbf{s}_i) = \sum_{i=1}^N c_i \beta - \sum_{i=1}^N \ln \left[\sum_{\mathbf{y}: \mathbf{s}(\mathbf{y})=\mathbf{s}_i} \exp\{c(\mathbf{y}) \beta\} \right] \quad (17)$$

and

$$\ell^S(\mathbf{S}; \alpha, \beta) = \sum_{i=1}^N \ln \mathbb{P}(\mathbf{s}_i|\alpha_i, \beta) = \sum_{i=1}^N \ln \left[\sum_{\mathbf{y}: \mathbf{s}(\mathbf{y})=\mathbf{s}_i} \exp\{\mathbf{s}_i' g(\alpha_i) + c(\mathbf{y}) \beta\} \right] \quad (18)$$

Function $\ell^C(\mathbf{C}, \mathbf{S}; \beta)$ is the *conditional log-likelihood function*. It is a well-defined likelihood and it has the important properties that does not depend on the incidental parameters α and that the maximization of this function uniquely identifies the parameter of interest β . In fact, the (sample) conditional log-likelihood function is globally concave in β . Function $\ell^S(\mathbf{S}; \alpha, \beta)$ is the likelihood for the sufficient statistic \mathbf{s}_i . All the information in the sample about the incidental

²Strictly speaking, the vector of statistics $\mathbf{s}(\mathbf{y}_i)$ includes also the number of periods T . However, this is a constant that does not vary over the different histories such that w.l.o.g. we can consider that $\mathbf{s}(\mathbf{y}_i) = (y_{i1}, y_{iT}, \sum_{t=2}^T y_{it})$.

³Aguirregabiria, Gu, and Luo (2019) show that this result applies also to extended versions of this model that include duration dependence and where individuals' behavior is forward-looking.

parameters appears only in this function. Note that it depends on the data only through the sufficient statistics \mathbf{s}_i . Therefore, equation (16) implies that, given β , all the information in the data about the incidental parameters – and therefore, about the distribution of α_i – appears in the empirical distribution of the sufficient statistic \mathbf{s}_i . This result can be used for our main identification result that we describe below.

3.2 Identification of $AME^{(1)}$ in AR1 Model

Parameter β is identified from the maximization of the conditional likelihood function $\ell^C(\mathbf{C}, \mathbf{S}; \beta)$. Therefore, to study identification of $AME^{(1)}$ we can treat β as known. The parameter $AME^{(1)}$ depends on β and on the probability distribution of the incidental parameters α_i . As shown above, given β , the probability distribution of \mathbf{s}_i – denoted as $p_s(\mathbf{s})$ – contains all the information in the sample about the incidental parameters α_i . Therefore, the parameter $AME^{(1)}$ is point identified if and only if we can uniquely determine this parameter as a function of β and $p_s(\mathbf{s})$.

Our proof of the identification of $AME^{(1)}$ exploits a relationship between the individual effect $\Delta^{(1)}(\alpha_i)$ and the transition probabilities of the model. Define the (individual specific) transition probabilities $\pi(\alpha_i, y) \equiv \mathbb{P}(y_{it+1} = 1 | \alpha_i, y_{it} = y)$. By definition – and without the Logit assumption – the individual effect $\Delta^{(1)}(\alpha_i)$ is equal to $\pi(\alpha_i, 1) - \pi(\alpha_i, 0)$. The following Lemma 2 establishes other relationship between the AME and the transition probabilities that holds for the Logit model.

LEMMA 2. In the AR1 Logit model, we have that:

$$\pi(\alpha_i, 0) [1 - \pi(\alpha_i, 1)] = \Delta^{(1)}(\alpha_i) \frac{1}{\exp\{\beta\} - 1} \quad \blacksquare \quad (19)$$

Proof of Lemma 2. By definition, we have that: $\pi(\alpha_i, 0) = \exp\{\alpha_i\} [1 + \exp\{\alpha_i\}]^{-1}$; and $1 - \pi(\alpha_i, 1) = [1 + \exp\{\alpha_i + \beta\}]^{-1}$, such that $\pi(\alpha_i, 0) [1 - \pi(\alpha_i, 1)] = \exp\{\alpha_i\} [1 + \exp\{\alpha_i\}]^{-1} [1 + \exp\{\alpha_i + \beta\}]^{-1}$. We also have that $\Delta^{(1)}(\alpha_i) = \pi(\alpha_i, 1) - \pi(\alpha_i, 0) = [\exp\{\alpha_i + \beta\} - \exp\{\alpha_i\}] [1 + \exp\{\alpha_i + \beta\}]^{-1} [1 + \exp\{\alpha_i\}]^{-1}$. Combining these expressions, we have that $\pi(\alpha_i, 0) [1 - \pi(\alpha_i, 1)] / \Delta^{(1)}(\alpha_i) = [\exp\{\beta\} - 1]^{-1}$. \blacksquare

Proposition 1 establishes the identification of $AME^{(1)}$.

PROPOSITION 1. (A) If $T \geq 4$, then the parameter β is uniquely identified from the maximization of the conditional likelihood function $\ell^C(\mathbf{C}, \mathbf{S}; \beta)$. (B) Given β , if $T \geq 3$, then the parameter $AME^{(1)}$ is identified as:

$$AME^{(1)} = [\exp\{\beta\} - 1] [\mathbb{P}(0, 1, 0) + \mathbb{P}(1, 0, 1)] \quad (20)$$

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

Proof of Proposition 1: Part (A) was proved in Chamberlain (1985). In particular, it can be shown that $\beta = \ln \mathbb{P}(0, 0, 1, 1) - \ln \mathbb{P}(0, 1, 0, 1)$. For part (B), w.l.o.g. we consider that $T = 3$.⁴ We have that:

$$\begin{cases} \mathbb{P}(0, 1, 0) &= \int p^*(0|\alpha_i) \pi(\alpha_i, 0) [1 - \pi(\alpha_i, 1)] f_\alpha(\alpha_i) d\alpha_i \\ \mathbb{P}(1, 0, 1) &= \int p^*(1|\alpha_i) \pi(\alpha_i, 0) [1 - \pi(\alpha_i, 1)] f_\alpha(\alpha_i) d\alpha_i \end{cases} \quad (21)$$

Applying Lemma 2, we have that:

$$\begin{cases} \mathbb{P}(0, 1, 0) &= \frac{1}{\exp\{\beta\} - 1} \int p^*(0|\alpha_i) \Delta^{(1)}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \\ \mathbb{P}(1, 0, 1) &= \frac{1}{\exp\{\beta\} - 1} \int p^*(1|\alpha_i) \Delta^{(1)}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \end{cases} \quad (22)$$

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. ■

Remark 1.1. With $T = 3$, each value of the sufficient statistic \mathbf{s}_i corresponds to only one value of the choice history \mathbf{y}_i . Therefore, the representation of $AME^{(1)}$ in equation (20) in terms of probabilities $\mathbb{P}(\mathbf{y}_i)$ is equivalent to the representation in terms of $\mathbb{P}(\mathbf{s}_i)$.

Remark 1.2 (Estimation). Equation (20) provides a simple analog or plug-in estimator for $AME^{(1)}$. In a first step, we estimate β using CML and probabilities $\mathbb{P}(0, 1, 0)$ and $\mathbb{P}(1, 0, 1)$ using a frequency estimator. Then, we plug these estimates in equation (20) to obtain an estimate of $AME^{(1)}$. This estimator is root-N consistent and asymptotically normal. When $T = 3$, this estimator is also efficient because equation (20) contains all the information in the distribution of $\mathbb{P}(\mathbf{s}_i)$ about the $AME^{(1)}$. However, when $T \geq 4$, this estimator is not efficient because the model imposes additional restrictions on $AME^{(1)}$. That is, with $T \geq 4$ this AME is over-identified. In section 3.5, we describe a general procedure to obtain this AME as a function of β and the distribution $\mathbb{P}(\mathbf{s}_i)$ for higher values of T using all the information available. For instance, for $T = 4$, we show that:

$$AME^{(1)} = \frac{\exp\{\beta\} - 1}{2} [p_s(0, 0, 1) + p_s(1, 1, 2)] + \frac{\exp\{\beta\} - 1}{\exp\{\beta\} + 1} [p_s(0, 1, 2) + p_s(1, 0, 1)] \quad (23)$$

where $p_s(s_1, s_2, s_3) \equiv \mathbb{P}\left((y_{i1}, y_{iT}, \sum_{t=2}^T y_{it}) = (s_1, s_2, s_3)\right)$.

⁴Given identification with $T = 3$, it is obvious that there is also identification for any value of T greater than 3. We can always take subhistories with three periods.

3.3 Identification of $AME^{(n)}$ in AR1 Model

Our proof of the identification of $AME^{(n)}$ builds on Lemma 1 and Lemma 2. First, Lemma 1 establishes that $\Delta^{(n)}(\alpha_i) = [\Delta^{(1)}(\alpha_i)]^n$. Combining this with Lemma 2, we have that for any integer $n \geq 1$:

$$[\pi(\alpha_i, 0)]^n [1 - \pi(\alpha_i, 1)]^n = \Delta^{(n)}(\alpha_i) \frac{1}{[\exp\{\beta\} - 1]^n} \quad (24)$$

We use this relationship in our proof of the following Proposition 2.

PROPOSITION 2. *Let n be any positive integer. If $T \geq 2n + 1$ and β is known (identified), then parameter $AME^{(n)}$ is identified as:*

$$AME^{(n)} = [\exp\{\beta\} - 1]^n [\mathbb{P}(A) + \mathbb{P}(B)] \quad (25)$$

where $\mathbb{P}(A)$ and $\mathbb{P}(B)$ are the probabilities of choice histories A and B , over $2n + 1$ periods, with $A = (0, 1, 0, 1, \dots, 0)$ and $B = (1, 0, 1, 0, \dots, 1)$. For instance: for $n = 2$, $A = (0, 1, 0, 1, 0)$ and $B = (1, 0, 1, 0, 1)$; for $n = 3$, $A = (0, 1, 0, 1, 0, 1, 0)$ and $B = (1, 0, 1, 0, 1, 0, 1)$; and so on. ■

Proof of Proposition 2: W.l.o.g. we consider that $T = 2n + 1$. Given the definition of histories A and B in Proposition 2, it is straightforward to see that:

$$\begin{cases} \mathbb{P}(A) = \int p^*(0|\alpha_i) [\pi(\alpha_i, 0)]^n [1 - \pi(\alpha_i, 1)]^n f_\alpha(\alpha_i) d\alpha_i \\ \mathbb{P}(B) = \int p^*(1|\alpha_i) [\pi(\alpha_i, 0)]^n [1 - \pi(\alpha_i, 1)]^n f_\alpha(\alpha_i) d\alpha_i \end{cases} \quad (26)$$

Applying equation (24) (Corollary of Lemmas 1 and 2), we have that:

$$\begin{cases} \mathbb{P}(A) = \frac{1}{[\exp\{\beta\} - 1]^n} \int p^*(0|\alpha_i) \Delta^{(n)}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \\ \mathbb{P}(B) = \frac{1}{[\exp\{\beta\} - 1]^n} \int p^*(1|\alpha_i) \Delta^{(n)}(\alpha_i) f_\alpha(\alpha_i) d\alpha_i \end{cases} \quad (27)$$

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 $AME^{(n)} = [\exp\{\beta\} - 1]^n [\mathbb{P}(A) + \mathbb{P}(B)]$ such that $AME^{(n)}$ is identified. ■

3.4 Identification of ATE_1 and ATE_0 in AR1 Model

In Example 1, we introduce two average treatment effects that can be particularly relevant to the applied researcher because they are closely related to counterfactual policy experiments. The

average treatment effect $ATE_{1,t}$ is the effect of a treatment that (exogenously) makes $y_{it} = 1$ for the individuals in the experimental group and leaves y_{it} as it is for the individuals in the control group. Average treatment effect $ATE_{0,t}$ has very similar interpretation but treatment in the experimental group is $y_{it} = 0$.

PROPOSITION 3. Given β , if $T \geq 3$, then parameter $ATE_{1,t}$ is identified as:

$$ATE_{1,t} = \exp\{\beta\} \mathbb{P}(0, 1, 0) + \mathbb{P}(0, 1, 1) + \mathbb{P}(1, 1, 0) + \mathbb{P}(1, 1, 1) - \mathbb{P}(y_{it} = 1|t) \quad (28)$$

where $\mathbb{P}(y_{it} = 1|t)$ is the empirical probability of $y_{it} = 1$ at period t . By definition, $ATE_{0,t} = ATE_{1,t} - AME^{(1)}$, such that $ATE_{0,t}$ is identified from the identification of $ATE_{1,t}$ and $AME^{(1)}$.

■

Proof of Proposition 3: W.l.o.g. we consider that $T = 3$. By definition, ATE_1 is difference between the expected value of y_{it} for the experimental group and for the control group. Since the control group does not receive any treatment, we have that

has two parts, the second part is

$$E(y_{it}|i \in Control) = \int \Lambda(\alpha_i + \beta y_{it-1}) f_{(\alpha,y)_{t-1}}(\alpha_i, y_{it-1}) d(\alpha_i, y_{it-1}) = \mathbb{P}(y_{it} = 1|t) \quad (29)$$

such that this term is identified. Therefore, it remains to prove the identification of the term:

$$E(y_{it}|i \in Experimental) = \int \Lambda(\alpha_i + \beta) f_{\alpha}(\alpha_i) d\alpha_i \quad (30)$$

We have that:

$$\begin{cases} \mathbb{P}(0, 1, 0) &= \int p^*(0|\alpha_i) \Lambda(\alpha_i) [1 - \Lambda(\alpha_i + \beta)] f_{\alpha}(\alpha_i) d\alpha_i \\ \mathbb{P}(0, 1, 1) &= \int p^*(0|\alpha_i) \Lambda(\alpha_i) \Lambda(\alpha_i + \beta) f_{\alpha}(\alpha_i) d\alpha_i \\ \mathbb{P}(1, 1, 0) &= \int p^*(1|\alpha_i) \Lambda(\alpha_i + \beta) [1 - \Lambda(\alpha_i + \beta)] f_{\alpha}(\alpha_i) d\alpha_i \\ \mathbb{P}(1, 1, 1) &= \int p^*(1|\alpha_i) \Lambda(\alpha_i + \beta) \Lambda(\alpha_i + \beta) f_{\alpha}(\alpha_i) d\alpha_i \end{cases} \quad (31)$$

Note that $\exp(\beta) \Lambda(\alpha_i) [1 - \Lambda(\alpha_i + \beta)] = \Lambda(\alpha_i + \beta) (1 - \Lambda(\alpha_i))$, such that

$$\exp\{\beta\} \mathbb{P}(0, 1, 0) + \mathbb{P}(0, 1, 1) = \int p^*(0|\alpha_i) \Lambda(\alpha_i + \beta) f_{\alpha}(\alpha_i) d\alpha_i \quad (32)$$

Furthermore, we have that:

$$\mathbb{P}(1, 1, 0) + \mathbb{P}(1, 1, 1) = \int p^*(1|\alpha_i) \Lambda(\alpha_i + \beta) f_{\alpha}(\alpha_i) d\alpha_i \quad (33)$$

Adding up (32) and (33), and taking into account that $p^*(0|\alpha_i) + p^*(1|\alpha_i) = 1$, we have that $\int \Lambda(\alpha_i + \beta) f_\alpha(\alpha_i) d\alpha_i = \exp\{\beta\} \mathbb{P}(0, 1, 0) + \mathbb{P}(0, 1, 1) + \mathbb{P}(1, 1, 0) + \mathbb{P}(1, 1, 1)$. This, together with (29), implies that $ATE_1 = \exp\{\beta\} \mathbb{P}(0, 1, 0) + \mathbb{P}(0, 1, 1) + \mathbb{P}(1, 1, 0) + \mathbb{P}(1, 1, 1) - \mathbb{P}(y_{it} = 1)$, such that ATE_1 is identified. ■

3.5 Identification of AMEs in AR1X model

Under the conditions in Assumption 1, Honoré and Kyriazidou (2000) show that the parameters β and γ in AR1X model are identified as long as $T \geq 4$. Note that we can take a subhistory of four periods, condition on $x_{i1} = x_{i2} = x_{i3} = x_{i4}$, and apply the sufficient statistics CML approach described above to identify β . In the same spirit, we first show that the identification results for the AMEs that we have presented above can be extended to the model with exogenous explanatory variables when we restrict the AMEs to the subpopulation where the x variables are constant over time. Then, we show the identification of other AMEs which do not impose this restriction.

PROPOSITION 4. *Under the conditions of Assumption 1: (A) If $T \geq 4$, then the parameters β and γ are uniquely identified. (B) Given β , if $T \geq 3$, then the parameter $AME_x^{(1)}(x^0)$ defined in (10) is identified for any value x^0 in the support of x_{it} . More specifically, given $\mathbf{x}^0 = (x^0, x^0, \dots, x^0)$, we have that:*

$$AME^{(1)}(x^0) = [\exp\{\beta\} - 1] [\mathbb{P}(0, 1, 0 | \mathbf{x}^0) + \mathbb{P}(1, 0, 1 | \mathbf{x}^0)]. \quad \blacksquare \quad (34)$$

Proof of Proposition 4: Part (A) was proved in Honoré and Kyriazidou (2000; section 2.1). For part (B), define the transition probabilities $\pi(\alpha_i, x^0, y) \equiv \mathbb{P}(y_{it} = 1 | \alpha_i, x_{it} = x^0, y_{it} = y)$. By definition, the individual effect $\Delta^{(1)}(\alpha_i, x^0)$ is equal to $\pi(\alpha_i, x^0, 1) - \pi(\alpha_i, x^0, 0)$. It is straightforward to show that Lemma 2 applies also to this model with x variable such that:

$$\pi(\alpha_i, x^0, 0) [1 - \pi(\alpha_i, x^0, 1)] = \Delta^{(1)}(\alpha_i, x^0) \frac{1}{\exp\{\beta\} - 1} \quad (35)$$

W.l.o.g. we consider that $T = 3$. We have that:

$$\begin{cases} \mathbb{P}(0, 1, 0 | \mathbf{x}^0) &= \int p^*(0|\alpha_i, x^0) \pi(\alpha_i, x^0, 0) [1 - \pi(\alpha_i, x^0, 1)] f_{\alpha|x}(\alpha_i|\mathbf{x}^0) d\alpha_i \\ \mathbb{P}(1, 0, 1 | \mathbf{x}^0) &= \int p^*(1|\alpha_i, x^0) \pi(\alpha_i, x^0, 0) [1 - \pi(\alpha_i, x^0, 1)] f_{\alpha|x}(\alpha_i|\mathbf{x}^0) d\alpha_i \end{cases} \quad (36)$$

Applying equation (35), we have that:

$$\begin{cases} \mathbb{P}(0, 1, 0 \mid \mathbf{x}^0) &= \frac{1}{\exp\{\beta\} - 1} \int p^*(0 \mid \alpha_i, x^0) \Delta^{(1)}(\alpha_i, x^0) f_{\alpha \mid x}(\alpha_i \mid \mathbf{x}^0) d\alpha_i \\ \mathbb{P}(1, 0, 1 \mid \mathbf{x}^0) &= \frac{1}{\exp\{\beta\} - 1} \int p^*(1 \mid \alpha_i, x^0) \Delta^{(1)}(\alpha_i, x^0) f_{\alpha \mid x}(\alpha_i \mid \mathbf{x}^0) d\alpha_i \end{cases} \quad (37)$$

Adding up these two equations, multiplying the resulting equation times $\exp\{\beta\} - 1$, and taking into account that $p^*(0 \mid \alpha_i, x^0) + p^*(1 \mid \alpha_i, x^0) = 1$, we have that:

$$AME_x^{(1)}(x^0) = [\exp\{\beta\} - 1] [\mathbb{P}(0, 1, 0 \mid \mathbf{x}^0) + \mathbb{P}(1, 0, 1 \mid \mathbf{x}^0)] \quad (38)$$

such that $AME_x^{(1)}(x^0)$ is identified. ■

We can also extend the identification of the n periods forward AME to the AR1X model.

PROPOSITION 5. *Let n be any positive integer. If $T \geq 2n + 1$ and β is known (identified), then the parameter $AME_x^{(n)}(x^0)$ is identified as:*

$$AME_x^{(n)}(x^0) = [\exp\{\beta\} - 1]^n [\mathbb{P}(A \mid \mathbf{x}^0) + \mathbb{P}(B \mid \mathbf{x}^0)] \quad (39)$$

where $\mathbb{P}(A \mid \mathbf{x}^0)$ and $\mathbb{P}(B \mid \mathbf{x}^0)$ are the probabilities of choice histories A and B , over $2n + 1$ periods, with $A = (0, 1, 0, 1, \dots, 0)$ and $B = (1, 0, 1, 0, \dots, 1)$, conditional $(x_{i1}, x_{i2}, \dots, x_{i,2n+1}) = \mathbf{x}^0 = (x^0, x^0, \dots, x^0)$. ■

Proof of Proposition 5: It follows from applying the same arguments as in Proposition 2.

The AMEs in Propositions 4 and 5 apply only to the subpopulation of individuals where the exogenous variable is constant over the sample and equal to x^0 . Since we can identify this AME for any value of x^0 , it is clear that we can obtain an integrated AME over all the values of x . However, that integrated AME is still imposing the restriction that the exogenous variable is constant over time, and therefore, it is an AME for that subpopulation of individuals. We would like to obtain an AME that does not impose this restriction. This type of AME corresponds to $AME_{x,t}^{(1)}$ that we have defined in equation (11). Proposition 6 establishes the identification of $AME_{x,t}^{(1)}$.

PROPOSITION 6. *Under the conditions of Assumption 1: (A) If $T \geq 4$, then the parameters β and γ are uniquely identified. (B) Given β and γ , if $T \geq 3$, then the parameter $AME_{x,t}^{(1)}$ defined in (11) is identified for any period $t \geq 3$ in the sample. More specifically, define $\mathbf{y}_i^{(t-2,t)} = (y_{it-2}, y_{it-1}, y_{it})$*

and $\mathbf{x}_i^{(t-2,t)} = (x_{it-2}, x_{it-1}, x_{it})$, and let $\mathcal{X}^{(t-2,t)}$ be the support of $\mathbf{x}_i^{(t-2,t)}$. Then, we have that:

$$AME_{x,t}^{(1)} = \sum_{\mathbf{x} \in \mathcal{X}^{(t-2,t)}} \mathbb{P}(\mathbf{x}_i^{(t-2,t)} = \mathbf{x}) \left[\begin{array}{l} w(0,0,1; \mathbf{x}) \mathbb{P}(\mathbf{y}_i^{(t-2,t)} = (0,0,1) | \mathbf{x}) \\ + w(0,1,0; \mathbf{x}) \mathbb{P}(\mathbf{y}_i^{(t-2,t)} = (0,1,0) | \mathbf{x}) \\ + w(1,0,1; \mathbf{x}) \mathbb{P}(\mathbf{y}_i^{(t-2,t)} = (1,0,1) | \mathbf{x}) \\ + w(1,1,0; \mathbf{x}) \mathbb{P}(\mathbf{y}_i^{(t-2,t)} = (1,1,0) | \mathbf{x}) \end{array} \right] \quad (40)$$

where the weights $w(y_1, y_2, y_3; x_1, x_2, x_3)$ are:

$$\left\{ \begin{array}{l} w(0,0,1; \mathbf{x}) = \frac{\exp(\gamma x_2) - \exp(\gamma x_3)}{\exp(\gamma x_3)} \\ w(0,1,0; \mathbf{x}) = \frac{\exp(\beta + \gamma x_3) - \exp(\gamma x_2)}{\exp(\gamma x_2)} \\ w(1,0,1; \mathbf{x}) = \frac{\exp(\beta + \gamma x_2) - \exp(\gamma x_3)}{\exp(\gamma x_3)} \\ w(1,1,0; \mathbf{x}) = \frac{\exp(\gamma x_3) - \exp(\gamma x_2)}{\exp(\gamma x_2)}. \quad \blacksquare \end{array} \right. \quad (41)$$

Proof of Proposition 6: W.l.o.g. we consider $T = 3$ and $t = 3$ such that $\mathbf{y}_i^{(t-2,t)} = \mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3})$ and $\mathbf{x}_i^{(t-2,t)} = \mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$. We first obtain the expression for the probabilities $\mathbb{P}(\mathbf{y}_i | \alpha_i, \mathbf{x}_i)$ for $\mathbf{y}_i = (0,0,1)$, $(0,1,0)$, $(1,0,1)$, and $(1,1,0)$.

$$\left\{ \begin{array}{l} \mathbb{P}(0,0,1 | \alpha_i, \mathbf{x}_i) = p^*(0 | \alpha_i, \mathbf{x}_i) \frac{1}{1 + \exp(\alpha_i + \gamma x_{i2})} \frac{\exp(\alpha_i + \gamma x_{i3})}{1 + \exp(\alpha_i + \gamma x_{i3})} \\ \mathbb{P}(0,1,0 | \alpha_i, \mathbf{x}_i) = p^*(0 | \alpha_i, \mathbf{x}_i) \frac{\exp(\alpha_i + \gamma x_{i2})}{1 + \exp(\alpha_i + \gamma x_{i2})} \frac{1}{1 + \exp(\alpha_i + \beta + \gamma x_{i3})} \\ \mathbb{P}(1,0,1 | \alpha_i, \mathbf{x}_i) = p^*(1 | \alpha_i, \mathbf{x}_i) \frac{1}{1 + \exp(\alpha_i + \beta + \gamma x_{i2})} \frac{\exp(\alpha_i + \gamma x_{i3})}{1 + \exp(\alpha_i + \gamma x_{i3})} \\ \mathbb{P}(1,1,0 | \alpha_i, \mathbf{x}_i) = p^*(1 | \alpha_i, \mathbf{x}_i) \frac{\exp(\alpha_i + \beta + \gamma x_{i2})}{1 + \exp(\alpha_i + \beta + \gamma x_{i2})} \frac{1}{1 + \exp(\alpha_i + \beta + \gamma x_{i3})} \end{array} \right.$$

Now, consider that we multiply each probability $\mathbb{P}(y_1, y_2, y_3 | \alpha_i, \mathbf{x}_i)$ by the corresponding weighting value $w(y_1, y_2, y_3; \mathbf{x}_i)$ as defined in (41). We have that:

$$\left\{ \begin{array}{l} w(0,0,1; \mathbf{x}_i) \mathbb{P}(0,0,1 | \alpha_i, \mathbf{x}_i) = p^*(0 | \alpha_i, \mathbf{x}_i) \frac{\exp(\alpha_i + \gamma x_{i2}) - \exp(\alpha_i + \gamma x_{i3})}{[1 + \exp(\alpha_i + \gamma x_{i2})][1 + \exp(\alpha_i + \gamma x_{i3})]} \\ w(0,1,0; \mathbf{x}_i) \mathbb{P}(0,1,0 | \alpha_i, \mathbf{x}_i) = p^*(0 | \alpha_i, \mathbf{x}_i) \frac{\exp(\alpha_i + \beta + \gamma x_{i3}) - \exp(\alpha_i + \gamma x_{i2})}{[1 + \exp(\alpha_i + \gamma x_{i2})][1 + \exp(\alpha_i + \beta + \gamma x_{i3})]} \\ w(1,0,1; \mathbf{x}_i) \mathbb{P}(1,0,1 | \alpha_i, \mathbf{x}_i) = p^*(1 | \alpha_i, \mathbf{x}_i) \frac{\exp(\alpha_i + \beta + \gamma x_{i2}) - \exp(\alpha_i + \gamma x_{i3})}{[1 + \exp(\alpha_i + \beta + \gamma x_{i2})][1 + \exp(\alpha_i + \gamma x_{i3})]} \\ w(1,1,0; \mathbf{x}_i) \mathbb{P}(1,1,0 | \alpha_i, \mathbf{x}_i) = p^*(1 | \alpha_i, \mathbf{x}_i) \frac{\exp(\alpha_i + \beta + \gamma x_{i3}) - \exp(\alpha_i + \beta + \gamma x_{i2})}{[1 + \exp(\alpha_i + \beta + \gamma x_{i2})][1 + \exp(\alpha_i + \beta + \gamma x_{i3})]} \end{array} \right.$$

To obtain the expression in the right hand side of equation (40) we first add these four terms. After some operations and taking into account that $p^*(0 | \alpha_i, \mathbf{x}_i) + p^*(1 | \alpha_i, \mathbf{x}_i) = 1$, we get:

$$\sum_{\mathbf{y}_i} w(\mathbf{y}_i; \mathbf{x}_i) \mathbb{P}(\mathbf{y}_i | \mathbf{x}_i) = \Lambda(\alpha_i + \beta + \gamma x_{i3}) - \Lambda(\alpha_i + \gamma x_{i3}) = \Delta^{(1)}(\alpha_i, x_{i3}) \quad (42)$$

In this equation (42), the LHS depends on (x_{i1}, x_{i2}, x_{i3}) while the RHS depends on (α_i, x_{i3}) . This is the case for any possible value of (α_i, x_{i3}) . Therefore, integrating the LHS over the distribution of (x_{i1}, x_{i2}, x_{i3}) is equivalent to integrating the RHS over the distribution of (α_i, x_{i3}) such that:

$$\sum_{\mathbf{x} \in \mathcal{X}} \mathbb{P}(\mathbf{x}_i = \mathbf{x}) \begin{bmatrix} w(0, 0, 1; \mathbf{x}) \mathbb{P}(0, 0, 1 | \mathbf{x}) \\ + w(0, 1, 0; \mathbf{x}) \mathbb{P}(0, 1, 0 | \mathbf{x}) \\ + w(1, 0, 1; \mathbf{x}) \mathbb{P}(1, 0, 1 | \mathbf{x}) \\ + w(1, 1, 0; \mathbf{x}) \mathbb{P}(1, 1, 0 | \mathbf{x}) \end{bmatrix} = AME_{x,3}^{(1)} \quad (43)$$

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

Remark 6.1. The notation in the enunciate and in the proof of Proposition 6 implicitly assumes that x_{it} is a discrete variable. However, this identification result trivially extends to the case of continuous x variable by simply replacing the sum over the probability function $\mathbb{P}(\mathbf{x}_i)$ with the integral over the density of \mathbf{x}_i .

Remark 6.2. Proposition 6 does not impose any restriction on the stochastic process of x_{it} , other than this variable is strictly exogenous with respect the transitory shock ε_{it} .

Remark 6.3. There is a relationship between the identification of $AME_{x,t}^{(1)}$ in Proposition 5 and the identification of $AME_x^{(1)}(x^0)$ in Proposition 3. By definition, these two AMEs are the same if, with probability one, x_{it} is constant for every individual in the sample. Under this condition, the (sub)population of individuals with constant x_{it} is simply the population of all the individuals. Using the expressions for the weights in equation (41), we can particularize these weights for the case where $x_{i1} = x_{i2} = x_{i3} = x^0$, to show that the expression for $AME_{x,t}^{(1)}$ in equation (40) becomes the same as the expression for $AME_x^{(1)}(x^0)$ in equation (34). That is:

$$\begin{cases} w(0, 0, 1; \mathbf{x}^0) = \frac{\exp(\gamma x^0) - \exp(\gamma x^0)}{\exp(\gamma x^0)} = 0 \\ w(0, 1, 0; \mathbf{x}^0) = \frac{\exp(\beta + \gamma x^0) - \exp(\gamma x^0)}{\exp(\gamma x^0)} = \exp(\beta) - 1 \\ w(1, 0, 1; \mathbf{x}^0) = \frac{\exp(\beta + \gamma x^0) - \exp(\gamma x^0)}{\exp(\gamma x^0)} = \exp(\beta) - 1 \\ w(1, 1, 0; \mathbf{x}^0) = \frac{\exp(\gamma x_3) - \exp(\gamma x_2)}{\exp(\gamma x_2)} = 0, \end{cases} \quad (44)$$

such that $AME_{x,3}^{(1)} = [\exp(\beta) - 1] [\mathbb{P}(0, 1, 0|\mathbf{x}^0) + \mathbb{P}(1, 0, 1|\mathbf{x}^0)]$, which is the expression for the identification of $AME_x^{(1)}(x^0)$ in Proposition 4.

Remark 6.4. The identification result in Proposition 5 may look like kind of "magical". It triggers some relevant questions. How were the correct weights obtained? Is there a general method to

obtain these weights? Are these weights unique, or there are other (linearly independent) weights that also provide identification of this AME? These are relevant questions not only to have a better understanding of this identification approach, but also for the efficient estimation of the AMEs. In the next section, we present a constructive approach to obtain, for any value of T , the weights that provide identification of the different AMEs that we study in this paper.

3.6 A constructive approach to derive (over)identifying restrictions on AMEs

3.6.1 Preliminaries

In Proposition 1 we show the identification of $AME^{(1)}$ using 3 periods of observed choices in the AR1 model. If T is larger, we can always take a sub-history with 3 periods. However, for the estimation of AMEs this approach does not deliver an efficient estimator when $T \geq 4$ because it does not exploit all the information that the data contains about the AME. In this section, we describe a procedure to derive additional restrictions that the model implies on $AME^{(1)}$ when $T \geq 4$. We are particularly interested in deriving these restrictions for the efficient estimation of the AMEs.

The probability distribution of the choice history \mathbf{y} contains all the information that the data has about $AME^{(1)}$. The model give us the functional relationship between a probability $\mathbb{P}(\mathbf{y})$ and the parameter β , the distribution of the incidental parameters, f_α , and the probability of the initial conditions, $p^*(y_1|\alpha_i)$. To represent this relationship, it is convenient to define the function,

$$\lambda(\mathbf{y}, \beta, \alpha_i) \equiv \mathbb{P}([y_{i2}, \dots, y_{iT}] = [y_2, \dots, y_T] \mid y_{i1} = y_1, \beta, \alpha_i) \quad (45)$$

For instance, with $T = 4$ and $\mathbf{y} = (0, 0, 1, 1)$, we have that $\lambda([0, 0, 1, 1], \beta, \alpha_i) = [1 - \Lambda(\alpha_i)] \Lambda(\alpha_i) \Lambda(\alpha_i + \beta)$. Using this function, we can write as follows the system of equations that relate the probabilities $\mathbb{P}(\mathbf{y})$ and the parameters of the model:

$$\mathbb{P}(\mathbf{y}) = \int \lambda(\mathbf{y}, \beta, \alpha_i) p^*(y_1|\alpha_i) f_\alpha(\alpha_i) d\alpha_i \quad (46)$$

Note that the right hand side in this equation is a linear operator in $p^*(0|\alpha_i) f_\alpha(\alpha_i)$ or $p^*(1|\alpha_i) f_\alpha(\alpha_i)$. For any value of \mathbf{y} , we can represent these equations or restrictions as follows:

$$\mathbb{P}(\mathbf{y}) = \mathbf{f}_\alpha^*(y_1)' \lambda_\alpha(\mathbf{y}, \beta) \quad \text{for any value of } \mathbf{y}, \quad (47)$$

where $\mathbf{f}_\alpha^*(0)$ and $\mathbf{f}_\alpha^*(1)$ are the infinite dimensional vector $\{p^*(y_1|\alpha_i) f_\alpha(\alpha_i) : \alpha_i \in \mathbb{R}\}$ for $y_1 = 0$ and $y_1 = 1$, respectively; and $\lambda_\alpha(\mathbf{y}, \beta)$ is the infinite dimensional vector with elements $\{\lambda(\mathbf{y}, \beta, \alpha_i) : \alpha_i \in \mathbb{R}\}$. Function $\lambda_\alpha(\mathbf{y}, \beta)$ is known to the researcher.

We also have that, by definition, $AME^{(1)} = \int [\Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)] f_\alpha(\alpha_i) d\alpha_i$, or equivalently:

$$AME^{(1)} = \int [\Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)] [p^*(0|\alpha_i) f_\alpha(\alpha_i) + p^*(1|\alpha_i) f_\alpha(\alpha_i)] d\alpha_i. \quad (48)$$

The right hand side of this equation is also a linear operator in $\mathbf{f}_\alpha^*(0)$ and $\mathbf{f}_\alpha^*(1)$. We can represent in a compact vector form as:

$$AME^{(1)} = [\mathbf{f}_\alpha^*(0) + \mathbf{f}_\alpha^*(1)]' \mathbf{\Delta}_\alpha(\beta) \quad (49)$$

where $\mathbf{\Delta}_\alpha(\beta)$ is the infinite dimensional vector with elements $\{\Lambda(\alpha_i + \beta) - \Lambda(\alpha_i) : \alpha_i \in \mathbb{R}\}$.

Given the system of equations in (47) and (49), we are interested in obtaining all the restrictions that this system implies on $AME^{(1)}$ and that have the following form:

$$h(AME^{(1)}, \beta, \mathbf{P}_\mathbf{y}) = 0 \quad (50)$$

where $\mathbf{P}_\mathbf{y}$ is the vector with the probabilities $\mathbb{P}(\mathbf{y})$ for any possible choice history \mathbf{y} , and function $h(AME^{(1)}, \beta, \mathbf{P}_\mathbf{y})$ – potentially vector valued – is known to the researcher, and in particular it does not depend on the incidental parameters $\{\mathbf{f}_\alpha^*(0), \mathbf{f}_\alpha^*(1)\}$.

We propose a procedure that provides simple analytical expressions of the restrictions $h(AME^{(1)}, \beta, \mathbf{P}_\mathbf{y}) = 0$. Our procedure exploits the linearity in incidental parameters of the functions $\mathbf{f}_\alpha^*(y_1)'$ $\lambda_\alpha(\mathbf{y}, \beta)$ and $[\mathbf{f}_\alpha^*(0) + \mathbf{f}_\alpha^*(1)]' \mathbf{\Delta}_\alpha(\beta)$. Combining equations (47) and (49), we have that, for any history \mathbf{y} :

$$AME^{(1)} - \mathbb{P}(\mathbf{y}) = [\mathbf{f}_\alpha^*(0) + \mathbf{f}_\alpha^*(1)]' \mathbf{\Delta}_\alpha(\beta) - \mathbf{f}_\alpha^*(y_1) \lambda_\alpha(\mathbf{y}, \beta) \quad (51)$$

The following Proposition 7 establishes a necessary and sufficient condition for the derivation of any restriction of the form $h(AME^{(1)}, \beta, \mathbf{P}_\mathbf{y}) = 0$. It also provides a characterization of these restrictions.

PROPOSITION 7. (A) A necessary and sufficient condition to have a restriction with the form $h(AME^{(1)}, \beta, \mathbf{P}_\mathbf{y}) = 0$ – where function $h(\cdot)$ is known to the researcher, does not depend on the incidental parameters, and holds for any possible value of \mathbf{f}_α^ – is that there is a vector of known weights $\mathbf{w}(\beta) \equiv \{w(\mathbf{y}, \beta) : \text{for any } \mathbf{y}\}$ such that the weights $w(\mathbf{y}, \beta)$ do not depend on the incidental parameters and satisfy the following restrictions. Being $\mathbf{y} = (y_1, \mathbf{y}_{-1})$, for $y_1 = 0, 1$:*

$$\sum_{\mathbf{y}_{-1}} w(y_1, \mathbf{y}_{-1}, \beta) \lambda(y_1, \mathbf{y}_{-1}, \beta, \alpha_i) = \Lambda(\alpha_i + \beta) - \Lambda(\alpha_i), \quad (52)$$

for every $\alpha_i \in \mathbb{R}$. (B) All the restrictions $h(AME^{(1)}, \beta, \mathbf{P}_y) = 0$ have the following form:

$$AME^{(1)} = \sum_{\mathbf{y}_{-1}} w(0, \mathbf{y}_{-1}, \beta) \mathbb{P}(0, \mathbf{y}_{-1}) - \sum_{\mathbf{y}_{-1}} w(1, \mathbf{y}_{-1}, \beta) \mathbb{P}(1, \mathbf{y}_{-1}). \quad \blacksquare \quad (53)$$

Proof of Proposition 7. (A) Sufficient condition. Consider equation (52) for the pair of histories $(0, \mathbf{y}_{-1})$ and $(1, \mathbf{y}_{-1})$, where the first element in y_1 . Multiplying times $p^*(0|\alpha_i) f_\alpha(\alpha_i)$ both sides of the equation for $(0, \mathbf{y}_{-1})$ and integrating over α_i , we get:

$$\sum_{\mathbf{y}_{-1}} w(0, \mathbf{y}_{-1}, \beta) \mathbb{P}(0, \mathbf{y}_{-1}) = \int [\Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)] p^*(0|\alpha_i) f_\alpha(\alpha_i) d\alpha_i \quad (54)$$

Similarly, multiplying times $p^*(1|\alpha_i) f_\alpha(\alpha_i)$ both sides of the equation for $(1, \mathbf{y}_{-1})$ and integrating over α_i , we get:

$$\sum_{\mathbf{y}_{-1}} w(1, \mathbf{y}_{-1}, \beta) \mathbb{P}(1, \mathbf{y}_{-1}) = \int [\Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)] p^*(1|\alpha_i) f_\alpha(\alpha_i) d\alpha_i \quad (55)$$

Adding equations (54) and (55) and applying the definition of $AME^{(1)}$, we obtain:

$$\sum_{\mathbf{y}_{-1}} w(0, \mathbf{y}_{-1}, \beta) \mathbb{P}(0, \mathbf{y}_{-1}) + \sum_{\mathbf{y}_{-1}} w(1, \mathbf{y}_{-1}, \beta) \mathbb{P}(1, \mathbf{y}_{-1}) = AME^{(1)} \quad (56)$$

that has the form $h(AME^{(1)}, \beta, \mathbf{P}_y) = 0$.

(A) *Necessary condition.* The proof of the necessary condition has two parts: (i) proving that to obtain $h(AME^{(1)}, \beta, \mathbf{P}_y) = 0$ it is necessary to use a weighted sum over multiple values of \mathbf{y} of the equations $AME^{(1)} - \mathbb{P}(\mathbf{y}) = [\mathbf{f}_\alpha^*(0) + \mathbf{f}_\alpha^*(1)]' \Delta_\alpha(\beta) - \mathbf{f}_\alpha^*(y_1) \lambda_\alpha(\mathbf{y}, \beta)$; and (ii) that the weights should satisfy condition (52).

(i) To be proved.

(ii) The proof is by contradiction. Suppose that: (a) the restriction $AME^{(1)} - \sum_{\mathbf{y}_{-1}} w(0, \mathbf{y}_{-1}, \beta) \mathbb{P}(0, \mathbf{y}_{-1}) + \sum_{\mathbf{y}_{-1}} w(1, \mathbf{y}_{-1}, \beta) \mathbb{P}(1, \mathbf{y}_{-1})$ holds for any possible value of the incidental parameters f_α and p^* in the DGP; and (b) there is (at least) a value of α_i and a value of y_1 – say α_1 and w.l.o.g. $y_1 = 0$, respectively – such that $\sum_{\mathbf{y}_{-1}} w(0, \mathbf{y}_{-1}, \beta) \lambda(0, \mathbf{y}_{-1}, \beta, \alpha_1) \neq \Lambda(\alpha_1 + \beta) - \Lambda(\alpha_1)$. Now, we show that given (b) there always exist a distribution function f_α (in fact, a continuum of distribution functions) such that the condition (a) does not hold. W.l.o.g. consider distributions of α_i with only two points support, α_1 and α_2 , such that $f_\alpha(\alpha_1) > 0$, $f_\alpha(\alpha_2) > 0$, and $f_\alpha(\alpha_1) > 0 + f_\alpha(\alpha_2) = 1$. Define $d_1(y_1) \equiv \sum_{\mathbf{y}_{-1}} w(y_1, \mathbf{y}_{-1}, \beta) \lambda(y_1, \mathbf{y}_{-1}, \beta, \alpha_1) - \Lambda(\alpha_1 + \beta) + \Lambda(\alpha_1)$; and similarly, define $d_2(y_1) \equiv \sum_{\mathbf{y}_{-1}} w(y_1, \mathbf{y}_{-1}, \beta) \lambda(y_1, \mathbf{y}_{-1}, \beta, \alpha_2) - \Lambda(\alpha_2 + \beta) + \Lambda(\alpha_2)$. By condition (b), we have that

$d_1(0) \neq 0$. The values of $d_1(1)$, $d_2(0)$, and $d_2(1)$ can be either zero or different than zero. Applying the same operations as in the proof of the sufficient condition above, we obtain that:

$$\begin{aligned} & \sum_{\mathbf{y}_{-1}} w(0, \mathbf{y}_{-1}, \beta) \mathbb{P}(0, \mathbf{y}_{-1}) + \sum_{\mathbf{y}_{-1}} w(1, \mathbf{y}_{-1}, \beta) \mathbb{P}(1, \mathbf{y}_{-1}) - AME^{(1)} \\ &= f_\alpha(\alpha_1) [p^*(0|\alpha_1) d_1(0) + p^*(1|\alpha_1) d_1(1)] + f_\alpha(\alpha_2) [p^*(0|\alpha_2) d_2(0) + p^*(1|\alpha_1) d_2(1)] \end{aligned} \quad (57)$$

By condition (a), this expression should be equal to zero for any distribution f_α . By definition, the values of $d_1(y_1)$ and $d_2(y_1)$ do not depend on the distribution function f_α . Therefore, there always exist (a continuum of) values of $f_\alpha(\alpha_1)$ such that the right hand side of equation (57) is different to zero such that condition (a) does not hold.

(B) is a direct implication of (A). ■

3.6.2 Obtaining the weights $w(\mathbf{y})$

We now derive the general expression for the weights $w(\mathbf{y}, \beta)$ that satisfy condition (52) in Proposition 7.

For any \mathbf{y} define the statistics n_{00} , n_{01} , n_{10} , and n_{11} as the number of times that (y_{t-1}, y_t) is equal to $(0, 0)$, $(0, 1)$, $(1, 0)$, and $(1, 1)$, respectively. Given these statistics and the definition of the transition probabilities $\pi(\alpha_i, 0)$ and $\pi(\alpha_i, 1)$, we have that:

$$\lambda(\mathbf{y}, \beta, \alpha_i) = [1 - \pi(\alpha_i, 0)]^{n_{00}} \pi(\alpha_i, 0)^{n_{01}} [1 - \pi(\alpha_i, 1)]^{n_{10}} \pi(\alpha_i, 1)^{n_{11}} \quad (58)$$

By Lemma 1, we know that $\pi(\alpha_i, 0) [1 - \pi(\alpha_i, 1)] = \Delta^{(1)}(\alpha_i) / (\exp\{\beta\} - 1)$. We also know that $\pi(\alpha_i, 1) [1 - \pi(\alpha_i, 0)] = \Delta^{(1)}(\alpha_i) \exp\{\beta\} / (\exp\{\beta\} - 1)$. Taking into account the expression, we can write $\lambda(\mathbf{y}, \beta, \alpha_i)$ as follows:

$$\begin{aligned} \lambda(\mathbf{y}, \beta, \alpha_i) &= \left[\frac{1}{\exp\{\beta\} - 1} \Delta^{(1)}(\alpha_i) \right]^{n_{10}} \pi(\alpha_i, 0)^{n_{01} - n_{10}} \left[\frac{\exp\{\beta\}}{\exp\{\beta\} - 1} \Delta^{(1)}(\alpha_i) \right]^{n_{00}} \pi(\alpha_i, 1)^{n_{11} - n_{00}} \\ &= \frac{\exp\{n_{00} \beta\}}{[\exp\{\beta\} - 1]^{n_{10} + n_{00}}} [\Delta^{(1)}(\alpha_i)]^{n_{10} + n_{00}} \pi(\alpha_i, 0)^{n_{01} - n_{10}} \pi(\alpha_i, 1)^{n_{11} - n_{00}} \end{aligned} \quad (59)$$

Given this expression, a sufficient condition to obtain $\Delta^{(1)}(\alpha_i)$ and a weighted sum of λ 's is the following. Given a history y , consider the following condition (C*):

$$\{n_{01}(\mathbf{y}) = n_{10}(\mathbf{y}) = 1 \ \& \ n_{00}(\mathbf{y}) = n_{11}(\mathbf{y}) = 0\} \text{ OR } \{n_{01}(\mathbf{y}) = n_{10}(\mathbf{y}) = 0 \ \& \ n_{00}(\mathbf{y}) = n_{11}(\mathbf{y}) = 1\}$$

Under this condition, we have that:

$$\lambda(\mathbf{y}, \beta, \alpha_i) = \frac{\exp\{n_{00} \beta\}}{\exp\{\beta\} - 1} \Delta^{(1)}(\alpha_i) \quad (60)$$

Suppose that we construct weights as follows. If a history (y_1, y_{-1}) satisfies condition (C^*) , then the weight $w(y_1, y_{-1}, \beta)$ is non-zero and equal to:

$$w(y_1, \mathbf{y}_{-1}, \beta) = \frac{\exp\{\beta\} - 1}{\exp\{n_{00}(\mathbf{y})\beta\}} \frac{1}{|C^*(y_1)|} \quad (61)$$

where $|C^*(y_1)|$ represents the number of histories with initial condition y_1 satisfying condition (C^*) . Otherwise, a history that does not satisfy condition (C^*) has weight $w(y, \beta)$ equal to zero.

EXAMPLE. When $T = 3$, condition (C^*) is satisfied only by two choice histories $(0, 1, 0)$ and $(1, 0, 1)$. Both histories have $n_{01} = n_{10} = 1$ and $n_{00} = n_{11} = 0$. Therefore, both histories have weights equal to $\exp\{\beta\} - 1$, such that:

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

This is exactly the expression that we have obtained in our identification result in Proposition 1. Now, we can see that when $T = 3$ the model does not provide additional restrictions on $AME^{(1)}$.

■

However, condition (C^*) is far of being necessary to characterize all the restrictions that satisfy condition (52) in Proposition 7. For instance, when $T = 4$ it is straightforward to verify that there is not any history that satisfies condition (C^*) . However, we show below that there are histories that satisfy condition (52). We illustrate this in the following subsections. A key feature is that histories with the the same value of the vector of sufficient statistics $\mathbf{s} \equiv (y_1, y_T, \sum_{t=2}^T y_t)$ have the same weight.

3.6.3 Case with T=4

Consider the case when $T = 4$. When $y_1 = 0$, the vector of statistics $\mathbf{s} \equiv (y_1, y_T, \sum_{t=2}^T y_t)$ can take 6 possible values: $(0, 0, 0)$, $(0, 0, 1)$, $(0, 1, 1)$, $(0, 0, 2)$, $(0, 1, 2)$, and $(0, 1, 3)$. Each of the values of \mathbf{s} equal to $(0, 0, 0)$, $(0, 1, 1)$, $(0, 0, 2)$, and $(0, 1, 3)$ is associated with only one value of the choice history. Therefore, according to Lemma 3(B) these values of \mathbf{s} have a weight equal to zero. We are left only with two values of \mathbf{s} : $(0, 0, 1)$ and $(0, 1, 2)$. We are looking for values for the weights $w(0, 0, 1)$ and $w(0, 1, 2)$ such that:

$$w(0, 0, 1) \mathbb{P}(\mathbf{s}_i = (0, 0, 1) \mid \alpha_i) + w(0, 1, 2) \mathbb{P}(\mathbf{s}_i = (0, 1, 2) \mid \alpha_i) = \Lambda(\alpha_i + \beta) - \Lambda(\alpha_i). \quad (62)$$

We have that,

$$\begin{aligned}\mathbb{P}(\mathbf{s}_i = (0, 0, 1) \mid \alpha_i) &= \mathbb{P}(\mathbf{y}_i = (0, 0, 1, 0) \mid \alpha_i) + \mathbb{P}(\mathbf{y}_i = (0, 1, 0, 0) \mid \alpha_i) \\ &= 2 (1 - \Lambda(\alpha_i + \beta)) \Lambda(\alpha_i) (1 - \Lambda(\alpha_i)) p^*(0 \mid \alpha_i)\end{aligned}\quad (63)$$

And,

$$\begin{aligned}\mathbb{P}(\mathbf{s}_i = (0, 1, 2) \mid \alpha_i) &= \mathbb{P}(\mathbf{y}_i = (0, 0, 1, 1) \mid \alpha_i) + \mathbb{P}(\mathbf{y}_i = (0, 1, 0, 1) \mid \alpha_i) \\ &= (\exp(\beta) + 1) \Lambda(\alpha_i)^2 (1 - \Lambda(\alpha_i + \beta)) p^*(0 \mid \alpha_i)\end{aligned}\quad (64)$$

where the last equality uses that $\Lambda(\alpha_i + \beta) (1 - \Lambda(\alpha_i)) = \exp(\beta) \Lambda(\alpha_i) * (1 - \Lambda(\alpha_i + \beta))$. Replacing (63) and (64) in the left-hand-side of equation (62), we obtain that:

$$\begin{aligned}w(0, 0, 1) 2 (1 - \Lambda(\alpha_i + \beta)) \Lambda(\alpha_i) (1 - \Lambda(\alpha_i)) + w(0, 1, 2) (\exp(\beta) + 1) \Lambda(\alpha_i)^2 (1 - \Lambda(\alpha_i + \beta)) \\ = (1 - \Lambda(\alpha_i + \beta)) \Lambda(\alpha_i) [w(0, 0, 1) 2 (1 - \Lambda(\alpha_i)) + w(0, 1, 2) (\exp(\beta) + 1) \Lambda(\alpha_i)] p^*(0 \mid \alpha_i)\end{aligned}\quad (65)$$

From Lemma 2, $(1 - \Lambda(\alpha_i + \beta)) \Lambda(\alpha_i) (\exp(\beta) - 1) = \Lambda(\alpha_i + \beta) - \Lambda(\alpha_i)$. Therefore, equation (62) holds up to the initial condition $p^*(0 \mid \alpha_i)$ if:

$$\frac{1}{\exp(\beta) - 1} [w(0, 0, 1) 2 (1 - \Lambda(\alpha_i)) + w(0, 1, 2) (\exp(\beta) + 1) \Lambda(\alpha_i)] = 1 \quad (66)$$

for any value of α_i . This requires that:

$$w(0, 0, 1) = \frac{\exp(\beta) - 1}{2} \quad \text{and} \quad w(0, 1, 2) = \frac{\exp(\beta) - 1}{\exp(\beta) + 1} \quad (67)$$

Now, we proceed similarly with the values of $\mathbf{s} \equiv (y_1, y_T, \sum_{t=2}^T y_t)$ with $y_1 = 1$. There are also 6 possible values but only two associated with more than one choice history: $\mathbf{s}_i = (1, 1, 2)$ and $\mathbf{s}_i = (1, 0, 1)$. We can proceed similarly as for the values with $y_1 = 0$ to show that:

$$w(1, 1, 2) = \frac{\exp(\beta) - 1}{2} \quad \text{and} \quad w(1, 0, 1) = \frac{\exp(\beta) - 1}{\exp(\beta) + 1} \quad (68)$$

Combining these two results, we have that, when $T = 4$:

$$\begin{aligned}AME^{(1)} &= \frac{\exp(\beta) - 1}{2} [\mathbb{P}(\mathbf{s}_i = (0, 0, 1)) + \mathbb{P}(\mathbf{s}_i = (1, 1, 2))] \\ &+ \frac{\exp(\beta) - 1}{\exp(\beta) + 1} [\mathbb{P}(\mathbf{s}_i = (0, 1, 2)) + \mathbb{P}(\mathbf{s}_i = (1, 0, 1))]\end{aligned}\quad (69)$$

3.6.4 T larger than 4 in AR1

Proceeding similarly with some higher T we have obtained the following weights with $y_{i1} = 0$.

$\mathbf{s}_i = (y_{i1},$ $y_{iT}, \sum_{t=2}^T y_{it})$	$w(\mathbf{s})$			
	$T = 4$	$T = 5$	$T = 6$	$T = 7$
(0, 0, 0)	0	0	0	0
(0, 0, 1)	$\frac{1}{2} (\exp\{\beta\} - 1)$	$\frac{1}{3} (\exp(\beta) - 1)$	$\frac{1}{4} (\exp(\beta) - 1)$	$\frac{1}{5} (\exp(\beta) - 1)$
(0, 1, 1)	0	0	0	0
(0, 0, 2)	0	$\frac{\exp(\beta) - 1}{1 + 2 \exp(\beta)}$	$\frac{2(\exp(\beta) - 1)}{3 + 3 \exp(\beta)}$	$\frac{3(\exp(\beta) - 1)}{6 + 4 \exp(\beta)}$
(0, 1, 2)	$\frac{\exp\{\beta\} - 1}{1 + \exp\{\beta\}}$	$\frac{\exp(\beta) - 1}{2 + \exp(\beta)}$	$\frac{\exp(\beta) - 1}{3 + \exp(\beta)}$	$\frac{\exp(\beta) - 1}{4 + \exp(\beta)}$
(0, 0, 3)	Not possible	0	$\frac{\exp(\beta) - 1}{2 + 2 \exp(\beta)}$	$\frac{(\exp(\beta) - 1)(1 + 2 \exp(\beta))}{1 + 6 \exp(\beta) + 3 \exp(\beta)^2}$
(0, 1, 3)	0	$\frac{\exp(\beta) - 1}{2 + \exp(\beta)}$	$\frac{(\exp(\beta) - 1)(1 + \exp(\beta))}{1 + 4 \exp(\beta) + \exp(\beta)^2}$	$\frac{(\exp(\beta) - 1)(2 + \exp(\beta))}{3 + 6 \exp(\beta) + \exp(\beta)^2}$
(0, 0, 4)	Not possible	Not possible	0	$\frac{\exp(\beta) - 1}{3 + 2 \exp(\beta)}$
(0, 1, 4)	Not possible	0	$\frac{\exp(\beta) - 1}{3 + \exp(\beta)}$	$\frac{(\exp(\beta) - 1)(2 + \exp(\beta))}{3 + 6 \exp(\beta) + \exp(\beta)^2}$
(0, 0, 5)	Not possible	Not possible	Not possible	0
(0, 1, 5)	Not possible	Not possible	0	$\frac{\exp(\beta) - 1}{4 + \exp(\beta)}$
(0, 1, 6)	Not possible	Not possible	Not possible	0

And with $y_{i1} = 1$, which is symmetric to the case with $y_{i1} = 0$.

$\mathbf{s}_i = (y_{i1}, \dots, y_{iT}, \sum_{t=2}^T y_{it})$	$w(\mathbf{s})$			
	$T = 4$	$T = 5$	$T = 6$	$T = 7$
(1, 0, 0)	0	0	0	0
(1, 0, 1)	$\frac{\exp\{\beta\} - 1}{1 + \exp\{\beta\}}$	$\frac{\exp\{\beta\} - 1}{2 + \exp\{\beta\}}$	$\frac{\exp\{\beta\} - 1}{3 + \exp\{\beta\}}$	$\frac{\exp\{\beta\} - 1}{4 + \exp\{\beta\}}$
(1, 1, 1)	0	0	0	0
(1, 0, 2)	0	$\frac{\exp\{\beta\} - 1}{2 + \exp\{\beta\}}$	$\frac{(\exp\{\beta\} - 1)(1 + \exp(\beta))}{1 + 4\exp(\beta) + \exp(\beta)^2}$	$\frac{(\exp\{\beta\} - 1)(2 + \exp(\beta))}{3 + 6\exp(\beta) + \exp(\beta)^2}$
(1, 1, 2)	$\frac{\exp\{\beta\} - 1}{2}$	$\frac{\exp\{\beta\} - 1}{1 + 2\exp\{\beta\}}$	$\frac{\exp\{\beta\} - 1}{2 + 2\exp\{\beta\}}$	$\frac{\exp\{\beta\} - 1}{3 + 2\exp\{\beta\}}$
(1, 0, 3)	Not possible	0	$\frac{\exp\{\beta\} - 1}{3 + \exp\{\beta\}}$	$\frac{(\exp\{\beta\} - 1)(2 + \exp(\beta))}{3 + 6\exp(\beta) + \exp(\beta)^2}$
(1, 1, 3)	0	$\frac{\exp\{\beta\} - 1}{3}$	$\frac{2(\exp\{\beta\} - 1)}{3 + 3\exp\{\beta\}}$	$\frac{(\exp\{\beta\} - 1)(1 + 2\exp(\beta))}{1 + 6\exp(\beta) + 3\exp(\beta)^2}$
(1, 0, 4)	Not possible	Not possible	0	$\frac{\exp\{\beta\} - 1}{4 + \exp\{\beta\}}$
(1, 1, 4)	Not possible	0	$\frac{\exp\{\beta\} - 1}{4}$	$\frac{3(\exp\{\beta\} - 1)}{6 + 4\exp\{\beta\}}$
(1, 0, 5)	Not possible	Not possible	Not possible	0
(1, 1, 5)	Not possible	Not possible	0	$\frac{\exp\{\beta\} - 1}{5}$
(1, 1, 6)	Not possible	Not possible	Not possible	0

3.6.5 Obtaining the weights for the AR1X

Here we answer the question raised in Remark 6.4 about how to obtain the weights in Proposition 6. Consider $T = 3$ and $t = 3$ such that $\mathbf{y}_i^{(t-2,t)} = \mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3})$ and $\mathbf{x}_i^{(t-2,t)} = \mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})$. For shortness, $w(0, 0, 0; \mathbf{x}_i)$ is written as w_{1i} , $w(0, 0, 1; \mathbf{x}_i)$ as w_{2i} , $w(0, 1, 0; \mathbf{x}_i)$ as w_{3i} , and so on until $w(1, 1, 1; \mathbf{x}_i) \equiv w_{8i}$. We start by finding w_{1i} , w_{2i} , w_{3i} , and w_{4i} such that

$$\begin{aligned}
& w_{1i} \mathbb{P}(\mathbf{y}_i = (0, 0, 0) | \mathbf{x}_i, \alpha_i, y_{i1} = 0) + w_{2i} \mathbb{P}(\mathbf{y}_i = (0, 0, 1) | \mathbf{x}_i, \alpha_i, y_{i1} = 0) \\
& + w_{3i} \mathbb{P}(\mathbf{y}_i = (0, 1, 0) | \mathbf{x}_i, \alpha_i, y_{i1} = 0) + w_{4i} \mathbb{P}(\mathbf{y}_i = (0, 1, 1) | \mathbf{x}_i, \alpha_i, y_{i1} = 0) \\
& = \Lambda(\alpha_i + \beta + \gamma x_{i3}) - \Lambda(\alpha_i + \gamma x_{i3}) \equiv \Delta^{(1)}(\alpha_i, x_{i3})
\end{aligned}$$

Replacing the probabilities by their expression:

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

Operating to undo the fractions in both sides and simplifying:

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

To have w_{1i} , w_{2i} , w_{3i} , and w_{4i} that do not depend on α_i , all the terms in the polynomial of $\exp(\alpha_i)$ in both sides must be the same. This implies the following solution for the weights:

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

We can proceed in the same way to obtain the values of the weights w_{5i} , w_{6i} , w_{7i} , and w_{8i} that satisfy the following equation for any value of α_i :

$$\begin{aligned}
& w_{5i} \mathbb{P}(\mathbf{y}_i = (1, 0, 0) | \mathbf{x}_i, \alpha_i, y_{i1} = 1) + w_{6i} \mathbb{P}(\mathbf{y}_i = (1, 0, 1) | \mathbf{x}_i, \alpha_i, y_{i1} = 1) \\
& + w_{7i} \mathbb{P}(\mathbf{y}_i = (1, 1, 0) | \mathbf{x}_i, \alpha_i, y_{i1} = 1) + w_{8i} \mathbb{P}(\mathbf{y}_i = (0, 1, 1) | \mathbf{x}_i, \alpha_i, y_{i1} = 1) \\
& = \Lambda(\alpha_i + \beta + \gamma x_{i3}) - \Lambda(\alpha_i + \gamma x_{i3})
\end{aligned}$$

We obtain:

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

which are the weights given in Proposition 6.

4 Conclusion

Average marginal effects (AMEs) are convenient way to represent causal effects at the population level. They are commonly used in econometric applications. These causal parameters depend on

the structural parameters of the model but also on the distribution of the unobserved heterogeneity. In fixed effects nonlinear panel data models – using 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 this paper, we prove the identification of AME associated with an exogenous change in the lagged dependent variable. This parameter has a clear economic interpretation as the effect of a counterfactual policy experiment on the average (or aggregate) value of the dependent variable. We also prove the identification of n-periods forward version of this AME, such that the researcher can identify the impulse response function of the counterfactual experiment.

Our proofs of the identification results are constructive and provide very simple closed-form expressions for the AMEs in terms of frequencies of choice histories that can be obtained from the data.

References

- [1] Aguirregabiria, V., J. Gu, Y. Luo (2019): "Sufficient Statistics for Unobserved Heterogeneity in Dynamic Structural Logit Models," *Journal of Econometrics*. Forthcoming.
- [2] Andersen, E (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society, Series B*, 32, 283-301.
- [3] Arellano, M., and S. Bonhomme (2017): "Nonlinear Panel Data Methods for Dynamic Heterogeneous Agent Models, *Annual Review of Economics*, forthcoming.
- [4] Arellano, M., and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in J. Heckman and E. Leamer (eds.) *Handbook of Econometrics*, Volume 5, Chapter 53, North-Holland, 3229-3296.
- [5] Bonhomme, S. (2012): "Functional differencing," *Econometrica*, 80(4), 1337-1385.
- [6] Browning, M., and J. M. Carro (2010): "Heterogeneity in dynamic discrete choice models," *Econometrics Journal*, 13(1), 1-39.
- [7] Browning, M., and J. M. Carro (2014): "Dynamic binary outcome models with maximal heterogeneity," *Journal of Econometrics*, 178(2), 805-823.
- [8] Carro, J. M. (2007): "Estimating dynamic panel data discrete choice models with fixed effects", *Journal of Econometrics*, 140, 503-528.
- [9] Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47(1), 225-238.
- [10] Chamberlain, G. (1984): "Panel Data", in Griliches, Z. and M.D. Intriligator (eds.) *Handbook of Econometrics*, vol. 2, Elsevier Science, Amsterdam.
- [11] Chamberlain, G. (1985): "Heterogeneity, Omitted Variable Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, edited by J. J. Heckman and B. Singer. Cambridge: Cambridge University Press.
- [12] Chamberlain, G. (2010): "Binary response models for panel data: Identification and information," *Econometrica*, 78(1), 159-168.
- [13] Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81(2), 535-580.
- [14] Cox, D. (1958): "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2), 215-242.

- [15] Heckman, J. (1981): "The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process," in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press.
- [16] Honoré, B. , and E. Kyriazidou (2000): "Panel data discrete choice models with lagged dependent variables," *Econometrica*, 68(4), 839-874.
- [17] Honoré, B. , and E. Kyriazidou (2019): "Identification in Binary Response Panel Data Models: Is Point-Identification More Common Than We Thought?," *Annals of Economics and Statistics*, 134, 207-226.
- [18] Honoré, B., and E. Tamer (2006): "Bounds on parameters in panel dynamic discrete choice models," *Econometrica*, 74(3), 611-629.
- [19] Lancaster, T. (2000): "The incidental parameter problem since 1948," *Journal of Econometrics*, 95(2), 391-413.
- [20] Magnac, T. (2004): "Panel Binary Variables and Sufficiency: Generalizing Conditional Logit," *Econometrica*, 72(6), 1859-1876.
- [21] Neyman, J. and E. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16(1), 1-32.
- [22] Rasch, G. (1961): "On General Laws and the Meaning of Measurement in Psychology," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4., 321-333. Berkeley: University of California Press.