Sufficient Statistics for Unobserved Heterogeneity in Structural Dynamic Logit Models

Victor Aguirregabiria*
University of Toronto and CEPR

Jiaying Gu*
University of Toronto

Yao Luo*
University of Toronto

December 28th, 2017 [Preliminary and incomplete]

Abstract

We study the identification and estimation of panel data structural dynamic logit models with a nonparametric specification of the joint distribution of time-invariant unobserved heterogeneity and observable state variables, i.e., a fixed-effects structural dynamic logit model. We consider multinomial models with two endogenous state variables: the lagged decision variable, and the duration in the last choice. This class of models includes as particular cases important economic applications such as models of market entry-exit, occupational choice, machine replacement, inventory and investment decisions, or demand of differentiated storable products. The main challenge is to find a sufficient statistic that (i) controls for the contribution of the fixed-effect not only to current utility but also to the continuation values in the forward-looking decision, and (ii) still leaves information on the parameters of interest. We characterize the minimum sufficient statistics for the structural parameters. Based on our identification results, we propose a Conditional Maximum Likelihood estimator. We apply this estimator to the bus engine replacement data in Rust (1987) and to the consumer scanner data on demand of a differentiated storable product in Erdem, Imai, and Keane (2003).

Keywords: Panel data discrete choice models; Dynamic structural models; Fixed effects; Unobserved heterogeneity; Structural state dependence; Identification; Sufficient statistic.

JEL: C23; C25; C41; C51; C61.

Victor Aguirregabiria. Address: 150 St. George Street. Toronto, ON, M5S 3G7, Canada.
E-mail: `victor.aguirregabiria@utoronto.ca`

Jiaying Gu. Address: 150 St. George Street. Toronto, ON, M5S 3G7, Canada.
E-mail: `jiaying.gu@utoronto.ca`

Yao Luo. Address: 150 St. George Street. Toronto, ON, M5S 3G7, Canada.
E-mail: `yao.luo@utoronto.ca`

*We have benefited from comments and conversations with Bo Honoré, Bob Miller, Pedro Mira, Whitney Newey, and Elie Tamer, and from participants in seminars at Harvard-MIT and Princeton, and at the UCL *Conference on Implementation of Structural Dynamic Models: Methodology and Applications*. We thank Tulin Erdem, Susumu Imai, and Michael Keane for sharing with us the working sample used in their paper.

1 Introduction

Persistent unobserved heterogeneity is pervasive in the empirical analysis using panel data of individuals, households, or firms. A key econometric issue in dynamic panel data models of economic behavior consists of distinguishing between state dependence ("true dynamics") and spurious dynamics due to unobserved heterogeneity (Heckman, 1981).[1] There are two general approaches to deal with this issue: random effects and fixed effects models/methods. Random-effects models impose restrictions on the distribution of unobserved heterogeneity (e.g., parametric, finite mixture), and on the joint distribution of these unobservables and the initial conditions of the observable explanatory variables. In contrast, fixed-effects methods are very attractive because they are fully nonparametric in the specification of the joint distribution of unobserved heterogeneity and exogenous or predetermined explanatory variables. Fixed effects are more robust than random effects methods.

There are different methods within the class fixed-effects estimators. From a conceptual point of view, the dummy variables estimator is the simplest of these methods: fixed effects are treated as any other parameter and are estimated jointly with the parameters of interest. Unfortunately, due to the incidental parameters problem, in most nonlinear panel data models this estimator of the structural parameters is inconsistent when the number of time periods is fixed (Neyman and Scott, 1948, Lancaster, 2000). Bias reduction methods, both analytical and simulation-based, have been proposed to deal with this problem (Hahn and Newey, 2004, Hahn and Kuersteiner, 2011). A different type of fixed effects methods is based on a transformation of the model that eliminates the fixed effects. For nonlinear panel data models, Manski's maximum score estimator of the panel data static binary choice model is an important example (Manksi, 1987). However, this estimator is not consistent in the dynamic binary choice model. Finally, other type of fixed effects methods is based on the derivation of sufficient statistics for the fixed effects and of a conditional maximum likelihood estimator of the structural parameters. This approach was pioneered by Andersen (1970) and extended by Chamberlain (1980). This paper focuses on the fixed effects - sufficient statistics approach and studies its applicability to dynamic discrete choice models where agents are forward-looking and maximize the expected and discounted value of the stream of current and future utilities. In this paper, we denote these models as *structural*.[2]

[1] See Arellano and Honoré (2001), and Arellano and Bonhomme (2012, 2017) for surveys on the econometrics of nonlinear panel data models.

[2] In contrast, we denote as *non-structural* to those dynamic models where agents are myopic (not forward-looking). Admittedly, the myopic model could be also structural.

Unfortunately, there is a wide class of nonlinear panel data models where it is not possible to derive sufficient statistics for the fixed effects. For instance, in the context of binary choice models, Chamberlain (1993, 2010) shows that a necessary and sufficient condition to have sufficient statistics for the fixed effects is that the distribution of the time-varying unobservable is logistic.[3] Similarly, these sufficient statistics do not exist in discrete choice models where the indexes that define the model are not additively separable between the fixed effects and the observable explanatory variables. This has important implications for structural dynamic discrete choice models. In these models, even if the fixed effect is additively separable in the one-period utility function, the solution of the structural model implies that this unobserved variable appears in the continuation value function interacting non-additively with the observable state variables. This interaction between the fixed-effect and the endogenous state variables typically makes sufficient-statistics for the fixed effects unfeasible.

For non-structural (i.e., myopic) dynamic logit models, Chamberlain (1985) and Honoré and Kyriazidou (2000) have derived sufficient statistics for the fixed effects, and have proposed consistent conditional maximum likelihood estimators. In contrast, all the methods and applications for structural dynamic discrete choice models have considered random-effects models with a finite mixture distribution, e.g., Keane and Wolpin (1997), Aguirregabiria and Mira (2007), Kasahara and Shimotsu (2009), Arcidiacono and Miller (2011), among many others. This random-effects approach imposes important restrictions: the number of points in the support of the unobserved heterogeneity is finite and is typically reduced to a small number of points, and the dependence between unobserved heterogeneity and the initial conditions of the observable state variables includes exclusion restrictions.

In this paper, we revisit the applicability of fixed-effects methods to the estimation of structural dynamic discrete choice models. We follow the sufficient statistics approach to study the identification of payoff function parameters in structural dynamic logit models with a fixed-effects specification of the time-invariant unobserved heterogeneity. We consider multinomial models with two different types of endogenous state variables: the lagged decision variable, and the duration in the last choice. The main challenge for the derivation of sufficient statistics is that, in general, the continuation values of the forward-looking decision problem depend on both the fixed-effect and the observable state variables in a non-additive way. Despite this property of the model, we

[3] Chamberlain (1993, 2010) considers the model where the time-varying unobservables are independently and identically distributed. Magnac (2004) studies a two-period model where the two time-varying unobservables have a general joint distribution.

show that there are pairs of choice histories such that the ratio between the probabilities of the two histories does not depend on the fixed effect but still depends on structural parameters, i.e., we can derive sufficient statistics for the fixed-effects. Based on our identification results, we propose a conditional maximum likelihood estimator. We apply this estimator to the bus engine replacement data in Rust (1987) and to the consumer scanner data on demand of a differentiated storable product in Erdem, Imai, and Keane (2003).

Performing counterfactual experiments and comparative statistics are important motivations in most applications of structural models. Though the fixed-effects approach delivers more robust estimates of the structural parameters, only by itself does not provide estimates counterfactuals and more generally of the marginal effects of parameters and state variables on choice probabilities. The estimation of these marginal effects requires the identification of the distribution of the fixed-effects. In our model, this distribution is only partially identified (see Chernozhukov et al., 2013). We describe a method to obtain a set (interval) estimate for the distribution of fixed effects and for the marginal effects and counterfactuals. We also show how to point-identify these objects under the restriction that the distribution of the fixed-effects has a finite mixture structure.

This paper contributes to the literature on structural dynamic discrete choice models. The structure of the payoff function and of the endogenous state variables that we consider in this paper includes as particular cases important economic applications in the literature of dynamic discrete choice structural models, such as models of market entry and exit either binary (Roberts and Tybout, 1997, Aguirregabiria and Mira, 2007) or multinomial (Sweeting, 2013; Caliendo et al, 2015); occupational choice models (Miller, 1984; Keane and Wolpin, 1997); machine replacement models (Rust, 1987; Das, 1992; Kennet, 1993; and Kasahara, 2009); inventory and investment decision models (Aguirregabiria 1999; Ryan, 2013; Kalouptsidi, 2014); demand of differentiated products with consumer brand switching costs (Erdem, Keane, and Sun, 2008) or storable products (Erdem, Imai, and Keane, 2003; Hendel and Nevo, 2006); and dynamic pricing models with menu costs (Willis, 2006), or with duration dependence due to inflation or other form of depreciation (Slade, 1998; Aguirregabiria, 1999; Kano, 2013); among others.[4] Our paper also contributes to the literature on nonlinear dynamic panel data models by providing new identification results of fixed effects dynamic logit models with duration dependence (Frederiksen, Honoré, and Hu, 2007).

The rest of the paper is organized as follows. Section 2 describes the class of models that we

[4]Note that most of the empirical applications cited above in this paragraph do not allow for time-invariant unobserved heterogeneity. This is still a common approach in empirical applications. The exceptions, within the cited papers, are Keane and Wolpin (1997), Erdem, Imai, and Keane (2003), Willis (2006), Aguirregabiria and Mira (2007), and Erdem, Keane, and Sun (2008).

study in this paper. Section 3 presents our identification results. Section 4 deals with estimation and inference. In section 5, we illustrate our methods in the context of two empirical applications. Section 6 summarizes and concludes.

2 Model

Time is discrete and indexed by t that belongs to $\{1, 2, ..., \infty\}$.[5] Agents are indexed by i. Every period t, agent i chooses a value of the discrete variable $y_{it} \in \mathcal{Y} = \{0, 1, ..., J\}$ to maximize her expected and discounted intertemporal utility $\mathbb{E}_t \left[\sum_{j=0}^{\infty} \delta_i^j \; U_{i,t+j}(y_{i,t+j}) \right]$, where $\delta_i \in (0, 1)$ is agent i's time discount factor, and $U_{it}(y)$ is her one-period utility if she chooses action y. This utility is a function of four types of state variables which are known to the agent at period t:

$$U_{it}(y) = \alpha\left(y, \eta_i, \mathbf{z}_{it}\right) + \beta\left(y, \mathbf{x}_{it}, \mathbf{z}_{it}\right) + \varepsilon_{it}(y). \tag{1}$$

$\mathbf{z}_{it}$ and $\mathbf{x}_{it}$ are observable to the researcher, and ε_{it} and η_i are unobservable. The vector $\mathbf{z}_{it}$ contains exogenous state variables and it follows a Markov process with transition probability function $f_{\mathbf{z}}(\mathbf{z}_{i,t+1}|\mathbf{z}_{it})$. The vector $\mathbf{x}_{it}$ contains endogenous state variables. We describe below the nature of these endogenous state variables and their transition rules. Both $\mathbf{z}_{it}$ and $\mathbf{x}_{it}$ have discrete supports $\mathcal{Z}$ and $\mathcal{X}$, respectively. The unobservable variables $\{\varepsilon_{it}(y) : y \in \mathcal{Y}\}$ are *i.i.d.* over (i, t, y) with an extreme value type I distribution. The vector η_i represents time-invariant unobserved heterogeneity from the point of view of the researcher. Let $\theta_i \equiv (\eta_i, \delta_i)$ represent the unobserved heterogeneity from individual i. The probability distribution of θ_i conditional on the history of observable state variables $\{\mathbf{z}_{it}, \mathbf{x}_{it} : t = 1, 2, ...\}$ is unrestricted and nonparametrically specified, i.e., fixed effects model. Functions $\alpha\left(y, \eta, \mathbf{z}\right)$ and $\beta\left(y, \mathbf{x}, \mathbf{z}\right)$ are nonparametrically specified but they are bounded.

This specification is closely related to *Rust model* (Rust, 1987, 1994). In particular, the time-varying unobservables $\varepsilon_{it}(y)$ satisfy conditions of *additive separability* and *conditional independence*, and they have a extreme value distribution. However, the model also relaxes some important conditions in Rust model. The inclusion of the unobservable η_i, through the term $\alpha\left(y, \eta_i, \mathbf{z}_{it}\right)$, implies relaxing Rust's assumptions of *additive separability* and *conditional independence* of the unobservables. There is time-invariant unobserved heterogeneity that can interact, in an unrestricted way, with the exogenous state variables $\mathbf{z}_{it}$ and the choice y_{it}. This specification of the utility function represents a pretty general fixed effects, semiparametric logit model. Furthermore, the model allows for unobserved heterogeneity in the discount factor δ_i.

[5] The time horizon of the decision problem is infinite.

The assumption of additive separability between η_i and the endogenous state variables in $\mathbf{x}_{it}$ is key for the identification results and estimation methods in this paper. This condition does not imply that the conditional-choice value functions, that describe the solution of the dynamic model, are additive separability between η_i and $\mathbf{x}_{it}$. In general, the solution of the dynamic programming problem implies a value function that is not additively separable in η_i and $\mathbf{x}_{it}$ even when the utility function is additive in these variables.

The model can accommodate two types of endogenous state variables that correspond to two different types of state dependence, $\mathbf{x}_{it} = (y_{i,t-1}, d_{it})$: (a) dependence on the the lagged decision variable, $y_{i,t-1}$; and (b) *duration dependence*, where $d_{it} \in \{1, 2, ..., \infty\}$ is the number of periods since the last change in choice. The lagged decision has the obvious transition rule. The transition rule for the duration variable is:

$$d_{i,t+1} = f_d(y_{it}, \mathbf{x}_{it}) \equiv 1\left\{y_{it} = y_{i,t-1}\right\} \; d_{it} + 1 \tag{2}$$

where $1\{.\}$ is the indicator function. We use the vector-valued function $f_x(y, \mathbf{x}_{it}) = [y \;,\; f_d(y, \mathbf{x}_{it})]$ to represent in vector form the transition rules of the vector $\mathbf{x}$ of endogenous state variables.[6]

The term $\beta\left(y, \mathbf{x}_{it}, \mathbf{z}_{it}\right)$ in the payoff function captures the dynamics, or structural state dependence, in the model. We distinguish in this function two additive components that correspond to the two forms of state dependence in the model:

$$\beta\left(y, \mathbf{x}_{it}, \mathbf{z}_{it}\right) = 1\{y = y_{i,t-1}\} \; \beta_d\left(y, d_{it}, \mathbf{z}_{it}\right) + 1\{y \neq y_{i,t-1}\} \; \beta_y\left(y, y_{i,t-1}, \mathbf{z}_{it}\right) \tag{3}$$

Function $\beta_d\left(y, d_{it}, \mathbf{z}_{it}\right)$ captures duration dependence. For instance, in an occupational choice model, this term captures the return on earnings of job experience in the current occupation. Function $\beta_y\left(y, y_{i,t-1}, \mathbf{z}_{it}\right)$ represents switching costs. In an occupational choice model, this term represents the cost of switching from occupation $y_{i,t-1}$ to occupation y. Without loss of generality, we set $\beta_y(y, y, \mathbf{z}_{it}) = 0$, i.e., the switching cost of no-switching is zero.[7] We also make two assumptions on function $\beta_d\left(y, d, \mathbf{z}_{it}\right)$ that play an important role in some of our identification results. First, we assume that there is not duration dependence in choice alternative $y = 0$, i.e., $\beta_d\left(0, d, \mathbf{z}_{it}\right) = 0$ for any value of d. Second, we assume that there is a finite value, $d^* < \infty$, for the duration variable

[6] Note that these endogenous state variables follow deterministic transition rules. This property of the model plays an important role in our identification results, and more specifically in our ability to control for unobserved heterogeneity in the continuation values of the optimal dynamic decision. We discuss this issue in section 3, as well as the possibility of extending our results to models with endogenous state variables that follow stochastic transition rules.

[7] Given the payoff function in equation (3), the parameter $\beta_y(y, y)$ is completely irrelevant for an individual's optimal decision. When $y_{it} = y_{i,t-1} = y$, we have that $\beta\left(y, \mathbf{x}_{it}\right) = \beta_d\left(y, d_{it}\right) + 0$ such that the term $\beta_y\left(y, y\right)$ never enters in the relevant payoff function. Therefore, $\beta_y\left(y, y\right)$ can be normalized to zero without loss of generality.

such that the marginal return of duration is zero for values greater that d^*:

$$\beta_d\left(y, d, \mathbf{z}\right) = \beta_d\left(y, d^*, \mathbf{z}\right) \quad \text{for any } d \geq d^* \tag{4}$$

For the moment, we assume that the researcher knows the value of d^*. Later we show that the value of d^* is identified from the data.

Assumption 1 summarizes all our conditions on the model. For the rest of the paper, we assume that this assumption holds.

ASSUMPTION 1. (A) The time horizon is infinite and $\delta_i \in (0,1)$. (B) The utility function has the form given by equations (1) and (3), and functions $\alpha\left(y, \eta, \mathbf{z}\right)$, $\beta_d\left(y, d, \mathbf{z}\right)$, and $\beta_y(y, y_{-1}, \mathbf{z})$ are bounded. (C) $\beta_y(y, y, \mathbf{z}) = 0$; $\beta_d\left(0, d, \mathbf{z}\right) = 0$; and there is a finite value of duration, $d^* < \infty$, such that $\beta_d\left(y, d, \mathbf{z}\right) = \beta_d\left(y, d^*, \mathbf{z}\right)$ for any $d \geq d^*$. (D) $\{\varepsilon_{it}(y) : y \in \mathcal{Y}\}$ are i.i.d. over (i, t, y) with a extreme value type I distribution. (E) $\mathbf{z}_{it}$ has discrete and finite support $\mathcal{Z}$ and follows a time-homogeneous Markov process. (F) The probability distribution of $\theta_i \equiv (\eta_i, \delta_i)$ conditional on $\{\mathbf{z}_{it}, \mathbf{x}_{it} : t = 1, 2, ...\}$ is nonparametrically specified and completely unrestricted. ■

The following are some examples of models within this class.

(a) Market entry-exit models. In the simpler version, there is only one market, and the choice variable is binary and represents a firm's decision of being active in the market ($y_{it} = 1$) or not ($y_{it} = 0$), e.g., Dunne et al. (2013). The only endogenous state variable is the lagged decision, $y_{i,t-1}$. The parameter $-\beta_y\left(1, 0\right)$ represents the cost of entry in the market. Similarly, the parameter $-\beta_y\left(0, 1\right)$ represents the cost of exit from the market. An extension of the basic entry model includes as an endogenous state variable the number of periods of experience since last entry in the market, d_{it}, that follows the transition rule $d_{i,t+1} = y_{it}\ d_{it} + 1$. The parameter $\beta_d\left(1, d\right)$ represents the effect of market experience on the firm's profit (Roberts and Tybout, 1997). The model can be extended to J markets, and then the two endogenous state variables are the index of the market where the firm was active at the last period ($y_{i,t-1}$) and the number of periods of experience in the current market (d_{it}). The parameter $\beta_y\left(y, y_{-1}\right)$ represents the cost of switching from market y_{-1} to market y (Sweeting, 2013; Caliendo et al, 2015).

(b) Occupational choice models. A worker chooses between J occupations and the choice alternative of not working ($y = 0$). There are costs of switching occupations. This implies that a worker's occupation at previous period, y_{it-1}, is a state variable of the model. The parameter $\beta_y\left(y, y_{-1}\right)$ is the cost of switching from occupation y_{-1} to occupation y. There is (passive) learning that increases productivity in the current occupation. Therefore, duration in the current occupation,

d_{it}, is a state variable (Miller, 1984; Keane and Wolpin, 1997). Parameter $\beta_d(y,d)$ represents the return of experience on the worker's utility (e.g., on earnings).

(c) Machine replacement models. The choice variable is binary and it represents the decision of keeping a machine ($y_{it} = 1$) or replacing it ($y_{it} = 0$). The only endogenous state variable is the number of periods since the last replacement, d_{it}, i.e., the machine age. The evolution of the machine age is $d_{i,t+1} = y_{it}\ d_{it} + 1$. The parameter $\beta_d(1,d)$ represents the effect of age on the firm's profit, e.g., productivity declines and maintenance costs increase with age (Rust, 1987; Das, 1992; Kennet, 1993; and Kasahara, 2009).[8] More generally, the class of models in this paper includes binary choice models of investment in capital, inventory, or capacity (Aguirregabiria 1999; Ryan, 2013; Kalouptsidi, 2014), as long as the depreciation of the stock is deterministic.

(d) Dynamic demand of differentiated products. A differentiated product has J varieties and a consumer chooses which one, if any, to purchase (no purchase is represented by $y = 0$). Brand switching costs imply that the brand in the last purchase is a state variable (Erdem, Keane, and Sun, 2008). For storable products, the duration since last purchase, d_{it}, represents (or proxies) the consumer's level of inventory that is an endogenous state variable (Erdem, Imai, and Keane, 2003; Hendel and Nevo, 2006). The parameter $\beta_d(y,d)$ captures the effect of inventory on the consumer's utility, and parameter $\beta_y(y, y_{-d})$ represents brand switching costs.

(e) Menu costs models of pricing. A firm sells a product and chooses its price to maximize intertemporal profits. Let p_{it} be the nominal log-price, and $p_{it} - \pi\ t$ is the real log-price, where π is the inflation rate in the economy. Every period the firm decides whether to keep its nominal price ($y_{it} = 0$) or to adjust the nominal price ($y_{it} = 1$) such that the real price becomes equal to a constant r^* (i.e., the nominal price becomes $r^* + \pi\ t$). Therefore, the real log-price at any period t can be represented as $r^* - \pi\ d_{it}$, where d_{it} represents the duration since the last price change. The firm's profit has two components: a variable profit that depends of the real price, $r^* - \pi\ d_{it}$; and a fixed menu cost that is paid only if the firm changes its nominal price. Duration since last price change is an endogenous state variable because inflation erodes the nominal price (Slade, 1998; Aguirregabiria, 1999; Willis, 2006; Kano, 2013). ■

We now derive the optimal decision rule and the conditional choice probabilities in this model. Agent i chooses y_{it} to maximize its expected and discounted intertemporal utility. In our model,

[8] In some versions of this model, such as Rust (1987), the endogenous state variable represents cumulative usage of the machine and it can follows a stochastic transition rule. In this paper, we do not study the identification of that model. We discuss this issue in section 3.

with infinite horizon and time-homogeneous utility and transition probability functions, Blackwell's Theorem establishes that the value function and the optimal decision rule are time-invariant (Blackwell, 1965). The optimal choice at period t can be represented as:

$$y_{it} \quad = \quad \arg\max_{y \in \mathcal{Y}} \left\{ \; \alpha\left(y, \eta_i, \mathbf{z}_{it}\right) + \beta\left(y, \mathbf{x}_{it}\right) + \varepsilon_{it}(y) + \delta_i \; \mathbb{E}_{\mathbf{z}_{i,t+1}|\mathbf{z}_{it}}\left[V_{\theta_i}\left(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{i,t+1}\right)\right] \; \right\} \qquad (5)$$

where $V_{\theta_i}(\mathbf{x}, \mathbf{z})$ is the integrated (or smoothed) value function for agent type θ_i, as defined by Rust (1994).[9] The operator $\mathbb{E}_{\mathbf{z}_{i,t+1}|\mathbf{z}_{it}}[.]$ represents the expectation over the distribution of the exogenous state variables $\mathbf{z}_{i,t+1}$ conditional to $\mathbf{z}_{it}$. The extreme value type 1 distribution of the unobservables ε implies the following Bellman equation for the integrated value function:

$$V_{\theta_i}\left(\mathbf{x}_{it}, \mathbf{z}_{it}\right) = \ln\left(\sum_{y \in \mathcal{Y}} \exp\left\{ \; \alpha\left(y, \eta_i, \mathbf{z}_{it}\right) + \beta\left(y, \mathbf{x}_{it}\right) + \delta_i \; \mathbb{E}_{\mathbf{z}_{i,t+1}|\mathbf{z}_{it}}\left[V_{\theta_i}\left(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{i,t+1}\right)\right] \; \right\}\right) \quad (6)$$

And the *conditional choice probability* (CCP) function has the following form:

$$P_{\theta_i}\left(y \mid \mathbf{x}_{it}, \mathbf{z}_{it}\right) \quad = \quad \frac{\exp\left\{ \; \alpha\left(y, \eta_i, \mathbf{z}_{it}\right) + \beta\left(y, \mathbf{x}_{it}\right) + v_{\theta_i}(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{it}) \; \right\}}{\sum_{j \in \mathcal{Y}} \exp\left\{ \; \alpha\left(j, \eta_i, \mathbf{z}_{it}\right) + \beta\left(j, \mathbf{x}_{it}\right) + v_{\theta_i}(f_x(j, \mathbf{x}_{it}), \mathbf{z}_{it}) \; \right\}} \qquad (7)$$

where $v_{\theta_i}(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{it})$ represents the continuation value $\delta_i \; \mathbb{E}_{\mathbf{z}_{i,t+1}|\mathbf{z}_{it}}\left[V_{\theta_i}\left(f_x(y, \mathbf{x}_{it}), \mathbf{z}_{i,t+1}\right)\right]$.

3 Identification

3.1 Preliminaries

The researcher has a panel dataset of N individuals over T periods of time, $\{y_{it}, \; \mathbf{x}_{it} \;, \; \mathbf{z}_{it} \; : \; i = 1, 2, ..., N \; ; \; t = 1, 2, ..., T\}$. We consider microeconometric applications where N is large and T is small. More precisely, our identification results and the asymptotic properties of the proposed estimators assume that N goes to infinity and T is small and fixed.[10] Sections 3.1 to 3.3 deal with the identification of the component of the utility function that represents dependence with respect to the endogenous state variables, i.e., functions $\beta_y\left(y, y_{-1}\right)$ and $\beta_d\left(y, d\right)$. In section 3.4., we study the identification of the distribution of time-invariant unobserved heterogeneity, θ_i.

For the rest of this section, we omit the individual subindex i in most of the expression, and instead we use θ as an index in those functions that depend on the time-invariant unobserved heterogeneity, i.e., $\alpha_\theta\left(y, \mathbf{z}\right)$ and $v_\theta\left(\mathbf{x}, \mathbf{z}\right)$.

[9] The integrated value function is defined as the integral of the value function.over the distribution of the i.i.d. unobservable state variables ε.

[10] Note that T represents the number of periods with data on the decision variable and the state variables for all the individuals. The set of observable state variables includes the endogenous state variables $y_{i,t-1}$ and d_{it}. Knowing the values of these state variables at the initial period $t = 1$ (i.e., y_{i0} and d_{i1}) may require data on the individual's choices for periods before $t = 1$. Therefore, the time dimension T may not correspond to the actual time dimension of the required panel dataset.

Let $\mathbf{y}^T = \{y_1, y_2, ..., y_T\}$ and $\mathbf{z}^T = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T\}$ be an individual's observed history of choices and exogenous state variables, respectively. The model implies that:

$$\mathbb{P}\left(\mathbf{y}^T \mid \mathbf{x}_1, \mathbf{z}^T, \theta\right) = \prod_{t=1}^{T} \frac{\exp\left\{\ \alpha_\theta\left(y_t, \mathbf{z}_t\right) + \beta\left(y_t, \mathbf{x}_t\right) + v_\theta\left(f_x(y_t, \mathbf{x}_t), \mathbf{z}_t\right)\ \right\}}{\sum\limits_{j \in \mathcal{Y}} \exp\left\{\ \alpha_\theta\left(j, \mathbf{z}_t\right) + \beta\left(j, \mathbf{x}_t\right) + v_\theta\left(f_x(j, \mathbf{x}_t), \mathbf{z}_t\right)\ \right\}} \tag{8}$$

Following Andersen (1970), Chamberlain (1980, 1985), and Honoré and Kyriazidou (2000), we look for a statistic, s, that is a deterministic function of the choice history $\mathbf{y}^T$ and that satisfies two conditions:[11]

(i) s is a sufficient statistic for θ, i.e., $\frac{\mathbb{P}\left(\mathbf{y}^T | \mathbf{x}_1, \mathbf{z}^T, \theta\right)}{\mathbb{P}\left(s | \mathbf{x}_1, \mathbf{z}^T, \theta\right)}$ does not depend on θ.

(ii) $\mathbb{P}\left(\mathbf{y}^T | \mathbf{x}_1, \mathbf{z}^T, s\right)$ still depends on (some of) the parameters of interest $\beta\left(y, \mathbf{x}\right)$.

The derivation of this sufficient statistic should deal with two issues that do not appear in the previous literature on sufficient statistics for fixed effects in non-structural (or myopic) nonlinear panel data models. First, we consider models with two types endogenous state variables, and in particular with duration dependence. Second, and more substantially, we should take into account that the fixed effect enters in the continuation value function, v_θ. This implies that we need a sufficient statistic not only for the fixed effect θ but also for the continuation values. This is challenging because, in general, these continuation values depend on the endogenous state variables. However, we cannot control for (or condition on) the value of the state variables because this implies controlling also for $\beta\left(y, \mathbf{x}\right)$ such that condition (ii) would not hold. Therefore, we need to find conditions under which it is possible to control for the continuation values without controlling for the current value of the endogenous state variables. In other words, we need conditions under which the continuation value does not depend on current state variables once we condition on current choices.

Let $\mathcal{Y}^T$ be the set of all the possible choice histories of length T. And let A and B be two choice histories in $\mathcal{Y}^T$. Without loss of generality (see section 3.3 below), we consider sufficient statistics with the form $s = 1\{\mathbf{y}^T \in A \cup B\}$. Therefore, the probability $\mathbb{P}\left(\mathbf{y}^T | \mathbf{x}_1, \mathbf{z}^T, \theta\right) / \mathbb{P}\left(\mathbf{y}^T | \mathbf{x}_1, \mathbf{z}^T, \theta\right)$ ratio in condition (i) above can be represented as $\mathbb{P}\left(A | \mathbf{x}_1, \mathbf{z}^T, \theta\right) / \mathbb{P}\left(A \cup B | \mathbf{x}_1, \mathbf{z}^T, \theta\right)$. Taking into account that A and B are mutually exclusive events such that $\mathbb{P}\left(A \cup B\right) = \mathbb{P}\left(A\right) + \mathbb{P}\left(B\right)$, we can represent conditions (i) and (ii) above as:

[11] A more standard definition of sufficient statistic in condition (i) is $\mathbb{P}\left(\mathbf{y}^T | \mathbf{x}_1, \mathbf{z}^T, \theta, s\right) = \mathbb{P}\left(\mathbf{y}^T | \mathbf{x}_1, \mathbf{z}^T, s\right)$. Note that, by Bayes' rule and taking into account that s is a deterministic function of $\mathbf{y}^T$, we have that condition (i) and condition $\mathbb{P}\left(\mathbf{y}^T | \mathbf{x}_1, \mathbf{z}^T, \theta, s\right) = \mathbb{P}\left(\mathbf{y}^T | \mathbf{x}_1, \mathbf{z}^T, s\right)$ are equivalent.

(i*) $\mathbb{P}\left(A|\mathbf{x}_1, \mathbf{z}^T, \theta\right)/\mathbb{P}\left(B|\mathbf{x}_1, \mathbf{z}^T, \theta\right)$ does not depend on θ.

(ii*) $\mathbb{P}\left(A|\mathbf{x}_1, \mathbf{z}^T, \theta\right)/\mathbb{P}\left(B|\mathbf{x}_1, \mathbf{z}^T, \theta\right)$ depends on the parameters of interest $\beta\left(y, \mathbf{x}\right)$.

As in Honoré and Kyriazidou (2000), our sufficient statistics include the condition that the exogenous state variables, $\mathbf{z}$, remains constant over several periods. We make this assumption explicit in our Propositions on identification, and we explain in section 4 how to deal with this condition in the implementation of the conditional maximum likelihood estimator. However, for notational simplicity, we omit $\mathbf{z}$ as an argument in most of the expressions for the rest of this section.

The presentation of our identification results tries to emphasize both the links and extensions with previous results in the literature. For this reason, we start with section 3.2 that presents our sufficient statistics in the binary choice model, that is the model more extensively studied in the literature of nonlinear dynamic panel data. Section 3.3 presents sufficient statistics for multinomial logit models. In these two sections, we provide examples of sufficient statistics that identify structural parameters but we do not characterize all the sufficient statistics that have information on the structural parameters. This important for efficient estimation. In section 3.4, we derive the set of all the sufficient statistics with information on the structural parameters.

Notation (statistics). It is convenient to define the statistic $S^{(y)}$ that represents the number of times that choice alternative y is visited in the choice history $\mathbf{y}^T$, i.e., $S^{(y)} \equiv \sum_{t=1}^{T} 1\{y_t = y\}$. The statistic $S^{(y,n)}$ represents the number of times that the pattern of n consecutive values of choice alternative y is observed in the history $(y_0, \mathbf{y}^T)$, e.g., $S^{(y,2)} \equiv \sum_{t=1}^{T} 1\{y_{t-1} = y_t = y\}$; $S^{(y,3)} \equiv \sum_{t=2}^{T} 1\{y_{t-2} = y_{t-1} = y_t = y\}$; and so on. Similarly, the statistic $S_{(-y_T)}^{(y,n)}$ has the same definition $S^{(y,n)}$ but it applies to the choice history $(y_0, y_1, ..., y_{T-1})$, i.e., excluding the last choice, y_T.

3.2 Binary choice models

Consider the binary choice version of the model characterized by Assumption 1. The optimal decision rule in this model is:

$$y_t = 1\left\{\begin{array}{l} \alpha_\theta(1) - \alpha_\theta(0) + \beta(1, y_{t-1}, d_t) - \beta(0, y_{t-1}, d_t) \\ \\ +v_\theta\left(f_x(1, y_{t-1}, d_t)\right) - v_\theta\left(f_x(0, y_{t-1}, d_t)\right) + \varepsilon_t(1) - \varepsilon_t(0) \geq 0 \end{array}\right\} \tag{9}$$

We now present sufficient statistics for the fixed effect θ in different versions of this model, starting with the myopic model without duration dependence that has been studied by Chamberlain (1985) and Honoré and Kyriazidou (2000).

3.2.1 Myopic dynamic logit without duration dependence

Consider the model in equation (9) under the restrictions of myopic behavior (i.e., $\delta = 0$) and no duration dependence (i.e., $\beta_d(y,d) = 0$). These restrictions imply that the difference in continuation values, $v_\theta\left(f_x(1, y_{t-1}, d_t)\right) - v_\theta\left(f_x(0, y_{t-1}, d_t)\right)$, becomes zero, and the term $\beta(1, y_{t-1}, d_t) - \beta(0, y_{t-1}, d_t)$ becomes equal to $\beta_y(1,0) - y_{t-1}\left[\beta_y(1,0) + \beta_y(0,1)\right]$. We can present this model using the more standard representation $y_t = 1\{\widetilde{\alpha}_\theta + \widetilde{\beta}_y \; y_{t-1} + \widetilde{\varepsilon}_t \geq 0\}$, with $\widetilde{\alpha}_\theta \equiv \alpha_\theta(1) - \alpha_\theta(0) + \beta_y(1,0)$, $\widetilde{\beta}_y \equiv -\beta_y(1,0) - \beta_y(0,1)$, and $\widetilde{\varepsilon}_t \equiv \varepsilon_t(1) - \varepsilon_t(0)$. In a model of market entry-exit, the parameter $\widetilde{\beta}_y$ represents the sum of the costs of entry and exit, or equivalently the sunk cost of entry. This is an important structural parameter.

Remember that $S^{(y)}$ represents the number of times that choice alternative y is visited in the choice history $\mathbf{y}^T$, and $S^{(y,n)}$ is the number of times that choice alternative y is visited n consecutive times over the history $(y_0, \mathbf{y}^T)$. For the binary choice model, we have that $S^{(1)} = \sum_{t=1}^T y_t$, $S^{(0)} = T - S^{(1)}$, and $S^{(1,2)} = \sum_{t=1}^T y_{t-1} y_t$. Define also the function $\sigma_\theta(y_{t-1}) \equiv \ln\left(1 + \exp\left\{\widetilde{\alpha}_\theta + \widetilde{\beta}_y y_{t-1}\right\}\right)$. Using this notation, the log-probability of the choice history $\mathbf{y}^T$ conditional on (y_0, θ) is:

$$
\begin{aligned}
\ln \mathbb{P}\left(\mathbf{y}^T \mid y_0, \theta\right) &= \sum_{t=1}^{T} y_t \left[\widetilde{\alpha}_\theta + \widetilde{\beta}_y y_{t-1}\right] - (1 - y_{t-1}) \; \sigma_\theta(0) - y_{t-1} \; \sigma_\theta(1) \\
&= S^{(1)} \widetilde{\alpha}_\theta + S^{(1,2)} \widetilde{\beta}_y - \left[S^{(0)} - y_0 + y_T\right] \sigma_\theta(0) - \left[S^{(1)} + y_0 - y_T\right] \sigma_\theta(1)
\end{aligned}
\tag{10}
$$

Let A and B be two possible realizations of the choice history $(y_0, \mathbf{y}^T)$. We are interested in characterizing the necessary and sufficient conditions on histories A and B such that $\ln \mathbb{P}\left(A \mid y_{A,0}, \theta\right) - \ln \mathbb{P}\left(B \mid y_{B,0}, \theta\right)$: (i*) does not depend on θ; and (ii*) depends on $\widetilde{\beta}_y$. From equation (10), we have that:

$$
\begin{aligned}
\ln \mathbb{P}\left(A | y_{A,0}, \theta\right) - \ln \mathbb{P}\left(B | y_{B,0}, \theta\right) = \left[S_A^{(1)} - S_B^{(1)}\right] \; \left[\widetilde{\alpha}_\theta + \sigma_\theta(0) - \sigma_\theta(1)\right] \\
+ \left[(y_{A,0} - y_{B,0}) - (y_{A,T} - y_{B,T})\right] \; \left[\sigma_\theta(0) - \sigma_\theta(1)\right] + \left[S_A^{(1,2)} - S_B^{(1,2)}\right] \widetilde{\beta}_y
\end{aligned}
\tag{11}
$$

Condition (i*) requires that the first two terms are zero for any value of θ. It is clear that this condition implies that $\{y_{A,0}, y_{A,T}, S_A^{(1)}\} = \{y_{B,0}, y_{B,T}, S_B^{(1)}\}$. Condition (ii*) requires $S_A^{(1,2)} \neq S_B^{(1,2)}$. That is, the pair of choice histories that define the sufficient statistic have the same values for the initial condition (y_0), the last choice (y_T), and number of times that each alternative is chosen, and they have different values for the number of times that alternative 1 is chosen at two consecutive periods. For instance, with $T = 3$, an example of a pair of choice histories $\{y_0, y_1, y_2, y_3\}$ satisfying conditions (i*) and (ii*) is $A = \{0, 0, 1, 1\}$ and $B = \{0, 1, 0, 1\}$.

Therefore, in the binary myopic model without duration dependence, the statistic $s = \{y_0, y_T, S^{(1)}\}$ is a sufficient statistic for θ in $\mathbb{P}\left(\mathbf{y}^{(T)} \mid y_0, \theta\right)$. Furthermore, there exit pairs of histories A and B with

$s_A = s_B = s$ and $S_A^{(1,2)} \neq S_B^{(1,2)}$ such that parameter $\widetilde{\beta}_y$ is identified as $\widetilde{\beta}_y = [\ln \mathbb{P}(A|s) - \ln \mathbb{P}(B|s)] / \left[S_A^{(1,2)} - S_B^{(1,2)}\right]$. This result was previously established by Chamberlain (1985).

3.2.2 Forward-looking dynamic logit without duration dependence

Consider a forward-looking version of the model in equation (9) but still without duration dependence. As in the previous model, we have that $\alpha_\theta(1) - \alpha_\theta(0) + \beta(1, y_{t-1}) - \beta(0, y_{t-1})$ can be represented as $\widetilde{\alpha}_\theta + \widetilde{\beta}_y y_{t-1}$, where $\widetilde{\alpha}_\theta$ and $\widetilde{\beta}_y$ have the same interpretation as before. Since the model is of forward-looking behavior, now we have the continuation values $v_\theta(f_x(1, y_{t-1}, d_t)) - v_\theta(f_x(0, y_{t-1}, d_t))$. Since there is not duration dependence, the only state variable is y_{t-1}, and the transition rule for this state variable is $f_y(y, y_{t-1}) = y$. Therefore, for this version of the model we have that $v_\theta(f_x(1, y_{t-1}, d_t)) - v_\theta(f_x(0, y_{t-1}, d_t)) = v_\theta(1) - v_\theta(0) \equiv \widetilde{v}_\theta$, i.e., continuation values depend on current choices but on the current state variable y_{t-1}. This is a key property for the derivation of sufficient statistics in this model of forward-looking behavior. We can represent this model using the expression, $y_t = 1\{\widetilde{\alpha}_\theta + \widetilde{v}_\theta + \widetilde{\beta}_y \ y_{t-1} + \widetilde{\varepsilon}_t \geq 0\}$. The only difference between this model and the myopic model is that now the fixed effect has two components: $\widetilde{\alpha}_\theta$ that comes from current profit, and $\widetilde{v}_\theta$ that comes from the continuation values. However, from the point of view of fixed-effects estimation, the two models are observationally equivalent.

A key feature of this model, that determines the observational equivalence with the myopic model, is the transition of the endogenous state variable, and in particular the property that the state variable at period $t+1$ depends on the choice at period t but not on the state variable at period t, i.e., $x_{t+1} = y_t$. Note that this property can be generalized to models where the transition rule contains unobserved heterogeneity or/and observable exogenous state variables, i.e., $x_{t+1} = f_\theta(y_t, \mathbf{z}_t)$.

Based on this equivalence between the myopic and the forward-looking models without duration dependence, Proposition 1 establishes necessary and sufficient conditions for the identification of $\widetilde{\beta}_y$ in the structural (forward-looking) dynamic binary logit model.

PROPOSITION 1. In the forward-looking binary choice model without duration dependence: (a) the statistic $s = \{y_0, y_T, S^{(1)}\}$ *is the minimum sufficient statistic for* θ *in* $\mathbb{P}\left(\mathbf{y}^T \mid y_0, \theta\right)$*; and (b) there exit pairs of histories, A and B, with* $s_A = s_B = s$ *and* $S_A^{(1,2)} \neq S_B^{(1,2)}$ *such that the parameter* $\widetilde{\beta}_y$ *is identified as* $\widetilde{\beta}_y = [\ln \mathbb{P}(A|s) - \ln \mathbb{P}(B|s)] / \left[S_A^{(1,2)} - S_B^{(1,2)}\right]$. ■

3.2.3 Myopic dynamic logit with duration dependence

Now, consider the myopic binary choice model with duration dependence. The continuation values are zero, and now

$$\begin{aligned} \beta(1, y_{t-1}, d_t) - \beta(0, y_{t-1}, d_t) &= [(1 - y_{t-1})\beta_y(1, 0) + y_{t-1}\beta_d(1, d_t)] - y_{t-1}\beta_y(0, 1) \\ &= \beta_y(1, 0) + \widetilde{\beta}_y \; y_{t-1} + \widetilde{\beta}_d(d_t) \; y_{t-1} \end{aligned} \tag{12}$$

with $\widetilde{\beta}_d(d_t) \equiv \beta_d(1, d_t)$. Therefore, we can present this model as $y_t = 1\{\widetilde{\alpha}_\theta + \widetilde{\beta}_y \; y_{t-1} + \widetilde{\beta}_d(d_t) \; y_{t-1} + \widetilde{\varepsilon}_t \geq 0\}$. For this model, the log-probability of the choice history $\mathbf{y}^T$ conditional on (y_0, d_1, θ) is:

$$\ln \mathbb{P}\left(\mathbf{y}^T \mid y_0, d_1, \theta\right) = \sum_{t=1}^{T} y_t \left[\widetilde{\alpha}_\theta + \widetilde{\beta}_y \; y_{t-1} + \widetilde{\beta}_d(d_t) \; y_{t-1}\right] - \sigma_\theta(y_{t-1}, d_t) \tag{13}$$

where $\sigma_\theta(y_{t-1}, d_t) \equiv \ln\left(1 + \exp\left\{\widetilde{\alpha}_\theta + \widetilde{\beta}_y \; y_{t-1} + \widetilde{\beta}_d(d_t) \; y_{t-1}\right\}\right)$. In order to emphasize that $\sigma_\theta(y_{t-1}, d_t)$ does not depend on d_t when $y_{t-1} = 0$, we use the notation $\sigma_\theta(0)$ to represent $\sigma_\theta(0, d)$. Without loss of generality (w.l.o.g.), we consider that the initial condition is $y_0 = 0$.[12] Then, we can rewrite the log probability in equation (13) as follows:

$$\begin{aligned} \ln \mathbb{P}\left(\mathbf{y}^T | \theta\right) &= S^{(1)}\widetilde{\alpha}_\theta - \left[S^{(0)} + y_T - y_0\right]\sigma_\theta(0) - \sum_{n=1}^{T-1}\left[S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)}\right]\sigma_\theta(1, n) \\ &+ S^{(1,2)}\widetilde{\beta}_y + \sum_{n=1}^{T-1}\left[S^{(1,n+1)} - S^{(1,n+2)}\right]\widetilde{\beta}_d(n) \end{aligned} \tag{14}$$

Looking at equation (14), we see that condition (i*) holds if and only if the vector of statistics $s = \{y_0, y_T, S^{(1)}, S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)} : n \geq 1\}$ is the same for the pair of histories A and B, such that the terms associated to $\widetilde{\alpha}_\theta$, $\sigma_\theta(0)$, and $\sigma_\theta(1, n)$ cancel in the log-probability difference $\ln \mathbb{P}(A) - \ln \mathbb{P}(B)$. It is clear that this sufficient statistic s always exist. Then, conditional on this statistic s, we have have that:

$$\ln\left[\frac{\mathbb{P}(A|s)}{\mathbb{P}(B|s)}\right] = \left[S^{(1,2)}_A - S^{(1,2)}_B\right]\widetilde{\beta}_y + \sum_{n=1}^{T-1}\left\{\left[S^{(1,n+1)}_A - S^{(1,n+2)}_A\right] - \left[S^{(1,n+1)}_B - S^{(1,n+2)}_B\right]\right\}\widetilde{\beta}_d(n) \tag{15}$$

Condition (ii*) requires that, conditional on s, we have that $S^{(1,2)}_A \neq S^{(1,2)}_B$ and/or $\left[S^{(1,n+1)}_A - S^{(1,n+2)}_A\right] \neq \left[S^{(1,n+1)}_B - S^{(1,n+2)}_B\right]$ for some integer $n \geq 1$. The following example shows that this pair of choice histories exists

EXAMPLE 1. Consider the pair of choice histories, A and B, from $t = 0$ to $t = T$: $A = \{0, 0, \mathbf{1}_d, 1\}$ and $B = \{0, \mathbf{1}_d, 0, 1\}$, where vector $\mathbf{1}_d$ represents a sequence of d consecutive $1's$. Note that

[12] Since the the initial duration, d_1, is observed, we know that before period $t = 1$ there were d_1 consecutive periods with $y = 1$ and $y_{-d_1} = 0$. Therefore, we can always construct a choice history where the initial condition is $y = 0$.

$d = T - 2$. First, we verify that these vector of statistics $s = \{y_0, y_T, S^{(1)}, S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)} : n \geq 1\}$. It is clear that $y_{0,A} = y_{0,B} = 0$, $y_{T,A} = y_{T,B} = 1$, and $S^{(1)}_A = S^{(1)}_B = d + 1$. Note also that removing the last period, $A_{(-y_T)} = \{0, 0, \mathbf{1}_d\}$ and $B_{(-y_T)} = \{0, \mathbf{1}_d, 0\}$, and these histories have the same values for $S^{(1,1)}_{(-y_T)} = d$, $S^{(1,2)}_{(-y_T)} = d - 1$, ..., $S^{(1,d)}_{(-y_T)} = 1$, and $S^{(1,n)}_{(-y_T)} = 0$ for $n > d$. Second, note that $S^{(1,2)}_A = d$ and $S^{(1,2)}_B = d - 1$, such that $\left[S^{(1,2)}_A - S^{(1,2)}_B\right] \widetilde{\beta}_y = \widetilde{\beta}_y$. Also, for $1 \leq n \leq d - 1$, the statistics $S^{(1,n+1)} - S^{(1,n+2)}$ are all equal to 1 in both histories. However, for $n = d$, we have that $S^{(1,n+1)}_A - S^{(1,n+2)}_A = 1$ and $S^{(1,n+1)}_B - S^{(1,n+2)}_B = 0$. Therefore, we have that $\ln \mathbb{P}(A|s) - \ln \mathbb{P}(B|s) = \widetilde{\beta}_y + \widetilde{\beta}_d(d)$. ■

PROPOSITION 2. In the myopic binary choice model with duration dependence: (a) the statistic $s = \{y_0,\ y_T,\ S^{(1)},\ S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)} : n \geq 1\}$ *is the minimum sufficient statistic for* θ *in* $\mathbb{P}\left(\mathbf{y}^T \mid y_0, d_1, \theta\right)$*; and (b) suppose that consider* $T \geq 3$ *and consider the pair of histories* $A = \{0, 0, \mathbf{1}_d, 1\}$ *and* $B = \{0, \mathbf{1}_d, 0, 1\}$ *with* $d \geq 1$*; then, parameter* $\gamma(d) \equiv \widetilde{\beta}_y + \widetilde{\beta}_d(d)$ *is identified as* $\gamma(d) = \ln \mathbb{P}(A|s) - \ln \mathbb{P}(B|s)$. ■

Given the parameters $\{\gamma(d) : d = 1, 2, ...T - 2\}$, we can identify the marginal returns to experience $\widetilde{\beta}_d(d) - \widetilde{\beta}_d(d - 1)$ as $\gamma(d) - \gamma(d - 1)$ for any value d between 2 and $T - 2$. Note that, in this binary choice model with both switching costs and duration dependence, it is not possible to separately identify the switching cost parameter , $\widetilde{\beta}_y$, and the return of the first period of experience, $\widetilde{\beta}_d(1)$.[13] However, knowledge of the structural parameters $\gamma(d)$ is sufficient to answer most relevant economic questions.

3.2.4 Forward-looking dynamic logit with duration dependence

Now, the optimal decision rule includes the difference of continuation values $v_\theta(1, d_t + 1) - v_\theta(0)$, where for choice $y = 0$ we use the $v_\theta(0)$ instead $v_\theta(0, d)$ to emphasize that there is not duration dependence when the state is $y = 0$. Therefore, we have the model:

$$y_t = 1\left\{\widetilde{\alpha}_\theta + \widetilde{\beta}_y\ y_{t-1} + \widetilde{\beta}_d(d_t)\ y_{t-1} + v_\theta(1, d_t + 1) - v_\theta(0) + \varepsilon_t \geq 0\right\} \tag{16}$$

For this model, the log-probability of the choice history $\mathbf{y}^T$ conditional on (y_0, d_1, θ) is:

$$\ln \mathbb{P}\left(\mathbf{y}^T \mid y_0, d_1, \theta\right) = \sum_{t=1}^{T} y_t \left[\widetilde{\alpha}_\theta + \widetilde{\beta}_y y_{t-1} + \widetilde{\beta}_d(d_t) y_{t-1} + v_\theta(1, d_t + 1)\right] - \sigma_\theta(y_{t-1}, d_t) \tag{17}$$

[13] This under-identification result is related to the under-identification of the autoregressive of order two model, AR(2), studied by Chamberlain (1985). In that model, we have $y_{it} = 1\{\widetilde{\alpha}_i + \beta_1\ y_{i,t-1} + \beta_2\ y_{i,t-2} + \widetilde{\varepsilon}_{it} \geq 0\}$. Chamberlain showed that the parameter β_2 is identified but the parameter β_1 is not.

where now $\widetilde{\alpha}_\theta \equiv \alpha_\theta(1) - \alpha_\theta(0) + \beta_y(1,0) - v_\theta(0)$, and $\sigma_\theta(y_{t-1}, d_t) \equiv \ln(1+ \exp\{\widetilde{\alpha}_\theta + \widetilde{\beta}_y y_{t-1} + \widetilde{\beta}_d(d_t) y_{t-1} + v_\theta(1, d_t+1)\})$. Again, to emphasize that $\sigma_\theta(y_{t-1}, d_t)$ does not depend on d_t when $y_{t-1} = 0$, we use the notation $\sigma_\theta(0)$ to represent $\sigma_\theta(0, d)$.

Similarly as we did in the myopic model with duration dependence, we consider that the initial condition is $y_0 = 0$. Then, we can obtain the following expression for the log-probability:

$$\begin{aligned}
\ln \mathbb{P}\left(\mathbf{y}^T | \theta\right) &= S^{(1)} \widetilde{\alpha}_\theta - \left[S^{(0)} + y_T - y_0\right] \sigma_\theta(0) - \sum_{n=1}^{T-1} \left[S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)}\right] \sigma_\theta(1,n) \\
&+ \sum_{n=1}^{T} \left[S^{(1,n)} - S^{(1,n+1)}\right] v_\theta(1,n) \qquad (18) \\
&+ S^{(1,2)} \widetilde{\beta}_y + \sum_{n=1}^{T-1} \left[S^{(1,n+1)} - S^{(1,n+2)}\right] \widetilde{\beta}_d(n)
\end{aligned}$$

This expression shows that to control for the fixed effect in $v_\theta(1,n)$ we need to control for the statistic $S^{(1,n)} - S^{(1,n+1)}$. Therefore, the minimum sufficient statistic for θ is $s = \{y_0,\ y_T,\ S^{(1)},\ S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)},\ S^{(1,n)} - S^{(1,n+1)} : n \geq 1\}$. Given a pair of histories A and B, then the log-probability difference $\ln \mathbb{P}(A|s_A) - \ln \mathbb{P}(B|s_B)$ does not depend on θ if and only if $s_A = s_B$, i.e., condition (i*). Furthermore, the log-probability difference depends on the structural parameters $\widetilde{\beta}_y$ and $\widetilde{\beta}_d(n)$ if and only if $S^{(1,2)}_A \neq S^{(1,2)}_B$ and $S^{(1,n+1)}_A - S^{(1,n+2)}_A \neq S^{(1,n+1)}_B - S^{(1,n+2)}_B$, respectively, i.e., condition (ii*).

Now, the key question is whether it is possible to find a pair of histories that satisfy conditions (i*) and (ii*).

EXAMPLE 2. Consider the pair of choice histories, A and B in Example 1 above: $A = \{0, 0, \mathbf{1}_d, 1\}$ and $B = \{0, \mathbf{1}_d, 0, 1\}$. Now, in the forward-looking model, we have that:

$$\ln \mathbb{P}(A|s_A) - \mathbb{P}(B|s_B) = \widetilde{\beta}_y + \widetilde{\beta}_d(d) + v_\theta(1,d) - v_\theta(1,1) \qquad (19)$$

In this model of forward-looking behavior, this pair of histories does not satisfy condition (i*), $s_A = s_B$, and log-probability difference depends on the fixed-effect through the continuation value $v_\theta(1,d) - v_\theta(1,1)$. ■

EXAMPLE 3. Consider now the pair of choice histories $A = \{0, \mathbf{1}_d, 0, \mathbf{1}_{d^*+1}\}$ and $B = \{0, \mathbf{1}_{d+1}, 0, \mathbf{1}_{d^*}\}$, where d^* is value defined in section 2 (i.e., $\widetilde{\beta}_d(d+1) - \widetilde{\beta}_d(d) = 0$ for any $d \geq d^*$), and d is a positive integer such that $d \leq d^* - 1$. It is straightforward to show that this pair of histories are such that:

$$\ln \left[\frac{\mathbb{P}(A|s_A)}{\mathbb{P}(B|s_B)}\right] = \widetilde{\beta}_d(d^*) - \widetilde{\beta}_d(d) + v_\theta(1, d^*+1) - v_\theta(1, d+1) - \sigma_\theta(1, d^*) + \sigma_\theta(1, d+1) \qquad (20)$$

Note that, for $d = d^* - 1$, we have that $\sigma_\theta(1, d^*) - \sigma_\theta(1, d+1) = 1$. Furthermore, by the definition of d^*, the continuation value function $v_\theta(1, d)$ becomes constant for durations greater of equal than d^*. Therefore, for $d = d^* - 1$, we have that $\ln \mathbb{P}(A|s_A) - \mathbb{P}(B|s_B) = \widetilde{\beta}_d(d^*) - \widetilde{\beta}_d(d^* - 1)$ such that the parameter $\widetilde{\beta}_d(d^*) - \widetilde{\beta}_d(d^* - 1)$ is identified. ■

PROPOSITION 3. In the forward-looking binary choice model with duration dependence: (a) the statistic $s = \{y_0,\, y_T,\, S^{(1)},\, S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)},\, S^{(1,n)} - S^{(1,n+1)} : n \geq 1\}$ *is the minimum sufficient statistic for* θ *in* $\mathbb{P}\left(\mathbf{y}^T \mid y_0, d_1, \theta\right)$*; and (b) consider the pair of histories* $A = \{0, \mathbf{1}_{d^*-1}, 0, \mathbf{1}_{d^*+1}\}$ *and* $B = \{0, \mathbf{1}_{d^*}, 0, \mathbf{1}_{d^*}\}$*; then, parameter* $\widetilde{\beta}_d(d^*) - \widetilde{\beta}_d(d^* - 1)$ *from* $\ln \mathbb{P}(A|s) - \ln \mathbb{P}(B|s)$. ■

Table 1 summarizes the identification results for the dynamic binary logit.

Table 1
Identification of Dynamic Binary Logit Models

Panel 1: Models without duration dependence

Myopic Model		*Forward-Looking Model*	
Sufficient Stat.	Identified parameters	Sufficient Stat.	Identified parameters
$\{y_0,\, y_T,\, S^{(1)}\}$	$\widetilde{\beta}_y$	$\{y_0,\, y_T,\, S^{(1)}\}$	$\widetilde{\beta}_y$

Panel 2: Models with duration dependence

Myopic Model		*Forward-Looking Model*	
Sufficient Stat.	Identified parameters	Sufficient Stat.	Identified parameters
$\{y_0,\, y_T,\, S^{(1)},$ $S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)} : n \geq 1\}$	$\widetilde{\beta}_y + \widetilde{\beta}_d(d)$ for $d \leq T - 2$	$\{y_0,\, y_T,\, S^{(1)},$ $S^{(1,n)}_{(-y_T)} - S^{(1,n+1)}_{(-y_T)},$ $S^{(1,n)} - S^{(1,n+1)} : n \geq 1\}$	$\widetilde{\beta}_d(d^*) - \widetilde{\beta}_d(d^* - 1)$ if $d^* \leq T - 2$

Identification of d^*

3.3 Multinomial choice models [Preliminary]

We now generalize the identification results of the forward-looking model in Examples 2 and 4 to the multinomial case with general T. Proposition 1 deals with the identification of the switching costs parameters $\beta_y(y, y_0)$, while Proposition 2 deals with the identification of the duration dependence parameters $\beta_d(y,d)$. In these Propositions we consider a model that includes both switching costs $\beta_y(y,y_0)$ and duration dependence $\beta_d(y,d)$.

PROPOSITION 4. Suppose that $T \geq 3$, *the conditions in Assumption 1 hold, and* $\mathbb{P}(\mathbf{z}_1 = \mathbf{z}_2 = \mathbf{z}_3) > 0$. *Let* y_0, a, b, *and* y_3 *be values in the choice set* $\mathcal{Y}$ *(not necessarily different), and let* $A = \{y_0,\ a,\ b,\ y_3\}$ *and* $B = \{y_0,\ b,\ a,\ y_3\}$ *represent the two possible choice paths to go from* y_0 *to* y_3 *visiting* a *and* b. *Let* $d_1 \in \{0,1,...,d^*\}$ *be the duration at period* $t=1$ *such the* $\mathbf{x}_1 = (y_0, d_1)$. *Consider the following restrictions on the values* d, y_0, a, b, *and* y_3*: (i)* $a \neq b$*; (ii) if* $y_0 \neq 0$ *and* $d_1 < d^*$, *then* $a \neq y_0$ *and* $b \neq y_0$*; and (iii) either* $y_3 = 0$ *or* $\{y_3 \neq a$ *and* $y_3 \neq b\}$. *Define* $\widetilde{\mathbf{y}} \equiv (y_0, y_1, y_2, y_3)$ *and let* $s \in \{0,1\}$ *be the binary statistic defined as,*

$$s = 1\left\{\mathbf{z}_t = \mathbf{z} \text{ for } t = 1,2,3;\ \widetilde{\mathbf{y}} \in A \cup B\right\} \tag{21}$$

Then, (1) the statistic s *is sufficient for* θ *such that the probability* $\mathbb{P}(\widetilde{\mathbf{y}} \mid s=1,\theta)$ *does not depend on* θ*; and (2)* $\ln\mathbb{P}(A \mid s=1) - \ln\mathbb{P}(B \mid s=1) = \Delta_y(y_0 \rightarrow y_3; a,b)$, *and*

$$\Delta_y(y_0 \rightarrow y_3; a,b) \equiv [\beta_y(a,y_0) + \beta_y(b,a) + \beta_y(y_3,b)] - [\beta_y(b,y_0) + \beta_y(a,b) + \beta_y(y_3,a)] \tag{22}$$

that represents the difference in total switching costs between choice paths A *and* B. *This parameter is identified and can be estimated consistently at a rate root-*N *as* N *goes to infinity and* T *is fixed.* ■

Proof: Conditional on $s=1$, there are only two possible choice histories, A or B. Therefore, $\mathbb{P}(\widetilde{\mathbf{y}} = A \mid s=1,\theta) = \mathbb{P}(A \mid \theta) / [\mathbb{P}(A \mid \theta) + \mathbb{P}(B \mid \theta)]$. Showing that $\mathbb{P}(\widetilde{\mathbf{y}} \mid s=1,\theta)$ does not depend on θ is equivalent to proving that $\mathbb{P}(A \mid \theta)/\mathbb{P}(B \mid \theta)$ does not depend on θ. Let $x_1 = \{y_0, d\}$. Under condition (ii) "if $y_0 \neq 0$ and $d_1 < d^*$, then $a \neq y_0$ and $b \neq y_0$" we have that the transition $f_d(a,y_0,d_1)$ can take only two possible values: either $a=0$ such that $f_d(a,y_0,d_1)=0$; or $a \neq 0$ and $a \neq y_0$ such that $f_d(a,y_0,d_1)=1$. Therefore, we have that $f_d(a,y_0,d_1) = 1\{a \neq 0\}$. Using the same argument, we have that $f_d(b,y_0,d_1) = 1\{b \neq 0\}$. Therefore, under these conditions, choice paths A and B imply the following paths for the endogenous state variables at periods $t=1,2,3$, where we use 1_a and 1_b to represent $1\{a \neq 0\}$ and $1\{b \neq 0\}$, respectively:

$$\begin{aligned} \widetilde{\mathbf{x}}_A &= \left\{ \begin{bmatrix} y_0 \\ d_1 \end{bmatrix}, \begin{bmatrix} a \\ 1_a \end{bmatrix}, \begin{bmatrix} b \\ 1_b \end{bmatrix} \right\} \\ \widetilde{\mathbf{x}}_B &= \left\{ \begin{bmatrix} y_0 \\ d_1 \end{bmatrix}, \begin{bmatrix} b \\ 1_b \end{bmatrix}, \begin{bmatrix} a \\ 1_a \end{bmatrix} \right\} \end{aligned} \tag{23}$$

These two paths visit the same set of states, but with different timing. Therefore, given that the denominator in the logit probabilities depends only on the state and not on the choice, we have that the denominators in the expressions for $\mathbb{P}(A|\theta)$ and $\mathbb{P}(B|\theta)$ are the same and they cancel in the ratio $\mathbb{P}(A|\theta)$ / $\mathbb{P}(B|\theta)$. To see this in more detail, consider the probabilities for the choice paths (where we have removed $\mathbf{z}$ as an argument for notational convenience):

$$\begin{aligned} \mathbb{P}(A|\theta) &= \frac{\exp\{\ \alpha_\theta(a) + \beta(a, y_0, d_1) + v_\theta(a, 1_a)\ \}}{\sum_{j\in\mathcal{Y}} \exp\{\ \alpha_\theta(j) + \beta(j, y_0, d_1) + v_\theta(j, f_d(j, y_0, d_1))\ \}} \\ &\times \frac{\exp\{\ \alpha_\theta(b) + \beta(b, a, 1_a) + v_\theta(b, 1_b)\ \}}{\sum_{j\in\mathcal{Y}} \exp\{\ \alpha_\theta(j) + \beta(j, a, 1_a) + v_\theta(j, f_d(j, a, 1_a))\ \}} \\ &\times \frac{\exp\{\ \alpha_\theta(y_3) + \beta(y_3, b, 1_b) + v_\theta(y_3, f_d(y_3, b, 1_b))\ \}}{\sum_{j\in\mathcal{Y}} \exp\{\ \alpha_\theta(j) + \beta(j, b, 1_b) + v_\theta(j, f_d(j, b, 1_b))\ \}} \end{aligned} \tag{24}$$

and,

$$\begin{aligned} \mathbb{P}(B|\theta) &= \frac{\exp\{\ \alpha_\theta(b) + \beta(b, y_0, d) + v_\theta(b, 1_b)\ \}}{\sum_{j\in\mathcal{Y}} \exp\{\ \alpha_\theta(j) + \beta(j, y_0, d) + v_\theta(j, f_d(j, y_0, d))\ \}} \\ &\times \frac{\exp\{\ \alpha_\theta(a) + \beta(a, b, 1_b) + v_\theta(a, 1_a)\ \}}{\sum_{j\in\mathcal{Y}} \exp\{\ \alpha_\theta(j) + \beta(j, b, 1_b) + v_\theta(j, f_d(j, b, 1_b))\ \}} \\ &\times \frac{\exp\{\ \alpha_\theta(y_3) + \beta(y_3, a, 1_a) + v_\theta(y_3, f_d(y_3, a, 1_a))\ \}}{\sum_{j\in\mathcal{Y}} \exp\{\ \alpha_\theta(j) + \beta(j, a, 1_a) + v_\theta(j, f_d(j, a, 1_a))\ \}} \end{aligned} \tag{25}$$

It is clear that the denominators are the same. Under condition (iii), we have that $f_d(y_3, a, 1_a) = f_d(y_3, b, 1_b) = 1\{y_3 \neq 0\}$. Thus, the ratio of these probabilities is:

$$\begin{aligned} \frac{\mathbb{P}(A|\theta)}{\mathbb{P}(B|\theta)} &= \frac{\begin{array}{c}\exp\{\ \alpha_\theta(a) + \beta(a, y_0, d_1) + v_\theta(a, 1_a)\ \}\ \exp\{\ \alpha_\theta(b) + \beta(b, a, 1_a) + v_\theta(b, 1_b)\ \} \\ \exp\{\ \alpha_\theta(y_3) + \beta(y_3, b, 1_b) + v_\theta(y_3, 1_{y_3})\ \}\end{array}}{\begin{array}{c}\exp\{\ \alpha_\theta(b) + \beta(b, y_0, d_1) + v_\theta(b, 1_b)\ \}\ \exp\{\ \alpha_\theta(a) + \beta(a, b, 1_b) + v_\theta(a, 1_a)\ \} \\ \exp\{\ \alpha_\theta(y_3) + \beta(y_3, a, 1_a) + v_\theta(y_3, 1_{y_3})\ \}\end{array}} \\ &= \exp\{\beta(a, y_0, d_1) - \beta(b, y_0, d_1) + \beta(b, a, 1_a) - \beta(a, b, 1_b) + \beta(y_3, b, 1_b) - \beta(y_3, a, 1_a)\} \end{aligned} \tag{26}$$

Now, remember the additive structure of function $\beta(y, y_{-1}, d) = 1\{y = y_{-1}\}\ \beta_d(y, d) + 1\{y \neq y_{-1}\}\ \beta_y(y, y_{-1})$, with $\beta(0, y_{-1}, d) = 0$, $\beta_y(y, y) = 0$, and $\beta_d(y, 0) = 0$. Given condition (i), $[a \neq b]$, we have that $\beta(b, a, 1_a) = \beta_y(b, a)$ and $\beta(a, b, 1_b) = \beta_y(a, b)$. Given condition (ii), [if $y_0 \neq 0$, then $a \neq y_0$ and $b \neq y_0$], we have that $\beta(a, y_0, d_1) = \beta_y(a, y_0)$ and $\beta(b, y_0, d_1) = \beta_y(b, y_0)$. Finally, given condition (iii), [either $y_3 = 0$ or $\{y_3 \neq a$ and $y_3 \neq b\}$], we have that $\beta(y_3, a, 1_a) = \beta_y(y_3, a)$ and

$\beta(y_3, b, 1_b) = \beta_y(y_3, b)$. Therefore,

$$\ln\left[\frac{\mathbb{P}(A)}{\mathbb{P}(B)}\right] = \Delta_y(y_0 \to y_3; a, b) \equiv [\beta_y(a, y_0) + \beta_y(b, a) + \beta_y(y_3, b)] - [\beta_y(b, y_0) + \beta_y(a, b) + \beta_y(y_3, a)] \tag{27}$$

We can obtain a root-N consistent estimator of $\Delta_y(y_0 \to y_3; a, b)$ by using the logarithm of the ratio between the sample frequencies of choice paths A and B. ■

Remark 1.1. [Trinomial Choice Model]. In the trinomial model ($J = 2$), suppose that $d_1 < d^*$, and consider the paths $A = \{0,\ 1,\ 2,\ 0\}$ and $B = \{0,\ 2,\ 1,\ 0\}$. These paths satisfy conditions (i) to (iii) in Proposition 1. From the sample frequencies of these histories we can identify the parameter:

$$\Delta_y(0 \to 0; 1, 2) = [\beta_y(1, 0) + \beta_y(2, 1) + \beta_y(0, 2)] - [\beta_y(2, 0) + \beta_y(1, 2) + \beta_y(0, 1)] \tag{28}$$

It is simple to verify that $\Delta_y(1 \to 1; 0, 2)$ and $\Delta_y(2 \to 2; 0, 1)$ are also identified and they should be equal to $\Delta_y(0 \to 0; 1, 2)$. Therefore, the model implies over-identifying restrictions. Consider the pair of paths $A = \{0,\ 0,\ 1,\ 2\}$ and $B = \{0,\ 1,\ 0,\ 2\}$. The frequencies of these paths identify the parameter:

$$\Delta_y(0 \to 2; 0, 1) = \beta_y(2, 1) - [\beta_y(0, 1) + \beta_y(2, 0)] \tag{29}$$

Similarly, we have that $\Delta_y(0 \to 1; 0, 2) = \beta_y(1, 2) - [\beta_y(0, 2) + \beta_y(1, 0)]$ is identified.

Remark 1.2. [Identification of parametric switching cost function]. Consider the Trinomial model above, and suppose that the switching cost function has the following parametric specification: $\beta_y(y_t, y_{t-1}) = \beta_1 |y_t - y_{t-1}| + \beta_2 (y_t - y_{t-1})^2$, where β_1 and β_2 are parameters. Given this specification and taking into account equation (29), we have that $\Delta_y(0 \to 2; 0, 1) = -2\beta_1 - 4\beta_2$. Furthermore, with $d_1 = d^*$, $\Delta_y(1 \to 2; 0, 1)$ is identified and this parameter is equal to $-2\beta_2$. Therefore, $\Delta_y(0 \to 2; 0, 1)$ and $\Delta_y(1 \to 2; 0, 1)$, together, can identify the parameters β_1 and β_2 in the switching cost function.

Remark 1.3. Using our notation, a dynamic structural model without duration dependence is equivalent assuming that $d^* = 0$. Under this restriction, condition (ii) in Proposition 1 becomes irrelevant such that we can use more choice paths to identify switching costs.

Remark 1.4. The identification result in Proposition 1 is similar to the identification in a non-structural (i.e., not forward-looking or myopic) dynamic logit model (Honoré and Kyriazidou, 2000). There is a key difference though and it comes from condition (iii). In the myopic model condition (iii) is not necessary and we can have $y_3 \neq 0$ and $\{y_3 = a \text{ or } y_3 = b\}$. This is possible in the myopic model because there are not continuation values and we should not have to control these values.

Interestingly, this difference provides a test of myopic versus forward-looking behavior that we present in section 3. The test is based on the following idea: in a myopic model we can estimate consistently the parameters β_y exploiting choice paths that do not satisfy condition (iii). In contrast, if individuals are forward-looking, using choice that do not satisfy condition (iii) generates inconsistent estimates of the same parameters. We use this idea to construct a Hausman test of the null hypothesis of myopic behavior.

Remark 1.5. For notational simplicity, in Proposition 1 we have considered that the choice paths A and B occur at periods $t = 0, 1, 2, 3$ in the sample. However, we can construct these choice paths using any four consecutive periods within the sample. Also, we have limited the discussion to choice paths with four periods (more precisely, three periods and an initial condition). It is possible to consider longer choice paths that have identification power.

Proposition 5 provides identification conditions for the duration dependence structural parameters $\beta^d(y, d)$.

PROPOSITION 5. Suppose that $T \geq d^* + 2$, *the conditions in Assumption 1 hold, and* $\mathbb{P}(\mathbf{z}_1 = \mathbf{z}_2 = ... = \mathbf{z}_{,d^*+2}) > 0$. *Suppose that the initial condition is* $\mathbf{x}_1 = (y, d^* - 1)$, *where* $y \in \mathcal{Y}$ *is an arbitrary choice. Let* $y\prime \neq y$ *be a different choice. Consider choice paths A and B, between periods* $t = 1$ *and* $t = d^* + 2$, *such that alternative y is chosen at every period expect at one period where this alternative is replaced by* $y\prime$: *for path A this replacement occurs at* $t = 1$, *and for path B it occurs at* $t = 2$, *i.e.,* $A = \{y\prime,\ y_t = y \text{ for } t = 2, ..., d^* + 2\}$, *and* $B = \{y,\ y\prime,\ y_t = y \text{ for } t = 3, ..., d^* + 2\}$. *Define* $\widetilde{\mathbf{y}} \equiv (y_0, y_1, ..., y_{d^*+2})$ *and let* $s \in \{0, 1\}$ *be the binary statistic defined as,*

$$s = 1\left\{\mathbf{z}_t = \mathbf{z} \text{ for } t = 1, 2, ..., d^* + 2;\ \mathbf{x}_1 = (y, d^* - 1);\ \widetilde{\mathbf{y}} \in A \cup B\right\} \tag{30}$$

Then, (1) the statistic s is sufficient for θ *such that the probability* $\mathbb{P}(\widetilde{\mathbf{y}} \mid s = 1, \theta)$ *does not depend on* θ; *and (2)* $\ln\mathbb{P}(A \mid s = 1) - \ln\mathbb{P}(B \mid s = 1) = \Delta_d(y)$, *and*

$$\Delta_d(y) \equiv \beta_d(y, d^*) - \beta_d(y, d^* - 1) \tag{31}$$

that represents the marginal return to duration at $d_t = d^*$. *This parameter is identified and can be estimated consistently at a rate root-N as N goes to infinity and T is fixed.* ■

Proof: Conditional on $s = 1$, there are only two possible choice paths, A or B. Therefore, we need to show that $\mathbb{P}(A|\theta)\ /\ \mathbb{P}(B|\theta)$ does not depend on θ. Under choice histories A and B, the paths of the endogenous state variables between period $t = 1$ and $t = d^* + 2$ are:

$$\begin{aligned} \widetilde{\mathbf{x}}_A &= \left\{ \begin{bmatrix} y \\ d^*-1 \end{bmatrix}, \begin{bmatrix} y\prime \\ 1 \end{bmatrix}, \begin{bmatrix} y \\ 1 \end{bmatrix}, ..., \begin{bmatrix} y \\ d^*-1 \end{bmatrix}, \begin{bmatrix} y \\ d^* \end{bmatrix} \right\} \\ \widetilde{\mathbf{x}}_B &= \left\{ \begin{bmatrix} y \\ d^*-1 \end{bmatrix}, \begin{bmatrix} y \\ d^* \end{bmatrix}, \begin{bmatrix} y\prime \\ 1 \end{bmatrix}, ..., \begin{bmatrix} y \\ d^*-2 \end{bmatrix}, \begin{bmatrix} y \\ d^*-1 \end{bmatrix} \right\} \end{aligned} \tag{32}$$

Note that the two histories visit the same states and the same number of times. Therefore, the denominators in the expressions for $\mathbb{P}(A|\theta)$ and $\mathbb{P}(B|\theta)$ are the same and they cancel in the ratio of probabilities $\mathbb{P}(A|\theta) \;/\; \mathbb{P}(B|\theta)$. The ratio of probabilities becomes:

$$\begin{aligned} \frac{\mathbb{P}(A|\theta)}{\mathbb{P}(B|\theta)} &= \frac{\exp\{\ (d^*+1)\ \alpha_\theta(y) +\ \alpha_\theta(y\prime) + \beta_y(y\prime, y) + \beta_y(y, y\prime)\ \}}{\exp\{\ (d^*+1)\ \alpha_\theta(y) +\ \alpha_\theta(y\prime) + \beta_y(y\prime, y) + \beta_y(y, y\prime)\ \}} \\ &\times \frac{\exp\{\ 0 + 0 + \beta_d(y,1) + ... + \beta_d(y, d^*-1) + \beta_d(y, d^*)\ \}}{\exp\{\ \beta_d(y, d^*-1) + 0 + 0 + \beta_d(y,1) + ... + \beta_d(y, d^*-1)\ \}} \\ &\times \frac{\exp\{\ v_\theta(y\prime,1) + v_\theta(y,1) + v_\theta(y,2) + ... + v_\theta(y,d^*) + v_\theta(y,d^*)\ \ \}}{\exp\{\ v_\theta(y,d^*) + v_\theta(y\prime,1) + v_\theta(y,1) + ... + v_\theta(y,d^*-1) + v_\theta(y,d^*)\ \ \}} \\ &= \exp\{\ \beta_d(y,d^*) - \beta_d(y, d^*-1)\ \} \end{aligned} \tag{33}$$

We can obtain a root-N consistent estimator of $\Delta_d(y) \equiv \beta_d(y, d^*) - \beta_d(y, d^*-1)$ by using the logarithm of the ratio between the sample frequencies (over the N individuals) of choice histories type A and type B. ■

Remark 2.1. A key condition for the identification result in Proposition 2 is that durations greater than d^* have zero marginal effect on the payoff function. If this assumption does not hold, then the probability ratio between choice paths A and B depend on the difference of continuation values $v_\theta(y, d^*+1) - v_\theta(y, d^*)$. This term depends on the unobserved heterogeneity θ such that s is no longer a sufficient statistic for this unobserved heterogeneity. However, the dynamic structural models in empirical applications impose restrictions on the maximum number of periods with returns to experience.

Remark 2.2. Proposition 2 considers only choice paths of length $d^* + 2$. When the dataset is such that $T > d^* + 2$, we can use more pairs of choice paths to construct sufficient statistics. In general, for $T \geq d^* + 2$, we can construct all the following pairs of choice paths A and B such that $\ln \mathbb{P}(A|\theta) / \mathbb{P}(B|\theta) = \Delta_d(y)$. For any $j = 1, ..., [T - d^* - 1]$, we can construct choice paths A and B of length $d^* + 1 + j$ with an initial condition $\mathbf{x}_1 = (y, d^* - j)$ and with $A = \{y\prime,\ y_t = y \text{ for } t = 2, ..., d^* + 1 + j\}$ and $B = \{y,\ y\prime,\ y_t = y \text{ for } t = 3, ..., d^* + 1 + j\}$. All this pairs of choice paths can be used to estimate consistently the parameter $\Delta_d(y)$.

Remark 2.3. Proposition 2 establishes that in a dynamic logit model with forward looking individuals and a fixed-effects specification of unobserved heterogeneity, the only structural parameter related to duration dependence that is identified is $\Delta_d(y) \equiv \beta_d(y, d^*) - \beta_d(y, d^* - 1)$. This result contrasts with the identification of the myopic version of the fixed-effects dynamic logit model. In the myopic model there are not heterogeneous continuation values that we should deal with.

Consider the myopic model. The initial condition is $\mathbf{x}_1 = (y, d_1)$, where $y \in Y$ is an arbitrary choice alternative and the duration d_1 is restricted to $T \geq d_1 + 3$. Let $y\prime \neq y$. Define the pair of choice paths A and B between $t = 1$ and $t = d_1 + 3$, with $A = \{y\prime,\ y_t = y \text{ for } t = 2, ..., d_1 + 3\}$ and $B = \{y,\ y\prime,\ y_t = y \text{ for } t = 3, ..., d_1 + 3\}$. These choice paths visit the same values of the endogenous state variables and with the same frequency. Therefore,

$$\begin{aligned} \frac{\mathbb{P}(A|\theta)}{\mathbb{P}(B|\theta)} &= \frac{\exp\{\ (d_1+2)\ \alpha_\theta(y) +\ \alpha_\theta(y\prime) + \beta_y(y\prime, y) + \beta_y(y, y\prime)\ \}}{\exp\{\ (d_1+2)\ \alpha_\theta(y) +\ \alpha_\theta(y\prime) + \beta_y(y\prime, y) + \beta_y(y, y\prime)\ \}} \\ &\times \frac{\exp\{\ 0 + 0 + \beta_d(y,1) + ... + \beta_d(y, d_1) + \beta_d(y, d_1+1)\ \}}{\exp\{\ \beta_d(y, d_1) + 0 + 0 + \beta_d(y,1) + ... + \beta_d(y, d_1)\ \}} \\ &= \exp\{\ \beta_d(y, d_1+1) - \beta_d(y, d_1)\ \} \end{aligned} \tag{34}$$

This implies that the whole function $\beta_d(y, d_1)$, for any $y \in Y$ and any $d_1 \leq \min\{d^*, T-3\}$, is identified. Define $\Delta_d^{myopic}(y, d_1) \equiv \beta_d(y, d_1+1) - \beta_d(y, d_1)$, that as shown above is identified for $d_1 \leq T - 3$. For $d_1 = 0$, given the normalization $\beta_d(y, 0) = 0$, we have that $\beta_d(y, 1) = \Delta_d^{myopic}(y, 1)$. Then, $\beta_d(y, 2) = \Delta_d^{myopic}(y, 1) + \Delta_d^{myopic}(y, 2)$, and so on $\beta_d(y, d_1) = \sum_{j=1}^{d_1} \Delta_d^{myopic}(y, j)$.

We see that, in a fixed-effects dynamic logit model, allowing for forward looking behavior has a cost in terms of identification of duration dependence.

4 Estimation and Inference

TBW

5 Empirical Applications

5.1 Bus replacement (Rust, 1987)

Here we revisit the model and data in the seminal article by Rust (1987). The model belongs to the class of *machine replacement models* that we have briefly described in section 2. The superintendent of maintenance at the Madison (Wisconsin) Metropolitan Bus Company has a fleet of N buses indexed by i. For every bus i and at every period t, the superintendent decides whether to keep the bus engine ($y_{it} = 1$) or to replace it ($y_{it} = 0$). In Rust's model, if the engine is replaced, the payoff is equal to $-RC + \varepsilon_{it}(0)$, where RC is a parameter that represents the replacement cost. If manager decides to keep the engine, the payoff is equal to $-c_0 - c_1(m_{it}) + \varepsilon_{it}(1)$, where m_{it} is a state variable that represents the engine cumulative mileage, and $c_0 + c_1(m_{it})$ is the maintenance cost. We incorporate two modifications in this model. First, we replace cumulative mileage m_{it} with duration since last replacement. We show below that the two variables are highly correlated. Second, we allow for time-invariant unobserved heterogeneity in the replacement cost, RC_i, and in the constant term in the maintenance cost function, c_{0i}. Using the notation in our paper, we have the payoff function is $\alpha_i(0) + \varepsilon_{it}(0)$ if $y_{it} = 0$, and $\alpha_i(1) + \beta_d(d_{it}) + \varepsilon_{it}(1)$ if $y_{it} = 1$, where $\alpha_i(0) = -RC_i$, $\alpha_i(1) = -c_{0i}$, and $\beta_d(d_{it}) = -c_1(d_{it})$.

Rust's full sample contains a total of 124 buses that are classified in eight groups according to the bus size, and the engine manufacturer, model and year. For the estimation of the structural model, Rust focuses on groups 1 to 4 that account for 104 buses: 15 buses in group 1; 4 buses in group 2; 48 buses in group 3; and 37 buses in group 4. For each bus engine, the choice history in the data includes the actual initial condition of the engine, i.e., the first month where the engine was installed.

For these 104 buses, the empirical distribution of the number of engine replacements per bus is the following: 0 engine replacements for 45 buses; 1 replacement for 58 buses; and 2 replacements for 1 bus. For our fixed effects estimation of the structural parameters β_d, choice histories with zero replacement do not contain any useful information. Therefore, for this first step in the estimation of the model, we have only 59 buses or choice histories, and 60 complete spells until replacement. The results below are based on the 58 buses (and duration spells) with only one replacement. For our analysis, we consider that the frequency of the superintendent's decisions is at the annual level. We observe replacement decisions for these 58 buses over 10 years (117 months).

Table 2 presents the empirical distribution of choice histories, and the CMLE of the parameter $\beta_d(d^*) - \beta_d(d^* - 1)$ for different possible values of d^*. Remember that for this model we have the following identification result. For $d^* \geq 2$:

$$\beta_d(d^*) - \beta_d(d^* - 1) = \ln \mathbb{P}\left(\{0, \mathbf{1}_{d^*}\} \mid d_1 = d^* - 1\right) - \ln \mathbb{P}\left(\{1, 0, \mathbf{1}_{d^*-1}\} \mid d_1 = d^* - 1\right) \tag{35}$$

We can obtain estimates for d^* equal to 4 and 3 years. Under the hypothesis that $d^* = 4$, the estimate of $\beta_d(4) - \beta_d(3)$ is equal to -0.205 with standard error 0.295 such that this estimate is not significantly different to zero (p-value = 0.4871). Following the approach described in section 3 to identify the value of d^*, we proceed by decreasing d^* by one unit. Under the hypothesis that $d^* = 3$, the estimate of $\beta_d(3) - \beta_d(2)$ is equal to -1.070 with standard error 0.121. This estimate is significantly different to zero with a p-value very close to zero. Note that $\beta_d(d)$ represents maintenance costs with negative value. Therefore, based on these estimates, we can conclude that the marginal maintenance costs of a 3 years old bus is significantly larger than the cost of a two-years old bus, but this marginal cost is insignificant after 3 years.

Table 2
Bus Engine Replacement (Rust, 1987)

Empirical Distribution of Choice Histories

Choice history	*Frequency* *Absolute*	*%*	*% cumulative*
1101111111	3	5.17	5.17
1110111111	11	18.96	24.13
1111011111	9	15.51	39.64
1111101111	18	31.03	70.67
1111110111	7	12.07	82.74
1111111011	5	8.62	91.36
1111111101	3	5.17	96.53
1111111110	2	3.45	100.00

CML Estimate of $\beta_d(\mathbf{d}^*) - \beta_d(\mathbf{d}^* - \mathbf{1})$

Parameter $\beta_d(d^*) - \beta_d(d^* - 1)$	*Estimate (s.e.) [p-value]*
with $d^* = 4$	-0.205 (0.295) *[0.4871]*
with $d^* = 3$	-1.070*** (0.121) *[0.0000]*

5.2 Demand of differentiated storable product [Preliminary]

- Consumer scanner data (A.C. Nielsen) on ketchup purchases.
  - Same dataset as in Pesendorfer (1998) and Erdem, Imai and Keane (2003), among others.
  - 2797 households over 123 weeks.
  - Three national brands (Heinz, Hunt's and Del Monte), and one store brand; these are choices $y = 1, 2, 3, 4$.
  - Outside option, $y = 0$, means "No purchase".
  - Duration since last purchase represents inventory depletion.
  - A consumer's choice sequence could look like: {1, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 1, 0, 0, ...}.

Table 3
Switching costs parameters (under symmetry)
Estimate (s.e.)

	Heinz	*Hunts*	*Del Monte*	*Store*
Heinz	-	1.052** (0.122)	1.711** (0.103)	2.199** (0.102)
Hunts		-	0.635 (0.099)	1.225** (0.095)
Del Monte			-	1.016** (0.092)
Store				-

Table 4
Structural duration dependence

Parameter $\beta_d(d^*) - \beta_d(d^* - 1)$	*Estimate*	*s.e.*
...		
with $d^* = 16$ weeks	-0.025	0.128
with $d^* = 15$ weeks	-0.124	0.151
with $d^* = 14$ weeks	-0.287	0.149
with $d^* = 13$ weeks	-0.304	0.159
with $d^* = 12$ weeks	**-0.516****	**0.136**

6 Conclusions

TO BE WRITTEN

References

[1] Aguirregabiria, V. (1999): "The dynamics of markups and inventories in retailing firms," *The Review of Economic Studies*, 66, 275-308.

[2] Aguirregabiria, V. (2010): "Another look at the identification of dynamic discrete decision processes: An application to retirement behavior," *Journal of Business & Economic Statistics*, 28(2), 201-218.

[3] Aguirregabiria, V. and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1-53.

[4] Andersen, E (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," *Journal of the Royal Statistical Society, Series B*, 32, 283 301.

[5] Arcidiacono, P., and R. Miller (2011): "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 79(6), 1823-1867.

[6] Arellano, M., and S. Bonhomme (2012): "Nonlinear Panel Data Analysis", *Annual Review of Economics*, 3, 395-424.

[7] Arellano, M., and S. Bonhomme (2017): "Nonlinear Panel Data Methods for Dynamic Heterogeneous Agent Models, *Annual Review of Economics*, forthcoming.

[8] Arellano, M., and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in J. J. Heckman and E. Leamer (eds.) *Handbook of Econometrics*, Volume 5, Chapter 53, North-Holland, 3229-3296.

[9] Blackwell, D. (1965): "Discounted dynamic programming," *The Annals of Mathematical Statistics*, 36(1), 226-235.

[10] Caliendo, L., M. Dvorkin, and F. Parro (2015): "Trade and Labor Market Dynamics," NBER WP No. 21149.

[11] Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47(1), 225-238.

[12] 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.

[13] Chamberlain, G. (1993): "Feedback in Panel Data Models," unpublished manuscript, No. 1656. Harvard-Institute of Economic Research.

[14] Chamberlain, G. (2010): “Binary response models for panel data: Identification and information,” *Econometrica*, 78(1), 159-168.

[15] Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81(2), 535-580.

[16] Chintagunta, P., E. Kyriazidou, and P. Perktold (2001): “Panel Data Analysis of Household Brand Choices,” *Journal of Econometrics*, 103(1), 111-153.

[17] Das, M. (1992): "A Micro-econometric Model of Capital Utilization and Retirement: The Case of the Cement Industry," *Review of Economic Studies*, 59, 277-297.

[18] Dunne, T., S. Klimek, M. Roberts, and D. Xu (2013): "Entry, Exit and the Determinants of Market Structure," *RAND Journal of Economics*, 44(3), 462-487.

[19] Erdem, T., S. Imai and M. P. Keane (2003): "Brand and Quantity Choice Dynamics under Price Uncertainty," *Quantitative Marketing and Economics*, 1, 5-64.

[20] Erdem, T., M. Keane, and B. Sun (2008): "A dynamic model of brand choice when price and advertising signal product quality," *Marketing Science*, 27(6), 1111-1125.

[21] Ericson, R. and A. Pakes (1995): "Markov Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62, 53-82.

[22] Frederiksen, A., B. Honoré, B., and L. Hu, L. (2007): "Discrete time duration models with group-level heterogeneity," *Journal of Econometrics*, 141(2), 1014-1043.

[23] Hahn, J.,and W. Newey (2004): "Jacknife and analytical bias reduction for nonlinear panel data models," *Econometrica*, 72(4), 1295-1319.

[24] Hahn, J. anf G. Kuersteiner (2011): "Bias reduction for dynamic nonlinear panel models with fixed effects," *Econometric Theory*, 27(6), 1152-1191.

[25] 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.

[26] Hendel, I., and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior. Econometrica 74, 1637-1674.

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

[28] Honoré, B. , and E. Kyriazidou (2017): "Panel Vector Autoregressions with Binary Data," manuscript. Princeton University.

[29] Kalouptsidi, M. (2014): "Time to build and fluctuations in bulk shipping," *American Economic Review*, 104(2), 564-608.

[30] Kalouptsidi, M., P. Scott, and E. Souza-Rodrigues (2015): "Identification of counterfactuals in dynamic discrete choice models," NBER working paper, No. 21527.

[31] Kano, K. (2013): "Menu Costs and Dynamic Duopoly," *International Journal of Industrial Organization*, 31(1), 102-118.

[32] Kasahara, H. (2009): "Temporary Increases in Tariffs and Investment: The Chilean Case," Journal of Business and Economic Statistics, 27(1), 113-127.

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

[34] Keane, M. and K. Wolpin (1997): "The career decisions of young men," *Journal of Political Economy*, 105, 473-522.

[35] Kennet, M. (1993) "Did Deregulation Affect Aircraft Engine Maintenance? An Empirical Policy Analysis," *The RAND Journal of Economics,* 24, 542-558.

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

[37] Magnac, T. (2004): "Panel Binary Variables and Sufficiency: Generalizing Conditional Logit," *Econometrica*, 72(6), 1859-1876.

[38] Magnac, T., and D. Thesmar (2002): "Identifying dynamic discrete decision processes," *Econometrica*, 70(2), 801-816.

[39] Manski, C. (1987): "Semiparametric analysis of random effects linear models from binary panel data," *Econometrica*, 55(2), 357-362.

[40] Miller, R. (1984): "Job Matching and Occupational Choice", *Journal of Political Economy*, 92, 1086-1120.

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

[42] Roberts, M. and J. Tybout (1997): "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87(4), 545-564.

[43] Rust, J. (1987): “Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher,” *Econometrica,* 55, 999-1033.

[44] Rust, J. (1994): “Structural estimation of Markov decision processes,” in R. E. Engle and McFadden (eds.) *Handbook of Econometrics Volume 4*, North-Holland. Amsterdam.

[45] Ryan, S. (2013) "The costs of environmental regulation in a concentrated industry," *Econometrica,* 80(3), 1019-1061.

[46] Slade, M. (1998): “Optimal Pricing with Costly Adjustment: Evidence from Retail Grocery Stores”, *Review of Economic Studies*, 65, 87-108.

[47] Sweeting, A. (2013): "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry," *Econometrica*, 81(5), 1763-1803.

[48] Willis, J. (2006): "Magazine prices revisited," *Journal of Applied Econometrics*, 21(3), 337-344.