Identifying Dynamic Games with Switching Costs*

Fabio A. Miessi Sanches†
University of São Paulo

Daniel Silva Junior‡
London School of Economics

Sorawoot Srisuma§
University of Surrey

February 14, 2015

Preliminary and Incomplete

Abstract

Most theoretical identification results for dynamic games with discrete choice focus on the (entire) payoff functions while taking other primitives as known. In practice, however, empirical researchers are often concerned about numerical costs and, when possible, use economic theory to reduce the dimensionality of the payoff functions to be estimated by dynamic game methods that are considered computationally expensive. Switching costs, such as entry costs, are recurring components of the payoffs seen in numerous empirical games modeled in practice. We show how natural exclusion restrictions that define switching costs can be exploited to obtain new identification results. Our identification strategy can be used to construct estimators that are simpler to compute and more robust than previously.

JEL Classification Numbers: C14, C25, C61

Keywords: Discrete Choice Games, Identification, Markovian Games.

*We are grateful Oliver Linton and Martin Pesendorfer for their advice and support. We also thank Kirill Evdokimov, Emmanuel Guerre, Arthur Lewbel, Aureo de Paula, Adam Rosen, Yuya Sasaki, Richard Spady and seminar participants at Johns Hopkins University, Queen Mary University of London, University of Cambridge, University of São Paulo, University of Southampton, University of Surrey, Bristol Econometrics Study Group, AMES (Taipei), EMES (Toulouse) and IAAE (London) for comments.

† *E-mail address*: fmiessi@gmail.com

‡ *E-mail address*: d.silva-junior@lse.ac.uk

§ *E-mail address*: s.srisuma@surrey.ac.uk

1 Introduction

The study of nonparametric identification in structural models is fundamentally important. It informs us whether or not the model under consideration can be consistently estimated from an ideal data set without introducing additional parametric or other restrictions. The model of interest in this paper is a class of dynamic discrete choice games that generalizes the single agent Markov decision models (Rust (1994)). Dynamic games provide a useful framework to study counterfactual experiments involving multiple economic agents making decisions over time.[1] Recent reviews of the identification and estimation of these games, and other related issues such as computational aspects, can be found in Aguirregabiria and Mira (2010) and Bajari, Hong and Nekipelov (2012). The primitives of the games we consider consist of players' payoff functions, discount factor, and Markov transition law of the variables in the model.

Most nonparametric identification results in this literature, following Magnac and Thesmar (2002), focus on identifying the payoff functions while taking other primitives of the model as known (Bajari, Chernozhukov, Hong and Nekipelov (2009), Pesendorfer and Schmidt-Dengler (2008)); also see Section 6 in Bajari, Hong and Nekipelov (2012).[2] These authors show that payoffs are generally not identified nonparametrically. They are underidentified. Positive identification results are typically obtained by imposing generic linear restrictions on the payoffs (such as equality and exclusion restrictions). The identification strategy along the line of Magnac and Thesmar is also constructive, and is related to the development of several general estimation methodologies.[3]

A common feature in the aforementioned works (on identification) aims to identify the entire payoff function for each player. However, the estimation strategies often employed in empirical work do not treat all components of the payoff function in the same way. In particular the estimation of dynamic games is considered a numerically demanding task, and the computational cost generally increases nontrivially with the cardinality of the state space as well as number of parameters to be estimated. Therefore, in the spirit of structural modeling, when possible empirical researchers use economic theory to estimate components of the payoff function directly without appealing to estimators developed specifically for dynamic games. In other words, some components of the payoff

[1] Examples of empirical applications include: Aguirregabiria and Mira (2007), Beresteanu, Ellickson and Misra (2010), Collard-Wexler (2013), Dunne, Klimek, Roberts and Xu (2013), Fan and Xiao (2012), Gowrisankaran, Lucarelli, Schmidt-Dengler and Town (2010), Lin (2012), Pesendorfer and Schmidt-Dengler (2003), Sanches, Silva Jr and Srisuma (2014), and Suzuki (2013).

[2] A notable exception is Norets and Tang (2012), who show in a single agent setting that without the distribution of the private values, generally payoff functions can only be partially identified.

[3] Examples of estimators in the literature include Aguirregabiria and Mira (2007), Bajari, Benkard and Levin (2007), Bajari et al. (2009), Pakes, Ostrovsky and Berry (2007), Pesendorfer and Schmidt-Dengler (2008), and Sanches, Silva and Srisuma (2013).

functions are treated as reduced forms; they are *structurally identified*.[4]

A recurring feature of many dynamic game models employed in practice involves costs that arise from players choosing different actions from the previous period (e.g. entry cost); see examples in footnote 1. Switching costs are (at least, part of) what sometimes called dynamic parameters of the model as they generally cannot be treated as reduced forms since economic theory rarely provides guidance on how they are determined. Dynamic parameters are typically estimated using dynamic game methods. Crucially, by definition, switching costs impose natural exclusion restrictions on the payoff functions. This paper explores how natural economic restrictions from switching costs can be exploited to improve the inference of dynamic games.

We show that, subject to a testable conditional independence assumption, and some normalization, switching costs can generally be nonparametrically identified independently of the discount factor and other components of the payoffs. Our identification strategy is also constructive and leads to a more robust and simpler to construct estimator than previously. In order to be more explicit about our contribution it will be helpful to introduce the main assumptions from the onset. In particular, let $\pi_i(a_{it}, a_{-it}, x_t, w_t)$ denote the per period payoff for player i at time t, where a_{it}, a_{-it}, x_t and w_t denote her own action choice, actions of other players, observed state variables and actions from the previous period respectively. We consider a payoff function that admits the following decomposition:

$$\pi_i(a_{it}, a_{-it}, x_t, w_t) = \mu_i(a_{it}, a_{-it}, x_t) + \phi_i(a_{it}, x_t, w_t; \eta_i) \cdot \eta_i(a_{it}, x_t, w_t). \tag{1}$$

We offer one economic interpretation for the above equation as follows. μ_i captures the static payoff from each period's competition or participation from the game. ϕ_i represents player's specific switching cost function. η_i is a *known* function that indicates whether a switch occurs (its purpose is solely to determine the domain of ϕ_i, hence the notation $\phi_i(;\eta_i)$). The key exclusion restrictions are: (i) past actions do not directly affect static payoff (w_t does not enter μ_i); and, (ii) only player i's own action determines whether a switching cost is incurred (a_{-it} does not enter ϕ_i and η_i). Equation (1) encompasses numerous payoff functions employed in practice. Some detailed examples will be given below. In addition, the testable independence assumption we require is that x_{t+1} is independent of w_t conditional on x_t and a_t. Such independence assumption is also implicitly assumed in numerous empirical work.

We provide conditions when ϕ_i can be identified independent of μ_i and the discount factor, denoted by β, and written in *closed-form* in terms of the transition and conditional choice probabilities

[4] For example, in an empirical model of an oligopolistic competition, firms' data on prices and quantities can be used to construct the variable profits by building a demand system and solving a particular model of competition (see Berry and Haile (2010,2012)). Another example is when bids data are available and the auction format is known so the expected revenue can be estimated nonparametrically (Athey and Haile (2002,2007), Guerre, Perrigne and Vuong (2000)).

observed from the data. The implication of our results depends on the empirical problem at hand and data availability.

1. The best case scenario arises when μ_i can be (structurally) identified directly from the data. Our result on ϕ_i implies that π_i can be identified independently of the discount factor. In this case we also give a condition to identify the discount factor.

2. Otherwise the identification of μ_i will rely on existing methods in the literature, particularly also assuming β, where the knowledge of ϕ_i can be used to reduce the dimensionality of the nonparametric components in π_i.

Our identification results can also be used directly to construct estimators. The numerical aspect of estimating dynamic games can present a non-trivial challenge in practice (e.g. see Egesdal, Lai and Su (2014) and Sanches, Silva and Srisuma (2013) for recent discussions). We propose a simple estimator for ϕ_i that is invariant to the value of the discount factor and any specification of μ_i; based on the closed-form expression of ϕ_i. Furthermore, if μ_i is also estimable directly from the data then we can estimate π_i independently of β without relying on existing methods to estimate games. In any case, we offer a practical way to simplify the computational problem and reduce some sensitivity to the specification of the payoff function.

The discount factor is a primitive of the model that is generally assumed to be known for the purpose of identification in dynamic games. Perhaps relatedly, most estimation methodologies and empirical applications in the literature do not estimate β but assign a fixed value for it. We provide a sufficient condition to identify the discount factor when μ_i can be identified independently of β. Our result shares some similarities with Proposition 4 in Magnac and Thesmar (2002), who give conditions for a positive identification result of the discount factor in a two-period model with a single decision maker when the period payoff function satisfies a particular exclusion restriction.[5] Our identification result for the discount factor is also constructive and can lead to a natural estimator.

From the mathematical standpoint, our identification strategy (for ϕ_i) also differs from the earlier results. The insight of Magnac and Thesmar (2002) reduces the identification problem to whether the normalized expected payoffs identified from the data (Hotz and Miller (1993)) can be uniquely generated by the model primitives. When all primitives apart from π_i are known, the expected payoffs can be written as a linear transform of the payoffs so the condition for identification is equivalent to whether some linear equation in π_i has a unique solution. However, if β is also part of the unknown terms, then the expected payoffs are no longer linear in these primitives. We provide conditions where a linear structure can be restored for the switching costs and used for identification.

[5]See Restriction R in Magnac and Thesmar (2002, page 809).

The decomposition of payoff functions and imposing nonparametric structures have also been used to identify other structural microeconomic models, e.g. see Berry and Haile (2010,2012) and Lewbel and Tang (2013). The only other paper we are aware of that considers identification while decomposing the payoff function for dynamic games is the recent work of Aguirregabiria and Suzuki (2013) on entry games. However, the content and motivation of our work and theirs are substantially different.[6] Their main concern is on identifying and interpreting certain counterfactuals for the purpose of policy analysis, rather than identifying and estimating the model primitives.

Throughout this paper we assume the most basic setup of a game with independent private values under the usual conditional independence, and we anticipate the data to have been generated from a single equilibrium.[7] Our results can be extended to games with unobserved heterogeneity, which has been used to accommodate a simple form of multiple equilibria, as long as nonparametric choice and transition probabilities can be identified (see Aguirregabiria and Mira (2007), Kasahara and Shimotsu (2009), Hu and Shum (2012)). The research on how to perform inference on a more general data structure is an important area of future research, which is outside the scope of our work.

The remainder of the paper is organized as follows. Section 2 illustrates the mathematical idea behind our identification strategy of the switching costs using a simple two-player entry game in Pesendorfer and Schmidt-Dengler (2003,2008). We define the theoretical model and modeling assumptions in Section 3. We give our identification results in Section 4. Section 5 provides a discussion on how our identification strategy can be used for estimation that we apply to data in Section 6. Section 7 concludes.

2 Preview of Identification Strategy

Consider a two-player repeated entry game in Pesendorfer and Schmidt-Dengler (2003,2008). At time t, each player i makes a decision, a_{it}, to play 1 (enter the market) or 0 (not enter) based on the status of market entrants from the previous period, $w_t = (a_{it-1}, a_{-it-1})$, and a private shock $\varepsilon_{it} = (\varepsilon_{it}(0), \varepsilon_{it}(1))$. The expected payoff from choosing action a_i is $v_i(a_i, w_t) + \varepsilon_{it}(a_i)$, where

$$\begin{aligned} v_i(a_i, w_t) &= E\left[\pi_i(a_i, a_{-it}, w_t)|w_t\right] + \beta E\left[m_i(w_{t+1})|\, w_t, a_{it} = a_i\right], \\ m_i(w_t) &= \sum_{\tau=0}^{\infty} \beta^{\tau} E\left[\left.\begin{array}{c} \pi_i(a_{it+\tau}, a_{-it+\tau}, w_{t+\tau}) \\ +\varepsilon_{it+\tau}(1)\cdot \mathbf{1}[a_{it+\tau}=1] + \varepsilon_{it+\tau}(0)\cdot \mathbf{1}[a_{it+\tau}=0] \end{array}\right| w_t\right]. \end{aligned} \tag{2}$$

[6] For example, Aguirregabiria and Suzuki (2013) assume the discount factor to be known throughout their paper; see their second paragraph of Section 3.1 (page 10).

[7] The test of Otsu, Pesendorfer and Takahashi (2014) can be used to detect multiple equilibria in the data.

In equilibrium $a_{it} = \alpha_i(w_t, \varepsilon_{it})$ for all i, t, where α_i denotes the player's Opt strategy, so that for any w_t, ε_{it}:

$$\alpha_i(w_t, \varepsilon_{it}) = \mathbf{1}[\Delta v_i(w_t) \geq \varepsilon_{it}(0) - \varepsilon_{it}(1)],$$

where $\Delta v_i(w_t) = v_i(1, w_t) - v_i(0, w_t)$. Given the distribution of ε_{it}, Δv_i can be recovered directly from the choice probabilities observable from the data. We can also relate Δv_i directly to the primitives from (2), as m_i can be written as some linear combination of π_i, where the linear scalar coefficients depend on the discount factor, conditional choice and transition probabilities; in particular, $E[\varepsilon_{it+\tau}(1) \cdot \mathbf{1}[a_{it+\tau} = 1] + \varepsilon_{it+\tau}(0) \cdot \mathbf{1}[a_{it+\tau} = 0] | w_t]$ can also be written in terms of choice probabilities (Hotz and Miller (1993)). Since the action space is finite, then we can summarize the relation between Δv_i and π_i, as identified by the data and implied by the model, by a matrix equation:

$$\mathbf{r}_i = \mathbf{T}_i \boldsymbol{\pi}_i, \tag{3}$$

where $\boldsymbol{\pi}_i$ is a vector of $\{\pi_i(a_i, a_{-i}, w)\}_{a_i, a_{-i}, w}$, and both $\mathbf{r}_i$ and $\mathbf{T}_i$ are known functions of β, and the conditional choice and transition probabilities. Following Magnac and Thesmar (2002), the expected discounted payoff represents the reduced form for this class of dynamic games. Bajari et al. (2009), Pesendorfer and Schmidt-Dengler (2008) then show the identifiability of π_i comes down to the ability to find a unique solution to equation (3), which can be written down in terms of rank conditions.

Now we impose more structure on π_i, in particular:

$$\pi_i(a_{it}, a_{-it}, w_t) = \mu_i(a_{it}, a_{-it}) + EC_i \cdot a_{it}(1 - a_{it-1}) + SV_i \cdot (1 - a_{it}) a_{it-1},$$

so that μ_i denotes the profit determines only by present period's actions (e.g. takes value zero if player i does not enter, otherwise it represents either a monopoly or duopoly profit depending on the number of players in the market), and $\theta_i = (EC_i, SV_i)$ consists of the switching costs parameters. From (2), it follows that

$$\begin{aligned} \Delta v_i(w_t) &= E[\mu_i(1, a_{-it}) + \beta m_i(1, a_{-it}) | w_t] + EC_i \cdot (1 - a_{it-1}) \\ &\quad - (E[\mu_i(0, a_{-it}) + \beta m_i(0, a_{-it}) | w_t] + SV_i \cdot a_{it-1}) \end{aligned}$$

Let $\Delta\mu_i(a_{-it}) = \mu_i(1, a_{-it}) - \mu_i(0, a_{-it})$, and define $\Delta m_i(a_{-it})$ similarly. Note that m_i itself also depends on β as well as π_i, therefore Δv_i is clearly not linear in (β, π_i). In order to restore the linearity in β, using $\Delta\mu_i$ and Δm_i, we define a nuisance function $\lambda_i(a_{-it}) = \Delta\mu_i(a_{-it}) + \beta\Delta m_i(a_{-it})$, so we can write

$$\Delta v_i(w_t) = E[\lambda_i(a_{-it}) | w_t] + EC_i \cdot (1 - a_{it-1}) - SV_i \cdot a_{it-1}. \tag{4}$$

By construction λ_i is a composite function consisting of *all* primitives in the model. However, the contribution of the entry cost from the present period is now additively separable from the other flow profits. Since the support of w_t is finite, which is $\{(0,0), (0,1), (1,0), (1,1)\}$, $\{\Delta v_i (w)\}_w$ can be represented by a matrix equation:

$$\Delta \mathbf{v}_i = \mathbf{Z}_i \boldsymbol{\lambda}_i + \mathbf{D}_i \theta_i, \text{ such that} \tag{5}$$

$$\begin{bmatrix} \Delta v_i ((0,0)) \\ \Delta v_i ((0,1)) \\ \Delta v_i ((1,0)) \\ \Delta v_i ((1,1)) \end{bmatrix} = \begin{bmatrix} P_{-i}(0|0,0) & P_{-i}(1|0,0) \\ P_{-i}(0|0,1) & P_{-i}(1|0,1) \\ P_{-i}(0|1,0) & P_{-i}(1|1,0) \\ P_{-i}(0|1,1) & P_{-i}(1|1,1) \end{bmatrix} \begin{bmatrix} \lambda_i (0) \\ \lambda_i (1) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} EC_i \\ SV_i \end{bmatrix},$$

where we use $P_i (a_i|w)$ to denote $\Pr [a_{it} = a_i | w_t = w]$.

Let $\mathbf{M}_{\mathbf{Z}_i}$ be a projection matrix whose null space is $\mathcal{CS}(\mathbf{Z}_i)$, and $\mathbf{D}_i = [\mathbf{d}_i^1 : \mathbf{d}_i^2]$. Note that the direction of projection does not matter. If $\mathbf{d}_i^k \notin \mathcal{CS}(\mathbf{Z}_i)$ then

$$\begin{aligned} EC_i &= \left(\mathbf{d}_i^{1\top} \mathbf{M}_{\mathbf{Z}_i} \mathbf{d}_i^1\right)^{-1} \mathbf{d}_i^{1\top} \mathbf{M}_{\mathbf{Z}_i} \left(\Delta \mathbf{v}_i - \mathbf{d}_i^2 SV_i\right), \\ SV_i &= \left(\mathbf{d}_i^{2\top} \mathbf{M}_{\mathbf{Z}_i} \mathbf{d}_i^2\right)^{-1} \mathbf{d}_i^{2\top} \mathbf{M}_{\mathbf{Z}_i} \left(\Delta \mathbf{v}_i - \mathbf{d}_i^1 EC_i\right). \end{aligned} \tag{6}$$

I.e., we can identify either the entry cost or scrap in terms of observables subjected to a normalization, in closed-form. The need to normalize in this context is not unfamiliar in empirical work. We delay a fuller discussion regarding normalization and other intuition in subsequent sections.

The sample counterparts of (6) provide a simple estimator for each θ_i^k that has a closed-form. However, such estimator is inefficient. More generally, a class of closed-form estimators for θ_i^k can be defined based on:

$$\begin{aligned} \mathbf{M}_{\mathbf{Z}_i} \Delta \mathbf{v}_i &= \mathbf{M}_{\mathbf{Z}_i} \mathbf{d}_i^k \theta_i^k, \text{ and} \\ \mathbf{S}\left(\theta_i^k; \mathbf{W}\right) &= \left(\mathbf{M}_{\mathbf{Z}_i} \Delta \mathbf{v}_i - \mathbf{M}_{\mathbf{Z}_i} \mathbf{d}_i^k \theta_i^k\right)^\top \mathbf{W} \left(\mathbf{M}_{\mathbf{Z}_i} \Delta \mathbf{v}_i - \mathbf{M}_{\mathbf{Z}_i} \mathbf{d}_i^k \theta_i^k\right), \\ \theta_i^k &= \left(\mathbf{d}_i^{k\top} \mathbf{M}_{\mathbf{Z}_i}^\top \mathbf{W} \mathbf{M}_{\mathbf{Z}_i} \mathbf{d}_i^k\right)^{-1} \mathbf{d}_i^{k\top} \mathbf{M}_{\mathbf{Z}_i}^\top \mathbf{W} \mathbf{M}_{\mathbf{Z}_i} \Delta \mathbf{v}_i, \end{aligned}$$

for some positive definite $\mathbf{W}$, as an *asymptotic least squares estimator* (Gourieroux and Monfort (1995)) in the spirit of Pesendorfer and Schmidt-Dengler (2008) and Sanches, Silva and Srisuma (2013). Then the asymptotic variance of such estimator is determined by the weighting matrix.

The constructive identification strategy above can be generalized considerably. Our results are applicable to non-entry games, for instance to games with multinomial actions (allocation or pricing problems, e.g. Marshall (2013)), or sequential decision problems (dynamic auction or investment games, e.g. Groeger (2013) and Ryan (2012)), as well as games with absorbing states (e.g. permanent market exit).

3 Model and Assumptions

We consider a game with I players, indexed by $i \in \mathcal{I} = \{1, \ldots, I\}$, over an infinite time horizon. The variables of the game in each period are action and state variables. The action set of each player is $A = \{0, 1, \ldots, K\}$. Let $a_t = (a_{1t}, \ldots, a_{It}) \in A^I$. We will also occasionally abuse the notation and write $a_t = (a_{it}, a_{-it})$ where $a_{-it} = (a_{1t}, \ldots, a_{i-1t}, a_{i+1t} \ldots, a_{It}) \in A^I$. Player i's information set is represented by the state variables $s_{it} \in S$, where $s_{it} = (x_t, w_t, \varepsilon_{it})$ such that $(x_t, w_t) \in X \times A^I$, for some compact set $X \subseteq \mathbb{R}^{d_X}$ and for simplicity we let $w_t = a_{t-1}$ and suppose x_t does not contain other past actions, are common knowledge to all players and $\varepsilon_{it} = (\varepsilon_{it}(0), \ldots, \varepsilon_{it}(K)) \in \mathbb{R}^{K+1}$ denotes private information only observed by player i. For notational simplicity, we delay the discussion of including lagged actions as state variables at the end of Section 4.1. We define $s_t = (x_t, w_t)$ and $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{It})$. Future states are uncertain. Players' actions and states today affect future states. The evolution of the states is summarized by a Markov transition law $P(s_{t+1}|s_t, a_t)$. Each player has a payoff function, $u_i : A^I \times S \to \mathbb{R}$, which is time separable. Future period's payoffs are discounted at the rate $\beta \in (0, 1)$.

The setup described above, and the following assumptions, which we shall assume throughout the paper, are standard in the modeling of dynamic discrete games.

Assumption M1 (Additive Separability): *For all $i, a_i, a_{-i}, x, w, \varepsilon_i$:*

$$u_i(a_i, a_{-i}, x, w, \varepsilon_i) = \pi_i(a_i, a_{-i}, x, w) + \sum_{a'_i \in A} \varepsilon_i(a'_i) \cdot \mathbf{1}[a_i = a'_i].$$

Assumption M2 (Conditional Independence I): *The transition distribution of the states has the following factorization for all $x', w', \varepsilon', x, w, \varepsilon, a$:*

$$P(x', w', \varepsilon'|x, w, \varepsilon, a) = Q(\varepsilon') G(x'|x, w, a),$$

where Q is the cumulative distribution function of ε_t and G denotes the transition law of x_{t+1} conditioning on x_t, w_t, a_t.

Assumption M3 (Independent Private Values): *The private information is independently distributed across players, and each is absolutely continuous with respect to the Lebesgue measure whose density is bounded on $\mathbb{R}^{K+1}$ with unbounded support.*

Assumption M4 (Discrete Public Values): *The support of x_t is finite so that $X = \{x^1, \ldots, x^J\}$ for some $J < \infty$.*

At time t every player observes s_{it}, each then chooses a_{it} simultaneously. We consider a Markovian framework where players' behaviors are stationary across time and players are assumed to play pure strategies. More specifically, for some $\alpha_i : S \to A$, $a_{it} = \alpha_i(s_{it})$ for all i, t, so that whenever $s_{it} = s_{i\tau}$ then $\alpha_i(s_{it}) = \alpha_i(s_{i\tau})$ for any τ. The beliefs are also time invariant. Player i's beliefs, σ_i, is a distribution of $a_t = (\alpha_1(s_{1t}), \ldots, \alpha_I(s_{It}))$ conditional on x_t for some pure Markov strategy profile $(\alpha_1, \ldots, \alpha_I)$. The decision problem for each player is to solve, for any s_i,

$$\begin{aligned} \max_{a_i \in \{0,1\}} \{E[u_i(a_{it}, a_{-it}, s_i)\,|s_{it} &= s_i, a_{it} = a_i] + \beta E[V_i(s_{it+1})\,|s_{it} = s_i, a_{it} = a_i]\}, \\ \text{where } V_i(s_i) &= \sum_{\tau=0}^{\infty} \beta^{\tau} E[u_i(a_{it+\tau}, a_{-it+\tau}, w_{t+\tau})\,|s_{it} = s_i]. \end{aligned} \tag{7}$$

The expectation operators in the display above integrate out variables with respect to the probability distribution induced by the equilibrium beliefs and Markov transition law. V_i denotes the value function. Note that the transition law for future states is completely determined by the primitives and the beliefs. Any strategy profile that solves the decision problems for all i and is consistent with the beliefs satisfies is an equilibrium strategy. Pure strategies Markov perfect equilibria have been shown to exist for such games (e.g. Aguirregabiria and Mira (2007), Pesendorfer and Schmidt-Dengler (2008)).

We consider identification based on the joint distribution of the observables, namely (a_t, x_t, w_t, x_{t+1}), consistent with a single equilibrium play. The primitives of the game under this setting are $(\{\pi_i\}_{i=1}^{I}, \beta, Q, G)$. Throughout the paper we shall also assume G and Q to be known (the former can be identified from the data). Thus far, our framework is familiar from the literature (e.g. Aguirregabiria and Mira (2007), Bajari, Benkard and Levin (2007), Pakes, Ostrovsky and Berry (2007), Pesendorfer and Schmidt-Dengler (2008)).

We now formally introduce the specific structures for the purpose of identification using past actions alluded in the introduction. In what follows, let $W_{\eta_i}^{d}(a_i, x) \equiv \{w \in A^I : \eta_i(a_i, x, w) = d\}$ for $d = 0, 1$. In addition to M1 - M4, we assume N1 - N2 hold for the remainder of this section.

Assumption N1 (Decomposition of Profits): *For all i, a_i, a_{-i}, x, w:*

$$\pi_i(a_i, a_{-i}, x, w) = \mu_i(a_i, a_{-i}, x) + \phi_i(a_i, x, w; \eta_i) \cdot \eta_i(a_i, x, w),$$

for some known function $\eta_i : A \times X \times A^I \to \{0, 1\}$ such that for any a_i, $\phi_i(a_i, x, w; \eta_i) = 0$ for all x when $w \in W_{\eta_i}^{0}(a_i, x)$.

Assumption N2 (Conditional Independence II): *The distribution of x_{t+1} conditional on a_t and x_t is independent of w_t.*

Assumption N1 assumes the period payoff function can be decomposed into two components with distinct exclusion restrictions; as alluded in the Introduction. First is μ_i that does not depend on w_t. Since η_i is a function chosen by the researcher that indicates a switching cost, we normalize ϕ_i to be zero whenever η_i takes value zero. When η_i takes value one, an exclusion restriction is imposed so that a_{-it} does not enter ϕ_i. Intuitively, N1 restricts us to consider payoffs that, for each player in any single time period, come from two separate sources: one comes from the interaction with the other players at the stage game, and the other is determined by her action relative to the previous period. This does not mean, however, that variables from the past cannot affect μ_i since x_t can contain lagged values, including past actions. Although our method relies on the restriction that other players' actions cannot contemporaneously affect ϕ_i when player i chooses a_{it}.

N2 imposes that knowing actions from the past does not help predict future state variables when the present actions are known. Note that N2 is not implied by M2. Therefore when x_t contains lagged actions N2 can be weakened to allow for dependence of other state variables with past actions. In any case, N2 is a restriction made on the observables so it can be tested directly from the data.

Both N1 and N2 are quite general and are implicitly assumed in many empirical models used in the literature. Here we provide detailed examples of $\phi_i \cdot \eta_i$ and $W^d_{\eta_i}(a_i, x)$.

Example 1 (Entry Cost): Suppose $K = 1$, then the switching cost at time t is

$$\phi_i\left(a_{it}, x_t, w_t; \eta_i\right) \cdot \eta_i(a_{it}, x_t, w_t) = EC_i\left(x_t, a_{-it-1}\right) \cdot a_{it}\left(1 - a_{it-1}\right).$$

So that $W^1_{\eta_i}(1, x) = \left\{w = (0, a_{-i}) : a_{-i} \in A^{I-1}\right\}$ and $W^0_{\eta_i}(1, x) = \left\{w = (1, a_{-i}) : a_{-i} \in A^{I-1}\right\}$, and $W^d_{\eta_i}(0, x) = \varnothing$ for all x.

Example 2 (Scrap Value): Suppose $K = 1$, then the switching cost at time t is

$$\phi_i\left(a_{it}, x_t, w_t; \eta_i\right) \cdot \eta_i(a_{it}, x_t, w_t) = SV_i\left(x_t, a_{-it-1}\right) \cdot \left(1 - a_{it}\right) a_{it-1}.$$

So that $W^d_{\eta_i}(1, x) = \varnothing$ and, $W^1_{\eta_i}(0, x) = \left\{w = (1, a_{-i}) : a_{-i} \in A^{I-1}\right\}$ and $W^0_{\eta_i}(0, x) = \left\{w = (0, a_{-i}) : a_{-i} \in A^{I}\right.$ for all x.

Example 3 (General Switching Costs): Suppose $K \geq 1$, then the switching cost at time t is

$$\phi_i\left(a_{it}, x_t, w_t; \eta_i\right) \cdot \eta_i(a_{it}, x_t, w_t) = \sum_{a_i', a_i'' \in A} SC_i\left(a_i', a_i'', x_t, a_{-it-1}\right) \cdot \mathbf{1}\left[a_{it} = a_i', a_{it-1} = a_i'', a_i' \neq a_i''\right].$$

So that, prior to imposing any normalizations, $W^1_{\eta_i}(a_i, x) = \left\{w = (a_i', a_{-i}) : a_i' \in A \backslash \{a_i\}, a_{-i} \in A^{I-1}\right\}$ and $W^0_{\eta_i}(a_i, x) = \left\{w = (a_i, a_{-i}) : a_{-i} \in A^{I-1}\right\}$ for all x.

Note that Examples 1 and 2 are just special cases of Example 3 when $K = 1$, with an additional normalization of zero scrap value and entry cost respectively.

In order to provide an intuitive explanation behind why the assumptions above are useful for identification, it will be useful to consider a single agent decision problem. Omitting the i subscript, the expected payoff for choosing action $a > 0$ under M1 to M4 is, cf. (9),

$$v(a,x,w) = \pi(a,x,w) + \beta E\left[m(x_{t+1},w_{t+1})\,|a_t = a, x_t = x, w_t = w\right],$$

where $m(x,w)$ denotes the ex-ante (also known as integrated) value function, $E[V(s_t)\,|x_t = x, w_t = w]$. N1 imposes separability and exclusion restrictions so the contribution of the payoff related to the past action can be isolated within a single period, as it has the following structure

$$\pi(a,x,w) = \gamma_1(a,x) + \gamma_2(a,x,w).$$

Under N2, the past action is also excluded in the future expected payoff, as $E[m(x_{t+1},w_{t+1})\,|a_t,x_t,w_t] = E[m(x_{t+1},w_{t+1})\,|a_t,x_t]$. Therefore

$$v(a,x,w) = \widetilde{\gamma}_1(a,x) + \gamma_2(a,x,w),$$

where $\widetilde{\gamma}_1(a,x) \equiv \gamma_1(a,x) + \beta E[m_i(x_{t+1},w_{t+1})\,|a_{it} = a_i, x_t = x]$. It is now clear that variations in expected payoff (net of the private shock) with respect to w for any given a,x can be traced only to the contribution from γ_2. Therefore $\{\gamma_2(a,x,w) - \gamma_2(0,x,w)\}_{w\in A}$ can be identified upto a location normalization by differencing $\{v(a,x,w) - v(0,x,w)\}_{w\in A}$ over the support of w that eliminates the free nuisance function $\widetilde{\gamma}_1(a,x) - \widetilde{\gamma}_1(0,x)$ for $a > 0, x$.

The combination of exclusion and independence assumptions are classic tools in the study of identification in econometrics; also see the recent works of Blevins (2013) and Chen (2013) who also rely on somewhat similar conditions in order to identify the distribution of normalized unobserved state variables in related single agent dynamic decision models. The idea above may first appear less transparent in a game environment since the present and future expected payoffs for each player are complicated by the beliefs formation of other players' actions that also depend on past actions. However, we can define an analogous nuisance function that can still be *differenced out* by considering a particular linear combination of the expected payoffs, which can be formalized by a projection, to identify the switching costs upto some normalizations.[8] We provide precise conditions for what can be identified from ϕ_i in the next Section.

[8] Mathematically, for fixed a,x, our identification strategy in a single agent case leads to: $g_1(w) = c + g_2(w)$. In the case of a game we have $g_1(w) = \int c(x)\,h(x|w)\,dx + g_2(w)$. In a linear functional notation: $g_1 = Ac + g_2$.

4 Main Results

We first present our identification results first without assuming the discount factor, then the identification of the discount factor.

4.1 Identification without the Discount Factor

We begin by introducing some additional notations and representation lemmas. For any x, w, we denote the ex-ante expected payoffs by $m_i(x,w) = E[V_i(x_t, w_t, \varepsilon_{it}) | x_t = x, w_t = w]$, where V_i is the value function defined in (7), that can also be defined recursively through

$$\begin{aligned} m_i(x,w) &= E[\pi_i(a_t, x_t, w_t) | x_t = x, w_t = w] + E[\sum_{a_i' \in A} \varepsilon_{it}(a_i') \cdot \mathbf{1}[a_{it} = a_i'] | x_t = x, w_t = w] \\ &\quad + \beta E[m_i(x_{t+1}, w_{t+1}) | x_t = x, w_t = w], \end{aligned} \tag{8}$$

and the choice specific expected payoffs for choosing action a_i prior to adding the period unobserved state variable is

$$\begin{aligned} v_i(a_i, x, w) &= E[\pi_i(a_{it}, a_{-it}, x_t, w_t) | a_{it} = a_i, x_t = x, w_t = w] \\ &\quad + \beta E[m_i(x_{t+1}, w_{t+1}) | a_{it} = a_i, x_t = x, w_t = w]. \end{aligned} \tag{9}$$

Both m_i and v_i are familiar quantities in this literature. Under Assumption N2, $E[m_i(x_{t+1}, w_{t+1}) | a_{it}, x_t, w_t]$ can be simplified further to $E[\widetilde{m}_i(a_{it}, a_{-it}, x_t) | a_{it}, x_t, w_t]$, where for all i, a_i, a_{-i}, x, $\widetilde{m}_i(a_i, a_{-i}, x)$ is defined as $E[m_i(x_{t+1}, a_{it}, a_{-it}) | a_{it} = a_i, a_{-it} = a_{-i}, x_t = x]$. Then, for $a_i > 0$, we define $\Delta v_i(a_i, x, w) \equiv v_i(a_i, x, w) - v_i(0, x, w)$, $\Delta \mu_i(a_i, a_{-i}, x) \equiv \mu_i(a_i, a_{-i}, x) - \mu_i(0, a_{-i}, x)$, and $\Delta \widetilde{m}_i(a_i, a_{-i}, x) \equiv \widetilde{m}_i(a_i, a_{-i}, x) - \widetilde{m}_i(0, a_{-i}, x)$ for all i, a_{-i}, x. Furthermore, since the action space is finite, the conditions imposed on $\phi_i \cdot \eta_i$ by N1 ensures for each $a_i > 0$ we can always write the normalized switching cost as

$$\phi_i(a_i, x, w; \eta_i) \cdot \eta_i(a_i, x, w) - \phi_i(0, x, w; \eta_i) \cdot \eta_i(0, x, w) = \sum_{w' \in W_{\eta_i}^{\Delta}(a_i, x)} \phi_{i,\eta_i}(a_i, x, w') \cdot \mathbf{1}[w = w'], \tag{10}$$

for $\phi_{i,\eta_i}(a_i, x, w) \equiv \phi_i(a_i, x, w; \eta_i) - \phi_i(0, x, w; \eta_i)$ that is only defined only on the set $W_{\eta_i}^{\Delta}(a_i, x) \equiv W_{\eta_i}^1(a_i, x) \cup W_{\eta_i}^1(0, x)$. To illustrate, we briefly return to Examples 1 - 3.

Example 1 (Entry Cost, Cont.): Here the only $a_i > 0$ is $a_i = 1$. Since $W_{\eta_i}^1(0, x)$ is empty $W_{\eta_i}^{\Delta}(1, x) = W_{\eta_i}^1(1, x)$, and for any $w = (0, a_{-i})$, $\phi_{i,\eta_i}(1, x, w) = EC_i(x, a_{-i})$ for all i, a_{-i}, x.

Example 2 (Scrap Value, Cont.): Similarly to the above, $W_{\eta_i}^{\Delta}(1, x) = W_{\eta_i}^1(0, x)$, and for any $w = (1, a_{-i})$, $\phi_{i,\eta_i}(1, x, w) = -SV_i(x, a_{-i})$ for all i, a_{-i}, x.

Example 3 (General Switching Costs, Cont.): For any $a_i > 0$, prior to imposing any additional normalizations, $W^1_{\eta_i}(a_i, x)$ and $W^1_{\eta_i}(0, x)$ can both be non-empty, so for all i, a_{-i}, x such that $a'_i \neq a_i$:

$$
\begin{aligned}
\phi_{i,\eta_i}(a_i, x, w) &= SC_i(a_i, 0, x, a_{-i}) \text{ when } w = (0, a_{-i}), \\
\phi_{i,\eta_i}(a_i, x, w) &= -SC_i(0, a_i, x, a_{-i}) \text{ when } w = (a_i, a_{-i}), \\
\phi_{i,\eta_i}(a_i, x, w) &= SC_i(a_i, a'_i, x, a_{-i}) - SC_i(0, a'_i, x, a_{-i}) \text{ when } w = (a'_i, a_{-i}) \text{ for } a'_i \neq a_i.
\end{aligned}
$$

Note that $SC_i(a'_i, a''_i, x, a_{-i})$ can be recovered for any $a_i \neq a'_i$ by taking some linear combination from $\{\phi_{i,\eta_i}(a_i, x, a'_i, a_{-i})\}_{a_i, a'_i \in A \times A}$.

The following lemmas generalize respectively equations (4) and (5) in Section 2.

Lemma 1: *Under M1 - M4 and N1 - N2, we have for all $i, a_i > 0$ and a_{-i}, x, w:*

$$
\Delta v_i(a_i, x, w) = E[\lambda_i(a_i, a_{-it}, x_t) | x_t = x, w_t = w] + \sum_{w' \in W^{\Delta}_{\eta_i}(a_i, x)} \phi_{i,\eta_i}(a_i, x, w') \cdot \mathbf{1}[w = w'], \tag{11}
$$

where

$$
\lambda_i(a_i, a_{-i}, x) \equiv \Delta\mu_i(a_i, a_{-i}, x) + \beta \Delta \widetilde{m}_i(a_i, a_{-i}, x). \tag{12}
$$

Proof of Lemma 1: Using the law of iterated expectation, under M3 $E[V_i(s_{it+1}) | a_{it} = a_i, x_t, w_t] = E[m_i(x_{t+1}, w_{t+1}) | a_{it} = a_i, x_t, w_t]$, which simplifies further, after another application of the law of iterated expectation and N2, to $E[\widetilde{m}_i(a_i, a_{-it}, x_t) | x_t, w_t]$. The remainder of the proof of Lemma 1 then follows from the definitions of the terms defined in the text.■

Lemma 1 says that the normalized choice specific expected payoffs can be decomposed into a sum of the fixed profits at time t and a conditional expectation of a nuisance function of λ_i consisting of composite terms of the primitives. In particular the conditional law for the expectation in (11), which is that of a_{-it} given (x_t, w_t), is identifiable from the data. Since a conditional expectation operator is a linear operator, and the support of w_t is a finite set with $(K+1)^I$ elements, we can then represent (11) by a matrix equation.

Lemma 2: *Under M1 - M4 and N1 - N2, we have for all $i, a_i > 0$ and x:*

$$
\Delta \mathbf{v}_i(a_i, x) = \mathbf{Z}_i(x) \boldsymbol{\lambda}_i(a_i, x) + \mathbf{D}_i(a_i, x) \boldsymbol{\phi}_{i,\eta_i}(a_i, x), \tag{13}
$$

where $\Delta \mathbf{v}_i(a_i, x)$ denotes a $(K+1)^I -$dimensional vector of normalized expected discounted payoffs, $\{\Delta v_i(a_i, x, w)\}_{w \in A^I}$, $\mathbf{Z}_i(x_t)$ is a $(K+1)^I$ by $(K+1)^{I-1}$ matrix of conditional probabilities,

$\{\Pr[a_{-it}=a_{-i}|x_t=x,w_t=w]\}_{(a_{-i},w)\in A^{I-1}\times A^I}$, $\boldsymbol{\lambda}_i(a_i,x)$ *denotes a* $(K+1)^{I-1}$ *by* 1 *vector of* $\{\lambda_i(a_i,a_{-i},x)\}_{a_{-i}\in A^{I-1}}$, $\mathbf{D}_i(a_i,x)$ *is a* $(K+1)^I$ *by* $\left|W^{\Delta}_{\eta_i}(a_i,x)\right|$ *matrix of ones and zeros, and* $\boldsymbol{\phi}_{i,\eta_i}(a_i,x)$ *is a* $\left|W^{1}_{\eta_i}(a_i,x)\right|$ *by* 1 *vector of* $\{\phi_{i,\eta_i}(a_i,x,w)\}_{w\in W^{\Delta}_{\eta_i}(a_i,x)}$.

Proof of Lemma 2: Immediate.■

Let $\rho(Z)$ denote the rank of matrix Z, and $\mathbf{M}_Z$ denotes a projection matrix whose null space is the column space of Z. We can now state first result that generalizes equations (6) in Section 2.

Theorem 1: *Under M1 - M4 and N1 - N2, for each* $i, a_i>0$ *and* x, *if (i)* $\mathbf{D}_i(a_i,x)$ *has full column rank, and, (ii)* $\rho(\mathbf{Z}_i(x))+\rho(\mathbf{D}_i(a_i,x))=\rho([\mathbf{Z}_i(x):\mathbf{D}_i(a_i,x)])$, *then* $\mathbf{D}_i(a_i,x)^{\top}\mathbf{M}_{\mathbf{Z}_i(x)}\mathbf{D}_i(a_i,x)$ *is non-singular, and*

$$\boldsymbol{\phi}_{i,\eta_i}(a_i,x)=(\mathbf{D}_i(a_i,x)^{\top}\ \mathbf{M}_{\mathbf{Z}_i(x)}\mathbf{D}_i(a_i,x))^{-1}\mathbf{D}_i(a_i,x)^{\top}\ \mathbf{M}_{\mathbf{Z}_i(x)}\Delta\mathbf{v}_i(a_i,x). \tag{14}$$

Proof: The full column rank condition of $\mathbf{D}_i(a_i,x)$ is a trivial assumption, the no perfect collinearity condition makes sure there is no redundancy in the modeling of the switching costs. The rank condition (ii) ensures $\mathbf{M}_{\mathbf{Z}_i(x)}\mathbf{D}_i(a_i,x)$ preserves the rank of $\mathbf{D}_i(a_i,x)$, therefore $\mathbf{D}_i(a_i,x)^{\top}\mathbf{M}_{\mathbf{Z}_i(x)}\mathbf{D}_i(a_i,x)$ is non-singular.[9] The proof can be completed by projecting the vectors on both sides of equation (13) by $\mathbf{M}_{\mathbf{Z}_i(x)}$and solve for $\boldsymbol{\phi}_{i,\eta_i}(a_i,x)$.■

In order for condition (ii) in Theorem 1 to hold, it is necessary to impose some normalizations on the switching costs. Before commenting further, it will be informative to again revisit Examples 1 - 3. For notational simplicity we shall assume $I=2$, so that $w_t\in\{(0,0),(0,1),(1,0),(1,1)\}$. And since $A=\{0,1\}$ in Examples 1 and 2, we shall also drop a_i from $\Delta\mathbf{v}_i(a_i,x)=\{\Delta v_i(a_i,x,w)\}_{w\in A^I}$ and $\boldsymbol{\lambda}_i(a_i,x)=\{\lambda_i(a_i,a_{-i},x)\}_{a_{-i}\in A^{I-1}}$.

Example 1 (Entry Cost, Cont.): Equation (13) can be written as

$$\begin{bmatrix}\Delta v_i(x,(0,0))\\ \Delta v_i(x,(0,1))\\ \Delta v_i(x,(1,0))\\ \Delta v_i(x,(1,1))\end{bmatrix}=\begin{bmatrix}P_{-i}(0|x,(0,0)) & P_{-i}(1|x,(0,0))\\ P_{-i}(0|x,(0,1)) & P_{-i}(1|x,(0,1))\\ P_{-i}(0|x,(1,0)) & P_{-i}(1|x,(1,0))\\ P_{-i}(0|x,(1,1)) & P_{-i}(1|x,(1,1))\end{bmatrix}\begin{bmatrix}\lambda_i(0,x)\\ \lambda_i(1,x)\end{bmatrix}+\begin{bmatrix}1&0\\0&1\\0&0\\0&0\end{bmatrix}\begin{bmatrix}EC_i(x,0)\\ EC_i(x,1)\end{bmatrix},$$

[9]Otherwise the columns of $\mathbf{M}_{\mathbf{Z}_i}(x)\,\mathbf{D}_i(a_i)$ is linearly dependent, and some linear combination of the columns in $\mathbf{D}_i(a_i)$ must lie in the column space of $\mathbf{Z}_i(x)$, which violates the rank condition.

where $P_{-i}(a_{-i}|x,w) \equiv \Pr[a_{-it}=a_{-i}|x_t=x,w_t=w]$. A simple sufficient condition that ensures condition (ii) in Theorem 1 to hold is when the lower half of $\mathbf{Z}_i(x)$ has full rank, i.e. when $P_{-i}(0|x,(1,0)) \neq P_{-i}(0|x,(1,1))$.

Example 2 (Scrap Value, Cont.): Equation (13) can be written as

$$\begin{aligned}
\begin{bmatrix} \Delta v_i(x,(0,0)) \\ \Delta v_i(x,(0,1)) \\ \Delta v_i(x,(1,0)) \\ \Delta v_i(x,(1,1)) \end{bmatrix} &= \begin{bmatrix} P_{-i}(0|x,(0,0)) & P_{-i}(1|x,(0,0)) \\ P_{-i}(0|x,(0,1)) & P_{-i}(1|x,(0,1)) \\ P_{-i}(0|x,(1,0)) & P_{-i}(1|x,(1,0)) \\ P_{-i}(0|x,(1,1)) & P_{-i}(1|x,(1,1)) \end{bmatrix} \begin{bmatrix} \lambda_i(0,x) \\ \lambda_i(1,x) \end{bmatrix} \\
&+ \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -SV_i(x,0) \\ -SV_i(x,1) \end{bmatrix}.
\end{aligned}$$

An analogous sufficient condition that ensures condition (ii) in Theorem 1 to hold in this case is $P_{-i}(0|x,(0,0)) \neq P_{-i}(0|x,(0,1))$.

Example 3 (General Switching Costs, Cont.): Suppose $K=2$, we consider $\Delta\mathbf{v}_i(2,x) =$

$\{\Delta v_i(2,x,w)\}_{w\in A^I}$,

$$\begin{bmatrix} \Delta v_i(2,x,(0,0)) \\ \Delta v_i(2,x,(0,1)) \\ \Delta v_i(2,x,(0,2)) \\ \Delta v_i(2,x,(1,0)) \\ \Delta v_i(2,x,(1,1)) \\ \Delta v_i(2,x,(1,2)) \\ \Delta v_i(2,x,(2,0)) \\ \Delta v_i(2,x,(2,1)) \\ \Delta v_i(2,x,(2,2)) \end{bmatrix} = \begin{bmatrix} P_{-i}(0|x,(0,0)) & P_{-i}(1|x,(0,0)) & P_{-i}(2|x,(0,0)) \\ P_{-i}(0|x,(0,1)) & P_{-i}(1|x,(0,1)) & P_{-i}(2|x,(0,1)) \\ P_{-i}(0|x,(0,2)) & P_{-i}(1|x,(0,2)) & P_{-i}(2|x,(0,2)) \\ P_{-i}(0|x,(1,0)) & P_{-i}(1|x,(1,0)) & P_{-i}(2|x,(1,0)) \\ P_{-i}(0|x,(1,1)) & P_{-i}(1|x,(1,1)) & P_{-i}(2|x,(1,1)) \\ P_{-i}(0|x,(1,2)) & P_{-i}(1|x,(1,2)) & P_{-i}(2|x,(1,2)) \\ P_{-i}(0|x,(2,0)) & P_{-i}(1|x,(2,0)) & P_{-i}(2|x,(2,0)) \\ P_{-i}(0|x,(2,1)) & P_{-i}(1|x,(2,1)) & P_{-i}(2|x,(2,1)) \\ P_{-i}(0|x,(2,2)) & P_{-i}(1|x,(2,2)) & P_{-i}(2|x,(2,2)) \end{bmatrix} \begin{bmatrix} \lambda_i(2,0,x) \\ \lambda_i(2,1,x) \\ \lambda_i(2,2,x) \end{bmatrix} \tag{15}$$

$$+ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} SC_i(2,0,x,0) \\ SC_i(2,0,x,1) \\ SC_i(2,0,x,2) \\ SC_i(2,1,x,0) - SC_i(0,1,x,0) \\ SC_i(2,1,x,1) - SC_i(0,1,x,1) \\ SC_i(2,1,x,2) - SC_i(0,1,x,2) \\ -SC_i(0,2,x,0) \\ -SC_i(0,2,x,1) \\ -SC_i(0,2,x,2) \end{bmatrix}.$$

Clearly the required rank condition of Theorem 1 cannot hold without any normalization on the switching costs. If $\rho(\mathbf{Z}_i(x)) = 3$, then the maximum number of elements in $\boldsymbol{\phi}_{i,\eta_i}(2,x)$ that can be identified using Lemma 2 is 6 given that we have 9 equations. Therefore we need to normalize three parameters, which can naturally be interpreted as normalizing one type of switching costs. Ideally available data or other prior knowledge can be used so known switching costs can be removed from the right hand side (RHS) of equation (15), as done in Section 2 (see equation (6)). Otherwise the most natural normalizations that can be employed include exclusions, e.g. zero switching cost from action 2 to 0 (or vice versa), equality, e.g. same magnitude of switching to and from actions 0 and 2. More specifically, for any x suppose $SC_i(0,2,x,a_{-i}) = 0$ for all a_{-i}, then similar to the two previous examples, a sufficient condition for condition (ii) in Theorem 1 to hold can be given in the form that ensures the lower third of $\mathbf{Z}_i(x)$ to have full rank, which is equivalent to the determinant of $\begin{pmatrix} P_{-i}(0|x,(2,0)) & P_{-i}(1|x,(2,0)) & P_{-i}(2|x,(2,0)) \\ P_{-i}(0|x,(2,1)) & P_{-i}(1|x,(2,1)) & P_{-i}(2|x,(2,1)) \\ P_{-i}(0|x,(2,2)) & P_{-i}(1|x,(2,2)) & P_{-i}(2|x,(2,2)) \end{pmatrix}$ is non-zero. Analogous conditions and comments apply for $\Delta\mathbf{v}_i(1,x)$.

Comments on Theorem 1:

Order Condition. Notice that our identification result is obtained pointwise for each $i, a_i > 0$ and x. In order to apply Theorem 1 some necessary order condition must be met. Firstly, $\rho(\mathbf{Z}_i(x))$ always takes value between 1 and $(K+1)^{I-1}$; the latter is the number of columns in $\mathbf{Z}_i(x)$ that equals the cardinality of the action space of all other players other than i. A necessary order condition based on the number of rows of the matrix equation in equation (13) can be obtained from : $\rho(\mathbf{Z}_i(x)) + \rho(\mathbf{D}_i(a_i, x)) \leq (K+1)^I$, so that (the number of switching cost parameters one wish to identify is $\left|W^{\Delta}_{\eta_i}(a_i, x)\right| =$) $\rho(\mathbf{D}_i(a_i, x)) \leq (K+1)^I - 1$. In the least favorable case, in terms of applying Theorem 1, the previous inequality can be strengthened by using the maximal rank of $\mathbf{Z}_i(x)$, which is $(K+1)^{I-1}$, so $\rho(\mathbf{D}_i(a_i, x))$ is bounded above by $K(K+1)^{I-1}$.

Underidentification. We argue that normalization of switching costs in this context is necessary. In order to see this, for the moment suppose we also know β so that we can apply the identification strategy along the line of Magnac and Thesmar (2002). Then for each x, without any a priori restrictions, there are $(K+1)^{2I}$ parameters from $\{\pi_i(a, x, w)\}_{a,w \in A^I \times A^I}$ that satisfy $K(K+1)^I$ equations from $\{\Delta v_i(a_i, x, w)\}_{a_i,w \in A \backslash \{0\} \times A^I}$; cf. equation (3) in Section 2. Therefore π_i is underidentified (cf. Proposition 2 in Pesendorfer and Schmidt-Dengler (2008)). Suppose π_i satisfies N1 with $K^2(K+1)^{I-1}$ of unknown switching cost parameters, which equals the maximum number of switching costs we can identify from the least favorable necessary order condition from Theorem 1 (there are $K(K+1)^{I-1}$ parameters for each $a_i > 0$). Since the number of parameters from $\{\mu_i(a, x)\}_{a \in A^I}$ for each x is $(K+1)^I$, the total number of payoff parameters under N1 is $(K+1)^I + K^2(K+1)^{I-1} = K(K+1)^I + (K+1)^{I-1}$, which is still more than the total number of equations. Therefore π_i remains underidentified under N1 even if the discount factor is known. Since we treat μ_i nonparametrically, as well as unknown β, we cannot hope to identify more than $K(K+1)^{I-1}$ switching costs parameters associated with $a_i > 0$.

Notice that N1 does not impose a priori restrictions on π_i beyond the implicit assumption that there is no switching cost payoffs when actions do not change over time, namely $\pi_i(a_i, a_{-i}, x, w_i, w_{-i}) = 0$ for all $a_i = w_i$. we impose only $K+1$ restrictions on π_i that is still underidentified.

Note that imposing switching cost structure alone still leads to underidentification of π_i even when β is known, in particular

$$\begin{aligned} f(a_1, a_2, w_1, w_2) &= g(a_1, a_2) + h(a_1 \neq w_1, w_2) \\ (2)^4 &= (2)^3 \end{aligned}$$

Normalization. Based on the above argument, concerning the identification of the switching costs, the effective degree of underidentification is $(K+1)^{I-1} = (K+1)^I - K(K+1)^{I-1}$. Note that $(K+1)^{I-1}$ equals also the cardinality of A^{I-1}, so one reasonable (and effective) approach is to normalize with respect to a particular action choice, or equality between two actions, generalizing the description for Example 3 above. Notably it will be adequate to assume that cost of switching to action 0 from other action is zero (or known), which is a weaker condition than a familiar normalization of the outside option for the entire payoff function (e.g. Proposition 2 of Magnac and Thesmar (2002) as well as Assumption 2 of Bajari et al. (2009)).

Interpretability. Recall that $\boldsymbol{\phi}_{i,\eta_i}$ is a vector of primitives of the game that have structural interpretations. Equation (13) gives the decomposition of the expected discounted payoffs in terms of $\boldsymbol{\phi}_{i,\eta_i}$ and other primitives of the game contained in $\boldsymbol{\lambda}_i$ (cf. (3)). If the rank conditions of Theorem 1 are satisfied, then $\boldsymbol{\phi}_{i,\eta_i}$ can be written in terms of just the choice probabilities that are reduced form parameters of the game, see (14). However, it is generally difficult to give a direct interpretation describing the relation between the primitive and reduced form parameter (also see Magnac and Thesmar (2002), Pesendorfer and Schmidt-Dengler (2008) and Bajari et al. (2009)).

Generally we can also formally impose prior knowledge restrictions on $\boldsymbol{\phi}_{i,\eta_i}$, then the rank requirement on $\mathbf{D}_i$ can be relaxed further. For instance, empirical work often assume firms' entry costs or scrap values do not vary with other players' past entry decisions (e.g. see the example in Section 2), or in a general switching cost framework certain costs may be known to be equal. We next show how to incorporate equality restrictions.

Assumption R1 (Equality Restrictions): *For all i, x, there exists a matrix $K(K+1)^I$ by κ matrix $\widetilde{\mathbf{D}}_i(x)$ with full column rank and a κ by 1 vector of functions $\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)$ so that $\widetilde{\mathbf{D}}_i(x)\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)$ represents a vector of functions that satisfy some equality constraints imposed on $\{\mathbf{D}_i(a_i,x)\boldsymbol{\phi}_{i,\eta_i}(a_i,x)\}_{a_i\in A}$.*

The matrix $\widetilde{\mathbf{D}}_i(x)$ can be constructed from $diag\{\mathbf{D}_i(1,x),\ldots,\mathbf{D}_i(K,x)\}$, and merging the columns of the latter matrix, by simply adding columns that satisfy the equality restriction together. Redundant components of $\{\boldsymbol{\phi}_{i,\eta_i}(a_i,x)\}_{a_i\in A}$ are then removed to define $\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)$. One example for $\widetilde{\mathbf{D}}_i(x)$ can be found in Section 2, where we consider a fixed cost function that does not depend on other players' past actions, also see Example 4 below. The following lemma gives the matrix representation of the expected payoffs in this case (cf. Lemma 2).

Lemma 3: *Under M1 - M4, N1 - N2 and R1, we have for all i, x:*

$$\Delta\mathbf{v}_i(x) = (I_K \otimes \mathbf{Z}_i(x))\,\boldsymbol{\lambda}_i(x) + \widetilde{\mathbf{D}}_i(x)\,\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x), \tag{16}$$

where $\Delta\mathbf{v}_i(x)$ denotes a $K(K+1)^I$ $-$dimensional vector of normalized expected discounted payoffs, $\{\Delta\mathbf{v}_i(a_i,x)\}_{a_i\in A\backslash\{0\}}$, $\mathbf{Z}_i(x)$ is a $(K+1)^I$ by $(K+1)^{I-1}$ matrix of conditional probabilities,

$\{\Pr[a_{-it}=a_{-i}|x,w_t=w]\}_{(a_{-i},w)\in A^{I-1}\times A^I}$, *$I_K$ is an Idn matrix of size K, $\otimes$ denotes the Kronecker product, $\boldsymbol{\lambda}_i(x)$ denotes a $K(K+1)^{I-1}$ by 1 vector of $\{\boldsymbol{\lambda}_i(a_i,x)\}_{a_i\in A\backslash\{0\}}$, $\widetilde{\mathbf{D}}_i(x)$ and $\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)$ are described in Assumption R1.*

Proof of Lemma 3: Immediate.■

Using Lemma 3, our next result generalizes Theorem 1 by allowing for the equality restrictions across actions.

Theorem 2: *Under M1 - M4, N1 - N2 and R1, for each i,x, if (i) $\widetilde{\mathbf{D}}_i(x)$ has full column rank and, (ii) $\rho(I_K\otimes\mathbf{Z}_i(x))+\rho(\widetilde{\mathbf{D}}_i(x))=\rho([I_K\otimes\mathbf{Z}_i(x):\widetilde{\mathbf{D}}_i(x)])$, then $\widetilde{\mathbf{D}}_i^\top(x)\mathbf{M}_{I_K\otimes\mathbf{Z}_i(x)}\widetilde{\mathbf{D}}_i(x)$ is non-singular, and*

$$\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)=(\widetilde{\mathbf{D}}_i^\top(x)\mathbf{M}_{I_K\otimes\mathbf{Z}_i(x)}\widetilde{\mathbf{D}}_i(x))^{-1}\widetilde{\mathbf{D}}_i^\top(x)\mathbf{M}_{I_K\otimes\mathbf{Z}_i(x)}\Delta\mathbf{v}_i(x).$$

Proof of Theorem 2: Same as the proof of Theorem 1.■

Our previous comments on Theorem 1 are also relevant for Theorem 2. However, the ability to relax the necessary order condition may not always be sufficient for identification. In particular, consider the following special case of Example 3 when $K=1$.

Example 4 (Entry Game with Entry Cost and Scrap Value): The period payoff at time t is

$$\begin{aligned}\pi_i(a_{it},a_{-it},x_t,w_t) &= \mu_i(a_{it},a_{-it},x_t)+EC_i(x_t)\cdot a_{it}(1-a_{it-1})\\ &\quad +SV_i(x_t)\cdot(1-a_{it})a_{it-1}.\end{aligned}$$

I.e. we have imposed the equality restrictions on the entry costs and scrap values for each player only depend on each her own actions. Then, for all i,x, the content of equation (16) (in Lemma 3) is

$$\begin{aligned}\begin{bmatrix}\Delta v_i(x,(0,0))\\ \Delta v_i(x,(0,1))\\ \Delta v_i(x,(1,0))\\ \Delta v_i(x,(1,1))\end{bmatrix} &= \begin{bmatrix}P_{-i}(0|x,(0,0)) & P_{-i}(1|x,(0,0))\\ P_{-i}(0|x,(0,1)) & P_{-i}(1|x,(0,1))\\ P_{-i}(0|x,(1,0)) & P_{-i}(1|x,(1,0))\\ P_{-i}(0|x,(1,1)) & P_{-i}(1|x,(1,1))\end{bmatrix}\begin{bmatrix}\lambda_i(0,x)\\ \lambda_i(1,x)\end{bmatrix}\\ &\quad+\begin{bmatrix}1&0\\1&0\\0&1\\0&1\end{bmatrix}\begin{bmatrix}EC_i(x)\\ -SV_i(x)\end{bmatrix}.\end{aligned}\tag{17}$$

Note that the order condition is now satisfied. However, condition (ii) in Theorem 2 still does not hold in this case since a vector of ones is contained in both $\mathcal{CS}(\mathbf{Z}_i(x))$ and $\mathcal{CS}(\widetilde{\mathbf{D}}_i(x))$.

The failure to apply our results in Example 4 is due to the fact that $\mathbf{Z}_i(x)$ is a stochastic matrix, whose rows each sums to one. Otherwise the finding itself may not be too surprising given normalizations of switching costs are fairly common in empirical work. For instance Aguirregabiria and Suzuki (2013) imply the nonidentification of entry cost and scrap value in a related decision problem with entry, while Pesendorfer and Schmidt-Dengler (2008) assigned a particular value for the scrap value. Thus, analogously, for our Example 4, if either EC_i or SV_i is normalized, and taken to the LHS of equation (17), then we can apply our Theorem 2 analogous to equation (6)).

We emphasize that our Theorems 1 and 2 only provide sufficient conditions for identification of ϕ_i without assuming either β or μ_i. The failure to apply our theorems does not mean ϕ_i cannot be identified with additional assumptions. For instance, if one assumes the knowledge of β then existing results in Bajari et al. (2009) and Pesendorfer and Schmidt-Dengler (2008) can be used to identify jointly both μ_i and ϕ_i if additional assumptions are imposed on π_i.

We end this subsection by commenting that all of our results thus far hold without modification if we re-define w_t to be $a_{t-\varsigma}$ for any finite $\varsigma \geq 1$, and then replace x_t by $\widetilde{x}_t = (x_t, a_{t-1}, \ldots, a_{t-\varsigma+1})$. The inclusion of such state variable does not violate assumption N2, and thus still allows us to define analogous nuisance function that can be projected away as shown in Theorems 1 and 2. It is also in the case of including lagged actions in the observed states that we naturally have $W^d_{\eta_i}(a_i, x) \neq W^d_{\eta_i}(a_i, x')$ for $x \neq x'$ since the principal interpretation of switching costs generally will depend on a_{t-1}.

4.2 Discount Factor

If μ_i is assumed to be known then, using Theorems 1 or 2, π_i can be identified without the knowledge of β. We now consider the identification of β and take all other primitives of the model as known (i.e. assume $(\{\pi_i\}_{i=1}^{I}, Q, G)$). The result in this subsection is not specific to games involving switching costs. Therefore we do not impose Assumptions N1 and N2 here, and henceforth we omit w_t.

The parameter space for the model is now $\mathfrak{B} \subseteq (0,1)$ and we are interested in the discount factor that is consistent with the data generating process, which we denote by β_0. We begin with an updated expression for the choice specific expected payoffs for choosing action a_i prior to adding the period unobserved state variable, where we now explicitly denote the dependence on the parameter β, so that for any i, a_i and x (cf. equation (9)):

$$v_i(a_i, x; \beta) = E\left[\pi_i(a_{it}, a_{-it}, x_t) \,|a_{it} = a_i, x_t = x\right] + \beta g_i(a_i, x; \beta), \tag{18}$$

where, similar to previously, $g_i(a_i, x; \beta) = E\left[V_i(s_{it+1}; \beta) \,|a_{it} = a_i, x_t = x\right]$, and $V_i(s_i; \beta) = \sum_{\tau=t}^{\infty} \beta^{\tau-t} E\left[u_i(a_\tau,\right.$ Note that the expectations are taken with respect to the observed choice and transition probabilities that are consistent with β_0. We consider the relative payoffs in (18) with action 0 as the base, so that for all $i, a_i > 0$ and x:

$$\Delta v_i (a_i, x; \beta) = E [\Delta \pi_i (a_i, a_{-it}, x_t) | x_t = x] + \beta \Delta g_i (a_i, x; \beta), \tag{19}$$

where $\Delta v_i (a_i, x; \beta) = v_i (a_i, x; \beta) - v_i (0, x; \beta), \Delta \pi_i (a_i, a_{-i}, x) = \pi_i (a_i, a_{-i}, x) - \pi_i (0, a_{-i}, x_t)$ for all a_{-i}, and $\Delta g_i (a_i, x; \beta) = g_i (a_i, x; \beta) - g_i (0, x; \beta)$. Since $\Delta v_i (a_i, x; \beta_0)$ is identified from the data for all a_i, x, we take each β to be a *structure* of the (pseudo-)model and its implied expected payoffs, denoted by $\mathcal{V}_\beta = \{\Delta v_i (a_i, x; \beta)\}_{i, a_i, x \in \mathcal{I} \times A \times X}$, to be a *reduced form*.[10,11] We can then define identification using the notion of observational equivalence in terms of the expected payoffs (cf. Magnac and Thesmar (2002)).

Definition I1 (Observational Equivalence): Any distinct β and β' in $\mathfrak{B}$ are *observationally equivalent* if and only if $\mathcal{V}_\beta = \mathcal{V}_{\beta'}$.

Definition I2 (Identification): An element in $\mathfrak{B}$, say β, is *identified* if and only if β' and β are not observationally equivalent for all $\beta' \neq \beta$ in $\mathfrak{B}$.

By inspecting equation (19), since the term involving π_i does not depend on β, identification is determined by $\beta \Delta g_i (\cdot, \cdot; \beta)$. The following lemma expresses $\{\Delta g_i (a_i, x; \beta)\}_{a_i, x \in A \setminus \{0\} \times X}$ in terms of β and other components that can be identified from the choice and transition probabilities. In what follows, for any $i, a_i > 0$ and x, we let: $\Delta H_i^{a_i}(x)$ denote a J by 1 vector of $\{\Pr [x_{t+1} = x' | x_t = x, a_{it} = a_i] - \Pr [x_{t+1} = x' | x_t = x, a_{it} = 0]\}_{x' \in X}$, $\mathbf{L}$ be a J by J stochastic matrix of transition probabilities of x_{t+1} conditioning on x_t, $\mathbf{R}$ is a J by J matrix of conditional choice probabilities such that $\mathbf{R} \boldsymbol{\pi}_i$ represents a J by 1 vector of $\{E [\pi_i (a_t, x_t) | x_t = x']\}_{x' \in X}$, and $\Delta \underline{v}_i (a_i, x; \beta_0) = \Delta H_i^{a_i}(x) \times (I - \beta \mathbf{L})^{-1} \underline{\mathbf{r}}_i$ where $\underline{\mathbf{r}}_i$ represents a J by 1 vector of $\left\{E \left[\sum_{a' \in A} \varepsilon_{it} (a') \mathbf{1} [a_{it} = a'] | x_t = x'\right]\right\}_{x' \in X}$.

Lemma 4: *Under M1 - M4, we have for all $i, a_i > 0$ and x:*

$$\Delta g_i (a_i, x; \beta) = \Delta H_i^{a_i}(x) \times (I - \beta \mathbf{L})^{-1} \mathbf{R} \boldsymbol{\pi}_i + \Delta \underline{v}_i (a_i, x; \beta_0). \tag{20}$$

Proof of Lemma 4: Immediate.■

Note that $\Delta v_i (a_i, x; \beta_0)$ and $\Delta \underline{v}_i (a_i, x; \beta_0)$ are identifiable from the observed data using Hotz and Miller's inversion. Therefore β_0 is identifiable if for any $\beta \neq \beta'$, there exists some i and a_i, x such that $\beta \Delta g_i (a_i, x; \beta) \neq \beta' \Delta g_i (a_i, x; \beta')$. The relation in (20) can be written in a matrix for across possible

[10] This is a pseudo-model in the sense that we only work with the equilibrium beliefs that generate the data.

[11] It is equivalent to define the reduced forms in terms of expected payoffs is equivalent to defining them in terms of conditional choice probabilities (Hotz and Miller (1993), Matzkin (1991), Norets and Takahashi (2013)).

values of x_t. Let $\Delta\mathbf{H}_i^{a_i}$ denote $\left[\Delta H_i^{a_i}(x^1)^\top, \ldots, \Delta H_i^{a_i}(x^J)^\top\right]^\top$, a J by J matrix, and $\Delta\mathbf{g}_i^{a_i}(\beta)$ denote $\{\Delta g_i(a_i, x; \beta)\}_{x\in X}$, a J by 1 vector, and similarly $\Delta\underline{\mathbf{v}}_i^{a_i}(\beta)$ denotes $\{\Delta\underline{v}_i(a_i, x; \beta)\}_{x\in X}$.

LEMMA 5: *Under M1 - M4, we have for all $i, a_i > 0$:*

$$\Delta\mathbf{g}_i^{a_i}(\beta) = \Delta\mathbf{H}_i^{a_i}(I - \beta\mathbf{L})^{-1}\mathbf{R}\boldsymbol{\pi}_i + \Delta\underline{\mathbf{v}}_i^{a_i}(\beta_0). \tag{21}$$

PROOF OF LEMMA 5: Immediate.■

Therefore β_0 is identified if and only if there is no other β' in $\mathfrak{B}$ such that $\beta'\Delta\mathbf{H}_i^{a_i}(I - \beta'\mathbf{L})^{-1}\mathbf{R}\boldsymbol{\pi}_i$ equals $\beta_0\Delta\mathbf{H}_i^{a_i}(I - \beta_0\mathbf{L})^{-1}\mathbf{R}\boldsymbol{\pi}_i$. Our next result gives one such sufficient condition.

THEOREM 3 (IDENTIFICATION OF DISCOUNT FACTOR): *Under M1 - M4, if $\mathbf{R}\boldsymbol{\pi}_i \neq \mathbf{0}$ and $\Delta\mathbf{H}_i^{a_i}$ is invertible for some i, a_i, then β_0 is identified.*

PROOF OF THEOREM 3: Take any $\beta, \beta' \in (0,1)$ such that $\beta \neq \beta'$, using equation (21) in Lemma 5 we obtain the following relation:

$$\beta\Delta\mathbf{g}_i(\beta) - \beta'\Delta\mathbf{g}_i(\beta') = \left(\beta\Delta\mathbf{H}_i \times (I - \beta\mathbf{L})^{-1} - \beta'\Delta\mathbf{H}_i \times (I - \beta'\mathbf{L})^{-1}\right)\mathbf{R}\boldsymbol{\pi}_i.$$

We consider the terms in the parenthesis on the RHS of the equation above,

$$\begin{aligned}
& \beta\Delta\mathbf{H}_i \times (I - \beta\mathbf{L})^{-1} - \beta'\Delta\mathbf{H}_i \times (I - \beta'\mathbf{L})^{-1} \\
= \; & (\beta - \beta')\Delta\mathbf{H}_i \times (I - \beta\mathbf{L})^{-1} + \beta'\Delta\mathbf{H}_i\left((I - \beta\mathbf{L})^{-1} - (I - \beta'\mathbf{L})^{-1}\right) \\
= \; & (\beta - \beta')\Delta\mathbf{H}_i \times (I - \beta\mathbf{L})^{-1} + \beta'(\beta - \beta')\Delta\mathbf{H}_i \times (I - \beta'\mathbf{L})^{-1}\mathbf{L}(I - \beta\mathbf{L})^{-1} \\
= \; & (\beta - \beta')\Delta\mathbf{H}_i \times \left(I + \beta'(I - \beta'\mathbf{L})^{-1}\mathbf{L}\right)(I - \beta\mathbf{L})^{-1} \\
= \; & (\beta - \beta')\Delta\mathbf{H}_i \times (I - \beta'\mathbf{L})^{-1}(I - \beta\mathbf{L})^{-1},
\end{aligned}$$

so that

$$\beta\Delta\mathbf{g}_i(\beta) - \beta'\Delta\mathbf{g}_i(\beta') = (\beta - \beta')\Delta\mathbf{H}_i \times (I - \beta'\mathbf{L})^{-1}(I - \beta\mathbf{L})^{-1}\mathbf{R}\boldsymbol{\pi}_i.$$

If $\mathbf{R}\boldsymbol{\pi}_i \neq \mathbf{0}$, then $(I - \beta'\mathbf{L})^{-1}(I - \beta\mathbf{L})^{-1}\mathbf{R}\boldsymbol{\pi}_i \neq \mathbf{0}$ since both $(I - \beta'\mathbf{L})^{-1}$ and $(I - \beta\mathbf{L})^{-1}$ are non-singular by the dominant diagonal theorem. Therefore $\Delta\mathbf{H}_i(I - \beta'\mathbf{L})^{-1}(I - \beta\mathbf{L})^{-1}\mathbf{R}\boldsymbol{\pi}_i$ cannot be a zero vector if $\Delta\mathbf{H}_i$ has full column rank, hence $\beta\Delta g_i(x; \beta)$ must differ from $\beta'\Delta g_i(x; \beta')$ for some x in X.■

The conditions in Theorem 3 are stated in terms of objects that are identified from the data therefore they are easy to check. Note that it is also evident that our argument to identify the discount factor allows for individual specific discount rate by simply replacing β by β_i everywhere.

5 Asymptotic Least Squares Estimation

Our identification results are constructive. For example, Theorems 1 and 2 provide closed-form expressions for ϕ_i that can be used for estimation by plugging in obvious empirical sample counterparts without any numerical optimization. However, such estimator is generally not efficient. This section we provide a brief discussion for constructing a class of asymptotic least squares estimators for ϕ_i and β. We shall consider the two cases separately since it is generally possible to construct a closed-form estimator for ϕ_i but not for β. Our exposition in this section shall we brief. We refer the reader to Sanches, Silva and Srisuma (2013) for further details regarding the estimation methodology and asymptotic results.

Estimation of the Switching Cost

Under the conditions of Theorems 1 and 2, using Lemmas 2 and 3 we have respectively for all $x \in X$:

$$\begin{aligned} \mathbf{M}_{\mathbf{Z}_i(x)}\Delta\mathbf{v}_i\left(a_i,x\right) &= \mathbf{M}_{\mathbf{Z}_i(x)}\mathbf{D}_i\left(a_i\right)\boldsymbol{\phi}_{i,\eta_i}\left(a_i,x\right), \\ \mathbf{M}_{I_K\otimes\mathbf{Z}_i(x)}\Delta\mathbf{v}_i\left(x\right) &= \mathbf{M}_{I_K\otimes\mathbf{Z}_i(x)}\widetilde{\mathbf{D}}_i\widetilde{\boldsymbol{\phi}}_{i,\eta_i}\left(x\right). \end{aligned}$$

since A and X are finite, we have a finite number of equality restrictions across i that can be vectorized in the form of

$$\mathcal{Y}^{sc} = \mathcal{X}^{sc}\theta \quad \text{when} \quad \theta = \theta_0.$$

So that θ_0 is data generating parameter of interest, and $\mathcal{X}^{sc}$ and $\mathcal{Y}^{sc}$ are smooth functions of the known model primitives that we denote by γ_0 (such as choice and transition probabilities, and also the payoff function in the case to estimate discount factor). Specifically, for any θ, $\mathcal{X}^{sc}$ and $\mathcal{Y}^{sc}$ equal $\mathcal{T}^{sc}_{\mathcal{X}}\left(\gamma^{sc}_0\right)$ and $\mathcal{T}^{sc}_{\mathcal{Y}}\left(\gamma^{sc}_0\right)$ respectively for some known functions $\mathcal{T}^{sc}_{\mathcal{X}}$ and $\mathcal{T}^{sc}_{\mathcal{Y}}$. Given a preliminary consistent estimator of γ^{sc}_0, denoted by $\widehat{\gamma}^{sc}$, we can define an estimation criterion where $\mathcal{X}^{sc}\left(\theta\right)$ and $\mathcal{Y}^{sc}$ are replaced by $\widehat{\mathcal{X}}^{sc} = \mathcal{T}^{sc}_{\mathcal{X}}\left(\widehat{\gamma}^{sc}\right)$ and $\widehat{\mathcal{Y}}^{sc} = \mathcal{T}^{sc}_{\mathcal{Y}}\left(\widehat{\gamma}^{sc}\right)$ respectively, so that

$$\widehat{\mathcal{S}}^{sc}(\theta;\widehat{\mathcal{W}}^{sc}) = (\widehat{\mathcal{Y}}^{sc} - \widehat{\mathcal{X}}^{sc}\theta)^\top\widehat{\mathcal{W}}^{sc}(\widehat{\mathcal{Y}}^{sc} - \widehat{\mathcal{X}}^{sc}\theta),$$

where $\widehat{\mathcal{W}}^{sc}$ is a positive definite matrix. We define our estimator, $\widehat{\theta}(\widehat{\mathcal{W}}^{sc})$, to be the minimizer of $\widehat{\mathcal{S}}^{sc}(\theta;\widehat{\mathcal{W}}^{sc})$ that has a closed-form weighted least squares expression (subject to some rank condition):

$$\begin{aligned} \widehat{\theta}(\widehat{\mathcal{W}}^{sc}) &= \arg\min_{\theta\in\Theta}\widehat{\mathcal{S}}^{sc}(\theta;\widehat{\mathcal{W}}^{sc}) \\ &= (\widehat{\mathcal{X}}^{sc\top}\widehat{\mathcal{W}}^{sc}\widehat{\mathcal{X}}^{sc})^{-1}\widehat{\mathcal{X}}^{sc\top}\widehat{\mathcal{W}}^{sc}\widehat{\mathcal{Y}}^{sc}. \end{aligned} \tag{22}$$

As usual, the choice of the weighting matrix will affect the relative efficiency of $\widehat{\theta}(\mathcal{W}^{sc})$ in the class of asymptotic least squares estimators indexed by the set of all positive definite matrices. In particular, the efficient weighting matrix converges in probability to the inverse of the asymptotic variance of $\sqrt{N}(\widehat{\mathcal{Y}}^{sc} - \widehat{\mathcal{X}}^{sc}\theta_0)$, which can be estimated using any preliminary estimator of θ_0, such as (the Idn weighted, ordinary least squares estimator) $(\widehat{\mathcal{X}}^{sc\top}\widehat{\mathcal{X}}^{sc})^{-1}\widehat{\mathcal{X}}^{sc\top}\widehat{\mathcal{Y}}^{sc}$.

Estimation of the Discount Factor

Rearranging equation (21) in Lemma 5 yields

$$\Delta\underline{\mathbf{v}}_i(\beta_0) = \Delta\mathbf{g}_i^{a_i}(\beta) - \Delta\mathbf{H}_i^{a_i}(I - \beta\mathbf{L})^{-1}\mathbf{R}\boldsymbol{\pi}_i.$$

Similar quantities across players can be vectorized in the form of

$$\mathcal{Y}^{df} = \mathcal{X}^{df}(\beta) \quad \text{when} \quad \beta = \beta_0,$$

where $\mathcal{X}^{df}(\beta)$ and $\mathcal{Y}^{df}$ are smooth functions of the known model primitives for every β. Similar to the previous case, for any β, $\mathcal{X}^{df}(\beta)$ and $\mathcal{Y}^{df}$ equal $\mathcal{T}_{\mathcal{X}}^{df}(\gamma_0;\beta)$ and $\mathcal{T}_{\mathcal{Y}}^{df}(\gamma_0)$ respectively for some known functions $\mathcal{T}_{\mathcal{X}}^{df}(\cdot;\beta)$ and $\mathcal{T}_{\mathcal{Y}}^{df}$. Given a preliminary consistent estimator of γ_0^{df}, denoted by $\widehat{\gamma}^{df}$, we can define an estimation criterion where $\mathcal{X}^{df}(\beta)$ and $\mathcal{Y}^{df}$ are replaced by $\widehat{\mathcal{X}}^{df}(\beta) = \mathcal{T}_{\mathcal{X}}^{df}(\widehat{\gamma};\beta)$ and $\widehat{\mathcal{Y}}^{df} = \mathcal{T}_{\mathcal{Y}}^{df}(\widehat{\gamma})$ respectively, so that

$$\widehat{\mathcal{S}}^{df}(\beta;\widehat{\mathcal{W}}^{df}) = (\widehat{\mathcal{Y}}^{df} - \widehat{\mathcal{X}}^{df}(\beta))^\top\widehat{\mathcal{W}}^{df}(\widehat{\mathcal{Y}}^{df} - \widehat{\mathcal{X}}^{df}(\beta)). \tag{23}$$

where $\widehat{\mathcal{W}}^{df}$ is a positive definite matrix. An asymptotic least square estimator can then be defined to minimize $\widehat{\mathcal{S}}^{df}(\beta;\widehat{\mathcal{W}}^{df})$. However, no closed-form estimator generally exists in this case. For efficient estimation, the weighting matrix needs to converge in probability to the inverse of the asymptotic variance of $\sqrt{N}(\widehat{\mathcal{Y}}^{df} - \widehat{\mathcal{X}}^{df}(\beta_0))$, which can be constructed from any consistent estimator of β_0.

6 Numerical Section

We illustrate the use of our proposed estimators for the switching cost and discount factor as described in the previous section.

We begin with a Monte Carlo study. Our simulation design is taken from Pesendorfer and Schmidt-Dengler (2008, Section 7). Consider a two-firm dynamic entry game. In each period t, each firm i has two possible choices, $a_{it} \in \{0,1\}$. Observed state variables are previous period's actions, $w_t = (a_{1t-1}, a_{2t-1})$. Firm $1'$s period payoffs are described as follows:

$$\pi_{1,\theta}(a_{1t}, a_{2t}, x_t) = a_{1t}(\mu_1 + \mu_2 a_{2t}) + a_{1t}(1 - a_{1t-1})F + (1 - a_{1t})a_{1t-1}W, \tag{24}$$

where $\theta = (\mu_1, \mu_2, F, W)$ denote respectively the monopoly profit, duopoly profit, entry cost and scrap value. Each firm also receives additive private shocks that are i.i.d. $\mathcal{N}(0,1)$. The game is symmetric and Firm's 2 payoffs are defined analogously.

We generate the data with $(\mu_1, \mu_2, F, W) = (1.2, -1.2, -0.2, 0.1)$ and set $\beta = 0.9$. There are three distinct equilibria for this game, one of which is symmetric. We generate the data using the symmetric equilibrium. W is assumed to be known, as done in Pesendorfer and Schmidt-Dengler (2008) W is assumed known as it cannot be identified jointly with F. We use equation (22) to estimate F, which has closed-form without any optimization. In order to estimate β we need estimators for μ_1 and μ_2 that do not depend on the discount factor. We denote the sample size by N. We use $\mu_1 + A_{1N}/\sqrt{N}$ and $\mu_2 + A_{2N}/\sqrt{N}$, where (A_{1N}, A_{2N}) are bivariate independent standard normal variables, the $\sqrt{N}-$scaling ensures the (sampling) errors converge to zero at a parametric rate as one would expect in empirical applications. We also report the estimates of F using the estimator in Sanches, Silva and Srisuma (2013) that require an assumption on β, for a range of value of β, to see the effect from assuming an incorrect discount factor and also to compare with our closed-form estimator when the true discount factor is known. We compute each estimator with an Idn and Opt weighting matrices. For each sample size $N = 1000, 10000, 100000$, using 1000 simulations. For the sake of space we only report the mean and standard deviation and the mean squared error (as these estimators are regular parametric estimators). Table 1 and Table 2 give the results for F and β respectively.

[Tables 1 and 2 about here]

7 Concluding Remarks

We show components of the payoff functions that can be interpreted as switching costs can be identified under weaker conditions than previously. Our identification strategy for the switching costs can also be applied to different setups, such as games with absorbing states (such as permanent exits) as well as switching costs from further periods into the past. When other components of the payoff functions can be identified independently elsewhere, the discount factor can also be identified. Our identification strategy also suggests a new way to estimate games, nonparametrically or otherwise, with attractive features that mimic the identification results.

Our results also immediately accommodate more general models with unobserved heterogeneity as long as the choice and transition probabilities can be nonparametrically identified (Kasahara and Shimotsu (2009)). And we expect the idea behind our identification results to be valid more generally when the observed state variables contains continuously distributed variables, by replacing various matrices with linear operators. However, sufficient conditions for identification become harder to check. Furthermore, the corresponding estimation problem also becomes more complicated as it involves estimating infinite dimensional parameters (e.g. see Bajari et al. (2009), and Srisuma and Linton (2012)).

Appendix

Absorbing States

Our strategy to identify switching costs also allows for models with absorbing states. For simplicity consider an entry game such that, if a player (potential entrant or incumbent) chooses to not enter a market at a particular time period she cannot enter ever after (i.e. model with an absorbing state). In this case we can simplify the notation slightly, with an abuse of notation, by writing $\mu_i(a_{it}, a_{-it}, x_t) = a_{it} \cdot \mu_i(a_{-it}, x_t)$. (We maintain assumptions M1 - M4 and N1 - N2.) Particularly the value function for a potential entrant becomes,

$$V_i(x,(0,a_{-i}),\varepsilon_i) = \max \begin{cases} E[\mu_i(a_{-it},x_t)\,|x_t = x, w_t = (0,a_{-i})] + EC(x,(0,a_{-i})) & \text{, if enter} \\ \quad +\beta E[V_i(x_{t+1},w_{t+1},\varepsilon_{it+1})\,|x_t = x, w_t = (0,a_{-i}), a_{it} = 1] + \varepsilon_i & \\ 0 & \text{, if not enter} \end{cases},$$

and for an incumbent,

$$V_i(x,(1,a_{-i}),\varepsilon_i) = \max \begin{cases} E[\mu_i(a_{-it},x_t)\,|x_t = x, w_t = (1,a_{-i})] & \text{, if stay in} \\ \quad +\beta E[V_i(x_{t+1},w_{t+1},\varepsilon_{it+1})\,|x_t = x, w_t = (1,a_{-i}), a_{it} = 1] + \varepsilon_i & \\ SV(x,(1,a_{-i})) & \text{, if exit} \end{cases}.$$

The above specification of the value function allows our argument in the main to proceed with no modification. Particularly, the corresponding Opt decision rule depends on whether the player has already entered the market. So that, for a potential entrant,

$$\begin{aligned} v_i(a_i,x,(0,a_{-i})) &= E[a_{it} \cdot \mu_i(a_{-it},x_t)\,|a_{it} = a_i, x_t = x, w_t = (0,a_{-i})] + a_{it} \cdot EC(x,(0,a_{-i})) \\ &\quad +\beta E[\widetilde{m}_i(a_{it},a_{-it},x_t)\,|a_{it} = a_i, x_t = x, w_t = (0,a_{-i})], \end{aligned}$$

and analogously for the incumbent,

$$\begin{aligned} v_i(a_i,x,(1,a_{-i})) &= E[a_{it} \cdot \mu_i(a_{-it},x_t)\,|a_{it} = a_i, x_t = x, w_t = (1,a_{-i})] - (1-a_{it}) \cdot SV(x,(1,a_{-i})) \\ &\quad +\beta E[\widetilde{m}_i(a_{it},a_{-it},x_t)\,|a_{it} = a_i, x_t = x, w_t = (1,a_{-i})], \end{aligned}$$

where, in both cases, $\widetilde{m}_i$ is defined as previously (see Section 4.1). So that

$$\begin{aligned} \Delta v_i(x,(0,a_{-i})) &= E[\mu_i(a_{-it},x_t)\,|x_t = x, w_t = (0,a_{-i})] + EC(x,(0,a_{-i})) \\ &\quad +\beta E[\Delta\widetilde{m}_i(a_{-it},x_t)\,|x_t = x, w_t = (0,a_{-i})]. \end{aligned}$$

Analogously, for an incumbent

$$\begin{aligned}\Delta v_i\left(x,\left(1,a_{-i}\right)\right) &= E\left[\mu_i\left(a_{-it},x_t\right)|x_t=x,w_t=\left(1,a_{-i}\right)\right]-SV\left(x,\left(0,a_{-i}\right)\right)\\ &\quad+\beta E\left[\Delta\widetilde{m}_i\left(a_{-it},x_t\right)|x_t=x,w_t=\left(1,a_{-i}\right)\right].\end{aligned}$$

Therefore we can define $\lambda_i\left(a_{-i},x\right)=\mu_i\left(a_{-i},x\right)+\beta\Delta\widetilde{m}_i\left(a_{-i},x\right)$, then vectorize, and form an identical equation (17) in Example 4.

References

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

[2] Aguirregabiria, V. and J. Suzuki (2014): "Identification and Counterfactuals in Dynamic Models of Market Entry and Exit," *Quantitative Marketing and Economics*, **12**, 267-304.

[3] Athey, S. and P. Haile (2002): "Identification of Standard Auction Models," *Econometrica*, **70**, 2107-2140

[4] Athey, S. and P. Haile (2007): "Nonparametric Approaches to Auctions," *Handbook of Econometrics*, vol. 6A, eds. J. Heckman and E. Leamer. North-Holland.

[5] Bajari, P., C.L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, **75**, 1331-1370.

[6] Bajari, P., V. Chernozhukov, H. Hong and D. Nekipelov (2009): "Identification and Efficient Estimation of a Dynamic Discrete Game," *Working Paper, University of Minnesota.*

[7] Beresteanu, A., P. Ellickson and S. Misra (2010): "The Dynamics of Retail Oligopolies," *Working Paper, University of Rochester, Simon School of Business.*

[8] Berry, S. T. and P. A. Haile (2010): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," *Cowles Foundation Discussion Paper No. 1718.*

[9] Berry, S. T. and P. A. Haile (2012): "Identification in Differentiated Products Markets Using Market Level Data," *Cowles Foundation Discussion Paper No. 1744R.*

[10] Blevins, J. (2013): "Nonparametric Identification of Dynamic Decision Processes with Discrete and Continuous Choices," *forthcoming in Quantitative Economics.*

[11] Collard-Wexler, A. (2013): "Demand Fluctuations and Plant Turnover in the Ready-Mix Concrete Industry," *forthcoming in Econometrica.*

[12] Dunne, T., S. Klimek, M. J. Roberts and Y. D. Xu (2013): "Entry, Exit, and the Determinants of Market Structure," *forthcoming in RAND Journal of Economics..*

[13] Fan, Y. and M. Xiao (2012): "Competition and Subsidies in the Deregulated U.S. Local Telephone Industry," *Working Paper, University of Michigan, Department of Economics.*

[14] Frisch, R. and F.V. Waugh (1933): "Partial Time Regressions as Compared with Individual Trends," *Econometrica*, **1**, 387-401.

[15] Gourieroux, C. and A. Monfort (1995): *Statistics and Econometric Models: Volume 1, General Concepts, Estimation, Prediction and Algorithms*, Themes in Modern Econometrics, Cambridge University Press.

[16] Gowrisankaran, G., C. Lucarelli, P. Schmidt-Dengler and R. Town (2010): "Government Policy and the Dynamics of Market Structure: Evidence from Critical Access Hospitals," *Eleventh CEPR Conference on Applied Industrial Organization, Toulouse School of Economics.*

[17] Guerre, E., I. Perrigne and Q. Vuong (2000): "Opt Nonparametric Estimation of First-Price Auctions," *Econometrica*, **60**, 525-574.

[18] Hotz, V., and R.A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, **60**, 497-531.

[19] Hu, Y. and Y. Sasaki (2013): "Closed-form Estimation of Dynamic Discrete Choice Models with Unobserved State Variables," *Working Paper, Johns Hopkins University.*

[20] Igami, M. (2012): "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry," *Working Paper, Yale University.*

[21] Kasahara, H. and K. Shimotsu (2009), "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, **77**, 135-175.

[22] Lin, H. (2012): "Quality Choice and Market Structure: A Dynamic Analysis of Nursing Home Oligopolies," *Working Paper, Indiana University, Kelley School of Business.*

[23] Magnac, M. and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, **70**, 801-816.

[24] Marshall, G. (2013): "Search and Wholesale Price Discrimination," *Working Paper, Northwestern University.*

[25] Matzkin, R.L. (1991): "Semiparametric Estimation of Monotone and Concave Utility Functions for Polychotomous Choice Models," *Econometrica*, **59**, 1315–1327.

[26] Norets, A. and S. Takahashi (2013): "On the Surjectivity of the Mapping Between Utilities and Choice Probabilities," *Quantitative Economics*, **4**, 149–155.

[27] Norets, A. and X. Tang (2013): "Semiparametric Inference in Dynamic Binary Choice Models," *forthcoming in Review of Economic Studies.*

[28] Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Example)," *RAND Journal of Economics*, **38**, 373-399.

[29] Pesendorfer, M. and P. Schmidt-Dengler (2003): "Identification and Estimation of Dynamic Games," *NBER Working Paper.*

[30] Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimator for Dynamic Games," *Review of Economics Studies*, **75**, 901-928.

[31] Ryan, S. P. (2012): "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, **80**, 1019-1061.

[32] Sanches, F.M., D. Silva Jr and S. Srisuma (2013): "An Alternative Asymptotic Least Squares Estimator for Dynamic Games," *mimeo.*

[33] Sanches, F.M., D. Silva Jr and S. Srisuma (2014): "Bank Privatization and the Supply of Banking Services: Evidence from a Dynamic Structural Model," *Working Paper, LSE.*

[34] Snider, C. (2009): "Predatory Incentives and Predation Policy: The American Airlines Case," *Working Paper, UCLA, Department of Economics.*

[35] Srisuma, S. and O.B. Linton (2012): "Semiparametric Estimation of Markov Decision Processes with Continuous State Space," *Journal of Econometrics*, **166**, 320-341.

[36] Suzuki, J. (2013): "Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry," *forthcoming in International Economic Review.*

[37] Sweeting, A. (2013): "Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry," *forthcoming in Econometrica.*