Identifying Dynamic Games with Switching Costs*

Fabio A. Miessi Sanches† Daniel Silva Junior‡
University of São Paulo London School of Economics

Sorawoot Srisuma§
University of Surrey

September 15, 2015

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. When possible, in the spirit of structural estimation, economic theory can be used 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, exit, or other generic adjustment 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. As an illustration we use the data from Ryan (2012) to estimate a version of dynamic game played by firms that produce Portland cement over the period that spans the implementation of the 1990 Clean Air Amendments Act (1990 CAAA). Our finding supports his result that the entry barrier following the 1990 CAAA has increased.

JEL Classification Numbers: C14, C25, C61

Keywords: Dynamic Discrete Choice Games, Identification, Estimation, Switching Costs

---

*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 CREATES, 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

A structural study involves modeling the economic problem of interest base on some primitives that govern an economic model. The empirical goal is to estimate them for counterfactual analysis. The structural model of interest in this paper is a class of dynamic discrete choice games that generalizes the single agent Markov decision models (Rust (1994)). These games have been used to study several interesting counterfactual experiments involving multiple economic agents making decisions over time.[1] A key aspect to the modeling decision that precedes estimation involves the issue of identification. Nonparametric identification of a structural model informs us whether or not the parameter of interest can be consistently estimated from an ideal data set without introducing additional parametric or other restrictions. Recent reviews of the identification and estimation of these games, as well as 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 constructive, and is related to the development of several general estimation methodologies.[3]

A common feature of the aforementioned works on identification aims to identify the entire payoff function. 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

[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).

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 functions are treated as reduced forms as they are identified by the data.[4]

A recurring feature of many empirical games employed in practice involves costs that arise from players choosing different actions from the past. Prominent examples include entry costs and scrap values in models with entry, as well as generic adjustment costs in capacity and pricing games and decision problems. We refer to these as *switching costs*. Switching costs are usually parts of what are often called dynamic parameters of the model. They generally cannot be treated as reduced forms since economic theory rarely provides guidance on how they are determined. Dynamic parameters have to be 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, switching costs can generally be nonparametrically identified independently of the discount factor and other components of the payoffs. Our identification strategy is constructive and it 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 two main assumptions from the onset. More specifically 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 shall consider a payoff function that admits the following decomposition:

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

The payoff structure above is in fact prevalent and it encompasses numerous payoff functions specified in practice. 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 sole purpose is 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

[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 the context of an auction. When bids data are available and the auction format is known, the expected revenue can be estimated nonparametrically; see Athey and Haile (2002), and Guerre, Perrigne and Vuong (2000).

η_i). In addition, we require that x_{t+1} is independent of w_t conditional on x_t and a_t. This conditional independence assumption is made only on the observables and hence is testable. Furthermore, such assumption is assumed in any empirical model that treats all observed state variables apart from past actions exogenously; e.g. see Pesendorfer and Schmidt-Dengler (2003) and Ryan (2012).

We provide conditions so that ϕ_i can be identified independently of μ_i and the discount factor, denoted by β. Furthermore, ϕ_i can be written in *closed-form* in terms of the transition and conditional choice probabilities that are observed from the data. The extent of the implication of our results depends on the empirical problem at hand and data availability.

1. The best case scenario is when μ_i can be identified directly from the observed data. Our identification result of ϕ_i then 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 strategy is constructive. The closed-form expression of the probabilities suggests switching costs can be estimated directly without any optimization. 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 (2014) 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 that can be computed using a closed-form expression. Furthermore, if μ_i is also identified and estimable directly from the data then we can estimate π_i independently of β. In any case the closed-form estimation of the switching costs offer a practical way to reduce the dimensionality of the estimation problem. Particularly, without any restrictions, the number of switching cost parameters for each player grows at the rate of the number of actions squared.

The discount factor is a primitive of the model that is traditionally assumed to be known in the study of identification in dynamic games. Consequently empirical work often simply assigns various numbers for this when it comes to estimation. One reason for this can perhaps be traced to the generic non-identification result of the discount factor for a single agent dynamic decision model described in Manski (1993). However, in the presence of additional structures on the payoff functions the discount factor can be identified; as Magnac and Thesmar (2002) illustrate for a two-period model. The caveat is that additional structures should be carefully motivated.[5] We build on

[5]One recent example can be found in Fang and Wang (2014), who use a particular exclusion restriction combined with a conditional independence assumption to identify the discount factor for a dynamic decision problem where economic agents use hyperbolic discounting.

our positive identification results of the switching costs, and provide a sufficient condition to identify β when μ_i can be identified independently of β. Our identification result is again constructive and we suggest a class of natural estimators for the discount factor.

The innovation of our work lies in the identification strategy of the switching costs. The seminal work of Magnac and Thesmar (2002) show in a single agent decision model that the identification of the model primitives can be analyzed from the normalized expected payoffs that are identified from the data (Hotz and Miller (1993)). Pesendorfer and Schmidt-Dengler (2008) and Bajari et al. (2009) extend this idea to a dynamic game setting. Particularly, when all primitives apart from π_i are known, the expected payoffs can be written as a linear transform of π_i 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 show that the interpretable and prevalent decomposition of the payoffs, coupled with conditional independence, can restore the linear structure for the switching costs that can then be used for identification. The combination of exclusion and independence restrictions is a classic tool used to identify structural econometric models (see Matzkin (2007) for others).[6] Particularly when the parameter of interest enters the identifying equation linearly, it can typically be identified by some form of differencing. We show the switching costs can be identified from a particular linear combination of equations, which can be characterized by a linear transformation in the form of a projection matrix.

The decomposition of payoffs and exploiting other nonparametric structures base on economic reasoning is a constructive way to identify structural models. Some recent explorations in this direction for other models can be found for example in Berry and Haile (2010,2012) and Lewbel and Tang (2013). The paper that is closest to ours in this regard is the recently published work by Aguirregabiria and Suzuki (2014) on single agent decision problems with entry. However, the content and motivation of our work and theirs are substantially different. They motivate their studies base on the notion that switching costs generally cannot all be jointly identified without making normalization assumptions. Their main concern is the identifiability and interpretation of certain counterfactual objects for the purpose of policy analysis under different normalization choice made on parts of the payoff functions. Our work on the other hand focuses on identification and estimation of the switching costs that can be identified, where normalization is taken as part of modeling decision, and the discount factor. Interestingly, despite their paper explicitly assuming the knowledge of the discount factor throughout, a careful inspection of their non-identification result (Proposition 2) will also suggest that the switching costs in their model can be identified independently of the discount

[6] Blevins (2013) and Chen (2014) also show how exclusion and independence restrictions can be used to identify the distribution of unobserved state variables in a closely related single agent decision problem.

value under some normalization (albeit by a different method).[7] This particular implication can be seen as a special case of our more general result.[8]

Some form of normalization (or restriction) is in general necessary to employ our identification strategy. We characterize the degree of underidentification of the switching costs when no other structure on ϕ_i is imposed beyond the definition of a switching cost. Our need to normalize may not be surprising since some form of normalization is prevalent in the empirical literature. In practice normalizations are often made on informal arguments that certain components of the payoff function cannot be identified. The only formal identification result in a related model that we are aware of is given by Aguirregabiria and Suzuki (2014, Proposition 2) that suggests entry costs and scrap values cannot be jointly identified when nothing else is known about the payoff function for a single agent's decision problem with entry. A common normalization choice in practice is to assign zero payoffs when a player chooses an "outside option". An analogous normalization in our framework would then be to impose that a switching cost associated with a player choosing the outside option is zero regardless of her previous action. Such normalization will be sufficient but not necessary. However, importantly, we wish to emphasize here that we are not advocating ad hoc normalizations for a computational gain, or a chance to estimate the discount factor that would otherwise likely have to be calibrated/normalized. As always the case with structural modeling, an ideal approach is to use economic reasoning and data (when available) to impose additional structures that are specific to each empirical application. For instance, certain switching costs may naturally be argued to be zero (such as manufacturers in a pricing game who bear the cost of introducing promotions but there is no cost reverting to the original price, see Myśliwski et al. (2015)). Alternatively other restrictions on switching costs will also suffice. E.g. equality of switching costs from one option to another and vice versa may be reasonable in some applications such as those with a traditional adjustment menu costs (see Slade (1998)).

We provide a small Monte Carlo study to show that our estimator is consistent and robust against the misspecification of the discount factor unlike some other existing estimators. We then use the dataset from Ryan (2012) to estimate a dynamic game played between firms in the Portland cement industry. In our version of the discrete game, firms choose whether to enter the market as well as decide on the capacity level of operation. Our model contains 25 switching cost parameters that we estimate without using any numerical optimization procedure. We assume firms compete in a

[7] We thank an anonymous referee for pointing this out.

[8] Beside the different focus, Aguirregabiria and Suzuki (2014) concentrate on single agent models with entry decisions that is a special case of a game with a general switching cost structure. Their results are also derived under an assumption that $\{x_t\}$ is a strictly exogenous (first order Markov) process. Specifically this implies x_{t+1} is independent of a_t conditional on x_t in addition to the conditional independence assumption that we impose.

capacity constrained Cournot competition that accounts for the remaining part of the payoffs. The Cournot profit is based on the demand and cost functions estimated in a static setting without assuming the knowledge of the discount factor. This enables us to estimate the discount factor. We estimate the switching costs and the discount factor of the model twice, once each using the data from before and after 1990 that coincides with the date of the 1990 Clean Air Act Amendments (1990 CAAA). Our switching costs estimates generally make economic sense in terms of the sign and relative magnitude. They show that firms entering the market that can operate at a higher capacity level incurs larger cost, and suggest that increasing capacity level is generally costly while a reduction can return some revenue. We also find that entry costs are generally much higher after the 1990 CAAA, which supports Ryan's key finding. Our estimate of the discount value is 0.64 in both periods, although lower than traditionally assumed values, suggesting that the firms do not change their discounting rule following the 1990 CAAA.

Throughout this work 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.[9] 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 idea behind our identification strategy of the switching costs and highlights key aspects of subsequent sections using a simple two-player entry game in Pesendorfer and Schmidt-Dengler (2008). We define the theoretical model and state the modeling assumptions in Section 3. Section 4 contains the identification results. Section 5 provides a discussion on how our identification strategy can be used for estimation. Section 6 is the numerical part of the paper that illustrates the use of our estimator with simulated and real data. Section 7 concludes.

2 Preview of Identification Strategy

Consider a two-player repeated entry game in Pesendorfer and Schmidt-Dengler (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 i.i.d. shock $\varepsilon_{it} = (\varepsilon_{it}(0), \varepsilon_{it}(1))$ that are independent across the players. In this model w_t serves as public information and is observed by the econometricians while ε_{it} is only observed by player i. Under

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

some standard conditions, the expected payoff from choosing action a_i is $v_i(a_i, w_t) + \varepsilon_{it}(a_i)$, where

$$v_i(a_i, w_t) = E[\pi_i(a_{it}, a_{-it}, w_t)|w_t, a_{it} = a_i] + \beta E[m_i(w_{t+1})|w_t, a_{it} = a_i], \text{ and} \tag{2}$$

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

In equilibrium $a_{it} = \alpha_i(w_t, \varepsilon_{it})$ for all i, t, where α_i denotes the player's optimal 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 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 be written in terms of choice probabilities (Hotz and Miller (1993)). Since the action space is finite, the relation between Δv_i and π_i can be summarized 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. The study of identification of games in Bajari et al. (2009) and Pesendorfer and Schmidt-Dengler (2008) then comes down to whether equation (3) has unique solution or not.

Next we impose some specific structure on the payoffs. The entry game of Pesendorfer and Schmidt-Dengler (2008) has switching costs components, 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. Since m_i denotes the expected discounted payoffs, it depends on β as well as π_i. Therefore Δv_i cannot be written as a linear function of both β and π_i. However, we can make the equation linear in the switching costs through an aid of a nuisance function defined as $\lambda_i(a_{-it}) = \Delta\mu_i(a_{-it}) + \beta\Delta m_i(a_{-it})$, 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 and scrap value from the present period are now additively separable from the other flow profits. Since the support of w_t is $\{(0,0),(0,1),(1,0),(1,1)\}$, $\{\Delta v_i(w)\}_w$ can be represented using 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 subject to a normalization in closed-form. The need to normalize in this context is familiar in empirical work. For instance Pesendorfer and Schmidt-Dengler (2003,2008) normalize SV_i to be zero. We delay a fuller discussion regarding normalization and other intuition in subsequent sections.

The sample counterpart of (6) provides a simple estimator for each θ_i^k that has a closed-form. However, such estimator is inefficient. In Section 5 we show such closed-form estimator is a member of a class of *asymptotic least squares estimators* in the sense described in Gourieroux and Monfort (1995). We also identify and describe how to estimate the efficient estimator of this class. 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, see the Appendix).

3 Model and Assumptions

We consider a game with I players, indexed by $i \in \mathcal{I} = \{1, \ldots, I\}$, who compete 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 we define $w_t \equiv a_{t-1}$. (x_t, w_t) are public information that are common knowledge to all players and observed by the econometrician, while $\varepsilon_{it} = (\varepsilon_{it}(0), \ldots, \varepsilon_{it}(K)) \in \mathbb{R}^{K+1}$ is private information only observed by player i. We define $s_t \equiv (x_t, w_t)$ and $\varepsilon_t \equiv (\varepsilon_{1t}, \ldots, \varepsilon_{It})$. 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. For examples, see Aguirregabiria and Mira (2007), Bajari, Benkard and Levin (2007), Pakes, Ostrovsky and Berry (2007), Pesendorfer and Schmidt-Dengler (2008).

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 $J < \infty$.*

The game procedes as follows. At time t, each player observes s_{it} and then chooses a_{it} simultaneously. Action and state variables at time t affects s_{it+1}. Upon observing their new states the players choose their actions again and so on. 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,

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

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 (see 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}), which is consistent with a single equilibrium play. The primitives of the game under this setting consists of $(\{\pi_i\}_{i=1}^{I}, \beta, Q, G)$. Throughout the paper we shall also assume G and Q to be known (the former can be identified from the data). Next, we formally introduce the specific structures of the payoffs and a conditional independence assumption alluded in the Introduction. 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 $w \in W_{\eta_i}^0(a_i, x)$, where $W_{\eta_i}^d(a_i, x) \equiv \{w \in A^I : \eta_i(a_i, x, w) = d\}$ for $d = 0, 1$.

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. First is μ_i that does not depend on w_t. η_i is a known function, chosen by the researcher, that indicates a switching cost. When switching cost is present, by definition, minimally for some a_i, $W^0_{\eta_i}(a_i,\cdot)$ will be non-empty since it contains $w \in A^I$ such that the action of player's i coincides with a_i, so it is possible to consider distinguishing μ_i from η_i. We define ϕ_i to be zero whenever η_i takes value zero. When η_i takes value one, indicating a presence of a switching cost, 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 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 actions and other state variables.

N2 imposes that knowing actions from the past does not help predict future state variables when the present action and (observable) state variables 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 addition, unlike M2, 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 studies in the literature. Here we provide some examples of $\phi_i \cdot \eta_i$ and $W^d_{\eta_i}$.

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

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

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

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

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

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

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

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

Here $SC_i(a'_i, a''_i, x_t, a_{-it-1})$ denotes a switching cost incurs to player i from choosing $a_{it} = a'_i$ when $a_{it-1} = a''_i$ when other states are x_t, a_{-it-1}. So for all x, using just the definition of a switching cost we can set $SC_i(a'_i, a'_i, x, a_{-i}) = 0$ for all a'_i by definition, $W^1_{\eta_i}(a_i, x) = \{w = (a'_i, a_{-i}) : a'_i \in A \backslash \{a_i\}, a_{-i} \in A^{I-1}\}$ and $W^0_{\eta_i}(a_i, x) = \{w = (a_i, a_{-i}) : a_{-i} \in A^{I-1}\}$ 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.

We end this section by providing an intuition as to why N1 and N2 are helpful for identifying the switching costs. The essence of our identification strategy is most transparent in a single agent decision problem. For the moment suppose $I = 1$. 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 integrated value function, $E\left[V(s_t)\,|x_t = x, w_t = w\right]$. N1 imposes separability and exclusion restrictions of the following type:

$$\pi(a,x,w) = \mu(a,x) + \phi(a,x,w;\eta)\cdot\eta(a,x,w),$$

where ϕ is a known indicator such that $\phi(a,x,w;\eta) = 0$ whenever $a \neq w$. Therefore the contribution from past action can be separated from the present one within a single time period. The direct effect of past action is also excluded from the future expected payoff under N2, as $E\left[m(x_{t+1},w_{t+1})\,|a_t,x_t,w_t\right]$ becomes $E\left[m(x_{t+1},a_t)\,|a_t,x_t\right]$. Therefore we can write

$$v(a,x,w) = \lambda(a,x) + \phi(a,x,w;\eta)\cdot\eta(a,x,w),$$

where $\lambda(a,x)$ is a nuisance function that equals to $\mu(a,x) + \beta E\left[m_i(x_{t+1},w_{t+1})\,|a_{it} = a_i, x_t = x\right]$. Any variation in $v(a,x,w)$ induced by changes in w while holding (a,x) fixed can be traced only to changes in $\eta(a,x,w)$. Since λ is a free parameter, the switching costs can be identified upto a location normalization by differencing over the support of w; e.g. through $(v(a,x,w) - v(0,x,w)) - (v(a,x,w_0) - v(0,x,w_0))$ for some reference point w_0. Our insight is this intuition can be generalized and applied to identify switching costs in dynamic games. However, the way to difference out the nuisance function becomes more complicated. Particularly the nuisance function will also vary for different past action profile since we have to integrate out other players' actions using the equilibrium beliefs that depends on past actions. Relatedly there are more degree of freedoms to be dealt with as the nuisance function contains more arguments. The precise form of differencing required can be formalized by a projection that enables the identification of the switching costs upto some normalizations.[10] We provide precise conditions for what can be identified from ϕ_i in the next section.

[10]Mathematically, for fixed a, x, our identification problem under N1 and N2 in a single agent case is equivalent to identifying g_2 that satisfies a linear relation:

$$g_1(w) = c + g_2(w),$$

4 Main Results

We first present our identification results of the switching costs that do not assume the knowledge of the discount factor. Then we provide the identification of the discount factor.

4.1 Identifying the Switching Costs

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, using the law of iterated expectation, $\widetilde{m}_i(a_i, a_{-i}, x) \equiv E[m_i(x_{t+1}, a_{it}, a_{-it}) | a_{it} = a_i, a_{-it} = a_{-i}, x_t = x]$. Then, for $a_i > 0$, let $\Delta v_i(a_i, x, w) \equiv v_i(a_i, x, w) - v_i(0, x, w)$, $\Delta\mu_i(a_i, a_{-i}, x) \equiv \mu_i(a_i, a_{-i}, x) - \mu_i(0, a_{-i}, x)$, 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 differences of switching costs as

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

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

for a *known* g_1 and an *unknown* constant c. In the case of a game, the equation above generalizes to

$$g_1(w) = \int c(x) h(dx|w) + g_2(w),$$

where the unknown constant is replaced by a generic linear transform (an expectaion) of some unknown function c.

Example 1 (Entry Cost, Cont.): Here the only $a_i > 0$ is $a_i = 1$. Since $W^1_{\eta_i}(0,x)$ is empty $W^{\Delta}_{\eta_i}(1,x) = W^1_{\eta_i}(1,x)$, 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, $W^{\Delta}_{\eta_i}(1,x) = W^1_{\eta_i}(0,x)$, 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 $a_i > 0$, based on the definition of a switching cost alone, both $W^1_{\eta_i}(a_i,x)$ and $W^1_{\eta_i}(0,x)$ can be non-empty. So for all i, a_{-i}, x such that $a'_i \neq a_i$:

$$
\begin{aligned}
\phi_{i,\eta_i}(a_i,x,w) &= SC_i(a_i,0,x,a_{-i}) \text{ when } w = (0,a_{-i}), \\
\phi_{i,\eta_i}(a_i,x,w) &= -SC_i(0,a_i,x,a_{-i}) \text{ when } w = (a_i,a_{-i}), \\
\phi_{i,\eta_i}(a_i,x,w) &= SC_i(a_i,a'_i,x,a_{-i}) - SC_i(0,a'_i,x,a_{-i}) \text{ when } w = (a'_i,a_{-i}) \text{ for } a'_i \neq a_i \text{ or } 0.
\end{aligned}
\tag{11}
$$

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 $\left\{\phi_{i,\eta_i}(a_i, x, a'_i, a_{-i})\right\}_{a_i, a'_i \in A \times A}$.

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*:*

$$
\Delta v_i(a_i,x,w) = E\left[\lambda_i(a_i,a_{-it},x_t)\,|x_t = x, w_t = w\right] + \sum_{w' \in W^{\Delta}_{\eta_i}(a_i,x)} \phi_{i,\eta_i}(a_i,x,w') \cdot \mathbf{1}\left[w = w'\right], \tag{12}
$$

where

$$
\lambda_i(a_i,a_{-i},x) \equiv \Delta\mu_i(a_i,a_{-i},x) + \beta\Delta\widetilde{m}_i(a_i,a_{-i},x). \tag{13}
$$

Proof of Lemma 1: Using the law of iterated expectation, under M3 $E\left[V_i(s_{it+1})\,|a_{it} = a_i, x_t, w_t\right] = E\left[m_i(x_{t+1}, w_{t+1})\,|a_{it} = a_i, x_t, w_t\right]$, which simplifies further, after another application of the law of iterated expectation and N2, to $E\left[\widetilde{m}_i(a_i,a_{-it},x_t)\,|x_t,w_t\right]$. 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 (differenced) 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 (12), 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 (12) by a matrix equation.

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

$$\Delta \mathbf{v}_i (a_i, x) = \mathbf{Z}_i (x) \boldsymbol{\lambda}_i (a_i, x) + \mathbf{D}_i (a_i, x) \boldsymbol{\phi}_{i,\eta_i} (a_i, x), \tag{14}$$

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* $\left\{ \phi_{i,\eta_i} (a_i, x, w) \right\}_{w \in W^{\Delta}_{\eta_i}(a_i, x)}$*.*

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 now state our first result.

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; (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*

$$\boldsymbol{\phi}_{i,\eta_i} (a_i, x) = (\mathbf{D}_i (a_i, x)^\top \ \mathbf{M}_{\mathbf{Z}_i(x)} \mathbf{D}_i (a_i, x))^{-1} \mathbf{D}_i (a_i, x)^\top \ \mathbf{M}_{\mathbf{Z}_i(x)} \Delta \mathbf{v}_i (a_i, x). \tag{15}$$

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) then ensures $\mathbf{M}_{\mathbf{Z}_i(x)} \mathbf{D}_i (a_i, x)$ preserves the rank of $\mathbf{D}_i (a_i, x)$. Therefore $\mathbf{D}_i (a_i, x)^\top \mathbf{M}_{\mathbf{Z}_i(x)} \mathbf{D}_i (a_i, x)$ must be non-singular. 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 $\mathbf{Z}_i (x)$, thus violating the assumed rank condition. The proof is then completed by projecting the vectors on both sides of equation (14) by $\mathbf{M}_{\mathbf{Z}_i(x)}$and solve for $\boldsymbol{\phi}_{i,\eta_i} (a_i, x)$.■

Equation (15) directly generalizes equation (6) in Section 2. In order for condition (ii) in Theorem 1 to hold, it is necessary for researchers to impose some a priori structures 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 (14) can be written as

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

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 (14) can be written as

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

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 $\Delta\mathbf{v}_i(2,x) =$

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

$$\begin{bmatrix} \Delta v_i(2,x,(0,0)) \\ \Delta v_i(2,x,(0,1)) \\ \Delta v_i(2,x,(0,2)) \\ \Delta v_i(2,x,(1,0)) \\ \Delta v_i(2,x,(1,1)) \\ \Delta v_i(2,x,(1,2)) \\ \Delta v_i(2,x,(2,0)) \\ \Delta v_i(2,x,(2,1)) \\ \Delta v_i(2,x,(2,2)) \end{bmatrix} = \begin{bmatrix} P_{-i}(0|x,(0,0)) & P_{-i}(1|x,(0,0)) & P_{-i}(2|x,(0,0)) \\ P_{-i}(0|x,(0,1)) & P_{-i}(1|x,(0,1)) & P_{-i}(2|x,(0,1)) \\ P_{-i}(0|x,(0,2)) & P_{-i}(1|x,(0,2)) & P_{-i}(2|x,(0,2)) \\ P_{-i}(0|x,(1,0)) & P_{-i}(1|x,(1,0)) & P_{-i}(2|x,(1,0)) \\ P_{-i}(0|x,(1,1)) & P_{-i}(1|x,(1,1)) & P_{-i}(2|x,(1,1)) \\ P_{-i}(0|x,(1,2)) & P_{-i}(1|x,(1,2)) & P_{-i}(2|x,(1,2)) \\ P_{-i}(0|x,(2,0)) & P_{-i}(1|x,(2,0)) & P_{-i}(2|x,(2,0)) \\ P_{-i}(0|x,(2,1)) & P_{-i}(1|x,(2,1)) & P_{-i}(2|x,(2,1)) \\ P_{-i}(0|x,(2,2)) & P_{-i}(1|x,(2,2)) & P_{-i}(2|x,(2,2)) \end{bmatrix} \begin{bmatrix} \lambda_i(2,0,x) \\ \lambda_i(2,1,x) \\ \lambda_i(2,2,x) \end{bmatrix} \tag{16}$$

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

Clearly the required rank condition of Theorem 1 cannot hold in this case. 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 at least three restrictions. For example by normalizing one type of switching costs to be zero. More specifically suppose $SC_i(0,a_i,x,a_{-i}) = 0$ for all $a_i > 0$, then $\mathbf{D}_i(2,x)\,\boldsymbol{\phi}_{i,\eta_i}(2,x)$ becomes

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

and 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. Such normalization is an example of an exclusion restriction. A preferred scenario would be to use economic or other prior knowledge to assign values so known switching costs can be removed from the right hand side (RHS) of equation (16); as done in Section 2 (see equation (6)). Other restrictions, such as equality of switch costs so that the costs from switching to and from actions that may be reasonable in capacity or pricing games can be used instead of a direct normalization. For instance suppose that $SC_i(a_i, a_i', x, a_{-i}) = SC_i(a_i', a_i, x, a_{-i})$ whenever $a_i \neq a_i'$, then $\mathbf{D}_i(2,x)\,\boldsymbol{\phi}_{i,\eta_i}(2,x)$ becomes

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

and we expect the rank condition to generally be satisfied. Analogous conditions and comments apply for $\Delta\mathbf{v}_i(1,x)$.

Comments on Theorem 1:

(i) Pointwise Identification. Our result is obtained pointwise for each $i, a_i > 0$ and x. Therefore the finite support assumption in M4 is not necessary. However, the theoretical and practical aspects of estimating models where the observable state has a continuous component becomes a semiparametric one and is more difficult. See Bajari et al. (2009) and Srisuma and Linton (2012).

(ii) Underidentification. In order to apply Theorem 1 a 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 (14) 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 the cardinality of $W^{\Delta}_{\eta_i}(a_i,x)$ equals) $\rho(\mathbf{D}_i(a_i,x)) \leq (K+1)^I - 1$. 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 $(K+1)^{I-1}$. Then $\rho(\mathbf{D}_i(a_i,x))$ is bounded above by $K(K+1)^{I-1}$. The order condition indicates the degree of underidentification if one aims to identify all switching costs without any other structure beyond the definition of a switching cost.

(iii) Normalization and Other Restrictions. The maximum number of parameters one can write down in equation (14) using the full generality of the definition of a switching cost is $(K+1)^I$; see (11). Therefore the previous comment suggests that $(K+1)^{I-1}$ restrictions will be required for a positive identification result if no further structure on the switching costs is known. One solution to this is normalization. Since $(K+1)^{I-1}$ equals also the cardinality of A^{I-1}, one convenient normalization restriction that will suffice here is to set values of switching cost associated with a single action. For instance the assumption that costs of switching to action 0 from any other action is zero will suffice. Note that such assumption 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)). Nevertheless an ad hoc normalization is not an ideal solution. A more optimal solution is to appeal to prior economic knowledge to impose additional structure on the switching costs such as equality restrictions as illustrated above with Example 3.

In practice researchers can impose prior knowledge restrictions directly on $\boldsymbol{\phi}_{i,\eta_i}$. This can be seen as part of the modeling decision. Next we show restrictions across all choice set can be used simultaneously.

Assumption R1 (Equality Restrictions): *For all i,x, there exists a $K(K+1)^I$ 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:*

$$\Delta\mathbf{v}_i(x) = (I_K\otimes\mathbf{Z}_i(x))\boldsymbol{\lambda}_i(x) + \widetilde{\mathbf{D}}_i(x)\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x), \tag{17}$$

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 identity 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 all 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, we caution that 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 $K = 1$ in the context of an entry game.

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 (17) (in Lemma 3) is

$$\begin{bmatrix}\Delta v_i(x,(0,0))\\ \Delta v_i(x,(0,1))\\ \Delta v_i(x,(1,0))\\ \Delta v_i(x,(1,1))\end{bmatrix} = \begin{bmatrix}P_{-i}(0|x,(0,0)) & P_{-i}(1|x,(0,0))\\ P_{-i}(0|x,(0,1)) & P_{-i}(1|x,(0,1))\\ P_{-i}(0|x,(1,0)) & P_{-i}(1|x,(1,0))\\ P_{-i}(0|x,(1,1)) & P_{-i}(1|x,(1,1))\end{bmatrix}\begin{bmatrix}\lambda_i(0,x)\\ \lambda_i(1,x)\end{bmatrix} + \begin{bmatrix}1 & 0\\ 1 & 0\\ 0 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}EC_i(x)\\ -SV_i(x)\end{bmatrix}. \tag{18}$$

Note that the order condition is now satisfied. However, condition (ii) in Theorem 1 does not hold since a vector of ones is contained in both $\mathcal{CS}(\mathbf{Z}_i(x))$ and $\mathcal{CS}(\mathbf{D}_i(x))$. Even if we go further and assume the entry cost and scrap value have the same magnitude (i.e. $EC_i(x) = -SV_i(x)$), the rank condition will still not be satisfied. In this case $\mathbf{D}_i(1,x)\,\boldsymbol{\phi}_{i,\eta_i}(1,x)$ becomes

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot EC_i(x).$$

Mathematically, the failure to apply our result in the example above can be traced to the fact that $\mathbf{Z}_i(x)$ is a stochastic matrix whose rows each sums to one. The inability to identify both entry cost and scrap value is not specific to our identification strategy. This issue is a familiar one in the empirical literature. We refer the readers to Aguirregabiria and Suzuki (2014) for a result relating to this as well as a list of references they provide of empirical works that make normalization assumptions on either one of these switching costs.[11]

We also wish to emphasize that our Theorems 1 and 2 only provide sufficient conditions for identification of ϕ_i *without the knowledge of either* β *or* μ_i. The failure to apply our theorems does not mean ϕ_i cannot be identified in the presence of additional information. In particular if one assumes the knowledge of β *as well as* μ_i, then existing results in Bajari et al. (2009) and Pesendorfer and Schmidt-Dengler (2008) may be used to identify the switching costs without the potential need to rely on normalzation or other restrictions.

We end this subsection by commenting that our results can be adapted to allow for effects from past actions beyond one period with little modification. Specifically, all results above hold if we redefine w_t to be $a_{t-\varsigma}$ for any finite $\varsigma \geq 1$, and then replace x_t by $\widetilde{x}_t = (x_t, a_{t-1}, \ldots, a_{t-\varsigma+1})$ everywhere. The inclusion of such state variable does not violate any of our assumptions, particularly assumption N2, and thus still allows us to define analogous nuisance function that can be projected away as shown in Theorems 1 and 2. In this case the interpretation of ϕ_i has to change accordingly and the switching cost parameters will be characterized according to $\widetilde{x}_t$; in such situation we naturally have $W^d_{\eta_i}(a_i, \widetilde{x}) \neq W^d_{\eta_i}(a_i, \widetilde{x}')$ for $\widetilde{x} \neq \widetilde{x}'$ since the principal interpretation of switching costs generally will depend on a_{t-1}.

[11] We commented in the introduction that a careful inspection of Proposition 2 in Aguirregabiria and Suzuki (2014) will suggest that either the entry cost or scrap value in their model can be identified independently of the discount value under some normalization. Further inspection (of their equation (21)) also reveals that assuming the entry cost and scrap value having the same magnitude will not help identification either.

4.2 Identifying the 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)):

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

where $g_i(a_i, x; \beta) \equiv E[V_i(s_{it+1}; \beta) | a_{it} = a_i, x_t = x]$ with $V_i(s_i; \beta) \equiv \sum_{\tau=0}^{\infty} \beta^{\tau} E[u_i(a_{it+\tau}, a_{-it+\tau}, w_{t+\tau}) | s_{it} = s_i$ 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 (19) with action 0 as the base, so that for all $i, a_i > 0$ and x:

$$\Delta v_i(a_i, x; \beta) = E[\Delta\pi_i(a_i, a_{-it}, x_t) | x_t = x] + \beta \Delta g_i(a_i, x; \beta), \tag{20}$$

where $\Delta v_i(a_i, x; \beta) \equiv v_i(a_i, x; \beta) - v_i(0, x; \beta), \Delta\pi_i(a_i, a_{-i}, x) \equiv \pi_i(a_i, a_{-i}, x) - \pi_i(0, a_{-i}, x)$ for all a_i, and $\Delta g_i(a_i, x; \beta) \equiv g_i(a_i, x; \beta) - g_i(0, x; \beta)$. Using Hotz-Miller's inversion, it follows that both $\Delta v_i(a_i, x; \beta_0)$ 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 $\mathcal{V}_{\beta} \equiv \{\Delta v_i(a_i, x; \beta)\}_{i, a_i, x \in \mathcal{I} \times A \times X}$, to be a corresponding *reduced form*.[12,13] 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 (20), since the first term 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

[12] This is a pseudo-model in the sense that we do not use different equilibria of the dynamic game for each β. We only consider the implied expected payoffs computed using the equilibrium beliefs that generate the data.

[13] 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)).

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 $\{\Pr[x_{t+1} = x'|x_t = x, a_{it} = a_i] - \Pr[x_{t+1} = x'|x_t$ $\mathbf{L}$ be a J by J stochastic matrix of transition probabilities of x_{t+1} conditioning on x_t, and $\mathbf{R}$ is a J by J matrix of conditional choice probabilities such that $\mathbf{Ru}_i$ represents a J by 1 vector of $\{E[u_i(a_t, s_{it})|x_t = x']\}_{x' \in X}$.

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

$$\Delta g_i(a_i, x; \beta) = \Delta H_i^{a_i}(x)(I - \beta\mathbf{L})^{-1}\mathbf{Ru}_i. \tag{21}$$

PROOF OF LEMMA 4: First note that $I - \beta\mathbf{L}$ is invertible for any $\beta \in \mathcal{B}$ by the dominant diagonal theorem (Taussky (1949)). Furthermore $(I - \beta\mathbf{L})^{-1}$ admits a Neumann series representation so that $(I - \beta\mathbf{L})^{-1}\mathbf{Ru}_i$ is precisely a vector of $\{[V_i(s_{it}; \beta)|x_t = x']\}_{x' \in X}$. The proof then follows since $\Delta H_i^{a_i}(x)$ is defined to be a vector that computes differences in conditional expectations of any functions of x_{t+1} given $x_t = x$ and $a_{it} = a_i$ and $a_{it} = 0$.■

It will be useful to collect $\Delta g_i(a_i, x; \beta)$ in a vector form. Let $\Delta\mathbf{H}_i^{a_i}$ denote a J by J matrix $\left[\Delta H_i^{a_i}(x^1)^\top, \ldots, \Delta H_i^{a_i}(x^J)^\top\right]^\top$, and $\Delta\mathbf{g}_i^{a_i}(\beta)$ denote a J by 1 vector $\{\Delta g_i(a_i, x; \beta)\}_{x \in X}$.

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

$$\Delta\mathbf{g}_i^{a_i}(\beta) = \Delta\mathbf{H}_i^{a_i}(I - \beta\mathbf{L})^{-1}\mathbf{Ru}_i. \tag{22}$$

PROOF OF LEMMA 5: Immediate.■

Our next result gives one sufficient condition for β_0 to be identified.

THEOREM 3 (IDENTIFICATION OF DISCOUNT FACTOR): *Under M1 - M4, if $\mathbf{Ru}_i \neq \mathbf{0}$ 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 $\beta, \beta' \in \mathcal{B}$ such that $\beta \neq \beta'$. From equation (20), $\Delta v_i(a_i, x; \beta)$ can differ from $\Delta v_i(a_i, x; \beta')$ if and only if $\beta\Delta g_i(a_i, x; \beta)$ differs from $\beta'\Delta g_i(a_i, x; \beta')$. Focusing on the latter, using Lemma 5, we have

$$\beta\Delta\mathbf{g}_i^{a_i}(\beta) - \beta'\Delta\mathbf{g}_i^{a_i}(\beta') = \left(\beta\Delta\mathbf{H}_i^{a_i}(I - \beta\mathbf{L})^{-1} - \beta'\Delta\mathbf{H}_i^{a_i}(I - \beta'\mathbf{L})^{-1}\right)\mathbf{Ru}_i.$$

Consider the terms in the parenthesis on the RHS of the equation above:

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

so that

$$\beta\Delta\mathbf{g}_i(\beta) - \beta'\Delta\mathbf{g}_i(\beta') = (\beta - \beta')\,\Delta\mathbf{H}_i^{a_i}(I - \beta'\mathbf{L})^{-1}(I - \beta\mathbf{L})^{-1}\mathbf{Ru}_i.$$

If $\mathbf{Ru}_i \neq \mathbf{0}$, then $(I - \beta'\mathbf{L})^{-1}(I - \beta\mathbf{L})^{-1}\mathbf{Ru}_i \neq \mathbf{0}$ since both $(I - \beta'\mathbf{L})^{-1}$ and $(I - \beta\mathbf{L})^{-1}$ are non-singular. Therefore if $\Delta\mathbf{H}_i^{a_i}$ has full column rank $\Delta\mathbf{H}_i^{a_i}(I - \beta'\mathbf{L})^{-1}(I - \beta\mathbf{L})^{-1}\mathbf{Ru}_i$ cannot be a zero vector. Hence $\beta\Delta g_i(a_i, x; \beta)$ must differ from $\beta'\Delta g_i(a_i, x; \beta')$ for some x in X. This in turns implies that β and β' that differ are not observationally equivalent. Thus β_0 is identified.■

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, so that β_{i0} can be identified if $\mathbf{Ru}_i \neq 0$ and $\Delta\mathbf{H}_i^{a_i}$ is invertible for some $a_i > 0$.

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 simply plugging in the sample counterparts of choice probabilities without any numerical optimization. However, such estimator is generally not efficient. In this section we provide a 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 the former but not the latter. Our exposition in this section shall we brief. We refer the reader to Sanches, Silva and Srisuma (2014) for further details regarding the estimation methodology and related asymptotic theorems.

Estimation of the Switching Cost

From Lemmas 2 and 3, we have:

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

Since A and X are finite, we have a finite number of identifying restrictions of the switching costs that can be vectorized across all players in the form of

$$\mathcal{Y}^{sc} = \mathcal{X}_1^{sc}\theta_1 + \mathcal{X}_2^{sc}\theta_2 \quad \text{when} \quad (\theta_1, \theta_2) = (\theta_{10}, \theta_{20}).$$

Let $\mathcal{X}^{sc} = [\mathcal{X}_1^{sc} : \mathcal{X}_2^{sc}]$ and $\theta_0 = (\theta_{10}, \theta_{20})$ such that θ_{10} consists of the nuisance functions and θ_{20} contains the switching costs. Note that $\mathcal{X}_2^{sc}$ is a deterministic matrix. $\mathcal{X}_1^{sc}$ and $\mathcal{Y}^{sc}$ are smooth functions of the choice and transition probabilities that we will denote by γ_0. Specifically it can be shown that $\mathcal{X}_1^{sc} = \mathcal{T}_{\mathcal{X}_1}^{sc}(\gamma_0)$ and $\mathcal{Y}^{sc} = \mathcal{T}_{\mathcal{Y}}^{sc}(\gamma_0)$ for some known functions $\mathcal{T}_{\mathcal{X}_1}^{sc}$ and $\mathcal{T}_{\mathcal{Y}}^{sc}$. Given a preliminary consistent estimator of γ_0, denoted by $\widehat{\gamma}$, we can define an estimation criterion where $\mathcal{X}_1^{sc}$ and $\mathcal{Y}^{sc}$ are replaced by $\widehat{\mathcal{X}}_1^{sc} = \mathcal{T}_{\mathcal{X}_1}^{sc}(\widehat{\gamma})$ and $\widehat{\mathcal{Y}}^{sc} = \mathcal{T}_{\mathcal{Y}}^{sc}(\widehat{\gamma})$ respectively, so that for $\theta = (\theta_1, \theta_2)$:

$$\widehat{\mathcal{S}}^{sc}(\theta; \widehat{\mathcal{W}}^{sc}) = (\widehat{\mathcal{Y}}^{sc} - \widehat{\mathcal{X}}_1^{sc}\theta_1 - \mathcal{X}_2^{sc}\theta_2)^\top \widehat{\mathcal{W}}^{sc}(\widehat{\mathcal{Y}}^{sc} - \widehat{\mathcal{X}}_1^{sc}\theta_1 - \mathcal{X}_2^{sc}\theta_2),$$

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

$$\begin{aligned} \widehat{\theta}(\widehat{\mathcal{W}}^{sc}) &= \arg\min_{\theta \in \Theta} \widehat{\mathcal{S}}^{sc}(\theta; \widehat{\mathcal{W}}^{sc}) \\ &= (\widehat{\mathcal{X}}^{sc\top} \widehat{\mathcal{W}}^{sc} \widehat{\mathcal{X}}^{sc})^{-1} \widehat{\mathcal{X}}^{sc\top} \widehat{\mathcal{W}}^{sc} \widehat{\mathcal{Y}}^{sc}, \end{aligned} \tag{23}$$

where $\widehat{\mathcal{X}}^{sc} = [\widehat{\mathcal{X}}_1^{sc} : \mathcal{X}_2^{sc}]$. However, we are primarily interested in θ_{20}. Its estimator can also be written in closed-form that takes an analogous expression to a (weighted) partition regression estimator:

$$\widehat{\theta}_2(\widehat{\mathcal{W}}^{sc}) = (\widehat{\mathcal{X}}^{sc\top} \widehat{\mathcal{M}}^{sc} \widehat{\mathcal{X}}^{sc})^{-1} \widehat{\mathcal{X}}^{sc\top} \widehat{\mathcal{M}}^{sc} \widehat{\mathcal{Y}}^{sc},$$

where $\widehat{\mathcal{M}}^{sc} = I - \widehat{\mathcal{X}}_1^{sc} \left(\widehat{\mathcal{X}}^{sc\top}_1 \widehat{\mathcal{W}}^{sc} \widehat{\mathcal{X}}_1^{sc} \right)^{-1} \widehat{\mathcal{X}}_1^{sc\top} \widehat{\mathcal{W}}^{sc}$ is an oblique projection matrix (e.g. see Davidson and MacKinnon (1993)). The choice of the weighting matrix will determine the relative efficiency of $\widehat{\theta}(\widehat{\mathcal{W}}^{sc})$ within the class of asymptotic least squares estimators indexed by the set of all positive definite matrices $\mathcal{W}^{sc}$. The efficient weighting matrix in this class is the one that converges in probability to the inverse of the asymptotic variance of $\sqrt{N}(\widehat{\mathcal{Y}}^{sc} - \widehat{\mathcal{X}}_1^{sc}\theta_{10} - \mathcal{X}_2^{sc}\theta_{20})$. The efficient estimator can then be constructed by using any preliminary consistent estimator of θ_0 (we need to both estimates for θ_{10} and θ_{20}). One such is the identity weighted estimator, which resembles an ordinary least squares estimator:

$$(\widehat{\mathcal{X}}^{sc\top} \widehat{\mathcal{X}}^{sc})^{-1} \widehat{\mathcal{X}}^{sc\top} \widehat{\mathcal{Y}}^{sc}. \tag{24}$$

In this case $\widehat{\mathcal{M}}^{sc}$ simplifies to $I - \widehat{\mathcal{X}}_1^{sc} \left(\widehat{\mathcal{X}}^{sc\top}_1 \widehat{\mathcal{X}}_1^{sc\top} \right)^{-1} \widehat{\mathcal{X}}_1^{sc\top}$, and the corresponding estimator of θ_{20} can equivalently be obtained by using the (plug-in) empirical counterparts of the expressions in Theorems 1 and 2.

Estimation of the Discount Factor

Rearranging equation (22) in Lemma 5 yields

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

These 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 choice and transition probabilities as well as the payoff parameters. Let us denote the latter by δ_0. Similar to the previous case, for any β, it can be shown that $\mathcal{X}^{df}(\beta) = \mathcal{T}^{df}_{\mathcal{X}_1}(\gamma_0, \delta_0; \beta)$ and $\mathcal{Y}^{df} = \mathcal{T}^{df}_{\mathcal{Y}}(\gamma_0)$ for some known functions $\mathcal{T}^{df}_{\mathcal{X}}$ and $\mathcal{T}^{df}_{\mathcal{Y}}$ respectively. Given preliminary consistent estimators of γ_0 and δ_0, say $\widehat{\gamma}$ and $\widehat{\delta}$, we can define an estimation criterion where $\mathcal{X}^{df}(\beta)$ and $\mathcal{Y}^{df}$ are replaced by $\widehat{\mathcal{X}}^{df}(\beta) = \mathcal{T}^{df}_{\mathcal{X}}\left(\widehat{\gamma}, \widehat{\delta}; \beta\right)$ and $\widehat{\mathcal{Y}}^{df} = \mathcal{T}^{df}_{\mathcal{Y}}(\widehat{\gamma})$ respectively, so that

$$\widehat{\mathcal{S}}^{df}(\beta; \widehat{\mathcal{W}}^{df}) = (\widehat{\mathcal{Y}}^{df} - \widehat{\mathcal{X}}^{df}(\beta))^\top \widehat{\mathcal{W}}^{df}(\widehat{\mathcal{Y}}^{df} - \widehat{\mathcal{X}}^{df}(\beta)). \tag{25}$$

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.

6.1 Monte Carlo Study

Our simulation design is taken from Pesendorfer and Schmidt-Dengler (2008, Section 7; also see Section 2 earlier in this paper). Consider a two-firm dynamic entry game. In each period t, each firm i has two possible choices, $a_{it} \in \{0, 1\}$. The observed state variables are previous period's actions, $w_t = (a_{1t-1}, a_{2t-1})$. Using their notation, firm $1'$s period payoffs are described as follows:

$$\pi_{1,\theta}(a_{1t}, a_{2t}, x_t) = a_{1t}(\mu_1 + \mu_2 a_{2t}) + a_{1t}(1 - a_{1t-1}) F + (1 - a_{1t}) a_{1t-1} W, \tag{26}$$

where $\theta = (\mu_1, \mu_2, F, W)$ containing respectively the monopoly profit, duopoly profit, entry cost and scrap value. The latter two components are switching costs. 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 set the payoff parameters to be $(\mu_{10}, \mu_{20}, F_0, W_0) = (1.2, -1.2, -0.2, 0.1)$ and $\beta_0 = 0.9$. There are three distinct equilibria for this game, one of which is symmetric. As an illustration, we only generate the data using the symmetric equilibrium and report the results using the identity weighted estimates. We take W_0 to be known since it cannot be identified jointly with F_0; see Pesendorfer and Schmidt-Dengler (2008). We estimate F_0 using the closed-form expression in (24). In order to estimate β_0 we need estimators for μ_{10} and μ_{20} that do not depend on the discount factor. For this we use $\mu_{10}+B_{1N}/\sqrt{N}$ and $\mu_{20}+B_{2N}/\sqrt{N}$ respectively, where N denotes the sample size and (B_{1N}, B_{2N}) are bivariate independent standard normal variables. The $\sqrt{N}-$scaling ensures the sampling errors converge to zero at the parametric rate as one would typically assumed in empirical applications. We also estimate F_0 using the estimator in Sanches, Silva and Srisuma (2014, hereafter SSS) that requires an assumption on the discount factor to illustrate the effect from assuming an incorrect discount factor and also to compare it to our closed-form estimator when the discount factor is correctly assumed to be known. For each sample size $N = 1000, 10000, 100000$, we perform 1000 simulations. We report the bias and standard deviation (in italics) for our estimators of F_0 and β_0 in Table 1, and analogous statistics for the estimator of F_0 using SSS in Table 2.

	F_0	β_0
$N = 1000$	0.004	0.002
	0.176	*0.753*
$N = 10000$	0.003	0.005
	0.054	*0.223*
$N = 100000$	0.002	0.001
	0.018	*0.081*

Table 1: Bias and standard deviation, the latter in italics, for the asymptotic least squares estimators of F and β using the estimator proposed in Section 5.

	β										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
$N = 1000$	0.549	0.507	0.457	0.396	0.325	0.247	0.167	0.083	0.022	0.002	0.039
	0.086	*0.078*	*0.083*	*0.079*	*0.080*	*0.085*	*0.084*	*0.087*	*0.092*	*0.097*	*0.104*
$N = 10000$	0.542	0.499	0.448	0.388	0.321	0.242	0.159	0.080	0.019	-0.002	0.039
	0.027	*0.028*	*0.027*	*0.026*	*0.027*	*0.025*	*0.025*	*0.027*	*0.027*	*0.028*	*0.029*
$N = 100000$	0.540	0.497	0.447	0.387	0.318	0.240	0.158	0.079	0.019	-0.001	0.041
	0.009	*0.008*	*0.008*	*0.009*	*0.008*	*0.009*	*0.008*	*0.009*	*0.008*	*0.009*	*0.009*

Table 2: Bias and standard deviation, the latter in italics, for the asymptotic least squares estimator of F_0 using the estimator of Sanches, Silva and Srisuma (2014) for different values of β.

Our estimators appear to be consistent as expected, while the estimators of SSS are not when the assumed value of β differs from β_0. However, our robust estimator of F_0 is less precise than SSS's. This is not surprising since the estimator in SSS explicitly makes use of other structure of model, particularly the remaining components of the profit function as well as the discount factor.

6.2 Empirical Illustration

We estimate a simplified version of an entry-investment game base on the model studied in Ryan (2012) using his data.[14] In what follows we provide a brief description of the data, highlight the main differences between the game we model and estimate with that of Ryan (2012). Then we present and discuss our estimates of the primitives.

Data (Ryan (2012))

The dataset contains aggregate data on quantities and outputs for the Portland cement industry in the United States from 1980 to 1998 as well as plant-level capacities and production quantities for all the plants. Data on plants includes the name of the firm that owns the plant, the location of the plant, the number of kilns in the plant and kiln characteristics (fuel type, process type and year of installation). Following Ryan (2012) we assume that the plant capacity equals the sum of the capacity of all kilns in the plant and that different plants are owned by different firms. We observe that plants' names and ownerships change frequently. This can be due to either mergers and acquisitions or to simple changes in the company name. We do not treat these changes as entry/exit movements. We check each observation in the sample using the kiln information (fuel type, process type, year of installation and plant location) installed in the plant. If a plant changes its name but keeps the same kiln characteristics, we assume that the name change is not associated to any entry/exit movement. This way of preparing the data enables us to replicate the summary statistics of plant-level data in Ryan; also see Section 5.2 in Otsu, Pesendorfer and Takahashi (2015).

Dynamic Game

Ryan (2012) models a dynamic game played between firms that own cement plants in order to measure the welfare costs of the 1990 Clean Air Act Amendments (1990 CAAA) on the US Portland cement industry. The decision for each firm is first whether to enter (or remain) in the market or exit, and if it is active in the market then how much to invest or divest. Firm's investment decisions is governed by its capacity level. The firm's profit is determined by variable payoffs from the competition in the product market with other firms, as well as switching costs from the entry and investment decisions. In Ryan's model, there are two action variables, one is a binary choice for entry and the other is a continuous level of investment. The only observed state variables that are endogenous in his game are past entry decisions and capacity levels. Other determinants of

[14] The dataset can be downloaded from the Econometrica webpage at https://www.econometricsociety.org/content/supplement costs-environmental-regulation-concentrated-industry-0.

the variable profits, namely aggregate prices and quantities that are used to construct the demand function, come from a different data source and are treated as exogenous.[15]

We consider a discrete game that fits the general model described in earlier Sections. The main departure from Ryan (2012) is we combine the entry decision along with the capacity level into a single discrete variable. We set the action space to be an ordinal set $\{0, 1, 2, 3, 4, 5\}$, where 0 represents exit/inactive, and the positive integers are ordered to denote entry/active with different capacity levels. The payoff for each firm has two additive separable components. One depends on the observables while the other is an unobserved shock. The observable component has two parts. One represents the variable profit, where firms compete in a capacity constrained Cournot game. The other consists of the switching costs that captures the essence of firms' entry and investment decisions. Lastly each firm receives unobserved profit shocks for each action with (standard) i.i.d. type-1 extreme value distributions with mean zero and variance $\pi^2/6$.

Estimation

We estimate the component of the payoff function that are functions of the observables. The Cournot profit is constructed from the same demand and cost functions estimated in Ryan's paper.[16] In particular this profit is zero when a firm chooses action 0 (as assumed in Ryan). For an active firm with $a_i > 0$, the profit is calculated using the $(a_i \times 20)$ −th percentile of the capacity level observed in the data. Therefore we estimate the Cournot profit function without appealing to the dynamic feature of the game. We assume switching costs for each firm is independently of what other firms do. We normalize switching cost of choosing action 0 to be zero, which is akin to normalizing the scrap value. We treat all firms to be symmetric. Therefore there are a total of 25 switching cost parameters to be estimated.[17] The 25 switching cost parameters are estimated using the closed-form expression in Section 5, see (24). In particular we only need the choice probabilities, which we estimate using a multinomial logit specification analogous to Ryan (2012).

The estimation of the discount factor takes the estimated payoff function as known. One practical issue when combining different parts of the payoff function that are estimated separately is the compatibility of scale. In our case the Cournot profit is estimated directly from the data (interpretable in a monetary unit), while the scale of the switching costs is determined by the distribution of the

[15] The data on entry and capacity are constructed using plant-level data that Ryan collected. The demand data come from the US Geological Survey's Mineral Yearbook.

[16] We use the specifications of demand and cost functions written in equations (1) and (2) of Ryan (2012) respectively. The estimated demand follows specification 1 in his Table 3, and the cost parameters are taken from his Table 4.

[17] Ryan (2012) models the switching costs differently. The fixed operating cost is normalized to be zero. Non-zero investment and divestment costs are drawn from two distinct independent normal distributions, whose means and variances are estimated using the methodology in Bajari, Benkard and Levin (2007).

private shocks; imposed in the application of Hotz-Miller's inversion. A standard solution in a dynamic game estimation is to include a scale parameter, such as the variance, for the private values to be estimated jointly with the switching costs. Since we can estimate the switching cost parameters without any optimization for any given distribution, such scale parameter can also be introduced in a least squares criterion and estimated along with the discount factor. However, repeatedly inverting probability distribution with many action choices is a very cumbersome task even with an addition of only a single unknown variance parameter. We instead use a computationally simpler alternative to correct the scale by inserting a multiplicative factor on the Cournot profit function, which we then estimate jointly with the discount factor.

We provide two sets of estimates of the switching costs and discount factors using the data from before and after the implementation of the 1990 CAAA.[18] Tables 3 and 4 contain the results for the switching costs using the data from the years 1980 to 1990 and 1991 to 1998 respectively. Tables 5 provides estimates for the discount factor from the two time periods. All of our estimates are based on the identity weighting and standard errors are computed using random resampling bootstrap with replacements.

	$a_{it-1}=0$	$a_{it-1}=1$	$a_{it-1}=2$	$a_{it-1}=3$	$a_{it-1}=4$	$a_{it-1}=5$
$a_{it}=1$	-2.84	-	2.56	5.37	8.09	3.82
	0.43	-	*0.55*	*0.92*	*1.05*	*1.20*
$a_{it}=2$	-10.56	-5.18	-	5.29	10.60	7.89
	1.47	*0.84*	-	*0.83*	*1.22*	*0.98*
$a_{it}=3$	-17.26	-15.45	-7.78	-	7.86	8.07
	0.84	*1.95*	*1.25*	-	*0.89*	*0.99*
$a_{it}=4$	-23.76	-23.52	-20.89	-10.48	-	2.68
	1.31	*1.35*	*2.32*	*1.33*	-	*1.32*
$a_{it}=5$	-25.71	-25.47	-25.01	-19.07	-7.16	-
	0.53	*0.50*	*0.64*	*1.27*	*1.43*	-

Table 3: Estimate and standard error, the latter in italics, for the switching cost parameters from a_{it-1} to a_{it} using data from the years 1980 to 1990. Standard errors are obtained from 50 bootstrap samples.

[18] Otsu, Pesendorfer and Takahashi (2015) recently suggest the data from Ryan (2012) between 1980 - 1990 should not be pooled across markets, while the data from 1991 - 1998 pass their poolability test.

	$a_{it-1}=0$	$a_{it-1}=1$	$a_{it-1}=2$	$a_{it-1}=3$	$a_{it-1}=4$	$a_{it-1}=5$
$a_{it}=1$	-4.78	-	6.09	10.85	9.83	3.74
	1.41	-	*1.32*	*2.08*	*1.19*	*1.44*
$a_{it}=2$	-14.73	-9.98	-	8.22	10.62	4.98
	2.18	*2.62*	-	*1.66*	*1.66*	*2.02*
$a_{it}=3$	-21.73	-22.19	-11.84	-	5.64	3.19
	2.72	*2.74*	*1.68*	-	*2.10*	*2.23*
$a_{it}=4$	-24.20	-24.77	-21.80	-10.43	-	2.91
	0.98	*1.11*	*1.07*	*2.09*	-	*1.16*
$a_{it}=5$	-22.77	-22.68	-20.44	-12.75	-8.48	-
	2.07	*2.27*	*2.04*	*2.61*	*1.84*	-

Table 4: Estimate and standard error, the latter in italics, for the switching cost parameters from a_{it-1} to a_{it} using data from the years 1991 to 1998. Standard errors are obtained from 50 bootstrap samples.

	Before 1990	After 1990
Discount factor	0.64	0.64
	0.03	*0.06*

Table 5: Estimate and standard error, the latter in italics, for the discount factor. Standard errors are obtained from 50 bootstrap samples.

The sign and relative magnitude of the estimated switching costs generally make plausible economic sense. Particularly, entry at higher capacity level should incur higher cost (negative payoff), and increasing the capacity level should be costly while divestment can return revenue for firms. We find the signs are uniformly correct, with positive estimates on the upper (right) triangular part of the tables and negative estimates for the lower triangular part. The relative magnitudes of the estimates also mostly conform to the economic intuition we expect. For instance, we should expect the estimates should be monotonically decreasing when reading down each column in Tables 3 and 4. One notable observation is that entry and investment costs are in general higher for the period after the 1990 CAAA has been implemented especially when entering at or switching into low to moderately high capacity level. Particularly the increase in entry costs is consistent with the finding in Ryan (2012) that is consistent with the tougher rules following the 1990 CAAA; such as new plants are required to undergo an additional certification procedure etc. Our estimates of the discount rate are lower than the usual range of assumed rate of discounting (e.g. Ryan takes β to be 0.9). The difference between the estimates using data from different time periods is negligible, suggesting the 1990 CAAA does not affect firms' costs of borrowing.

7 Conclusion

Many empirical dynamic games and decision problems naturally involve adjustment or switching costs from choosing different actions. We show components of the payoff functions that can be interpreted as switching costs can be identified in closed-form in terms of the observed choice and transition probabilities alone. Hence, when other components of the payoff function can also be identified independently elsewhere, the entire payoff function can be recovered without the knowledge of the discount factor. We show the discount factor can then be identified. Our identification strategy also suggests a new way to estimate games, nonparametrically or otherwise, with attractive computational features. We illustrate the scope of its applications in a Monte Carlo study and an empirical game using real data.

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

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

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

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

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

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

[10] Chen, L. (2014): "Identification of Discrete Choice Dynamic Programming Models with Nonparametric Distribution of Unobservables," *Working Paper, Institute of Economics, Academia Sinica.*

[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] Fang, H. and Y. Wang (2014): "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions," *forthcoming in International Economic Review.*

[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): "Optimal 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] Manski, C. (1993): "Dynamic Choice in Social Settings," *Journal of Econometrics*, **58**, 121-136.

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

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

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

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

[29] Otsu, T., M. Pesendorfer and Y. Takahashi (2015): "Pooling Data across Markets in Dynamic Markov Games," *Working Paper, LSE.*

[30] 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.

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

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

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

[34] Sanches, F.M., D. Silva Jr and S. Srisuma (2014): "Ordinary Least Squares Estimation for a Dynamic Game Model," *forthcoming, International Economic Review.*

[35] 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.*

[36] Slade, M.E. (1998): "Optimal Pricing with Costly Adjustment: Evidene from Retail-Grocery Prices," *Review of Economic Studies*, **65**, 87-107.

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

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

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

[40] 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.*

[41] Taussky, O. (1949): "A Recurring Theorem on Determinants," *American Mathematical Monthly*, **56**, 672-676.