Komarova, T., Sanches, F., Silva Junior, D. & Srisuma, S. (2016). Joint Analysis of the Discount Factor and Payoff Parameters in Dynamic Discrete Choice Games (Report No. 16/15). London: Department of Economics, City, University of London.

Department of Economics

Joint Analysis of the Discount Factor and Payoff Parameters in Dynamic Discrete Choice Games

Tatiana Komarova[1]
London School of Economics

Fabio Sanches[1]
University of São Paulo

Daniel Silva Junior[1]
City, University of London

Sorawoot Srisuma[1]
University of Surrey

Department of Economics
Discussion Paper Series
No. 16/15

[1] Corresponding authors: Email: t.komarova@lse.ac.uk; fmiessi@gmail.com; Daniel.Silva-Junior@city.ac.uk; s.srisuma@surrey.ac.uk

Joint Analysis of the Discount Factor and Payoff Parameters in Dynamic Discrete Choice Games*†

Tatiana Komarova
London School of Economics

Fabio Sanches
University of São Paulo

Daniel Silva Junior
City University of London

Sorawoot Srisuma
University of Surrey

August 23, 2016

Abstract

Most empirical models of dynamic games assume the discount factor to be known and focus on the estimation of the payoff parameters. However, the discount factor can be identified when the payoffs satisfy parametric or other nonparametric restrictions. We show when the payoffs take the popular linear-in-parameter specification, the joint identification of the discount factor and payoff parameters can be simplified to a one-dimensional model that is easy to analyze. We also show that switching costs (e.g. entry costs) that often feature in empirical work can be identified in closed-form, independently of the discount factor and other specification of the payoff function. Our identification strategies are constructive. They lead to easy to compute estimands that are global solutions. Estimating the discount factor permits direct inference on borrowing rate. Our estimates of the switching costs can be used for specification testing. We illustrate with a Monte Carlo study and the dataset from Ryan (2012).

JEL Classification Numbers: C14, C25, C61

Keywords: Discount Factor, Dynamic Games, Identification, Estimation, Switching Costs

*Earlier versions of this research were circulated under the title "Identifying Dynamic Games with Switching Costs". We thank Kirill Evdokimov, Emmanuel Guerre, Koen Jochmans, Arthur Lewbel, Oliver Linton, Aureo de Paula, Martin Pesendorfer, Yuya Sasaki, Richard Spady, Pasquale Schiraldi and seminar participants at CREATES, GRIPS, Johns Hopkins University, LSE, NUS, Queen Mary University of London, Stony Brook University, University of Cambridge, University of São Paulo, and participants at Bristol Econometrics Study Group (2014), AMES (Taipei), EMES (Toulouse) and IAAE (London) for comments.

†*E-mail address*: t.komarova@lse.ac.uk; fmiessi@gmail.com; d.silva-junior@lse.ac.uk; s.srisuma@surrey.ac.uk.

1 Introduction

A structural study involves modeling the economic problem of interest based on some primitives that govern an economic model. The primitives have a clear interpretation. The empirical goal is to estimate them, which can then be used for counterfactual analysis. Our paper studies some identification and estimation aspects for a stationary dynamic discrete game that generalizes the single agent Markov decision problem surveyed in Rust (1994). The primitives of the games we consider consist of players' (per-period) payoff functions, discount factor, and Markov transition law of the variables in the model.

There is anecdotal evidence from the literature on single agent models that implies that dynamic games are generally not identified nonparametrically. For example, Manski (1993) shows that the discount factor cannot be identified jointly with the payoff function that is nonparametric; Magnac and Thesmar (2002) show the payoff function cannot be identified even if all other primitives of the model are known; Norets and Tang (2014) show the payoff function can only be partially identified when the distributions of unobservable state variables are unknown with discrete observable states. But identification is possible with more structure on the model. For examples, see Pesendorfer and Schmidt-Dengler (2008), Bajari, Chernozhukov, Hong and Nekipelov (2009), Blevins (2014), Chen (2014), as well as Fang and Wang (2014). Hence, in spite of the under-identified nature of a general structural dynamic model, many fruitful empirical research can be, and has been, conducted based on these dynamic models using the theoretical results as guide.

Empirical applications of dynamic games often focus on the estimation of the parametric payoff functions and seemingly always assume the value of the discount factor to be known. An indiscriminating list of examples include: Beresteanu, Ellickson and Misra (2010), Collard-Wexler (2013), Dunne, Klimek, Roberts and Xu (2013), Gowrisankaran, Lucarelli, Schmidt-Dengler and Town (2010), Igami (2015), Lin (2012), Sanches, Silva Junior and Srisuma (2014), Snider (2009) and Suzuki (2013). There appears to be no formal justification as to why the discount factor has to be presumed known rather than estimated. Commonly cited reasons, if any is given at all, include precedence from the single agent literature, lack of identification, numerical difficulties (e.g. intractability or convergence failure) and post-estimation issues (e.g. implausible or imprecise estimates). The underlying sources for the first reasoning can also be traced to the other closely related, but distinct, issues.[1,2]

[1]Some noted estimation attempts in the single agent context include: Rust (1987, pp. 1023), who says "I was not able to precisely estimate the discount factor", while Slade (1998, pp. 102) also fixes the discount factor after "it was found that the objective function is fairly flat" over some range.

[2]Estimation of the discount factor in some related finite-time horizon models is more standard, and with more encouraging findings. E.g. see Keane and Wolpin (1997). But negative or other implausible estimates have also been

Identification is a property of the model. It is customary to translate the behavioral condition that defines (parametric, point-) identification into a loss function with a unique minimum for the purpose of estimation. There are often many candidates of loss functions. A positive identification for one is sufficient to identify the model. However, in general, verifying that a nonlinear function of several variables has a unique minimum point is a difficult mathematical task. The degree of difficulty can depend crucially on the choice of the loss function. This also relates directly to the practical aspects of computing the estimand.[3] Particularly it may not even be a trivial assumption to assume that one can always find the global minimum of a nonlinear loss function with many parameters in a dynamic game due to intractable components of the model. Therefore, in practice, implausible estimates may also arise due to a purely a numerical reason even if the model is correctly specified.[4]

Our paper aims to show that it is not necessary to assume the discount factor a priori in order to analyze empirical games. We consider two important special cases. First, we show that when payoffs take a linear parameterization, joint identification of the discount factor and payoff parameters can be analyzed as a one-parameter model irrespective of the number of payoff parameters. Second, for games with switching costs (such as entry costs and scrap values), we show the switching cost parameters can be identified in closed-form independently of the discount factor and specification of other parts of the payoff function. Our identification strategies are constructive. The corresponding estimands are easy to compute. An important feature is they aim to obtain global solutions to potentially complex optimization problems in a transparent manner. Then the estimates of the discount factor permit testing of borrowing costs and other dynamic considerations directly. Also the closed-form estimators for the switching costs can be used for specification testing. E.g. testing the mode of competition amongst firms, by comparing them with estimates from existing methods that explicitly specify the entire payoff function.

The non-identification argument in Manski (1993) does not preclude us from studying the identification of the discount factor since the payoff functions employed in practice satisfy a priori specified parametric and/or other nonparametric restrictions. However, even in a single agent model with a known discount factor, establishing that the parametric payoff parameters are identified is difficult due to the nonlinear nature of the model that contains an intractable value function. Furthermore in dynamic games there may be multiple equilibria, subsequently the model may be *incomplete* (Tamer (2003)). We proceed in the same way as Pesendorfer and Schmidt-Dengler (2008) and Bajari et

reported (e.g. see Hotz and Miller (1993)).

[3]For example, as Hotz, Miller, Sanders and Smith (1994) noted in their footnote 13 on pp. 280 that: *"There is nothing inherent in our method which precludes estimation of β [the discount factor] ... our primary reason for not estimating β was the intractability it presented for implementing the ML [a competing] estimator."*

[4]Since a structural model is interpreted as an approximation of the data generating process, misspecification here means that the data is not fitted well by the model with economically plausible parameter values.

al. (2009) and study identification using the implied expected discounted payoffs that generates the data based on the observed transition probabilities. More specifically we take the model to be the collection of implied expected discounted payoffs as a mapping from the parameter space. Such model reduces the degree of intractability of the model and circumvents the issue of incompleteness, and is the basis for all what is known as "two-step" estimation methods in the literature (e.g. Aguirregabiria and Mira (2007), Bajari, Benkard, Levin (2007), Pakes, Ostrovsky and Berry (2007), Pesendorfer and Schmidt-Dengler (2008)).

We first consider the linear-in-parameter payoff specification due to its overwhelmingly common usage in empirical work; examples include those in the list of applications above.[5] When the discount factor is known, the corresponding implied expected discounted payoff also takes the linear-in-parameter structure. Various computational exploits of this linear structure have been noted, e.g. see Miller, Sanders and Smith (1994), Bajari, Benkard, Levin (2007), Bajari et al. (2009), and Sanches, Silva Junior and Srisuma (2016). In particular Sanches et al. (2016) translate the identification condition for the linear payoff parameter in terms of the uniqueness of the minimum Euclidean norm between the observed and model implied expected discounted values. Their estimator has the familiar closed-form OLS expression and condition for identification can be given in terms of the full rank condition of a matrix. See Assumption B1 in Sanches et al. (2016). It is worth emphasizing that their Assumption B1 is also necessary for consistent estimation of any two-step estimator in that setting.

When the discount factor is unknown and taken as part of the parameter space the model becomes intrinsically nonlinear. Existing conditions that ensure identification in a nonlinear parametric model in econometrics can be hard to verify and the scope of applications is limited by stringent conditions; see Komunjer (2012) for recent results. Here we show that the identification for games with linear-in-parameter payoffs can be analyzed exhaustively even when the parameter space is large. We follow the approach in Sanches et al. (2016) and expand the parameter space to include the discount factor. The least squares framework enables us to simplify the problem by just considering a one-dimensional path of the parameter space. In particular, for any value of the discount factor, there exists a vector of payoff parameters that minimize the least squares that has a closed-form OLS expression. The profiled distance becomes a mapping from $[0, 1]$ to $\mathbb{R}$. Therefore an exhaustive analysis of identification for the discount factor reduces to simply evaluating a function with one argument over a small domain. Once the identification of the discount factor is established it can be taken as known. The payoff parameters is then identified if an analogous condition to Assumption

[5] Other specifications that have been employed are often motivated by the need to impose additional constraints. E.g. Fan and Xiao (2014) use a linear index in an exponential function to ensure non-negativity of their variable profits.

B1 in Sanches et al. (2016) holds.

When the parameterization of the payoff function is not linear we focus on reducing the parameter space instead of studying the joint identification of the discount factor and payoff parameters. Our approach reflects a common practice that not all components of the payoff function need to be treated in the same way. Parts of the payoff function, such as variable profits, can be estimated directly using economic theory if relevant data are available. These serve as exclusion restrictions (e.g. see Berry and Haile (2010, 2012)). The remaining components are *dynamic parameters* of the game that have to be estimated using the structure of the dynamic models. One of the most prevalent type of dynamic parameters arises from players choosing different actions from the previous period. Specific examples include entry cost and scrap value in games with entry, menu costs in pricing problems, as well as adjustment costs in investment decisions. We refer to these as *switching costs*. Switching costs, by definition, have built-in nonparametric structures that impose how they can appear in the payoff function.

We show that switching costs can be identified, in a closed-form, independently of the discount factor and specification of the remaining components of the payoff function. It may not come as a surprise that such result requires some restrictions on the payoffs as well as the dependence structure of the controlled Markov process. However, the conditions we impose can be motivated empirically and have been frequently assumed in the empirical literature. Specifically, we assume that, whether a player may incur a switching cost in each period is only determined by her own action. The state variables, such as past actions of all players, can otherwise affect today's switching costs in an arbitrary way. We also require that the remaining components of the payoff function do not depend on past actions (this can be relaxed to allow dependence of a finite time lag). The latter condition is satisfied by typical payoff components. E.g. variable profits that are determined by the competition between players depend only on those present in the game (for instance a Cournot or an auction game), as well as fixed operating costs. We also limit the feedback of past actions in the Markov process. We assume that the past actions do not affect the transition law of future states conditional on today's actions and states. Our conditional independence requirement is a testable assumption, and is weaker than the frequently assumed condition that state variables other than actions are strictly exogenous. Examples of empirical models that satisfy these assumptions can be found in the applications cited above amongst many others.

The classic combination of exclusion and independence restrictions is a powerful tool for establishing nonparametric identification in structural models; see Matzkin (2007, 2012). Examples of related models can be found in Blevins (2014) and Chen (2014), who use different exclusion and independence assumptions to identify the distribution of the unobserved state variables in a single agent setting. In our case, the proposed framework enables us to set up a linear system containing switching costs and some nuisance parameters that depend on all primitives of the game. In a single agent dynamic decision model, the switching costs can then be identified by simply differencing out the nuisance parameters. For a dynamic game, the nuisance terms can be eliminated by a projection that can be interpreted as a generalized difference. Therefore the switching costs can be identified up to some location normalizations that accounts for the nonparametric specification of the remaining components of the payoff function. Our approach to eliminate the nuisance term therefore shares some similarities with the pair-wise differencing approach that is useful for the estimation of complicated nonlinear models (e.g. see Honoré and Powell (2005)). Notably, the pair-wise difference estimator that Hong and Shum (2010) propose for a single agent dynamic investment model can also be computed without the knowledge of the discount factor.[6]

The estimation of dyamic games is generally considered to be a numerically challenging task. Analogous to the identification argument above, the choice of the estimation methodology can be crucial for practical analysis of dynamic games. Traditional approach in econometrics takes consistent estimation for granted and focuses on efficiency. However, even consistency of a sensible looking estimation procedure may be problematic in practice due to the complicated nature of dynamic games. E.g. see Appendix A in Srisuma (2013), and also a series of papers related to sequential estimation methods (Pesendorfer and Schmidt-Dengler (2010), Kasahara and Shimotsu (2012) and Egesdal, Lai and Su (2015)). In this paper we focus on the simplicity of implementation. We adopt the approach of Sanches et al. (2016). The contribution of that paper highlights the computational advantages that least squares criterions in expected payoffs have over its dual representation in terms of the choice probabilities; particularly as proposed by Pesendorfer and Schmidt-Dengler (2008).[7] Importantly they show the estimators are in fact asymptotically equivalent but the numerical efforts in computing the latter can be substantially higher.[8] It can be shown that these advantages are conserved when the parameter space expands to include the discount factor.

Our estimators can then be constructed according to our identification arguments. Our profiling estimator uses the closed-form OLS expression for the linear payoff parameters in terms of the

[6]The motivation behind Hong and Shum (2010)'s estimator is actually to avoid the computation of the value function rather than constructing a robust estimator. In particular they difference out the future discounted payoffs between two economic agents if their investment accumulations are (nearly) equal under a deterministic state transition rule.

[7]There are also other authors have also proposed estimators that minimize expected payoffs. In particular, under the linear-in-parameter assumption, the estimators of Miller, Sanders and Smith (1994) and Bajari et al. (2009) take and an IV form.

[8]The class of estimators proposed by Pesendorfer and Schmidt-Dengler (2008) has been well received. It includes the non-iterative estimator of Aguirregabiria and Mira (2007) and the moment estimator of Pakes, Ostrovsky and Berry (2008) as special cases.

discount factor. Therefore our joint estimation of the discount factor and the payoff parameters can be conducted by a simple and exhaustive one-dimensional search over the support of the discount factor. In games with switching costs, closed-form estimation of switching costs serves to reduce the number of parameters to be estimated. The dimensionality reduction can be substantial in a game with large dimensions; as the number of unrestricted switching costs for each player grow at a quadratic rate with respect to the number of possible actions, which then grows exponentially fast with the number of players for every state.

We provide a Monte Carlo study to analyze some basic statistical properties of our proposed estimators. We then use the dataset from Ryan (2012) to estimate a dynamic game played between firms in the US Portland cement industry. In our version of the game, firms choose whether to enter the market as well as decide on the capacity level of operation (five different levels). We assume firms compete in a capacity constrained Cournot game, so the period profit can be estimated directly from the data as done in Ryan. The remaining part of the payoff consists of fixed operating costs and 25 switching cost parameters. Other dynamic parameters we estimate include the discount factor and fixed operating cost. We estimate the model twice. Once using the data from before 1990, and once after 1990, which coincides with the date of the 1990 Clean Air Act Amendments (1990 CAAA). Our switching costs estimates generally appear sensible, having correct signs and relative magnitudes. They show that firms entering the market with a higher capacity level incur larger costs, and suggest that increasing capacity level is generally costly while a reduction can return some revenue. We also find that operating and entry costs are generally higher after the 1990 CAAA, which supports Ryan's key finding. We are also able to estimate the discount factor with reasonable precision.

The remainder of the paper is organized as follows. Section 2 defines the theoretical model and states the modeling assumptions. Section 3 considers the joint identification of discount factor and payoff parameters under the linear specification. Section 4 shows the closed-form identification of switching costs. Section 5 illustrates the use of our estimator with simulated and real data. Section 6 concludes.

2 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 $s_{it} = (x_t, \varepsilon_{it})$ such that $x_t \in X$, for some compact set $X \subseteq \mathbb{R}^{d_X}$. State x_t is public information, which is 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 $s_t \equiv (x_t, \varepsilon_t)$ 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, $u_i : A^I \times S \rightarrow \mathbb{R}$, 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 $i, a_i, a_{-i}, x, \varepsilon_i$:*

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

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

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

where Q is the cumulative distribution function of ε_t and G denotes the transition law of x_{t+1} conditioning on x_t, 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 proceeds 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, players choose their actions again and so on. We consider a Markovian framework where players' behavior is stationary across time and players are assumed to play pure strategies. More specifically, for some $\alpha_i : S \rightarrow A$, $a_{it} = \alpha_i(s_{it})$ 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. Beliefs are also time invariant. Player i's beliefs, σ_i, is a distribution of $a_t = (\alpha_1(s_{1t}), \ldots, \alpha_I(s_{It}))$

conditional on x_t for some pure Markov strategy profile $(\alpha_1, \ldots, \alpha_I)$. The decision problem for each player is to solve, for any s_i,

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

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 beliefs and primitives completely determine the transition law for future states. 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, x_{t+1}), which is consistent with a single equilibrium play. The ideal data set is therefore a long time series from a single market. Although more commonly, datasets used in empirical work have short panel from multiple markets, the joint distribution of the observables can still be identified if they are generated from the same equilibrium.[9] 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).

3 Identification with Linear-in-Parameter Payoffs

In this section we consider games where payoffs have a linear-in-parameter specification. Section 3.1 defines identification for the parameter of interest and provide some representation lemmas based on the linear payoff structure. Section 3.2 studies identification by profiling.

3.1 Definition of Identification and Some Representation Lemmas

We assume the following assumption holds throughout this section.

Assumption M5 (Linear-in-Parameter): *For all i, a_i, a_{-i}, x:*

$$\pi_i(a_i, a_{-i}, x; \theta) = \pi_{i0}(a_i, a_{-i}, x) + \theta^\top \pi_{i1}(a_i, a_{-i}, x),$$

where π_{i0} is a known real value function, π_{i1} is a known $p-$dimensional vector value function and θ belongs to $\mathbb{R}^p$.

[9] Otsu, Pesendorfer and Takahashi (2015) have recently proposed a test for the poolability of data across markets.

The role of π_{i0} is to represent the payoff components that are identifiable without the knowledge of the discount factor. In practice π_{i0} and possibly parts of π_{i1} may have to be estimated (e.g. see Section 5.2). For the purpose of identification they can be treated as known.

The primitives of interest belong to $\mathcal{B} \times \Theta$, where $\mathcal{B} = [0, 1)$ and $\Theta = \mathbb{R}^p$ for some non-negative integer p. We are interested in the data generating discount factor and payoff parameters, which we denote by β_0 and θ_0 respectively. We first define the choice specific expected payoffs for choosing action a_i prior to adding the period unobserved state variable, which is computed for different β and θ, for any i, a_i and x:

$$v_i(a_i, x; \beta, \theta) = E\left[\pi_i(a_i, a_{-it}, x_t; \theta) | x_t = x\right] + \beta g_i(a_i, x; \beta, \theta), \tag{2}$$

where $g_i(a_i, x; \beta, \theta) \equiv E\left[V_i(s_{it+1}; \beta, \theta) | a_{it} = a_i, x_t = x\right]$ with $V_i(s_i; \beta, \theta) \equiv \sum_{\tau=0}^{\infty} \beta^\tau E\left[u_i(a_{t+\tau}, s_{it+\tau}; \theta) | s_{it} =\right.$ and $u_i(a_t, s_{it}; \theta) \equiv \pi_i(a_t, x_t; \theta) + \sum_{a_i' \in A} \varepsilon_{it}(a_i') \cdot \mathbf{1}\left[a_{it} = a_i'\right]$. Note that the expectations here are taken with respect to the observed choice and transition probabilities that are consistent with β_0 and θ_0. We consider the relative payoffs in (2) with action 0 as the base, so that for all $i, a_i > 0$ and x:

$$\Delta v_i(a_i, x; \beta, \theta) = E\left[\Delta\pi_i(a_i, a_{-it}, x; \theta) | x_t = x\right] + \beta \Delta g_i(a_i, x; \beta, \theta), \tag{3}$$

where $\Delta v_i(a_i, x; \beta, \theta) \equiv v_i(a_i, x; \beta, \theta) - v_i(0, x; \beta, \theta)$, $\Delta\pi_i(a_i, a_{-i}, x; \theta) \equiv \pi_i(a_i, a_{-i}, x; \theta) - \pi_i(0, a_{-i}, x; \theta)$ for all a_i, and $\Delta g_i(a_i, x; \beta, \theta) \equiv g_i(a_i, x; \beta, \theta) - g_i(0, x; \beta, \theta)$. Using Hotz-Miller's inversion, it follows that $\Delta v_i(a_i, x; \beta_0, \theta_0)$ is identified from the data for all i, a_i, x. We take each pair (β, θ) to be a *structure* of the (empirical) model and its implied expected payoffs, denoted by $\mathcal{V}_{\beta,\theta} \equiv \{\Delta v_i(a_i, x; \beta, \theta)\}_{i, a_i, x \in \mathcal{I} \times A \times X}$, to be its corresponding *reduced form*.[10,11] We can then define identification using the notion of observational equivalence in terms of the expected payoffs.

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

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

The following lemma relates the parameters we want to identify to what can be observed.

[10]The empirical model is a pseudo-model. Because we do not use the equilibrium probabilities of the dynamic game corresponding to β, θ. We only consider the implied expected payoffs computed using the equilibrium beliefs that generate the data.

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

Lemma 1: *Under M1 - M5, we have for all $i, a_i > 0$, $\Delta v_i(a_i, x; \beta, \theta)$ can collected in the following vector form for all $(\beta, \theta) \in \mathcal{B} \times \Theta$:*

$$\begin{aligned} \Delta \mathbf{v}_i^{a_i}(\beta, \theta) &= \Delta \mathbf{R}_{i0}^{a_i} + \beta \Delta \mathbf{H}_i^{a_i} (I_J - \beta \mathbf{L})^{-1} \mathbf{R}_{i0} \\ &\quad + \left(\Delta \mathbf{R}_{i1}^{a_i} + \beta \Delta \mathbf{H}_i^{a_i} (I_J - \beta \mathbf{L})^{-1} \mathbf{R}_{i1} \right) \theta \\ &\quad + \beta \Delta \mathbf{H}_i^{a_i} (I_J - \beta \mathbf{L})^{-1} \boldsymbol{\epsilon}_i, \end{aligned} \tag{4}$$

where the elements in the equation above are collected and explained in Tables 1 and 2.

Matrix	Dimension	Representing
$\Delta \mathbf{R}_{i1}^{a_i}$	J by p	$E[\pi_{i1}(a_{-it}) \mid x_t = \cdot, a_{it} = a_i] - E[\pi_{i1}(a_{-it}) \mid x_t = \cdot, a_{it} = 0]$
$\mathbf{R}_{i1}$	J by p	$E[\pi_{i1}(a_{-it}) \mid x_t = \cdot]$
$\mathbf{L}$	J by J	$E[\psi(x_{t+1}) \mid x_t = \cdot]$
$\Delta \mathbf{H}_i$	J by J	$E[\psi(x_{t+1}) \mid x_t = \cdot, a_{it} = a_i] - E[\psi(x_{t+1}) \mid x_t = \cdot, a_{it} = 0]$

Table A. The matrices consist of (differences in) expected payoffs and probabilities. The latter represent conditional expectations for any function ψ of x_{t+1}.

Vector	Representing
$\boldsymbol{\epsilon}_i$	$E\left[\sum_{a_i' \in A} \varepsilon_{it}(a_i') \cdot \mathbf{1}[a_{it} = a_i'] \middle| x_t = \cdot \right]$
$\Delta \mathbf{R}_{i0}^{a_i}$	$E[\pi_{i0}(a_{it}, a_{-it}, x_t) \mid x_t = \cdot, a_{it} = a_i] - E[\pi_{i0}(a_{it}, a_{-it}, x_t) \mid x_t = \cdot, a_{it} = 0]$
$\mathbf{R}_{i0}$	$E[\pi_{i0}(a_t, x_t) \mid x_t = \cdot]$
$(I_J - \beta \mathbf{L})^{-1} \mathbf{R\Pi}_{ij}$	$\sum_{\tau=0}^{\infty} \beta^\tau E[\pi_{ij}(a_{t+\tau}, x_{t+\tau}) \mid x_t = \cdot]$
$\beta \Delta \mathbf{H}_i^{a_i} (I_J - \beta \mathbf{L})^{-1} \mathbf{R\Pi}_{ij}$	$\sum_{\tau=1}^{\infty} \beta^\tau E[\pi_{ij}(a_{t+\tau}, x_{t+\tau}) \mid x_t = \cdot, a_{it} = a_i]$ $- \sum_{\tau=1}^{\infty} \beta^\tau E[\pi_{ij}(a_{t+\tau}, x_{t+\tau}) \mid x_t = \cdot, a_{it} = 0]$
$\beta \Delta \mathbf{H}_i^{a_i} (I_J - \beta \mathbf{L})^{-1} \boldsymbol{\epsilon}_i$	$\sum_{\tau=1}^{\infty} \beta^\tau E\left[\sum_{a_i' \in A} \varepsilon_{it+\tau}(a_i') \cdot \mathbf{1}[a_{it+\tau} = a_i'] \middle| x_t = \cdot, a_{it} = a_i \right]$ $- \sum_{\tau=1}^{\infty} \beta^\tau E\left[\sum_{a_i' \in A} \varepsilon_{it+\tau}(a_i') \cdot \mathbf{1}[a_{it+\tau} = a_i'] \middle| x_t = \cdot, a_{it} = 0 \right]$

Table B. The J by 1 vectors represent (differences in) expected payoffs.

Proof: This is a slight variation of Lemma R in Sanches et al. (2016).■

Lemma 1 is simply a vectorization (across states) of the differences in discounted expected payoffs for player i from choosing action a_i relative to action 0. From the data we can identify $\Delta \mathbf{v}_i^{a_i}(\beta_0, \theta_0)$ for all $i, a_i > 0$. Hence, to identify (β_0, θ_0), it is enough to show that for all $(\beta, \theta) \neq (\beta_0, \theta_0)$, $\Delta \mathbf{v}_i^{a_i}(\beta, \theta) \neq \Delta \mathbf{v}_i^{a_i}(\beta_0, \theta_0)$ for some i and a_i. Our next lemma provides a characterization as to how changing β and θ can affect the expected payoffs.

Lemma 2: *Under M1 - M5, for any $i, a_i > 0$ and $(\beta, \theta), (\beta', \theta') \in \mathcal{B} \times \Theta$:*

$$\begin{aligned} \Delta \mathbf{v}_i^a(\beta, \theta) - \Delta \mathbf{v}_i^a(\beta, \theta') &= \left(\Delta \mathbf{R}_{i1}^{a_i} + \beta \Delta \mathbf{H}_i^{a_i}(I_J - \beta \mathbf{L})^{-1} \mathbf{R}_{i1}\right)(\theta - \theta'), & (5) \\ \Delta \mathbf{v}_i^a(\beta', \theta') - \Delta \mathbf{v}_i^a(\beta, \theta') &= (\beta - \beta') \Delta \mathbf{H}_i^{a_i}(I_J - \beta' \mathbf{L})^{-1}(I_J - \beta \mathbf{L})^{-1}(\mathbf{R}_{i0} + \mathbf{R}_{i1}\theta' + \boldsymbol{\epsilon}_i). & (6) \end{aligned}$$

And (β, θ) is identifiable if and only if there is no other (β', θ') such that for all $i, a_i > 0$:

$$\Delta \mathbf{v}_i^a(\beta', \theta') - \Delta \mathbf{v}_i^a(\beta, \theta') = \Delta \mathbf{v}_i^a(\beta, \theta) - \Delta \mathbf{v}_i^a(\beta, \theta').$$

Proof: Follows from some algebra based on equation (4).■

Lemma 2 illustrates the nature of the identification problem we have at hand. We highlight the following particulars:

(i) *If the discount rate is assumed to be known*, from (5), a sufficient condition for $\Delta \mathbf{v}_i^a(\beta_0, \theta) \neq \Delta \mathbf{v}_i^a(\beta_0, \theta')$ when $\theta \neq \theta'$ is that $\Delta \mathbf{R}_{i1}^{a_i} + \beta \Delta \mathbf{H}_i^{a_i}(I_J - \beta \mathbf{L})^{-1} \mathbf{R}_{i1}$ has full column rank for some $i, a_i > 0$.

(ii) *If the payoff function is assumed to be known*, from (6), a sufficient condition for $\Delta \mathbf{v}_i^a(\beta', \theta_0) \neq \Delta \mathbf{v}_i^a(\beta, \theta_0)$ when $\beta \neq \beta'$ is that $(\mathbf{R}_{i0} + \mathbf{R}_{i1}\theta' + \boldsymbol{\epsilon}_i) \neq 0$ and $\Delta \mathbf{H}_i^{a_i}$ is invertible some $i, a_i > 0$.

(iii) Suppose p is large relative to J. Then for any $i, a_i > 0$ such that $\Delta \mathbf{R}_{i1}^{a_i} + \beta \Delta \mathbf{H}_i^{a_i}(I_J - \beta \mathbf{L})^{-1} \mathbf{R}_{i1}$ has rank J, and for any $\theta', \beta \neq \beta'$ that $\Delta \mathbf{v}_i^a(\beta', \theta') \neq \Delta \mathbf{v}_i^a(\beta, \theta')$, by equating (5) and (6), we can always find θ such that $\Delta \mathbf{v}_i^a(\beta', \theta') = \Delta \mathbf{v}_i^a(\beta, \theta)$.

Point (i) shows that sufficient conditions for identification of the payoff parameters when the discount rate is assumed known can be easily stated and verified. More generally the sufficient condition for the identification of the payoff parameter can be stated in terms of the full column rank of the matrix that stacks together $\Delta \mathbf{R}_{i1}^{a_i} + \beta \Delta \mathbf{H}_i^{a_i}(I_J - \beta \mathbf{L})^{-1} \mathbf{R}_{i1}$ over all i and a_i. In the case we can identify the payoff function directly from the data, (ii) shows that the discount factor can also be identified and provide one type of sufficient conditions that can be readily checked. Point (iii) shares the intuition along the line of Manski (1993) that when the parameterization on the payoff function is too rich, (β, θ) may not identifiable in $\mathcal{B} \times \Theta$.

From Lemma 2, it is also apparent that we should be able to identify (β_0, θ_0) jointly when the change in the vector of expected payoffs from altering the discount factor moves in a different direction to the change caused by altering the payoff parameters.

3.2 Profiling

Profiling makes use of the fact that for each β the expected payoffs are linear in θ. We define $\mathbf{m}_i^{a_i}(\beta, \theta) \equiv \Delta \mathbf{v}_i^{a_i}(\beta_0, \theta_0) - \Delta \mathbf{v}_i^{a_i}(\beta, \theta)$, so that we can write:

$$\mathbf{m}_i^{a_i}(\beta, \theta) = \mathbf{a}_i^{a_i}(\beta) - \mathbf{B}_i^{a_i}(\beta)\theta,$$

where from (4),

$$\begin{aligned}\mathbf{a}_i^{a_i}(\beta) &= \Delta\mathbf{v}_i^{a_i}(\beta_0,\theta_0) - \Delta\mathbf{R}_{i0}^{a_i} - \beta\Delta\mathbf{H}_i^{a_i}(I_J-\beta\mathbf{L})^{-1}(\mathbf{R}_{i0}+\boldsymbol{\epsilon}_i),\\ \mathbf{B}_i^{a_i}(\beta) &= \Delta\mathbf{R}_{i1}^{a_i} + \beta\Delta\mathbf{H}_i^{a_i}(I_J-\beta\mathbf{L})^{-1}\mathbf{R}_{i1}.\end{aligned}$$

It is clear that for any given β, $\mathbf{m}_i^{a_i}(\beta,\theta)$ is linear in θ. The system of equations above can be expanded by stacking them across all i and a_i. In doing so we obtain the following vector value function, $\mathbf{m}:\mathcal{B}\times\Theta\rightarrow\mathbb{R}^{IJK}$:

$$\mathbf{m}(\beta,\theta)=\mathbf{a}(\beta)-\mathbf{B}(\beta)\theta,$$

where $\mathbf{a}(\beta)$ is a IJK by 1 vector and $\mathbf{B}(\beta)$ is a IJK by p matrix. Let $\mathcal{M}(\beta,\theta)\equiv\|\mathbf{m}(\beta,\theta)\|$, i.e. the Euclidean norm of $\mathbf{m}(\beta,\theta)$. Then by construction,

$$\mathcal{M}(\beta,\theta)=0 \text{ if } (\beta,\theta)=(\beta_0,\theta_0),$$

and any other (β,θ) such that $\mathcal{M}(\beta,\theta)=0$ is observationally equivalent to (β_0,θ_0) by the property of the norm. Next we profile out θ. Let † denotes the Moore-Penrose generalized inverse of a matrix. For each β, we define:

$$\theta^*(\beta)=\left(\mathbf{B}(\beta)^\top\mathbf{B}(\beta)\right)^\dagger\mathbf{B}(\beta)^\top\mathbf{a}(\beta),$$

so that $\theta^*(\beta)$ is a least squares solution to $\min_{\theta\in\Theta}\mathcal{M}(\beta,\theta)$. Then we define:

$$\mathcal{M}^*(\beta)=\mathcal{M}(\beta,\theta^*(\beta)).$$

By construction it also holds that

$$\mathcal{M}^*(\beta)=0 \text{ if } \beta=\beta_0.$$

In this way we can temporarily reduce the parameter space in the identification problem to a one-dimensional one. The reasoning is analogous to profiling in an estimation routine. Particularly we can ignore any θ that does not lie in $\arg\min_{\theta\in\Theta}\mathcal{M}(\beta,\theta)$ since necessarily,

$$\mathcal{M}(\beta,\theta)>\mathcal{M}(\beta,\theta^*(\beta))\geq 0.$$

Therefore (β_0,θ_0) is identified when $\mathcal{M}^*(\beta)$ has a unique minimum *and* $\min_{\theta\in\Theta}\mathcal{M}(\beta_0,\theta)$ has a unique solution.

THEOREM 1: (β_0,θ_0) *is identifiable if*

$$\mathcal{M}^*(\beta)=0 \text{ if and only if } \beta=\beta_0,$$

and $\mathbf{B}(\beta_0)$ *has full column rank.*

Proof: Suppose (β_0, θ_0) is identifiable. If there is $\beta' \neq \beta_0$ such that $\mathcal{M}^*(\beta') = 0$, then $\Delta \mathbf{v}_i^{a_i}(\beta_0, \theta_0) = \Delta \mathbf{v}_i^{a_i}(\beta', \theta^*(\beta'))$ for all i, a_i by the property of the norm. Since Θ is closed, by the projection theorem, $\theta^*(\beta')$ exists and is the unique element in Θ. This leads to a contradiction since (β_0, θ_0) and $(\beta', \theta^*(\beta'))$ are observationally equivalent. Next, suppose that $\mathbf{B}(\beta_0)$ does not have full column rank. Let θ' be another element in $\arg\min_{\theta \in \Theta} \mathcal{M}(\beta_0, \theta)$ that differs from θ_0. Since $\mathcal{M}(\beta_0, \theta) \geq 0$ for all $\theta \in \Theta$ and $\mathcal{M}(\beta_0, \theta_0) = 0$, $\mathcal{M}(\beta_0, \theta') = 0$. Thus (β_0, θ_0) and (β_0, θ') are observationally equivalent, also a contradiction.■

Comments on Theorem 1:

(i) High Level Assumptions. Conditions in Theorem 1 are high level as we do not relate them to the underlying primitives of the model. However, they are statements made on objects that observed or can be consistently estimated nonparametrically (as other conditions used in all of our theorems in this paper).

(ii) Feasible Check and Estimation. Since we have reduced the identification problem to a single-parameter that can reside only in a narrow range, there is no need to refer to complicated results for the identification of a general nonlinear model. Since it is possible to estimate $\mathcal{M}^*(\beta)$ consistently for all β, one can simply plot the sample counterpart of $\mathcal{M}^*$ over $\mathcal{B}$ for an exhaustive analysis of the problem. Once the minimum of $\mathcal{M}^*$ is found, the corresponding rank matrix can then be checked. This is indeed one way to estimate the discount factor, namely by grid search. We can detect an identification problem if the sample counterpart of $\mathcal{M}^*$ contains a flat region at the minimum, or when the sample counterpart of $\mathbf{B}(\beta_0)$ does not have full column rank.

4 Identification of the Switching Costs

In this section we consider games with switching costs. Section 4.1 introduces the specific structures of the payoff function and an additional conditional independence assumption. Section 4.2 derives the closed-form expressions for the switching costs. Throughout this section we do not need M5, but will continue to assume that M1 - M4 hold.

4.1 Games with Switching Costs

In what follows we distinguish past actions from other state variables. We denote actions from the previous period by w. So that, with a slight abuse of notation, at time t, there are two types of observed state variables, (x_t, w_t), where $w_t \equiv a_{t-1}$ and x_t is a vector of any other state variables. Actions from the past of more than one period can also be handled. We provide a discussion on this at the end of the section.

Assumptions M1 - M4 are now be updated accordingly by replacing x with (x, w) everywhere. In addition we need the following assumptions.

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

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

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

The components of the decomposition of π_i can be interpreted as follows. ϕ_i denotes the switching cost. η_i is an indicator function, modeled by the researcher, which takes value 1 if and only if a switching cost is present. We define ϕ_i to be zero whenever η_i takes value zero. In a model that contains switching costs, it must be the case that for some a_i, $W_{\eta_i}^0\left(a_i, \cdot\right)$ will be non-empty since it contains $w \in A^I$ such that the action of player's i coincides with a_i. Hence it is possible to consider distinguishing μ_i from ϕ_i. Then μ_i is to be interpreted as the residual of the payoffs that excludes the switching costs. Assumption N1 also imposes some distinct exclusion restrictions. Firstly, switching costs of each player are not affected by other players' actions in the same period. However, players' past actions and other state variables can have direct effects on switching costs. Secondly, past actions are excluded from μ_i. Typical components in μ_i that are often modeled to satisfy the required exclusion restrictions include payoff derived from interactions between players at the stage game, as well as other fixed costs such as fixed operating costs. Furthermore, this does not mean that variables from the past cannot affect μ_i since x_t can contain lagged actions and other state variables. N1 is assumed in many applications in the literature.

N2 imposes that knowing actions from the past does not help predict future state variables when the present action and other 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 many applications $\{x_t\}$ is assumed to be 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 N2. In any case, 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 $W_{\eta_i}^d$.

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

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

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

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

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

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

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

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

Here $SC_i\left(a_i', a_i'', x_t, a_{-it-1}\right)$ denotes a cost player i incurs from switching from action $a_{it-1} = a_i''$ to $a_{it} = a_i'$, at the state (x_t, a_{-it-1}). So for all x and a_{-i}, using just the definition of a switching cost we can set $SC_i\left(a_i', a_i', x, a_{-i}\right) = 0$ for all a_i'. Therefore without any further restrictions: $W_{\eta_i}^1(a_i, x) = \left\{w = (a_i', a_{-i}) : a_i' \in A \backslash \{a_i\}, a_{-i} \in A^{I-1}\right\}$ and $W_{\eta_i}^0(a_i, x) = \left\{w = (a_i, a_{-i}) : a_{-i} \in A^{I-1}\right\}$ for all x.

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

Before giving the formal results we briefly provide an intuition as to why N1 and N2 are helpful for identifying the switching costs.

Exclusion and Independence

The essence of our identification strategy is most transparent in a single agent decision problem under a two-period framework. For the moment suppose $I = 1$. Omitting the i subscript, the expected payoff for choosing action $a > 0$ under M1 to M4 is, cf. (8),

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

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 $E\left[\pi\left(a_{t+1},x_{t+1},w_{t+1}\right)|a_t,x_t,w_t\right]$ simplifies to $E\left[\pi\left(a_{t+1},x_{t+1},w_{t+1}\right)|a_t,x_t\right]$. Therefore we can write

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

where $\lambda\left(a,x\right)$ is a nuisance function that equals to $\mu\left(a,x\right)+\beta E\left[\pi\left(a_{t+1},x_{t+1},w_{t+1}\right)|a_t=a,x_t=x\right]$. Any variation in $v\left(a,x,w\right)$ induced by changes in w while holding (a,x) fixed can be traced only to changes in $\eta\left(a,x,w\right)$. Since λ is a free parameter, the switching costs can then be identified up to a location normalization by differencing over the support of w; e.g. through $\left(v\left(a,x,w\right)-v\left(0,x,w\right)\right)-\left(v\left(a,x,w_0\right)-v\left(0,x,w_0\right)\right)$ for some reference point w_0.

This simple argument can be generalized to identify switching costs in dynamic games. However, the way to difference out the nuisance function necessarily becomes more complicated. Particularly the nuisance function will then 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 also larger 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 up to some normalizations.[12]

4.2 Closed-Form Identification

We begin by introducing some additional notations and representation lemmas. For any x,w, we denote the ex-ante expected payoffs by $m_i\left(x,w\right)=E\left[V_i\left(x_t,w_t,\varepsilon_{it}\right)|x_t=x,w_t=w\right]$, where V_i is the value function defined in (1) that can also be defined recursively through

$$\begin{aligned} m_i\left(x,w\right) &= E\left[\pi_i\left(a_t,x_t,w_t\right)|x_t=x,w_t=w\right]+E[\sum_{a_i'\in A}\varepsilon_{it}\left(a_i'\right)\cdot\mathbf{1}\left[a_{it}=a_i'\right]|x_t=x,w_t=w] \\ &\quad +\beta E\left[m_i\left(x_{t+1},w_{t+1}\right)|x_t=x,w_t=w\right], \end{aligned} \quad (7)$$

[12] 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 the relation:

$$g_1\left(w\right)=c+g_2\left(w\right),$$

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

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

where the unknown constant is replaced by a linear transform (an expectation) of an unknown function.

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

$$\begin{aligned} v_i(a_i, x, w) &= E[\pi_i(a_{it}, a_{-it}, x_t, w_t) | a_{it} = a_i, x_t = x, w_t = w] \\ &\quad + \beta E[m_i(x_{t+1}, w_{t+1}) | a_{it} = a_i, x_t = x, w_t = w]. \end{aligned} \tag{8}$$

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

$$\phi_i(a_i, x, w; \eta_i) \cdot \eta_i(a_i, x, w) - \phi_i(0, x, w; \eta_i) \cdot \eta_i(0, x, w) = \sum_{w' \in W^{\Delta}_{\eta_i}(a_i, x)} \phi_{i,\eta_i}(a_i, x, w') \cdot \mathbf{1}[w = w'], \tag{9}$$

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.

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 $a'_i \neq a_i$:

$$\begin{aligned} \phi_{i,\eta_i}(a_i, x, w) &= SC_i(a_i, 0, x, a_{-i}) \text{ when } w = (0, a_{-i}), \\ \phi_{i,\eta_i}(a_i, x, w) &= -SC_i(0, a_i, x, a_{-i}) \text{ when } w = (a_i, a_{-i}), \\ \phi_{i,\eta_i}(a_i, x, w) &= SC_i(a_i, a'_i, x, a_{-i}) - SC_i(0, a'_i, x, a_{-i}) \text{ when } w = (a'_i, a_{-i}) \text{ for } a'_i \neq a_i \text{ or } 0. \end{aligned} \tag{10}$$

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 formalize the intuition at the end of Section 4.1 regarding how assumptions N1 and N2 allow us to isolate the (present period's) switching costs from other components of the payoffs.

LEMMA 3: *Under M1 - M4 and N1 - N2, we have for all* $i, a_i > 0$ *and* a_{-i}, x, w:

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

where

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

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

Lemma 3 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 (11), which is that of a_{-it} given (x_t, w_t), is identifiable from the data. Since a conditional expectation operator is a linear operator, and the support of w_t is a finite set with $(K+1)^I$ elements, we can then represent (11) by a matrix equation.

LEMMA 4: *Under M1 - M4 and N1 - N2, we have for all* $i, a_i > 0$ *and* x:

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

where $\Delta \mathbf{v}_i (a_i, x)$ *denotes a* $(K+1)^I-$*dimensional vector of normalized expected discounted payoffs,* $\{\Delta v_i (a_i, x, w)\}_{w \in A^I}$, $\mathbf{Z}_i (x_t)$ *is a* $(K+1)^I$ *by* $(K+1)^{I-1}$ *matrix of conditional probabilities,* $\{\Pr [a_{-it} = a_{-i} | x_t = x, w_t = w]\}_{(a_{-i},w) \in A^{I-1} \times A^I}$, $\boldsymbol{\lambda}_i (a_i, x)$ *denotes a* $(K+1)^{I-1}$ *by 1 vector of* $\{\lambda_i (a_i, a_{-i}, x)\}_{a}$ $\mathbf{Q}_i (a_i, x)$ *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: 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 2: *Under M1 - M4 and N1 - N2, for each* $i, a_i > 0$ *and* x, *if (i)* $\mathbf{Q}_i (a_i, x)$ *has full column rank; (ii)* $\rho (\mathbf{Z}_i (x)) + \rho (\mathbf{Q}_i (a_i, x)) = \rho([\mathbf{Z}_i (x) : \mathbf{Q}_i (a_i, x)])$, *then* $\mathbf{Q}_i (a_i, x)^\top \mathbf{M}_{\mathbf{Z}_i(x)} \mathbf{Q}_i (a_i, x)$ *is non-singular, and*

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

PROOF: The full column rank condition of $\mathbf{Q}_i(a_i, x)$ 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 $\mathbf{M}_{\mathbf{Z}_i(x)}\mathbf{Q}_i(a_i, x)$ preserves the rank of $\mathbf{Q}_i(a_i, x)$. Therefore $\mathbf{Q}_i(a_i, x)^\top \mathbf{M}_{\mathbf{Z}_i(x)}\mathbf{Q}_i(a_i, x)$ must be non-singular. Otherwise the columns of $\mathbf{M}_{\mathbf{Z}_i}(x)\mathbf{Q}_i(a_i)$ is linearly dependent, and some linear combination of the columns in $\mathbf{Q}_i(a_i)$ 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 (13) by $\mathbf{M}_{\mathbf{Z}_i(x)}$and solve for $\boldsymbol{\phi}_{i,\eta_i}(a_i, x)$.$\blacksquare$

In order for condition (ii) in Theorem 2 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 revisit Examples 1 - 3. For notational simplicity we shall assume $I = 2$, so that $w_t \in \{(0,0), (0,1), (1,0), (1,1)\}$. And since $A = \{0,1\}$ in Examples 1 and 2, we shall also drop a_i from $\Delta\mathbf{v}_i(a_i, x) = \{\Delta v_i(a_i, x, w)\}_{w\in A^I}$ and $\boldsymbol{\lambda}_i(a_i, x) = \{\lambda_i(a_i, a_{-i}, x)\}_{a_{-i}\in A^{I-1}}$.

EXAMPLE 1 (ENTRY COST, CONT.): Equation (13) can be written as

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

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 3 to hold is when the lower half of $\mathbf{Z}_i(x)$ has full rank, i.e. when $P_{-i}(0|x,(1,0)) \neq P_{-i}(0|x,(1,1))$.

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

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

An analogous sufficient condition that ensures condition (ii) in Theorem 3 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 $\Delta \mathbf{v}_i(2,x) = \{\Delta v_i(2,x,w)\}_{w \in A^I}$,

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

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

Clearly the required rank condition of Theorem 2 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 4 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 $\mathbf{Q}_i(2, x)\,\boldsymbol{\phi}_{i,\eta_i}(2, x)$ 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 2 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 (15). 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 $\mathbf{Q}_i(2, x)\,\boldsymbol{\phi}_{i,\eta_i}(2, x)$ 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 2:

(i) Pointwise Closed-form Identification. Our result is obtained pointwise for each $i, a_i > 0$ and x. Therefore the finite support assumption in M4 is not necessary. The closed-form expression in (14) also suggests that a closed-form estimator for the switching costs can be obtained by simply replacing the unknown probabilities and expected payoffs by their sample counterparts. 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 2 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 (13) can be obtained from: $\rho(\mathbf{Z}_i(x)) + \rho(\mathbf{Q}_i(a_i, x)) \leq (K+1)^I$, so that (the number of switching cost parameters one wish to identify is the cardinality of $W_{\eta_i}^{\Delta}(a_i, x)$ equals) $\rho(\mathbf{Q}_i(a_i, x)) \leq (K+1)^I - 1$. In the least favorable case, in terms of applying Theorem 2, the previous inequality can be strengthened by using the maximal rank of $\mathbf{Z}_i(x)$, which is $(K+1)^{I-1}$. Then $\rho(\mathbf{Q}_i(a_i, x))$ 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 (13) using the full generality of the definition of a switching cost is $(K+1)^I$; see (10). 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 restriction than a more common 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.[13] A preferable solution is to appeal industry specific knowledge to approximate certain costs, or use other prior economic impose additional structure on the switching costs. A natural example of the latter is the menu cost, or other adjustment costs in pricing games (Slade (1998)). Also see Myśliwski (2015) who uses the identification strategy proposed in this paper, where he imposes equality restrictions (cf. Example 3) on costs associated with supermarket discount decisions.

[13] There are recent studies focusing on the effects on counterfactuals from an incorrect normalization, for example see Aguirregabiria and Suzuki (2014).

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 N3 (Equality Restrictions): *For all i, x, there exists a $K(K+1)^I$ by κ matrix $\widetilde{\mathbf{Q}}_i(x)$ with full column rank and a κ by 1 vector of functions $\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)$ so that $\widetilde{\mathbf{Q}}_i(x)\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)$ represents a vector of functions that satisfy some equality constraints imposed on $\{\mathbf{Q}_i(a_i,x)\boldsymbol{\phi}_{i,\eta_i}(a_i,x)\}_{a_i \in A}$.*

The matrix $\widetilde{\mathbf{Q}}_i(x)$ can be constructed from $diag\{\mathbf{Q}_i(1,x),\ldots,\mathbf{Q}_i(K,x)\}$, 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 $\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)$. The following lemma gives the matrix representation of the expected payoffs in this case (cf. Lemma 4).

Lemma 5: *Under M1 - M4, N1 - N3, we have for all i, x:*

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

where $\Delta\mathbf{v}_i(x)$ denotes a $K(K+1)^I$ −dimensional vector of normalized expected discounted payoffs, $\{\Delta\mathbf{v}_i(a_i,x)\}_{a_i \in A\setminus\{0\}}$, $\mathbf{Z}_i(x)$ 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 $\{\boldsymbol{\lambda}_i(a_i,x)\}_{a_i \in A\setminus\{0\}}$, $\widetilde{\mathbf{Q}}_i(x)$ and $\widetilde{\boldsymbol{\phi}}_{i,\eta_i}(x)$ are described in Assumption N3.

Proof: Immediate.■

Using Lemma 5, our next result generalizes Theorem 3 by allowing for the equality restrictions across all actions.

Theorem 3: *Under M1 - M4, N1 - N3, for each i, x, if (i) $\widetilde{\mathbf{Q}}_i(x)$ has full column rank and, (ii) $\rho(I_K \otimes \mathbf{Z}_i(x)) + \rho(\widetilde{\mathbf{Q}}_i(x)) = \rho([I_K \otimes \mathbf{Z}_i(x) : \widetilde{\mathbf{Q}}_i(x)])$, then $\widetilde{\mathbf{Q}}_i^\top(x)\mathbf{M}_{I_K \otimes \mathbf{Z}_i(x)}\widetilde{\mathbf{Q}}_i(x)$ is non-singular, and*

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

Proof: Same as the proof of Theorem 2.■

Our previous comments on Theorem 2 are also relevant for Theorem 3. 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

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

I.e. we have imposed the equality restrictions on the entry costs and scrap values for each player only depend on each her own actions. Then, for all i, x, the content of equation (16) (in Lemma 5) is

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

Note that the order condition is now satisfied. However, condition (ii) in Theorem 3 does not hold since a vector of ones is contained in both $\mathcal{CS}\left(\mathbf{Z}_i\left(x\right)\right)$ and $\mathcal{CS}(\mathbf{Q}_i\left(x\right))$. Even if we go further and assume the entry cost and scrap value have the same magnitude (i.e. $EC_i\left(x\right) = -SV_i\left(x\right)$), the rank condition will still not be satisfied. In this case $\mathbf{Q}_i\left(1,x\right)\boldsymbol{\phi}_{i,\eta_i}\left(1,x\right)$ becomes

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

Mathematically, the failure to apply our result in the example above can be traced to the fact that $\mathbf{Z}_i\left(x\right)$ 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. Similar finding can be found for instance in Aguirregabiria and Suzuki (2014, equation (21)).[14] We refer the reader to their work for related results in a simpler setting as well

[14] Interestingly, although Aguirregabiria and Suzuki (2014) explicitly assume the knowledge of the discount factor throughout their paper, a careful inspection of their Proposition 2 will also suggest that either the entry cost or scrap value in their model can be identified independently of the discount factor under some normalization. Our Theorem 3 is a more general version of this particular implication; Aguirregabiria and Suzuki (2014) derive their result for a single agent model with binary choice and $\{x_t\}$ is assumed to be strictly exogenous. We thank a referee for pointing this out.

as a list of references they provide of empirical work that make normalization assumptions on either one of these switching costs. It is worth noting that the work of Aguirregabiria and Suzuki (2014) focuses on the effects normalizations may have on certain counterfactuals, unlike ours, which is only concerned with identification and estimation of the primitives. Quantifying these effects from a particular misspecification, whether by assuming an incorrect discount factor or imposing a wrong normalization on a switching cost, is an important issue but it is beyond the scope of our work.

The above results can be adapted to allow for effects from past actions beyond one period with little modification. Specifically, all results above hold if we re-define w_t to be $a_{t-\varsigma}$ for any finite $\varsigma \geq 1$, and then replace x_t by $\widetilde{x}_t = (x_t, a_{t-1}, \ldots, a_{t-\varsigma+1})$ 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 2 and 3. 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}.

5 Numerical Illustration

We illustrate the use of our identification strategies and implement the suggested estimators in the previous sections. Section 5.1 gives results from a Monte Carlo study taken from Pesendorfer and Schmidt-Dengler (2008). Section 5.2 estimates a discrete investment game using the data from Ryan (2012).

5.1 Monte Carlo Study

The simulation design is the two-firm dynamic entry game taken from Section 7 in Pesendorfer and Schmidt-Dengler (2008). In period t each firm i has two possible choices, $a_{it} \in \{0, 1\}$, with $a_{it} = 1$ denoting entry. The only 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:

$$\pi_{1,\theta}(a_{1t}, a_{2t}, x_t) = a_{1t}(\mu_1 + \mu_2 a_{2t}) + a_{1t}(1 - a_{1t-1}) F + (1 - a_{1t}) a_{1t-1} W, \tag{18}$$

where μ_1, μ_2, F and W are 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. The data generating parameters are set as: $(\mu_{10}, \mu_{20}, F_0, W_0) = (1.2, -1.2, -0.2, 0.1)$ and $\beta_0 = 0.9$. Pesendorfer and Schmidt-Dengler (2008) show there are three distinct equilibria for this game.

We take W_0 to be known since it cannot be identified jointly with F_0, and estimate there remaining parameters. Since the payoff function satisfies Assumption M5 there are two ways to estimate the model. One (Method A) is profiling out all the payoff parameters using the OLS expression and use grid search to estimate the discount factor. The other (Method B) is to estimate F_0 in closed-form independently first before profiling out the other payoff parameters. Both estimators are expected to be consistent since we know the correct model specification. Otherwise we can perform formal tests to see if they differ; see Section 5.2 below. We provide summary statistics for both methods. We consider all three equilibria as enumerated in Pesendorfer and Schmidt-Dengler (2008). We perform 10000 simulations with each sample size, N, of $100, 1000, 10000$ and 100000. We report the mean and standard deviation (in italics) for each estimator, as well as the square root of the aggregated mean square errors (in bold) for each estimation method.

Equ	*N*	*Met*	β_0		F_0		μ_{10}		μ_{20}		*RMSE*
1	100	A	0.822	*0.267*	-0.255	*0.283*	1.176	*0.303*	-1.120	*0.297*	**0.589**
		B	0.823	*0.265*	-0.299	*0.629*	1.158	*0.391*	-1.088	*0.508*	**0.952**
	1000	A	0.861	*0.182*	-0.204	*0.090*	1.210	*0.114*	-1.191	*0.095*	**0.254**
		B	0.861	*0.181*	-0.200	*0.148*	1.213	*0.118*	-1.195	*0.121*	**0.292**
	10000	A	0.899	*0.020*	-0.201	*0.028*	1.200	*0.030*	-1.199	*0.030*	**0.055**
		B	0.899	*0.026*	-0.200	*0.044*	1.200	*0.031*	-1.200	*0.036*	**0.069**
	100000	A	0.900	*0.000*	-0.200	*0.009*	1.200	*0.009*	-1.200	*0.009*	**0.016**
		B	0.900	*0.000*	-0.200	*0.014*	1.200	*0.009*	-1.200	*0.011*	**0.020**
2	100	A	0.833	*0.248*	-0.363	*0.415*	1.007	*0.409*	-0.856	*0.545*	**0.940**
		B	0.834	*0.247*	-0.206	*0.552*	1.097	*0.455*	-0.998	*0.658*	**1.031**
	1000	A	0.869	*0.161*	-0.223	*0.153*	1.187	*0.173*	-1.137	*0.218*	**0.363**
		B	0.869	*0.159*	-0.199	*0.150*	1.203	*0.167*	-1.161	*0.210*	**0.350**
	10000	A	0.900	*0.020*	-0.203	*0.049*	1.196	*0.042*	-1.193	*0.065*	**0.094**
		B	0.899	*0.026*	-0.200	*0.046*	1.198	*0.042*	-1.196	*0.062*	**0.092**
	100000	A	0.900	*0.000*	-0.200	*0.015*	1.200	*0.012*	-1.200	*0.020*	**0.028**
		B	0.900	*0.000*	-0.200	*0.015*	1.200	*0.011*	-1.200	*0.019*	**0.026**
3	100	A	0.834	*0.247*	-0.368	*0.422*	1.011	*0.412*	-0.857	*0.549*	**0.947**
		B	0.832	*0.250*	-0.211	*0.600*	1.104	*0.480*	-1.002	*0.710*	**1.110**
	1000	A	0.865	*0.171*	-0.223	*0.155*	1.192	*0.182*	-1.136	*0.220*	**0.375**
		B	0.865	*0.171*	-0.199	*0.160*	1.209	*0.181*	-1.161	*0.226*	**0.376**
	10000	A	0.900	*0.020*	-0.204	*0.050*	1.196	*0.042*	-1.193	*0.065*	**0.094**
		B	0.900	*0.018*	-0.201	*0.050*	1.197	*0.042*	-1.196	*0.065*	**0.094**
	100000	A	0.900	*0.000*	-0.200	*0.015*	1.200	*0.012*	-1.200	*0.020*	**0.028**
		B	0.900	*0.000*	-0.200	*0.015*	1.200	*0.012*	-1.200	*0.020*	**0.028**

Table 1: Summary statistics from estimating $\beta_0, F_0, \mu_{10}, \mu_{20}$ using data generated from equilibria 1 to 3 in Pesendorfer and Schmidt-Dengler (2008).

Our simulation study shows there is no reason why the discount factor cannot be consistently estimated along with other payoff parameters. It is also interesting to compare the statistical properties of the estimators obtained using method A and B. We find that the square root of the aggregated mean square error for the two estimators to be very similar across all three equilibria in large samples;

with method A performing marginally better in equilibria 1 and 3, and method B performing marginally better in equilibrium 2. With smaller sample size method B seems to do worse than method A. There also does not seem to be a dominating estimator for F_0. Recall that method B requires fewer assumptions on structure of the model, while method A correctly imposes the remaining parametric structure of the payoff function but also has more parameters to estimate simultaneously. Earlier versions of our paper show that when β_0 is correctly assumed then the OLS estimator of Sanches et al. (2016) performs better than method B in estimating F_0 (using the mean square criterion). We also find that the OLS estimator is inconsistent when incorrect guesses of the discount factor are used.

5.2 Empirical Illustration

We estimate a simplified version of an entry-investment game based on the model studied in Ryan (2012) using his data. 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

We download Ryan's data from the Econometrica webpage.[15] There are two sets of data. One contains aggregate prices and quantities for all the US regional markets from the US Geological Survey's Mineral Yearbook. The other contains the capacities of plants and plant-level information that Ryan has collected for the Portland cement industry in the United States from 1980 to 1998. 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. Following Ryan 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 match most of the summary statistics of plant-level data in Table 2 of Ryan. Any discrepancies most likely can be attributed to the way we treat the change in plants' names, which may differ to Ryan in a

[15] https://www.econometricsociety.org/content/supplement-costs-environmental-regulation-concentrated-industry-0.

small number of cases.

Dynamic Game

Ryan 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/divestment decisions. There are two action variables in Ryan's model. One is a binary choice for entry and the other is a continuous level of investment. Past actions are the only observed endogenous state variables in the game. The aggregate data that are used to construct variable profits, through a static Cournot game with capacity constraints between firms, are treated as exogenous.

We consider a discrete game that fits the general model description in Section 2. The main departure from Ryan (2012) is that 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 can be broken down to variable and fixed profits. We assume the variable profit is determined by the players competing 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 a standard i.i.d. type-1 extreme value distribution.

Estimation

The period expected payoff for each firm as a function of the observables consists of variable profits, operating costs and switching costs. The variable profit is derived from a capacity constrained Cournot game constructed from the same demand and cost functions estimated as in Ryan's paper. Operating cost enters the payoff function additively and is treated as a dynamic parameter to be estimated. These two components are non-zero if $a_{it} > 0$. For the switching costs we normalize the payoff for choosing action 0 to be zero. Therefore there are a total of 25 switching cost parameters to be estimated.[16]

[16] 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).

The payoff function in our empirical model satisfies Assumption M5 so we can profile out the payoff parameters. We estimate the model using methods A and B as described in Section 5.1. We also test if the two estimates of the switching costs statistically differ. Instead of using nonparametric estimator, similar to Ryan, we use a multinomial logit to estimate the choice and transition probabilities in the first stage. More specifically, method A profiles out the 26 linear coefficients and uses grid search to estimate the discount factor. Method B first estimates the 25 switching cost parameters in closed-form using the closed-form expression in Section 4, treat them as known, before profiling and performing the grid search. We estimate the standard errors, as well as computing the p-value of the Wald statistics to test if the switching costs estimators from methods A and B differ by bootstrapping. Our bootstrap sample is generated using the multinomial logit choice and transition probabilities for each player in each market in the same manner as a parametric bootstrap; cf. Kasahara and Shimotsu (2008) and Pakes, Ostrovsky and Berry (2007).

Results

We estimate the model twice. Once using the data from before and after the implementation of the 1990 CAAA. We assume the data are generated from different equilibria over the two time periods, but the same equilibrium is played in all markets within each time period.[17] Table 2 and 3 compiles the results from estimating switching costs using the data from the years 1980 to 1990 and 1991 to 1998 respectively. Tables 4 and 5 give the estimates for the discount factor and fixed operating cost using the data from the corresponding periods. Tables 4 and 5 report analogous results using the data from the years 1991 to 1998.

The signs and relative magnitudes of the estimated switching costs almost uniformly make sensible economic sense. E.g., by reading down the columns in Tables 2 and 3, we see that entering at higher capacity level generally implies higher cost (negative payoff), and increasing the capacity level should be costly while divestment can return revenue for firms. This is quite an impressive finding in particular for method B, which shows that the observed probabilities can generate switching costs estimates that capture reasonably well a key feature of a complicated structural model. The switching cost estimates from both methods A and B are similar. The Wald statistics do not find the two switching costs estimators to be statistically different.[18] Therefore we do not reject the capacity constrained Cournot game specification based on comparing the switching costs estimates.

[17] If the same equilibrium is played across markets then data can be pooled together. Otsu, Pesendorfer and Takahashi (2015) suggest the data in Ryan (2012) between 1980 and 1990 should not be pooled across markets, while the data from 1991 to 1998 pass their poolability test.

[18] Our test statistic takes a standard quardratic form of the difference between the switching costs estimates from methods A and B. Its asymptotic distribution under the null hypothesis (of no difference) is a Chi-squared random variable with 25 degree of freedoms.

Comparing Tables 2 and 3 also show the entry and switching costs increase after the implementation of 1990 CAAA. Higher entry costs is a key finding in Ryan's paper as new entrants face more stringent regulations than incumbents. Our finding of the increase in switching costs can also be partly attributed to the new plants using newer (or better maintained) equipment that requires with more certification and testing than previously. We also find the discount factor to be around the range that are usually used (between 0.9 and 0.95) apart from the estimate using method B before the 1990 CAAA that appears close to the boundary.[19] Although our estimates suggest firms face a lower borrowing rate than in Ryan, we do not reject the hypothesis that $\beta = 0.9$ as assumed in his paper. We also find a small increase in the fixed operating costs after the implemetation of 1990 CAAA, which can account in parts for the installation and maintenance of monitoring equipment for regular emission reporting needed to apply for operating permits. We refer the reader to Section 2 in Ryan for further details of the industry background.

[19] The infinite time expected discounted payoffs with respect to each action is unbounded with $\beta = 1$. However, the differences between diverge very slowly when we approximate them with a Neumann sum, and the objective function appears to be well-defined numerically even as β is very close to 1.

Method A						
	$a_{it-1}=0$	$a_{it-1}=1$	$a_{it-1}=2$	$a_{it-1}=3$	$a_{it-1}=4$	$a_{it-1}=5$
$a_{it}=1$	-3.300	-	2.265	5.080	7.956	10.770
	0.985	-	*0.680*	*0.707*	*0.766*	*0.929*
$a_{it}=2$	-10.502	-5.243	-	5.528	10.609	15.810
	0.937	*0.719*	-	*0.887*	*0.998*	*1.117*
$a_{it}=3$	-23.266	-15.439	-7.624	-	7.996	16.050
	1.405	*1.010*	*0.683*	-	*0.923*	*1.237*
$a_{it}=4$	-41.023	-30.620	-20.196	-9.808	-	11.648
	2.003	*1.850*	*1.430*	*1.094*	-	*1.442*
$a_{it}=5$	-52.879	-50.648	-39.027	-25.756	-11.949	-
	2.281	*2.585*	*2.041*	*1.395*	*1.537*	-

Method B						
	$a_{it-1}=0$	$a_{it-1}=1$	$a_{it-1}=2$	$a_{it-1}=3$	$a_{it-1}=4$	$a_{it-1}=5$
$a_{it}=1$	-2.776	-	2.540	5.333	8.014	11.696
	0.269	-	*0.333*	*0.567*	*0.967*	*1.113*
$a_{it}=2$	-10.483	-5.197	-	5.243	10.466	15.893
	0.689	*0.365*	-	*0.368*	*0.718*	*1.110*
$a_{it}=3$	-23.279	-15.427	-7.769	-	7.732	16.134
	1.339	*0.920*	*0.474*	-	*0.640*	*1.006*
$a_{it}=4$	-41.422	-31.007	-20.797	-10.416	-	10.852
	1.808	*1.594*	*1.078*	*0.682*	-	*0.864*
$a_{it}=5$	-54.378	-52.892	-41.874	-28.792	-16.091	-
	1.911	*2.232*	*1.844*	*1.659*	*1.835*	-

Specification	Test
Statistic	14.069
p-value	*0.961*

Table 2: Results from estimating switching costs using data from the years 1980 to 1990. Standard errors and p-value are obtained using 500 bootstrap samples (reported in italics).

Method A						
	$a_{it-1} = 0$	$a_{it-1} = 1$	$a_{it-1} = 2$	$a_{it-1} = 3$	$a_{it-1} = 4$	$a_{it-1} = 5$
$a_{it} = 1$	-6.962	-	4.449	9.881	15.125	20.264
	1.530	-	*1.514*	*1.501*	*1.689*	*1.634*
$a_{it} = 2$	-17.038	-8.291	-	9.872	18.531	26.722
	1.723	*1.364*	-	*1.714*	*1.860*	*1.527*
$a_{it} = 3$	-35.489	-23.412	-11.411	-	12.961	24.283
	2.444	*1.866*	*1.371*	-	*1.955*	*1.614*
$a_{it} = 4$	-51.544	-50.043	-33.220	-16.363	-	16.524
	3.061	*3.419*	*3.278*	*2.825*	-	*3.561*
$a_{it} = 5$	-64.018	-63.994	-61.481	-48.514	-24.374	
	4.514	*4.524*	*4.502*	*3.683*	*2.056*	

Method B						
	$a_{it-1} = 0$	$a_{it-1} = 1$	$a_{it-1} = 2$	$a_{it-1} = 3$	$a_{it-1} = 4$	$a_{it-1} = 5$
$a_{it} = 1$	-5.653	-	5.294	10.730	16.264	21.567
	0.726	-	*0.704*	*1.109*	*1.703*	*1.378*
$a_{it} = 2$	-17.746	-9.278	-	8.774	17.461	25.754
	1.379	*0.780*	-	*0.857*	*1.364*	*1.218*
$a_{it} = 3$	-36.098	-24.537	-11.950	-	11.862	23.489
	2.282	*1.767*	*1.128*	-	*1.221*	*1.401*
$a_{it} = 4$	-51.840	-50.425	-33.468	-16.760	-	16.753
	2.202	*2.649*	*2.397*	*1.904*	-	*2.025*
$a_{it} = 5$	-64.236	-64.355	-61.706	-48.272	-24.093	
	6.712	*6.771*	*6.713*	*5.695*	*3.389*	

Specification	Test
Statistic	13.196
p-value	*0.975*

Table 3: Results from estimating switching costs using data from the years 1991 to 1998. Standard errors and p-value are obtained using 500 bootstrap samples (reported in italics).

Method A	
Discount Factor	Operating Cost
0.956	-1.679
0.084	*0.489*
Method B	
Discount Factor	Operating Cost
0.999	-1.523
0.075	*0.649*

Table 4: Results from estimating the discount factor and fixed operating cost using data from the years 1980 to 1990. Standard errors are obtained using 500 bootstrap samples (reported in italics).

Method A	
Discount Factor	Operating Cost
0.938	-2.079
0.162	*1.10*
Method B	
Discount Factor	Operating Cost
0.946	-1.893
0.160	*0.948*

Table 5: Results from estimating the discount factor and fixed operating cost using data from the years 1991 to 1998. Standard errors are obtained using 500 bootstrap samples (reported in italics).

6 Conclusion

Studies of identification of dynamic games typically focus on the payoff parameters and take other primitives to be known. The value of the discount factor in empirical work is often assumed rather than estimated. Therefore the presumption on the value of the discount factor appears to be necessary for identification and estimation of the payoff parameters. We show that the analysis of joint identification and estimation of the discount factor and payoff parameters can be very simple when the payoffs have a linear-in-parameter specification. The complete analysis is equivalent to doing a grid search over the $[0, 1]$ interval. There is some novelty in our approach.[20] Our identification criterion makes full information on the underlying parameter on the empirical model and has implications for all two-step estimation methods that wish to estimate the discount factor. One can of course argue that analogous analysis can be performed with other loss functions. However, without profiling, an exhaustive search with other loss functions, e.g. those based on the choice probabilities (e.g. maximum likelihood, moments or asymptotic least squares) may not be feasible. Especially when there are many payoff parameters, it may not be trivial to locate the global minimizer even for a single candidate value of the discount factor. We also provide conditions when switching costs can be identified in closed-form under independently of the discount factor and specification of other payoff components (linear or otherwise). The latter gives a closed-form estimator for the switching costs that can be used for specification testing, which for example can use to test for the mode of competition between firms. We illustrate the scope of its applications in a Monte Carlo study and an empirical game using real data.

Our take away message that one should be able to identify the discount factor in dynamic games extends beyond discrete choice games. For example, the joint identification and estimation of the discount factor and payoff parameters in games with supermodular payoff functions should also be possible. See Bajari, Benkard and Levin (2007) and Srisuma (2013) for descriptions of other types of dynamic games. However, the practical implementation may be an issue since there is no obvious ways to reduce the parameter space even when the payoff functions take a linear-in-parameter structure as discussed in the previous paragraph.

Throughout the paper we assume the most basic setup of a game with independent private values under the usual conditional independence, and we anticipate the data to have been generated from a single equilibrium. Our results can in principle 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 with a more general data structure is an important area of future research.

[20] Appendix A provides a more analytical condition for identification. However, it is only sufficient and the failure to apply Theorem 4 (below) does not mean the model cannot be identified.

Appendix A

This appendix we attempt to give a more analytical approach that ensures identification of (β_0, θ_0). We assume the setup in Section 3 (i.e. assume Assumptions M1 to M5). We proceed by setting up a map that defines the parameter of interest as its fixed-point. We first introduce some additional notations and a characterization for the observationally equivalence of (β_0, θ_0).

For any $x = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$ and $y = (y_1, \ldots, y_{p+1})^\top \in \mathbb{R}^{p+1}$, let $\|x\|_{\alpha_1} = \max_{i=1,\ldots,p} |x_i|$ and $\|y\|_{\alpha_2} = \max_{i=1,\ldots,p} |y_i| + |y_{p+1}|$. Then for a class of $p+1$ by p real matrices, we denote the matrix norms induced by $\left(\|\cdot\|_{\alpha_1}, \|\cdot\|_{\alpha_2}\right)$ by $\|\cdot\|_{\alpha_1,\alpha_2}$. We comment that these are not standard induced matrix norms, however they have simple explicit bounds. In particular it is easy to verify that, for any matrix $p+1$ by p, $C = (c_{ij})$,

$$\|C\|_{\alpha_1,\alpha_2} \leq \max_{i=1,\ldots,p} \sum_{j=1}^{p} |c_{ij}| + \sum_{j=1}^{p} |c_{p+1,j}| \,.$$

We also need the parameter space to be compact. Let $\overline{\Theta} \equiv \left\{\theta \in \Theta : \max_{i=1,\ldots,p} |\theta_i| \leq \overline{k}\right\}$ and $\overline{\mathcal{B}} \equiv \left[0, \overline{b}\right]$ for some positive $\overline{k}$ and $\overline{b} \in (0,1)$. Next we introduce the following relation to study the identification of (β_0, θ_0).

Lemma 6: *Under M1 - M6, (β, θ) is observationally equivalent to (β_0, θ_0) if and only if (β, θ) satisfies*

$$\mathbf{c}_i^{a_i} - \mathbf{D}_i^{a_i}(\beta)\,\theta - \mathbf{E}_i(\beta) = \mathbf{F}_i^{a_i} \begin{pmatrix} \theta \\ \beta \end{pmatrix} \tag{19}$$

for all $i, a_i > 0$, where

$$\begin{aligned}
\mathbf{c}_i^{a_i} &= \Delta \mathbf{v}_i^{a_i}(\beta_0, \theta_0) - \Delta \mathbf{R}_{i0}^{a_i}, \\
\mathbf{D}_i^{a_i}(\beta) &= \beta \Delta \mathbf{H}_i^{a_i} (I_J - \beta \mathbf{L})^{-1} \mathbf{R}_{i1}, \\
\mathbf{E}_i(\beta) &= \beta^2 \Delta \mathbf{H}_i^{a_i} \mathbf{L} (I_J - \beta \mathbf{L})^{-1} (\mathbf{R}_{i0} + \boldsymbol{\epsilon}_i), \\
\mathbf{F}_i^{a_i} &= \left[\Delta \mathbf{R}_{i1}^{a_i} : \Delta \mathbf{H}_i^{a_i} (\mathbf{R}_{i0} + \boldsymbol{\epsilon}_i)\right].
\end{aligned}$$

Proof: Equation (19) is obtained by re-arranging equation (4), after applying the identity that $(I_J - \beta\mathbf{L})^{-1} = I_J + \beta\mathbf{L}(I_J - \beta\mathbf{L})^{-1}$ and replace $\Delta\mathbf{v}_i^{a_i}(\beta, \theta)$ by $\Delta\mathbf{v}_i^{a_i}(\beta_0, \theta_0)$. Therefore, by construction, (β, θ) satisfies (19) if and only if it is observationally equivalent to (β_0, θ_0).■

The following result provides one condition that is sufficient for the identification of (β_0, θ_0).

Theorem 4: *Assume that $J \geq p+1$ and M1 - M5 hold. Suppose there exists i, a_i such that: (i) the rank of $\mathbf{F}_i^{a_i}$ is $p+1$; (ii) there exists a $p+1$ by J matrix $\mathbf{A}_0$ such that $\mathbf{A}_0 \mathbf{F}_i^{a_i}$ is non-singular;*

and (iii) $\max\{\mathbf{g}_1, \mathbf{g}_2\} < 1$, *where*

$$\mathbf{g}_1 = \max_{\beta \in \overline{\mathcal{B}}} \left\| \left(\mathbf{A}_0 \mathbf{F}_i^{a_i}\right)^{-1} \mathbf{A}_0 \Delta \mathbf{H}_i^{a_i} \beta \left(I_J - \beta \mathbf{L}\right)^{-1} \mathbf{R}_{1i} \right\|_{\alpha_1, \alpha_2},$$

$$\mathbf{g}_2 = \max_{\beta, \beta' \in \overline{\mathcal{B}}, \theta \in \overline{\Theta}} \left\| \left(\mathbf{A}_0 \mathbf{F}_i^{a_i}\right)^{-1} \mathbf{A}_0 \Delta \mathbf{H}_i^{a_i} \begin{pmatrix} \left(I_J - \beta \mathbf{L}\right)^{-1} \left(I_J - \beta' \mathbf{L}\right)^{-1} \mathbf{R}_{1i} \theta \\ + \mathbf{L} \left(I_J - \beta \mathbf{L}\right)^{-1} \left(\left(\beta + \beta'\right) I_J - \beta \beta' \mathbf{L}\right) \left(I_J - \beta' \mathbf{L}\right)^{-1} \left(\mathbf{R}_{0i} + \boldsymbol{\epsilon}_i\right) \end{pmatrix} \right\|_{\alpha_1, \alpha_2}.$$

Then (β_0, θ_0) *is identifiable.*

Proof: First define $\mathcal{Q}_i^{a_i} : [0,1] \times \Theta_k \to \mathbb{R}^{p+1}$ as follows:

$$\mathcal{Q}_i^{a_i}(\beta, \theta) = \left(\mathbf{A}_0 \mathbf{F}_i^{a_i}\right)^{-1} \mathbf{A}_0 \mathbf{c}_i^{a_i} - \left(\mathbf{A}_0 \mathbf{F}_i^{a_i}\right)^{-1} \mathbf{A}_0 \mathbf{D}_i^{a_i}(\beta) \theta - \left(\mathbf{A}_0 \mathbf{F}_i^{a_i}\right)^{-1} \mathbf{A}_0 \mathbf{E}_i(\beta).$$

By construction, from (19), it is easy to see that (β_0, θ_0) is a fixed-point of $\mathcal{Q}$. Take any $(\beta, \theta), (\beta', \theta') \in \overline{\mathcal{B}} \times \overline{\Theta}$, then

$$\mathcal{Q}_i^{a_i}(\beta, \theta) - \mathcal{Q}_i^{a_i}(\beta', \theta') = -\left(\mathbf{A}_0 \mathbf{F}_i^{a_i}\right)^{-1} \mathbf{A}_0 \left(\mathbf{D}_i^{a_i}(\beta) \theta - \mathbf{D}_i^{a_i}(\beta') \theta' + \mathbf{E}_i(\beta) - \mathbf{E}_i(\beta')\right),$$

where

$$\begin{aligned} \mathbf{D}_i^{a_i}(\beta) \theta - \mathbf{D}_i^{a_i}(\beta') \theta' &= \Delta \mathbf{H}_i^{a_i} \left(\beta \left(I_J - \beta \mathbf{L}\right)^{-1} \mathbf{R}_{i1} \theta - \beta' \left(I_J - \beta' \mathbf{L}\right)^{-1} \mathbf{R}_{i1} \theta'\right) \\ &= \Delta \mathbf{H}_i^{a_i} \begin{pmatrix} \left(\beta - \beta'\right) \left(I_J - \beta \mathbf{L}\right)^{-1} \left(I_J - \beta' \mathbf{L}\right)^{-1} \mathbf{R}_{i1} \theta \\ + \beta' \left(I_J - \beta' \mathbf{L}\right)^{-1} \mathbf{R}_{i1} \left(\theta - \theta'\right) \end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{E}_i(\beta) - \mathbf{E}_i(\beta') &= \Delta \mathbf{H}_i^{a_i} \mathbf{L} \left(\beta^2 \left(I_J - \beta \mathbf{L}\right)^{-1} - \beta'^2 \left(I_J - \beta' \mathbf{L}\right)^{-1}\right) \left(\mathbf{R}_{i0} + \boldsymbol{\epsilon}_i\right) \\ &= \Delta \mathbf{H}_i^{a_i} \mathbf{L} \left(\left(\beta - \beta'\right) \left(I_J - \beta \mathbf{L}\right)^{-1} \left(\left(\beta + \beta'\right) I_J - \beta \beta' \mathbf{L}\right) \left(I_J - \beta' \mathbf{L}\right)^{-1}\right) \left(\mathbf{R}_{i0} + \boldsymbol{\epsilon}_i\right), \end{aligned}$$

which can be shown by making use of the following identities:

$$\begin{aligned} \beta \left(I_J - \beta \mathbf{L}\right)^{-1} - \beta' \left(I_J - \beta' \mathbf{L}\right)^{-1} &= \left(\beta - \beta'\right) \left(I_J - \beta \mathbf{L}\right)^{-1} \left(I_J - \beta' \mathbf{L}\right)^{-1}, \\ \beta^2 \left(I_J - \beta \mathbf{L}\right)^{-1} - \beta'^2 \left(I_J - \beta' \mathbf{L}\right)^{-1} &= \left(\beta - \beta'\right) \left(I_J - \beta \mathbf{L}\right)^{-1} \left(\left(\beta + \beta'\right) I_J - \beta \beta' \mathbf{L}\right) \left(I_J - \beta' \mathbf{L}\right)^{-1}. \end{aligned}$$

It then follows that

$$\begin{aligned} \left|\mathcal{Q}_i^{a_i}(\beta, \theta) - \mathcal{Q}_i^{a_i}(\beta', \theta')\right| &\leq \mathbf{g}_1 \left\|\theta - \theta'\right\|_{\alpha_1} + \mathbf{g}_2 \left|\beta - \beta'\right| \\ &\leq \max\{\mathbf{g}_1, \mathbf{g}_2\} \left\| \begin{pmatrix} \theta \\ \beta \end{pmatrix} - \begin{pmatrix} \theta' \\ \beta' \end{pmatrix} \right\|_{\alpha_2}. \end{aligned}$$

I.e. $\mathcal{Q}_i^{a_i}$ is a contraction, hence it has a unique fixed point. Now suppose (β_0, θ_0) is not identifiable. Then there exists some $(\beta, \theta) \neq (\beta_0, \theta_0)$ that is observationally equivalent to (β_0, θ_0). By an implication of Lemma 6 (β, θ) must also be a fixed point of $\mathcal{Q}_i^{a_i}$, which is a contradiction. Thus (β_0, θ_0) is identifiable.■

Comments on Theorem 4:

(i) Compact Domain. $\mathcal{B}$ cannot include 1 as the expected discounted returns would then be unbounded. Compactness is useful for showing existence of a fixed point. There is also a tradeoff in the choice of $\bar{b}$ and $\bar{k}$ in the definitions of $\overline{\mathcal{B}}$ and $\overline{\Theta}$ respectively. For example, smaller $\bar{b}$ and $\bar{k}$ means smaller $\max\{\mathbf{g}_1, \mathbf{g}_2\}$ but this is a restriction on the parameter space.

(ii) Choice of $\mathbf{A}_0$. The need to select $\mathbf{A}_0$ can be eliminated altogether by removing some rows in (19) so that we have exactly $p+1$ equations. In fact it is not necessary to take equations that only correspond to the states from a particular player i and a_i. Since the parametric structure in (19) is the same for all states we can select any $p+1$ equations from any i and a_i and compute the corresponding matrix norms for $\mathbf{g}_1$ and $\mathbf{g}_2$. This gives us different combinations of equations we can use, and we only need the analog of $\max\{\mathbf{g}_1, \mathbf{g}_2\}$ to be less than 1 for one of them to ensure (β_0, θ_0) is identifiable.

We provide some details for the vectors and matrices for Theorem 4 in the context of the game simulated in the Monte Carlo study in Section 5.1 as an illustration. However, we note that the conditions for contraction in Theorem 4 are not met in this particular case. In what follows we use the shorthand notation that $p_i(w_1, w_2) \equiv \Pr[a_{it} = 1 | w_{1t} = w_1, w_{2t} = w_2]$. Note that, from our definition, in a symmetric equilibrium $p_1(w_1, w_2) \neq p_2(w_1, w_2)$ but instead $p_1(w_1, w_2) = p_2(w_2, w_1)$ when $w_1 \neq w_2$. We need to vectorize these functions.

$$
\begin{aligned}
\pi_1(a_1, a_2, w_1, w_2) &= \pi_{10}(a_1, a_2, w_1, w_2) + \pi_{11}(a_1, a_2, w_1, w_2), \\
\pi_{10}(a_1, a_2, w_1, w_2) &= (1 - a_1) w_1 W_0, \\
\pi_{11}(a_1, a_2, w_1, w_2) &= a_1 \mu_{10} + a_1 a_2 \mu_{20} + a_1 (1 - w_1) F_0, \\
E[\pi_1(a_t, w_t) | w_t = w] &= p_1(w) \mu_{10} + p_1(w) p_2(w) \mu_{20} + p_1(w)(1 - w_1) F_0 + (1 - p_1(w)) \\
E[\pi_{10}(a_{1t}, a_{2t}, w_t) | w_t = w, a_{1t} = a_1] &= (1 - a_1) w_1 W_0, \\
E[\pi_{11}(a_{1t}, a_{2t}, w_t) | w_t = w, a_{1t} = a_1] &= a_1 \mu_{10} + a_1 p_2(w) \mu_{20} + a_1 (1 - w_1) F_0, \\
E[\varepsilon_{1t}(1) a_{1t} + \varepsilon_{1t}(0)(1 - a_{1t}) | w_t = w] &= -p_1(w) \phi(\Delta v(w)),
\end{aligned}
$$

where $\phi(\cdot)$ denotes the pdf of a standard normal variable. Ordering the four states in the following order $\{(0,0),(0,1),(1,0),(1,1)\}$, we have

$$
\begin{aligned}
\Delta \mathbf{v}_i^{a_i}(\beta_0,\theta_0) &= \begin{bmatrix} \Phi^{-1}(p_1(0,0)) \\ \Phi^{-1}(p_1(0,1)) \\ \Phi^{-1}(p_1(1,0)) \\ \Phi^{-1}(p_1(1,1)) \end{bmatrix}, \boldsymbol{\epsilon}_1 = - \begin{bmatrix} p_1(0,0)\,\phi(\Delta v(0,0)) \\ p_1(0,1)\,\phi(\Delta v(0,1)) \\ p_1(1,0)\,\phi(\Delta v(1,0)) \\ p_1(1,1)\,\phi(\Delta v(1,1)) \end{bmatrix}, \\
\mathbf{R}_{10} &= W_0 \times \begin{bmatrix} 0 \\ 0 \\ 1-p_1(1,0) \\ 1-p_1(1,1) \end{bmatrix}, \mathbf{R}_{11} = \begin{bmatrix} p_1(0,0) & p_1(0,0)\,p_2(0,0) & p_1(0,0) \\ p_1(0,1) & p_1(0,1)\,p_2(0,1) & p_1(0,1) \\ p_1(1,0) & p_1(1,0)\,p_2(1,0) & 0 \\ p_1(1,1) & p_1(1,1)\,p_2(1,1) & 0 \end{bmatrix}, \\
\Delta \mathbf{R}_{11}^1 &= \mathbf{R}_{11}^1 - \mathbf{R}_{11}^0, \quad \text{where} \\
\mathbf{R}_{11}^1 &= \begin{bmatrix} 1 & p_2(0,0) & 1 \\ 1 & p_2(0,1) & 1 \\ 1 & p_2(1,0) & 0 \\ 1 & p_2(1,1) & 0 \end{bmatrix}, \mathbf{R}_{11}^0 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.
\end{aligned}
$$

and the transition probability matrices are

$$
\begin{aligned}
\Delta \mathbf{H}_1^1 &= \mathbf{H}_1^1 - \mathbf{H}_1^0, \\
\mathbf{H}_1^1 &= \begin{bmatrix} 0 & 0 & 1-p_2(0,0) & p_2(0,0) \\ 0 & 0 & 1-p_2(0,1) & p_2(0,1) \\ 0 & 0 & 1-p_2(1,0) & p_2(1,0) \\ 0 & 0 & 1-p_2(1,1) & p_2(1,1) \end{bmatrix}, \mathbf{H}_1^0 = \begin{bmatrix} 1-p_2(0,0) & p_2(0,0) & 0 & 0 \\ 1-p_2(0,1) & p_2(0,1) & 0 & 0 \\ 1-p_2(1,0) & p_2(1,0) & 0 & 0 \\ 1-p_2(1,1) & p_2(1,1) & 0 & 0 \end{bmatrix}, \\
\mathbf{L} &= \begin{bmatrix} (1-p_1(0,0))(1-p_2(0,0)) & (1-p_1(0,0))\,p_2(0,0) & p_1(0,0)(1-p_2(0,0)) & p_1(0,0)\,p_2(0,0) \\ (1-p_1(0,1))(1-p_2(0,1)) & (1-p_1(0,1))\,p_2(0,1) & p_1(0,1)(1-p_2(0,1)) & p_1(0,1)\,p_2(0,1) \\ (1-p_1(1,0))(1-p_2(1,0)) & (1-p_1(1,0))\,p_2(1,0) & p_1(1,0)(1-p_2(1,0)) & p_1(1,0)\,p_2(1,0) \\ (1-p_1(1,1))(1-p_2(1,1)) & (1-p_1(1,1))\,p_2(1,1) & p_1(1,1)(1-p_2(1,1)) & p_1(1,1)\,p_2(1,1) \end{bmatrix}.
\end{aligned}
$$

So that $\mathbf{c}_i^{a_i}, \mathbf{D}_i^{a_i}(\beta), \mathbf{E}_i(\beta), \mathbf{F}_i^{a_i}, \mathbf{g}_1$ and $\mathbf{g}_2$ can be readily constructed.

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

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

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

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

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

[8] Blevins, J. (2014): "Nonparametric Identification of Dynamic Decision Processes with Discrete and Continuous Choices," *forthcoming in Quantitative Economics.*

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

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

[11] 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.*

[12] Egesdal, M., Z. Lai and C. Su (2014): "Estimating Dynamic Discrete-Choice Games of Incomplete Information," *forthcoming in Quantitative Economics*

[13] Fan, Y. and M. Xiao (2014): "Competition and Subsidies in the Deregulated U.S. Local Telephone Industry," *forthcoming, Rand Journal 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] Hotz, V., and R.A. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, **60**, 497-531.

[18] Hotz, V., R.A. Miller, S. Sanders and J. Smith (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, **61**, 265-289.

[19] Igami, M. (2015): "Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction in the Hard Disk Drive Industry," *forthcoming, Journal of Political Economy.*

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

[21] Kasahara, H. and K. Shimotsu (2012), "Sequential Estimation of Structural Models with a Fixed Point Constraint," *Econometrica*, **80**, 2303–2319.

[22] Komunjer, I. (2012): "Global Identification in Nonlinear Models with Moment Restrictions," Econometric Theory, **28**, 719-729.

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

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

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

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

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

[28] Matzkin, R.L. (2007): "Nonparametric Identification," *Handbook of Econometrics, Volume 6b.*

[29] Matzkin, R.L. (2012): "Nonparametric Identification in Structural Economic Models," *Annual Economic Reviews, Volume 5.*

[30] Myśliwski, M. (2015): "Dynamic Pricing and Consumer Loyalty in a Multiproduct Oligopoly," *Working Paper, University College London.*

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

[32] Norets, A. and X. Tang (2014): "Semiparametric Inference in Dynamic Binary Choice Models," *Review of Economic Studies*, **81**, 1229-1262.

[33] Otsu, T., M. Pesendorfer and Y. Takahashi (2015): "Pooling Data across Markets in Dynamic Markov Games," *forthcoming in Quantitative Economics.*

[34] 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.

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

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

[37] Pesendorfer, M. and P. Schmidt-Dengler (2010), "Sequential Estimation of Dynamic Discrete Games: a Comment," *Econometrica*, **78**, 833–842.

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

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

[40] Sanches, F., D. Silva Junior and S. Srisuma (2016): "Ordinary Least Sqaures Estimation of a Dynamic Game Model," *International Economic Review*, **57**, 623-634.

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

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

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

[44] Srisuma, S. (2013): “Minimum Distance Estimators for Dynamic Games,” *Quantitative Economics*, **4**, 549-583.

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

[46] 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.*

[47] Tamer, E. (2003): “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria,” *Review of Economic Studies*, **70**, 147-165.