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

Tatiana Komarova
London School of Economics

Fabio Sanches
PUC-Rio

Daniel Silva Junior
City University of London

Sorawoot Srisuma
University of Surrey

August 04, 2017

Abstract

Most empirical and theoretical econometric studies of dynamic discrete choice models assume the discount factor to be known. We show the knowledge of the discount factor is not necessary to identify parts, or all, of the payoff function. We show the discount factor can be generically identified jointly with the payoff parameters. It is known the payoff function cannot nonparametrically identified without any a priori restrictions. Our identification of the discount factor is robust to any normalization choice on the payoff parameters. In IO applications normalizations are usually made on switching costs, such as entry costs and scrap values. We also show that switching costs can be nonparametrically identified, in closed-form, independently of the discount factor and other parts of the payoff function. Our identification strategies are constructive. They lead to easy to compute estimands that are global solutions. We illustrate with a Monte Carlo study and the dataset from Ryan (2012).

JEL Classification Numbers: C14, C25, C51

Keywords: Discount Factor, Dynamic Discrete Choice Problem, 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

The stationary dynamic discrete decision model surveyed in Rust (1994) has been a subject of much research in econometric theory and empirical studies. The primitives of the model consist of the period payoff function, Markov transition law, and discount factor. A well-known characteristic of a dynamic decision model is that it is not identified. For example, Manski (1983) points out in general that the discount factor and payoff function cannot be jointly identified nonparametrically. Most positive identification results in the literature until recently focus on identifying payoff parameters while assuming other primitives to be known; e.g. see Magnac and Thesmar (2002), and also Pesendorfer and Schmidt-Dengler (2008) and Bajari et al. (2009). Meanwhile an empirical work typically parameterizes the payoff function, parameterizes at least part of the distribution of the variables, and assumes the discount factor to be known.

In this paper we are interested in identifying the discount factor jointly with the payoff function under the linear-in-parameter specification. This parametric model is the most commonly used specification in practice. When there are finite states the linear specification can represent any nonparametric function. Most empirical studies assume the value of the discount factor to be known without any formal justification. To the best of our knowledge we are not aware of any prior identification study involving the discount factor in a general parametric model. We provide conditions under which both the discount factor and payoff parameters can be identified, and propose an easy to compute estimator for them. Other positive identification of the discount factor in the literature use a nonparametric approach. They use exclusion restrictions in the form of variables affecting future utilities but not current utilities to identify the discount factor; see Dubé et al. (2014), Wang (2014), Fang and Wang (2015), and Ching and Osborne (2017). We do not rely on these assumptions.

A nonparametric payoff function without any restriction cannot be identified even if the discount factor is known. The fundamental identification characteristic in a discrete choice model can be traced to the static random utility model of McFadden (1974), where utility is ordinal and its level cannot be identified. Some form of normalization has to be made. Aguirregabiria and Suzuki (2014, AS hereafter) recently highlight the undesirable effects that an arbitrary normalization have on un-normalized parameters and counterfactual studies, and emphasize the importance of identifiable objects without any normalization; also see Kaloupsidi et al. (2016a, 2016b). An important question then is whether our identification result is robust to misspecifying the normalization choice.

We verify that our identification of the discount factor is robust against any normalization choice. The payoff parameters are not robust, but some of their meaningful combinations are. To this end we also contribute to the literature by providing a nonparametric framework to identify the payoff parameters that arise from changing in the actions of players between time periods. We call these *switching costs*[1]. For example, in an entry/exit model, they are entry cost and scrap value. Individually the *entry cost* and *scrap value* cannot be separately identified but their difference, the *sunk entry cost*, can be identified. We show that switching costs can be written explicitly in terms of the observed choice probabilities, independently of the discount factor as well as other (non switching costs) components of the payoff function. AS has already shown the sunk entry costs in several IO models can be identified in this fashion. We extend these results to sunk investment costs that can arise from firm investing and divesting, as well as individual switching costs themselves under other a priori restrictions.

A general discussion on the non-identification of the a dynamic model we consider can be found in Rust (1994). Positive identification is possible when more structures are imposed on the primitives. Magnac and Thesmar (2002) have shown the problem of identifying the payoff parameters nonparametrically when all other primitives of the model are assumed to be known can be reduced to a study of solutions to a linear system; also see Pesendorfer and Schmidt-Dengler (2008) and Bajari et al. (2009). We are interested in the payoff parameters as well as the discount factor. The discount factor enters the decision problem recursively and thereby introduces nonlinearity in the model.

Magnac and Thesmar (2002, Section 4.2) suggest that exclusion or parametric restrictions can be used to identify the discount factor. For the former, their Proposition 4 illustrates in a simple two-period model the discount factor is in fact typically overidentified. The identifying restriction is that: for some states utilities in the first period are the same but differ in the second period. This idea has been elaborated, and applied in different empirical contexts, by Dube et al. (2014), Wang (2014), Fang and Wang (2015), and Ching and Osborne (2017) amongst others. On the other hand, while it may be plausible to assume identification is possible in a parametric model we are not aware of any theoretical result that has verified this to be true. In particular establishing parametric identification in a general nonlinear model is a non-trivial task; see Komunjer (2012) for a recent illustration. We prove identification using a *pseudo-model* that is linear in the payoff parameters conditioning on the discount factor. We construct a one-dimensional criterion function to be used for identification. It exploits the conditional linear structure to profile out the payoff parameters and reduce the nonlinear nature of the problem to just one dimension. The criterion function we construct to establish identification has a sample counterpart that can be used for estimation.

In many IO applications, switching costs are often the essence of a dynamic decision problem and can even be the central object of the dynamic model itself (e.g. see Slade (1998), and also the general discussions in Ackerberg et al. (2007) and Pesendorfer (2010)). Our study on the switching costs takes

[1] We use the term switching costs that shares the same spirit as generic adjustment costs and other inertia. Examples of usages in various fields of economics and marketing include the cost to change in health insurance plan, changing of credit and other utility providers, and retailer's decisions on promotions.

a nonparametric approach. We identify combinations of the switching costs by exploiting empirically motivated exclusion and testable independence assumptions. A key step involves eliminating common future expected discounted payoffs that arise from different states. Our result does not depend on the discount factor and some other components of the payoff function. The robust identification of this nature has precedence in the literature but has not been highlighted.[2] For example, an inspection of Proposition 2 in Aguirregabiria and Suzuki (2014) will reveal that the same implication of our Theorem 2 has already been obtained for a binary action game of entry/exit[3]. We provide closed-form expressions for switching costs and their combinations in terms of only the observed choice probabilities. They can therefore be trivially estimated. They also suggest overidentification test can be constructed by comparing against other estimates of switching costs obtained under additional assumptions on the model primitives.

Throughout the paper our identification results are obtained using a *pseudo-model* under the assumption that the choice and transition probabilities are nonparametrically identified. These same probabilities are used to compute expected payoffs in a *pseudo-decision problem* for all values of the model parameters as opposed to the *actual (full-solution) model* where equilibrium probabilities are used. The pseudo-model is used because it is tractable. Indeed a pseudo-model is the basis for any *two-step* estimation procedures, following Hotz and Miller (1993), that are preferred on computational grounds over the full-solution nested fixed-point algorithm of Rust (1987). The estimator we propose in this paper will be based on the two-step approach of Sanches et al. (2016) with computational simplicity in mind. It is worth noting that, although consistent, a simple two-step estimator like ours tend to have larger finite sample bias and is less efficient than estimators that enforce the equilibrium restriction of the model. Equilibrium constraints can be imposed during estimation with additional computational cost, also without the need to solve out a dynamic optimization problem (cf. Rust (1987)). E.g. Aguirregabiria and Mira (2002, 2007) and Egesdal, Lai and Su (2015) have shown the fully efficient maximum likelihood estimator can be obtained in this way.

When the data come from a single time series, or when they are pooled across short panels of multiple homogeneous markets, the choice and transition probabilities are nonparametrically identified under weak conditions. In practice many datasets are short panels, where it would be more reasonable to assume some form of unobserved heterogeneity exists across markets. A flexible yet tractable way to model unobserved heterogeneity in this literature is to use a finite mixture model. For example Aguirregabiria and Mira (2007) suggest economic agents' payoffs have time-invariant

[2] In one instance, for a slightly different model with a mixed continuous-discrete decision variable, Hong and Shum (2010) rely on a deterministic state transition rule to define a *pairwise- difference* estimator that matches on, and thereby avoid computing, future expected discounted payoffs from different states.

[3] We thank an anonymous referee for pointing this out to us.

unobserved market specific component that is unobserved to the econometrician, therefore markets of different types have different equilibrium distributions on the observables. Kasahara and Shimotsu (2009) and Arcidiacono and Miller (2011) have given conditions so that the probabilities for each mixture type can be nonparametrically identified under different frameworks, thereby extending the scope of applying two-step estimation methods to models with unobserved heterogeneity. All identification results in our paper are valid in such setting as long as we can identify the type specific probabilities to be able to set up the corresponding pseudo-decision problem. Specifically the degree of overidentification on the model primitives increases proportionally to the number of mixture types.

The class of decision problems we consider is a special case of dynamic games described in Aguirregabiria and Nevo (2010) and Bajari, Hong and Nekipelov (2010). All of our intuition and results are applicable to these games. The most parts of this paper focus on the single agent model for notational simplicity and clarity of idea, and to abstract ourselves away from game specific issues (such as multiple equilibria). For the same reasoning given for models with unobserved heterogeneity, the portability of our results to dynamic games is immediate as long as the choice and transition probabilities can be consistently estimated nonparametrically. The numerical studies of our proposed estimators are in fact performed in a dynamic game setting. The details on extending our single agent's results to games can be found in the Appendix.

We perform a Monte Carlo study of our proposed estimators using the simulation design in Pesendorfer and Schmidt-Dengler (2008). 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 primitives we estimate are the discount factor, fixed operating cost, and 25 switching cost parameters. 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 estimates on switching costs 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 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 introduces the theoretical model and the basic modeling assumptions. Section 3 gives a joint identification result on the discount factor and the payoff parameters with a linear specification. Section 4 studies nonparametric identification of the switching costs. Section 5 illustrates the performance and use of our estimator with simulated and real data. Section 6 concludes. The Appendix contains details for extending our identification results to dynamic games.

Notations. We use $\rho(\mathbf{A}), CS(\mathbf{A}), \mathbf{A}^\top$ and $\mathbf{A}^\dagger$ to respectively denote the rank, column space, transpose and Moore-Penrose inverse of matrix $\mathbf{A}$. For any positive integers p, q, we let $\mathbf{I}_p$ and $\mathbf{0}_{p\times q}$ respectively denote the identity matrix of size p and a $p \times q$ matrix of zeros.

2 Basic Modelling Framework

We begin by describing an infinite time horizon dynamic discrete choice model as in Rust (1987, 1994).[4] Given our empirical examples and application below, we shall sometimes refer to our representative economic agent as a *firm* and her payoffs as *profits*. Let $t \in \{1, 2, \ldots, \infty\}$ denote time. The random variables in our model are: the action and state variables, denoted by a_t and s_t respectively. a_t takes values from a finite set of alternatives $A = \{0, 1, \ldots, J\}$. $s_t \equiv (x_t, \varepsilon_t) \in X \times \mathbb{R}^{J+1}$, where $X \subseteq \mathbb{R}$. x_t is public information to both the firm and the econometrician, while $\varepsilon_t \equiv (\varepsilon_t(0), \ldots, \varepsilon_t(J)) \in \mathbb{R}^{J+1}$ is private information only observed by the firm. Future states are uncertain. Today's action and states affect outcomes for states in the future. The evolution of the states is summarized by a Markov transition law $P(s_{t+1}|s_t, a_t)$. The firm's period payoff function is $u(a_t, s_t) \in \mathbb{R}$. Future period's payoffs are discounted at the rate $\beta \in [0, 1)$. At time t the firms observes s_t and chooses an action optimally. Specifically, $a_t \equiv \alpha(s_t)$ so that,

$$\begin{aligned} \alpha(s) &= \arg\max_{a\in A}\{u(a,s) + \beta E[V(s_{t+1})|s_t = s, a_t = a]\}, \quad (1) \\ \text{where } V(s) &= \max_{a\in A}\{u(a,s) + \beta E[V(s_{t+1})|s_t = s, a_t = a]\}. \end{aligned}$$

Using the optimal decision rule we can remove the max operator and write the value function as,

$$V(s) = E\left[\sum_{t=0}^{\infty} \beta^t u(a_t, s_t)|s_0 = s\right]. \quad (2)$$

The expectation operators in the displays above integrate out variables with respect to the probability distribution induced by the equilibrium choice probabilities and Markov transition law. As standard in the literature we assume the following assumptions.

Assumption M:
(i) (Additive Separability) For all a, x, ε:

$$u(a, x, \varepsilon) = \pi(a, x) + \varepsilon(a).$$

[4]The infinite time feature simplifies the notation. Our identification strategy is valid for finite time horizon models with/out absorbing state; e.g. Hotz and Miller (1993).

(ii) (Conditional Independence) The transition distribution of the states has the following factorization for all $x', \varepsilon', x, \varepsilon, a$:

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

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. Furthermore, ε_t has finite first moments, and a positive, continuous and bounded density on $\mathbb{R}^{J+1}$.

(iii) (Finite Observed State) $X = \{1, \ldots, K\}$.

The primitives of the model under this setting consist of (π, β, Q, G). Throughout the paper we shall assume (G, Q) to be known. G can be identified from the data when (a_t, x_t, x_{t+1}) are observed. Consistent estimation of the joint distribution of (a_t, x_t, x_{t+1}) holds under weak conditions with a single time series, as well as repeated observations from short panels when there is no other unobserved heterogeneity. Q is typically assumed known in most empirical applications. Conditions for the identification of Q exist using a large support type argument, e.g. see Aguirregabiria and Suzuki (2014, Proposition 1) and Chen (2014, Theorem 4). On the other hand our results do not depend on any continuity assumption to achieve identification as we take x_t to be a discrete random variable.

Our subsequent analysis will use as the starting point the fact that we can identify the choice probability from data, which in turn is informative about (π, β). More specifically, for any $a > 0$, let $\Delta v\left(a, x\right) \equiv v\left(a, x\right) - v\left(0, x\right)$, where $v\left(a, x\right)$ denotes the choice-specific value function that serves as the mean utility in a discrete choice modelling:

$$\begin{aligned} v\left(a, x\right) &= \pi\left(a, x\right) + \beta E\left[V\left(s_{t+1}\right) | x_t = x, a_t = a\right], \qquad (3) \\ \Pr\left[a_t = a | x_t = x\right] &= \Pr\left[\Delta v\left(a, x\right) - \Delta v\left(a', x\right) > \varepsilon_t\left(a'\right) - \varepsilon_t\left(a\right) \quad \text{for all} \quad a' \neq a\right]. \end{aligned}$$

By inverting the choice probabilities (Hotz and Miller (1993)) we can recover $\Delta v\left(a, x\right)$ for all $a > 0, x$.

3 Identifying the Discount Factor with Linear-in-Parameter Payoffs

In this section we assume the payoff function takes on a linear-in-parameter specification. Section 3.1 defines the identification concept for the discount factor and payoff parameters. Section 3.2 provides some representation lemmas that will be useful for defining a criterion function to study identification. Section 3.3 gives the identification result.

3.1 Definition of Parametric Identification

We will assume Assumption M and the following assumption throughout this section.

Assumption P (Linear-in-Parameter): *For all a, x:*

$$\pi(a, x; \theta) = \pi_0(a, x) + \theta^\top \pi_1(a, x),$$

where π_0 is a known real value function, π_1 is a known $p-$dimensional vector value function and θ belongs to $\mathbb{R}^p$.

Assumption P can be interpreted as nonparametric. For example it can represent an unrestricted nonparametric function of π by by assigning a parameter for each pair of a and x. However, such function is too rich and cannot be identified. We will maintain the parametric appearance for π as we will be exploiting any nonparametric restrictions in our study on identifying the discount factor.

The role of π_0 is to represent the payoff components that are identifiable without the knowledge of the discount factor. In practice π_0 and possibly parts of π_1 may have to be estimated (e.g. see Section 5.2). For the purpose of identification they can be treated as known. The primitives in this setting are (β, θ). They belong to $\mathcal{B} \times \Theta$ where $\mathcal{B} = [0, 1)$ and $\Theta = \mathbb{R}^p$. We are interested in the data generating discount factor and payoff parameters, which we denote by β_0 and θ_0 respectively.

We begin by defining the parametric choice-specific value function (cf. equation (3)):

$$v(a, x; \beta, \theta) \equiv \sum_{t=0}^{\infty} \beta^t E\left[\pi(a_t, x_t; \theta) + \varepsilon_t(a_t) | a_0 = a, x_0 = x\right]. \tag{4}$$

Then we define $\Delta v(a, x; \beta, \theta) \equiv v(a, x; \beta, \theta) - v(0, x; \beta, \theta)$. It is important to recall the sequence $\{a_t, x_t, \varepsilon_t\}_{t=0}^{\infty}$ in equation (4) follows an optimal controlled process consistent with (β_0, θ_0). Therefore, using Hotz-Miller's inversion, it follows that $\Delta v(a, x; \beta_0, \theta_0)$ is identified from the observed choice probabilities for all a, x.

We take each pair (β, θ) to be a *structure* of the empirical model and its implied choice-specific values, denoted by $\mathcal{V}_{\beta,\theta} \equiv \{\Delta v(a, x; \beta, \theta)\}_{a,x \in A \times X}$, to be its corresponding *reduced form.* We 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 $\mathcal{B} \times \Theta$.

For our identification study we define our statistical model to be $\{\mathcal{V}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$.[5] It is appropriate to call $\{\mathcal{V}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$ a *pseudo-model* in the sense that $\mathcal{V}_{\beta,\theta}$ characterizes the optimal decision rule when the random variables in equation (4) follow an optimal controlled process consistent with (β_0, θ_0) rather than (β, θ). I.e. the reduced form of a pseudo-model comes from an economic agent solving a *pseudo-decision problem.* All statements made on identification in Section 3 are within the context of the model described by $\{\mathcal{V}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$.

Note that we can also define a statistical model based on probability distributions as in the traditional studies of identification. Specifically, by letting

$$\mathcal{P}_{\beta,\theta} \equiv \{\Pr[\Delta v(a, x; \beta, \theta) - \Delta v(a', x; \beta, \theta) > \varepsilon(a') - \varepsilon(a) \text{ for all } a' \neq a]\}_{a,x\in A\times X}.$$

It is known there is a one-to-one relation between $\{\mathcal{V}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$ and $\{\mathcal{P}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$; see Matzkin (1991), Hotz and Miller (1993), and Norets and Takahashi (2013). Therefore identification for our decision problem can be equivalently established with either $\{\mathcal{V}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$ or $\{\mathcal{P}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$.

3.2 Some Representation Lemmas

The advantage of a pseudo-model is that $\mathcal{V}_{\beta,\theta}$ is mathematically tractable as a function of (β, θ). Under Assumptions M and P, it shall be useful to separate out the contributions of the expected discounted payoffs in (4) as follows:

$$\begin{aligned} v(a, x; \beta, \theta) &= \pi_0(a, x) + \beta \sum_{t=0}^{\infty} \beta^t E[\pi_0(a_t, x_t) | a_0 = a, x_0 = x] \\ &+ \beta \sum_{t=0}^{\infty} \beta^t E[\varepsilon_t(a_t) | a_0 = a, x_0 = x] \\ &+ \theta^\top (\pi_1(a, x) + \beta \sum_{t=0}^{\infty} \beta^t E[\pi_1(a_t, x_t) | a_0 = a, x_0 = x]). \end{aligned}$$

[5] The traditional literature on identification in econometrics uses probability distributions to describe a statistical model. Let

$$\mathcal{P}_{\beta,\theta} \equiv \{\Pr[\Delta v(a, x; \beta, \theta) - \Delta v(a', x; \beta, \theta) > \varepsilon(a') - \varepsilon(a) \text{ for all } a' \neq a]\}_{a,x\in A\times X}.$$

There is in fact a one-to-one relation between $\{\mathcal{V}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$ and $\{\mathcal{P}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$; see Matzkin (1991), Hotz and Miller (1993), and Norets and Takahashi (2013). Therefore identification for our decision problem can be equivalently established with either $\{\mathcal{V}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$ or $\{\mathcal{P}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$.

Subsequently, by defining $\Delta\pi_l(a,x) \equiv \pi_l(a,x) - \pi_l(0,x)$ for $l = 0, 1$, we have:

$$\begin{aligned}\Delta v(a,x;\beta,\theta) &= \Delta\pi_0(a,x) + \beta\sum_{t=0}^{\infty}\beta^t\left(E\left[\pi_0(a_t,x_t)|a_0 = a, x_0 = x\right] - E\left[\pi_0(a_t,x_t)|a_0 = 0, x_0 = x\right]\right) \\ &\quad +\beta\sum_{t=0}^{\infty}\beta^t\left(E\left[\varepsilon_t(a_t)|a_0 = a, x_0 = x\right] - E\left[\varepsilon_t(a_t)|a_0 = 0, x_0 = x\right]\right) \\ &\quad +\theta^\top(\Delta\pi_1(a,x) + \beta\sum_{t=0}^{\infty}\beta^t\left(E\left[\pi_1(a_t,x_t)|a_0 = a, x_0 = x\right] - E\left[\pi_1(a_t,x_t)|a_0 = 0, x_0 = x\right]\right)).\end{aligned}$$

The decomposition of Δv helps us distinguish how β and/or θ affect different parts of the per-period payoffs. Lemma 1 summarizes this in a matrix form.

LEMMA 1: *Under Assumptions M and P, for all $a > 0$, $\Delta v(a,x;\beta,\theta)$ can be collected in the following vector form for all $(\beta,\theta) \in \mathcal{B} \times \Theta$:*

$$\begin{aligned}\Delta\mathbf{v}^a(\beta,\theta) &= \Delta\mathbf{R}_0^a + \beta\Delta\mathbf{H}^a(\mathbf{I}_K - \beta\mathbf{L})^{-1}\mathbf{R}_0 \\ &\quad +\beta\Delta\mathbf{H}^a(\mathbf{I}_K - \beta\mathbf{L})^{-1}\boldsymbol{\epsilon} \\ &\quad +\left(\Delta\mathbf{R}_1^a + \beta\Delta\mathbf{H}^a(\mathbf{I}_K - \beta\mathbf{L})^{-1}\mathbf{R}_1\right)\theta,\end{aligned} \quad (5)$$

where the elements in the above display are collected and explained in Tables A and B.

Matrix	Dimension	Representing
$\Delta\mathbf{R}_1^a$	$K \times p$	$\Delta\pi_1(a,\cdot)$
$\mathbf{R}_1$	$K \times p$	$\pi_1(a,\cdot)$
$\mathbf{L}$	$K \times K$	$E[\psi(x_{t+1})\|x_t = \cdot]$
$\mathbf{H}^a$	$K \times K$	$E[\psi(x_{t+1})\|x_t = \cdot, a_t = a]$
$\Delta\mathbf{H}^a$	$K \times K$	$E[\psi(x_{t+1})\|x_t = \cdot, a_t = a] - E[\psi(x_{t+1})\|x_t = \cdot, a_t = 0]$

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
$\boldsymbol{\epsilon}$	$E[\varepsilon_t(a_t)\|x_t = \cdot]$
$\Delta\mathbf{R}_0^a$	$\Delta\pi_0(a,\cdot)$
$\mathbf{R}_0$	$E[\pi_0(a_t,x_t)\|x_t = \cdot]$
$\Delta\mathbf{H}^a(\mathbf{I}_K - \beta\mathbf{L})^{-1}\mathbf{R}_0$	$\sum_{t=0}^{\infty}\beta^t(E[\pi_0(a_t,x_t)\|a_0 = a, x_0 = x] - E[\pi_0(a_t,x_t)\|a_0 = 0, x_0 = x])$
$\Delta\mathbf{H}^a(\mathbf{I}_K - \beta\mathbf{L})^{-1}\mathbf{R}_1$	$\sum_{t=0}^{\infty}\beta^t(E[\pi_1(a_t,x_t)\|a_0 = a, x_0 = x] - E[\pi_1(a_t,x_t)\|a_0 = 0, x_0 = x])$
$\Delta\mathbf{H}^a(\mathbf{I}_K - \beta\mathbf{L})^{-1}\boldsymbol{\epsilon}$	$\sum_{t=0}^{\infty}\beta^t(E[\varepsilon_t(a_t)\|a_0 = a, x_0 = x] - E[\varepsilon_t(a_t)\|a_0 = 0, x_0 = x])$

Table B. The $K \times 1$ vectors represent (differences in) expected payoffs.

PROOF: This is a special case of Lemma R in Sanches et al. (2016).■

All vectors and matrices in Tables A and B are either known or estimable from the choice and transitional probabilities. The tables will serve as a useful reference for constructing the necessary components we use for defining the criterion function in Section 3.3.

Given that we can identify $\Delta\mathbf{v}^a(\beta_0,\theta_0)$ for all $a>0$, to identify (β_0,θ_0), it is sufficient to show that for all $(\beta,\theta)\neq(\beta_0,\theta_0)$, $\Delta\mathbf{v}^a(\beta,\theta)\neq\Delta\mathbf{v}^a(\beta_0,\theta_0)$ for some a. Our next lemma provides a characterization as to how changing β and θ can affect $\Delta\mathbf{v}^a$.

LEMMA 2: *Under Assumptions M and P, for any $a>0$ and $(\beta,\theta),(\beta',\theta')\in\mathcal{B}\times\Theta$:*

$$\Delta\mathbf{v}^a(\beta,\theta)-\Delta\mathbf{v}^a(\beta,\theta') = \left(\Delta\mathbf{R}_1^a+\beta\Delta\mathbf{H}^a(\mathbf{I}_K-\beta\mathbf{L})^{-1}\mathbf{R}_1\right)(\theta-\theta'), \tag{6}$$

$$\Delta\mathbf{v}^a(\beta',\theta')-\Delta\mathbf{v}^a(\beta,\theta') = (\beta-\beta')\Delta\mathbf{H}^a(\mathbf{I}_K-\beta'\mathbf{L})^{-1}(\mathbf{I}_K-\beta\mathbf{L})^{-1}(\mathbf{R}_0+\mathbf{R}_1\theta'+\boldsymbol{\epsilon}). \tag{7}$$

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

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

PROOF: Follows from some algebra based on equation (5).■

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 (6), a sufficient condition for $\Delta\mathbf{v}^a(\beta_0,\theta)\neq\Delta\mathbf{v}^a(\beta_0,\theta')$ when $\theta\neq\theta'$ is that $\Delta\mathbf{R}_1^a+\beta\Delta\mathbf{H}^a(\mathbf{I}_K-\beta\mathbf{L})^{-1}\mathbf{R}_1$ has full column rank for some $a>0$.

(ii) *If the payoff function is assumed to be known*, from (7), a sufficient condition for $\Delta\mathbf{v}^a(\beta',\theta_0)\neq\Delta\mathbf{v}^a(\beta,\theta_0)$ when $\beta\neq\beta'$ is that $(\mathbf{R}_0+\mathbf{R}_1\theta'+\boldsymbol{\epsilon})\neq 0$ and $\Delta\mathbf{H}^a$ is invertible some $a>0$.

(iii) Suppose p is large relative to K. Then for any $a>0$ such that $\Delta\mathbf{R}_1^a+\beta\Delta\mathbf{H}^a(\mathbf{I}_K-\beta\mathbf{L})^{-1}\mathbf{R}_1$ has rank K, and for any $\theta',\beta\neq\beta'$ that $\Delta\mathbf{v}^a(\beta',\theta')\neq\Delta\mathbf{v}^a(\beta,\theta')$, by equating (6) and (7), we can always find θ such that $\Delta\mathbf{v}^a(\beta',\theta')=\Delta\mathbf{v}^a(\beta,\theta)$.

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 $\Delta\mathbf{R}_1^a+\beta\Delta\mathbf{H}^a(\mathbf{I}_K-\beta\mathbf{L})^{-1}\mathbf{R}_1$ over a. In the case we are able to identify the payoff function outside of the dynamic model, (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.

3.3 Sum of Squares Criterion Function

The study of identification involving the discount factor is complicated due to the fact that $\mathcal{V}_{\beta,\theta}$ is nonlinear in (β, θ). However, for a given β, we can see from (5) that $\mathcal{V}_{\beta,\theta}$ is linear in θ. We use profiling to exploit the conditional linearity to simplify the identification problem for a nonlinear model with $p+1$ parameters to a one-dimensional problem.

Let $\mathbf{m}^a(\beta,\theta) \equiv \Delta\mathbf{v}^a(\beta_0,\theta_0) - \Delta\mathbf{v}^a(\beta,\theta)$. Then we can write, using (5):

$$\begin{aligned}
\mathbf{m}^a(\beta,\theta) &= \mathbf{a}^a(\beta) - \mathbf{B}^a(\beta)\,\theta, \\
\mathbf{a}^a(\beta) &\equiv \Delta\mathbf{v}^a(\beta_0,\theta_0) - \Delta\mathbf{R}_0^a - \beta\Delta\mathbf{H}^a(\mathbf{I}_K - \beta\mathbf{L})^{-1}(\mathbf{R}_0 + \boldsymbol{\epsilon}), \\
\mathbf{B}^a(\beta) &\equiv \Delta\mathbf{R}_1^a + \beta\Delta\mathbf{H}^a(\mathbf{I}_K - \beta\mathbf{L})^{-1}\mathbf{R}_1.
\end{aligned}$$

It is clear that $\mathbf{m}^a(\beta,\theta)$ is linear in θ for any given β. We can stack together the system of equations above across a. In doing so we obtain the following vector value function, $\mathbf{m} : \mathcal{B} \times \Theta \to \mathbb{R}^{KJ}$:

$$\mathbf{m}(\beta,\theta) = \mathbf{a}(\beta) - \mathbf{B}(\beta)\,\theta, \tag{8}$$

where $\mathbf{a}(\beta)$ is a $KJ \times 1$ vector and $\mathbf{B}(\beta)$ is a $KJ \times p$ matrix.

Let $\mathcal{M}(\beta,\theta) \equiv \|\mathbf{m}(\beta,\theta)\|$, i.e. $\mathcal{M}(\beta,\theta)$ is 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. Therefore $\mathcal{M}$ has the necessary property to serve as a criterion for identification.

Next we profile out θ in order to reduce the dimensionality on $\mathcal{M}$ by exploiting its least squares structure. For each β, run a regression of $\mathbf{a}(\beta)$ on $\mathbf{B}(\beta)$, we can define:

$$\theta^*(\beta) \equiv (\mathbf{B}(\beta)^\top \mathbf{B}(\beta))^\dagger \mathbf{B}(\beta)^\top \mathbf{a}(\beta).$$

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

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

By construction it also holds that

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

In this way we have reduced the parameter space in the identification problem to a one-dimensional one. Furthermore the domain of the parameter space is on a small interval: $[0,1)$. 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: *Under Assumptions M and P,* (β_0,θ_0) *is identifiable in* $\{\mathcal{V}_{\beta,\theta}\}_{\beta,\theta\in\mathcal{B}\times\Theta}$ *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 $\Delta\mathbf{v}^a(\beta_0,\theta_0) = \Delta\mathbf{v}^a(\beta',\theta^*(\beta'))$ for all a by the property of the norm. Since Θ is a closed set, 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 are observed or can be consistently estimated nonparametrically. In the Appendix we give a more detailed conditions for $\mathcal{M}^*$ to have a unique minimum.

(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. We can use the sample parts of components in Tables A and B to consistently estimate $\mathcal{M}^*(\beta)$ for all β. So one can 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 suggests one natural way to estimate the discount factor, namely by grid search. In practice 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.

By inspecting of the proof of Theorem 1 it is clear there are some separation between the identifiability of β_0 and θ_0. In particular we have defined $\theta^*(\beta)$ using a generalized inverse of the matrix $\mathbf{B}(\beta)^\top \mathbf{B}(\beta)$. Therefore β_0 can be identified even if θ_0 is not.

The full rank condition on $\mathbf{B}(\beta_0)$ is not an innocuous assumption when we view Assumption P as a representation of a nonparametric function. In practice this is often delivered by an exclusion assumption or more generally *normalization* of certain payoff parameters. Next section we will focus on payoff parameters that we call *switching costs*. We will revisit the question of identifiability of the discount factor under different normalization choice in Section 4.3.

4 Nonparametric Identification of Switching Costs

In this section we consider payoff functions under nonparametric restrictions that allow us to obtain closed-form expressions for the switching costs parameters. In Section 4.1 we define a switching cost function and explain the assumptions required for our identification result. Section 4.2 gives the identification result. Section 4.3 relates the identification of the discount factor under Assumption P to models with switching costs.

4.1 Switching Costs

The payoff function cannot be nonparametrically identified without any restrictions. Economic theory can help guide how to impose structures on the payoff function. A main consideration in making a dynamic discrete decision is how a change in one's action from the previous period immediately affect today's payoffs. Actions from the past are therefore often important components of the state variables.

In order to highlight the role of *switching costs* we distinguish past actions from other state variables. At time t we denote actions from the previous period by w_t, so that $w_t \equiv a_{t-1}$. We denote the switching cost from changing action from w to a by $SC^{w\to a}$. Subsequently, in this section we shall maintain an updated version of Assumption M where x_t is replaced with (w_t, x_t) everywhere. In addition we impose the following assumptions.

Assumption N

(i) (Decomposition of Profits): For all a, w, x:

$$\pi(a,w,x) = \mu(a,x) + \phi(a-w,w,x),$$

such that $\phi(0,w,x) = 0$.

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

The decomposition of π in N(i) may appear peculiar but it is typical in many empirical IO applications. We will give an interpretation of its components in the context of an IO model. The defining feature of μ is that it excludes past actions. μ can represent the firm's operational profit in the current period, such as variable profits and operational costs, which does not depend on actions from the past. ϕ is the *switching cost function* that takes non-zero values only when a change of action occurs. Note that, by construction, we have

$$\phi(a-w,w,x) = SC^{w\to a}(x) \cdot \mathbf{1}[w \neq a], \tag{9}$$

where $\mathbf{1}[\cdot]$ denotes the indicator function.

Assumption N(ii) 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 N(ii) is not implied by M(ii). In many applications $\{x_t\}$ is simply 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 N(ii). In any case, unlike M(ii), N(ii) is a restriction made on the observables so it can be tested directly from the data. Later we shall show how x_t can be modified to contain past actions so N(ii) can be weakened to allow for dependence of other state variables with past actions.

Even under Assumption N(i) identification issue persists (e.g. see the discussion in Aguirregabiria and Suzuki (2014)). $SC^{w\to a}$ cannot be identified for all $w \neq a$ without any further restrictions. Some of their differences can be identified. For example identification is possible if we *normalize* some baseline switching costs to be known. We will look at different restrictions that can be used to identify individual or combination of the switching costs. Before giving the formal result we provide an intuition as to why Assumption N is helpful for identifying the switching costs, and illustrate the key steps of our identification strategy.

Exclusion and Independence Restrictions

Consider a two-period entry/exit decision problem. Let $A = \{0,1\}$, where 0 denotes exit and 1 denotes entry. Then $SC^{0\to 1}$ and $SC^{1\to 0}$ respectively have interpretations of entry cost and scrap value. In this case we can write

$$\phi(a-w,w,x) = SC^{0\to 1}(x) \cdot a(1-w) + SC^{1\to 0}(x) \cdot (1-a)w. \tag{10}$$

The choice-specific value function (cf. (3)) in this model is:

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

Let $\Delta\nu(w,x) \equiv \nu(1,w,x) - \nu(0,w,x)$. At time t, a firm will enter if and only if $\Delta\nu(w,x) > \varepsilon_t(0) - \varepsilon_t(1)$. We can identify $\Delta\nu$ from the observed choice probabilities.

The role of our assumptions is to isolate today's switching costs from the remaining components in the choice-specific value function. Specifically, we apply N(i) to decompose the profit function in the current period and use N(ii) to simplify the expected future profits. We can then re-write the equation above as

$$\begin{aligned} \nu(a,w,x) &= \lambda(a,x) + \phi(a-w,w,x), \text{ where} \\ \lambda(a,x) &= \mu(a,x) + \beta E\left[\pi(a_{t+1},a,x_{t+1}) \,|a_t = a, x_t = x\right]. \end{aligned}$$

Crucially note that the conditional expectation on future profits in λ no longer depends on w_t under N(ii) by the law of iterated expectation. We treat λ as a nuisance parameter; it is a nonparametric object that depends on all primitives in the model. Let $\Delta\lambda(x) \equiv \lambda(1,x) - \lambda(0,x)$. Using equation (10) we have,

$$\Delta\nu(w,x) = \Delta\lambda(x) + SC^{0\to 1}(x)\cdot(1-w) - SC^{1\to 0}(x)\cdot w. \tag{11}$$

It is now clear we can identify a combination of the switching costs by differencing out $\Delta\lambda$ in the equation above:

$$\Delta\nu(1,x) - \Delta\nu(0,x) = -SC^{0\to 1}(x) - SC^{1\to 0}(x). \tag{12}$$

In an entry/exit game the quantity $-SC^{0\to 1} - SC^{1\to 0}$ represents the *sunk entry cost* that a firm cannot recover back once it decides to leave the market after entering. Equation (12) shows the sunk entry cost can be identified independently of β and μ. On the other hand it is well known that entry cost and scrap value cannot be nonparametrically identified separately in this particular model. In an empirical work an unidentified object gets normalized. It is clear from equation (12) that either the entry cost or scrap value can be identified if other parameter is assumed to be known. For example, a common assumption is to normalize the scrap value to be zero, the entry cost can be estimated conditionally on this value along with the other parameters.

The identification strategy above can be generalized substantially. Results for a more general decision model under M and N can be obtained with little modification. The extension to dynamic games is more complex. It requires more notations and the notion of a difference generalizes to projection of a matrix. We defer these details to the Appendix.

4.2 Closed-Form Identification

We start by providing an expression for the differences in choice-specific valuations that generalizes equation (11). For any $a > 0$, let $\Delta v(a,w,x) \equiv v(a,w,x) - v(0,w,x)$, $\Delta\lambda(a,x) \equiv \lambda(a,x) - \lambda(0,x)$, and $\Delta\phi(a,w,x) \equiv \phi(a-w,w,x) - \phi(-w,w,x)$. Lemma 3 generalizes equation (11).

Lemma 3: *Under Assumptions M and N, we have for all* $i, a > 0$ *and* w, x*:*

$$\Delta v(a, w, x) = \Delta\lambda(a, x) + \Delta\phi(a, w, x), \tag{13}$$

where

$$\begin{aligned} \Delta\lambda(a, x) &\equiv \mu(a, x) - \mu(0, x) + \beta(\widetilde{m}(a, x) - \widetilde{m}(0, x)), \\ \widetilde{m}(a, x) &\equiv E[m(a, x_{t+1})|a_t = a, x_t = x], \\ m(w, x) &\equiv E[V(s_t)|w_t = w, x_t = x]. \end{aligned}$$

Proof: Using the law of iterated expectation, the value function as defined in equation (2), satisfies $E[V(s_{t+1})|a_t, w_t, x_t] = E[m(w_{t+1}, x_{t+1})|a_t, w_t, x_t]$ under M(ii). $E[m(w_{t+1}, x_{t+1})|a_t, w_t, x_t]$ can be simplified further to $E[\widetilde{m}(a_t, x_t)|a_t, x_t]$ after another application of the law of iterated expectation and imposing N(ii). The remainder of the proof then follows from the definitions of the terms defined in the text.■

Components of Δv are $\Delta\lambda$ and $\Delta\phi$. We treat $\Delta\lambda$ as a nuisance parameter. $\Delta\phi$ contains the switching costs of interest, for any a, w, x:

$$\Delta\phi(a, w, x) = SC^{w\to a}(x) \cdot \mathbf{1}[w \neq a] - SC^{w\to 0}(x) \cdot \mathbf{1}[w \neq 0]. \tag{14}$$

As seen previously we can identify differences in $\Delta\phi$ by eliminating $\Delta\lambda$. This can be done by looking at differences of $\Delta v(a, w, x)$ across different w while holding (a, x) fixed.

Theorem 2: *Under Assumptions M and N, we have for all* $a > 0$ *and* x, w, w'*:*

$$\Delta\phi(a, w, x) - \Delta\phi(a, w', x) = \Delta v(a, w, x) - \Delta v(a, w', x). \tag{15}$$

Theorem 2 follows immediately from Lemma 3. Equation (15) tells us that we can always identify *some* combinations of the switching costs nonparametrically. Importantly the identified objects do not depend on β or μ.

Comments on Theorem 2.

(i) Certain differences in $\Delta\phi$ in equation (15) are economically meaningful. For example we have already introduced the sunk entry cost in the entry/exit model. The notion of sunk costs naturally generalizes when there is a varying degree of commitment. More specifically consider an investment or capacity game where it costs a firm to choose $a_t > a_{t-1}$, and conversely a firm can divest to recover some of these costs by choosing $a_t < a_{t-1}$. In this case $-SC^{a'\to a} - SC^{a\to a'}$ with $a > a'$ represents a sunk investment cost for a firm that increases its investment level from a' to a, and divest back to a'. Using equations (14) and (15), Corollaries 1 and 2 give closed-form expressions for identifying the sunk investment costs.

Corollary 1. For all $a > 0, x$:

$$-SC^{0\to a}(x) - SC^{a\to 0}(x) = \Delta v(a, a, x) - \Delta v(a, 0, x).$$

Corollary 2. For all $a, a' > 0, x$:

$$-SC^{a'\to a}(x) - SC^{a\to a'}(x) = \Delta v(a, a, x) + \Delta v(a', a', x) - \Delta v(a, a', x) - \Delta v(a', a, x).$$

(ii) We would prefer to identify the switching costs individually. However, without further information, they are not identified nonparametrically for this type of models; for example see Aguirregabiria and Suzuki (2014) for a thorough discussion. But identification can be achieved if we are willing to impose some constraints on the switching costs. One example is by assuming symmetry of switching costs between any two actions, which would be reasonable in applications with logistical or physical adjustment costs such as the traditional menu costs (e.g. see Slade (1998)). Corollary 3 shows that individual switching costs under symmetry are identified. It follows immediately from Corollaries 1 and 2.

Corollary 3. For all a, a', x, suppose that $SC^{a'\to a}(x) = SC^{a\to a'}(x)$, then for any $a, a' > 0$:

$$\begin{aligned} SC^{0\to a}(x) &= -(\Delta v(a, a, x) - \Delta v(a, 0, x))/2, \\ SC^{a\to a'}(x) &= -(\Delta v(a, a, x) + \Delta v(a', a', x) - \Delta v(a, a', x) - \Delta v(a', a, x))/2. \end{aligned}$$

(iii) In many applications components of the switching costs are taken to be known. Typically this is done by way of a normalization assumption. The most commonly used assumption is taking action zero yields zero payoff. For example, for an entry or investment game with entry, such assumption means a firm has no recovery value of assets upon leaving the market. In others, some institutional knowledge outside of the dynamic model are used. For example Kalouptsidi (2014) uses data on resale value of second hand ships to estimate scrap values and entry costs directly. Another example, in a study of promotion pricing decisions, is Myśliwski et al. (2017) who rely on anecdotal evidence to assume a cost is incurred to producers when a promotion is on while there is no costs for switching back to the regular price. In these cases we can identify individual switching costs directly as Corollary 4 shows.

Corollary 4. For all a', suppose $SC^{a'\to 0}(x) = \phi_0(w, x)$ then for any a, a', x:

$$SC^{a'\to a}(x) = \Delta v(a, a', x) - \Delta v(a, a, x) + \phi_0(a', x) - \phi_0(a, x). \tag{16}$$

It is important to highlight that assigning incorrect values to ϕ_0 generally leads to incorrect values of $SC^{w \to a}$. However, it is easy to verify that certain combinations of switching costs, including those in Corollaries 1 and 2, are robust against any choice of ϕ_0.

(iv) Generally Corollaries 1 and 2 can be informative on the validity of a particular normalization choice since they themselves are derived without normalization. For example, let us go back to the discussion on investment game at the end of our first comment where there is a divestment opportunity. In this context it would be natural to assume that $-SC^{a' \to a} - SC^{a \to a'} = c_0 > 0$ when $a > a'$. Then, given both $-SC^{a' \to a}$ and $SC^{a \to a'}$ are positive, it must be the case that $-SC^{a' \to a}$ is bounded below by c_0.

(v) When $A = \{0, 1\}$ our Theorem 1 implies the sunk entry cost can be identified without any normalization. Proposition 2 in Aguirregabiria and Suzuki (2014) has established the same result using a different argument.

The results of Theorem 2 and Corollaries 1 to 4 are constructive. We can replace the unknown Δv using the empirical choice probabilities. The sample analog estimators can be computed without any optimization. Given the empirical literature is concerned with the computational cost our closed-form identification result can substantially reduce the number of parameters to be estimated in a model. Such estimators will be consistent and asymptotically normal as long as the initial choice probabilities have these properties.

4.3 Identification and Normalization

We have emphasized that normalizations of switching costs are necessary in many situations. The validity of the identification of payoff parameters can depend directly on the normalization choice. Now we study the effect of normalizations on the identification of the discount factor.

In the empirical IO literature the discount factor is customarily assumed to be known while the focus on identification falls on which payoff parameters can (or cannot) be identified. A particular normalization choice is made by assigning a value to an unknown parameter as previously explained. The normalization assumption is made independent to the choice of the discount factor. The non-identifiability of the payoff parameters considered in practice therefore mathematically translates to the matrix $\mathbf{B}(\beta)$ in equation (8) being rank deficient for all β. For our result, we only need to consider rank deficiency at β_0.

Recall that $\mathbf{B}(\beta_0)$ is a $KJ \times p$ matrix. Suppose $\rho(\mathbf{B}(\beta_0)) = r < p$. I.e. the rank condition in Theorem 1 fails. Then without any loss of generality we can write:

$$\mathbf{B}(\beta_0) = [\mathbf{B}_1 : \mathbf{B}_2],$$

where $\mathbf{B}_1$ is a matrix consisting of the first r columns of $\mathbf{B}(\beta_0)$ such that $CS(\mathbf{B}_1) = CS(\mathbf{B}(\beta_0))$, and $\mathbf{B}_2$ is a matrix containing the last $(p-r)$ columns of $\mathbf{B}(\beta_0)$. Since $CS(\mathbf{B}_2) \subset CS(\mathbf{B}_1)$ there exists an $r \times (p-r)$ matrix $\mathbf{\Gamma}$ such that $\mathbf{B}_2 = \mathbf{B}_1\mathbf{\Gamma}$.

Now recall equation (8), where we define $\mathbf{m}(\beta,\theta) = \mathbf{a}(\beta) - \mathbf{B}(\beta)\theta$ to study identification. By construction $\mathbf{m}(\beta_0,\theta_0) = 0$, which implies $\mathbf{a}(\beta_0) = \mathbf{B}(\beta_0)\theta_0$. Let θ_{01} and θ_{02} respectively denote column vectors containing the first r elements and the last $(p-r)$ elements of θ_0. Then it is easy to verify that for any $\widetilde{\theta}_2 \in \mathbb{R}^{p-r}$ we can construct $\widetilde{\theta}_1 \in \mathbb{R}^r$ such that,

$$\begin{aligned}
\mathbf{a}(\beta_0) &= \mathbf{B}_1\theta_{01} + \mathbf{B}_2\theta_{02} \\
&= \mathbf{B}_1\widetilde{\theta}_1 + \mathbf{B}_2\widetilde{\theta}_2, \\
\text{by setting } \widetilde{\theta}_1 &= \theta_{01} + \mathbf{\Gamma}(\theta_{02} - \widetilde{\theta}_2)).
\end{aligned}$$

I.e. (β_0,θ_0) and $(\beta_0,\widetilde{\theta})$, where $\widetilde{\theta} = [\widetilde{\theta}_1^\top : \widetilde{\theta}_2^\top]^\top$, are observationally equivalent. Therefore: *if the discount factor can be identified, it can be identified for all normalization choices.* Our argument can be straightforwardly extended to other normalization types (e.g. by assigning a value to a combination of parameters) using basic linear algebra. An equivalent way to state this result is given in the following proposition.

Proposition 1: *Suppose $\mathcal{M}^*(\beta) = 0$ if and only if $\beta = \beta_0$, then (β_0,θ_0) and $(\beta_0,\theta_0+\theta)$ are observationally equivalent for all θ that belongs to the null space of $\mathbf{B}(\beta_0)$.*

The discount factor can therefore be identified independently of the normalization choice on switching costs. Our discussion here also leads to another empirical fact that may not be obvious a priori.

Suppose a particular π is chosen and it satisfies both P and N. Then there are two different ways to estimate the switching costs based on our parametric and nonparametric identification approaches. We have shown some combinations of the switching costs can be identified without any normalization using the nonparametric approach. We are interested to know if the parametric approach in Section 3 that relies on a possibly incorrect normalization choice can consistently estimate these combinations.

The answer is positive. Consider any combination of the switching costs, which can be written explicitly in terms of the differences in choice-specific valuations (e.g. sunk costs, see Corollaries 1 and 2). A vector of such combinations can be represented by $\mathbf{\Sigma a}_0$ for some matrix $\mathbf{\Sigma}$. Then for any $\widetilde{\theta}$ such that $(\beta_0,\widetilde{\theta})$ is observationally equivalent to (β_0,θ_0) we also have $\mathbf{\Sigma a}_0 = \mathbf{\Sigma B}(\beta_0)\theta_0 = \mathbf{\Sigma B}(\beta_0)\widetilde{\theta}$. I.e. the combinations of switching costs described by $\mathbf{\Sigma B}(\beta_0)$ identify the same objects.

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:

$$\pi_1 (a_{1t}, a_{2t}, x_t; \theta) = a_{1t} (\mu_1 + \mu_2 a_{2t}) + a_{1t} (1 - a_{1t-1}) F + (1 - a_{1t}) a_{1t-1} W, \tag{17}$$

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.

The model satisfies both Assumptions MN and MP in the Appendix. We consider two estimation methods. Method A profiles out all the payoff parameters using the OLS expression and use grid search to estimate the discount factor. Method B first estimates the entry cost in closed-form independently before profiling out the other payoff parameters and use grid search to estimate the discount factor. We are also interested to see how sensitive our estimates are with respect to the normalization choice.

For each equilibrium we perform 10000 simulations with sample sizes $N = 100, 1000, 10000$. Since the entry cost and scrap value cannot be jointly identified we estimate the model under different normalized values for W. We report: the bias and standard deviation (in *italics*) for $(\widehat{\beta}, \widehat{\mu}_1, \widehat{\mu}_2, \widehat{F})$ and the sunk entry cost $(\widehat{SUNK})$; we use the **bold** font to highlight the statistics that correspond to the correctly assumed choice of W. We estimate the sunk entry cost for Methods A and B by first estimating the entry cost and combine it with the assumed scrap value. In addition we also estimate the sunk entry cost without normalizing the scrap value according to Example 1 in the Appendix (also see Corollary 1). We label the columns of statistics for the sunk entry estimator with no normalization by N-N. Tables 1-3 below correspond to the data generated according to the three equilibria as enumerated in Pesendorfer and Schmidt-Dengler (2008) respectively.

The findings are in line with the theory part of the paper. First it shows the discount factor can be consistently estimated. The consistency property is robust against the normalization choice of the scrap value. The sunk entry cost can also be consistently estimated independently of the scrap value used. When the model is correctly specified in the sense we correctly assume $W = W_0$ all estimators are consistent. While misspecifying the scrap value cause biases to all estimators of the individual payoff parameters. The estimation results from Methods A and B, as well as N-N for the sunk entry, are qualitatively the same across all equilibria. The performances between estimation methods seem to depend on the equilibrium and sample size. Method A performs better in Equilibrium 1, and generally in smaller samples. We may be able to attribute the difference in smaller samples performance to the fact that Method A fully exploits the correctly specified parametric form of the payoff function while the others use nonparametric estimators. At larger sample sizes there appear to be no dominating estimation methods for Equilibria 2 and 3.

		Method A			Method B			N-N
N	W	0	0.1	0.2	0	0.1	0.2	-
100	$\widehat{\beta}$	-0.0809	**-0.0806**	-0.0799	-0.0752	**-0.0768**	-0.0738	-
		0.2697	***0.2691***	*0.2686*	*0.2619*	***0.2640***	*0.2596*	
	$\widehat{\mu}_1$	-0.0418	**-0.0253**	-0.0071	-0.0631	**-0.0450**	-0.0291	-
		0.2974	***0.3050***	*0.3150*	*0.3693*	***0.3774***	*0.3858*	
	$\widehat{\mu}_2$	0.0627	**0.0815**	0.0988	0.0963	**0.1141**	0.1313	-
		0.2970	***0.2991***	*0.3029*	*0.4779*	***0.4801***	*0.4831*	
	$\widehat{F}$	0.0446	**-0.0554**	-0.1552	-0.0019	**-0.1017**	-0.2021	-
		0.2836	***0.2835***	*0.2839*	*0.5692*	***0.5699***	*0.5702*	
	$\widehat{SUNK}$	0.0554	**0.0554**	0.0552	0.1019	**0.1017**	0.1021	0.0477
		0.2836	***0.2835***	*0.2839*	*0.5692*	***0.5699***	*0.5702*	*0.5935*
1000	$\widehat{\beta}$	-0.0356	**-0.0372**	-0.0380	-0.0328	**-0.0339**	-0.0343	-
		0.1741	***0.1790***	*0.1801*	*0.1677*	***0.1695***	*0.1715*	
	$\widehat{\mu}_1$	-0.0051	**0.0090**	0.0229	-0.0028	**0.0110**	0.0244	-
		0.1032	***0.1129***	*0.1251*	*0.1066*	***0.1152***	*0.1265*	
	$\widehat{\mu}_2$	-0.0046	**0.0091**	0.0231	-0.0084	**0.0050**	0.0185	-
		0.0934	***0.0946***	*0.0992*	*0.1190*	***0.1204***	*0.1246*	
	$\widehat{F}$	0.0958	**-0.0042**	-0.1042	0.1000	**0.0000**	-0.1000	-
		0.0901	**0.0901**	*0.0902*	*0.1480*	***0.1480***	*0.1480*	
	$\widehat{SUNK}$	0.0042	**0.0042**	0.0042	0.0001	**0.0001**	0.0001	-0.0132
		0.0901	***0.0901***	*0.0902*	*0.1480*	***0.1480***	*0.1480*	*0.1573*
10000	$\widehat{\beta}$	-0.0005	**-0.0003**	-0.0005	-0.0005	**-0.0007**	-0.0005	-
		0.0204	***0.0158***	*0.0204*	*0.0204*	***0.0238***	*0.0205*	
	$\widehat{\mu}_1$	-0.0104	**-0.0004**	0.0097	-0.0101	**0.0000**	0.0100	-
		0.0298	***0.0299***	*0.0309*	*0.0302*	***0.0310***	*0.0312*	
	$\widehat{\mu}_2$	-0.0093	**0.0007**	0.0108	-0.0098	**0.0003**	0.0103	-
		0.0297	***0.0298***	*0.0300*	*0.0355*	***0.0356***	*0.0358*	
	$\widehat{F}$	0.0992	**-0.0008**	-0.1008	0.0998	**-0.0002**	-0.1002	-
		0.0282	***0.0282***	*0.0282*	*0.0437*	***0.0437***	*0.0437*	
	$\widehat{SUNK}$	0.0008	**0.0008**	0.0008	0.0002	**0.0002**	0.0002	-0.0011
		0.0282	***0.0282***	*0.0282*	*0.0437*	***0.0437***	*0.0437*	*0.0454*

Table 1: Data generated from equilibrium 1 in Pesendorfer and Schmidt-Dengler (2008).

		Method A			Method B			N-N
N	W	0	0.1	0.2	0	0.1	0.2	-
100	$\widehat{\beta}$	-0.0675	**-0.0691**	-0.0704	-0.0667	**-0.0660**	-0.0684	-
		0.2501	***0.2523***	*0.2542*	*0.2493*	***0.2477***	*0.2513*	
	$\widehat{\mu}_1$	-0.2087	**-0.1899**	-0.1726	-0.1185	**-0.1027**	-0.0835	-
		0.3978	***0.4135***	*0.4286*	*0.4495*	***0.4572***	*0.4718*	
	$\widehat{\mu}_2$	0.3264	**0.3447**	0.3623	0.1847	**0.2025**	0.2196	-
		0.5430	***0.5454***	*0.5500*	*0.6563*	***0.6605***	*0.6641*	
	$\widehat{F}$	-0.0630	**-0.1632**	-0.2632	0.0942	**-0.0058**	-0.1058	-
		0.4166	***0.4161***	*0.4159*	*0.5515*	***0.5515***	*0.5515*	
	$\widehat{SUNK}$	0.1630	**0.1632**	0.1632	0.0058	**0.0058**	0.0058	-0.0455
		0.4166	***0.4161***	*0.4159*	*0.5515*	***0.5515***	*0.5515*	*0.5991*
1000	$\widehat{\beta}$	-0.0296	**-0.0302**	-0.0314	-0.0318	**-0.0306**	-0.0304	-
		0.1584	***0.1600***	*0.1625*	*0.1637*	***0.1603***	*0.1594*	
	$\widehat{\mu}_1$	-0.0275	**-0.0139**	0.0003	-0.0096	**0.0028**	0.0158	-
		0.1631	***0.1739***	*0.1872*	*0.1596*	***0.1691***	*0.1807*	
	$\widehat{\mu}_2$	0.0494	**0.0626**	0.0763	0.0267	**0.0394**	0.0523	-
		0.2108	***0.2159***	*0.2234*	*0.2047*	***0.2097***	*0.2162*	
	$\widehat{F}$	0.0767	**-0.0233**	-0.1233	0.1006	**0.0006**	-0.0994	-
		0.1526	**0.1526**	*0.1526*	*0.1495*	***0.1495***	*0.1495*	
	$\widehat{SUNK}$	0.0233	**0.0233**	0.0233	-0.0006	**-0.0006**	-0.0006	-0.0052
		0.1526	***0.1526***	*0.1526*	*0.1495*	***0.1495***	*0.1495*	*0.1638*
10000	$\widehat{\beta}$	-0.0001	**-0.0002**	-0.0004	-0.0002	**-0.0004**	-0.0002	-
		0.0093	***0.0127***	*0.0183*	*0.0130*	***0.0183***	*0.0128*	
	$\widehat{\mu}_1$	-0.0147	**-0.0046**	0.0056	-0.0127	**-0.0025**	0.0073	-
		0.0399	***0.0405***	*0.0425*	*0.0381*	***0.0398***	*0.0387*	
	$\widehat{\mu}_2$	-0.0036	**0.0064**	0.0166	-0.0063	**0.0039**	0.0138	-
		0.0639	***0.0642***	*0.0649*	*0.0608*	***0.0613***	*0.0610*	
	$\widehat{F}$	0.0968	**-0.0032**	-0.1032	0.0995	**-0.0005**	-0.1005	-
		0.0487	***0.0487***	*0.0487*	*0.0464*	***0.0464***	*0.0464*	
	$\widehat{SUNK}$	0.0032	**0.0032**	0.0032	0.0005	**0.0005**	0.0005	-0.0002
		0.0487	***0.0487***	*0.0487*	*0.0464*	***0.0464***	*0.0464*	*0.0508*

Table 2: Data generated from equilibrium 2 in Pesendorfer and Schmidt-Dengler (2008).

		Method A			Method B			N-N
N	W	0	0.1	0.2	0	0.1	0.2	-
100	$\widehat{\beta}$	-0.0649	**-0.0641**	-0.0658	-0.0695	**-0.0649**	-0.0663	-
		0.2459	***0.2427***	*0.2472*	*0.2526*	***0.2450***	*0.2471*	
	$\widehat{\mu}_1$	-0.2070	**-0.1907**	-0.1725	-0.1116	**-0.0986**	-0.0807	-
		0.3991	***0.4108***	*0.4261*	*0.4724*	***0.4804***	*0.4920*	
	$\widehat{\mu}_2$	0.3263	**0.3420**	0.3588	0.1801	**0.1961**	0.2130	-
		0.5460	***0.5484***	*0.5551*	*0.7092*	***0.7109***	*0.7158*	
	$\widehat{F}$	-0.0677	**-0.1676**	-0.2672	0.0897	**-0.0103**	-0.1103	-
		0.4224	***0.4227***	*0.4230*	*0.5987*	***0.5988***	*0.5988*	
	$\widehat{SUNK}$	0.1677	**0.1676**	0.1672	0.0103	**0.0103**	0.0103	-0.0370
		0.4224	***0.4227***	*0.4230*	*0.5987*	***0.5988***	*0.5988*	*0.6455*
1000	$\widehat{\beta}$	-0.0320	**-0.0322**	-0.0333	-0.0326	**-0.0324**	-0.0319	-
		0.1634	***0.1643***	*0.1666*	*0.1647*	***0.1648***	*0.1638*	
	$\widehat{\mu}_1$	-0.0237	**-0.0104**	0.0041	-0.0060	**0.0071**	0.0199	-
		0.1677	***0.1796***	*0.1932*	*0.1678*	***0.1790***	*0.1900*	
	$\widehat{\mu}_2$	0.0500	**0.0633**	0.0771	0.0251	**0.0383**	0.0511	-
		0.2130	***0.2188***	*0.2264*	*0.2174*	***0.2235***	*0.2305*	
	$\widehat{F}$	0.0766	**-0.0234**	-0.1234	0.1014	**0.0014**	-0.0986	-
		0.1549	**0.1550**	*0.1550*	*0.1604*	***0.1604***	*0.1604*	
	$\widehat{SUNK}$	0.0234	**0.0234**	0.0234	-0.0014	**-0.0014**	-0.0014	-0.0061
		0.1549	***0.1550***	*0.1550*	*0.1604*	***0.1604***	*0.1604*	*0.1785*
10000	$\widehat{\beta}$	-0.0003	**-0.0003**	-0.0003	-0.0001	**-0.0004**	-0.0002	-
		0.0159	***0.0158***	*0.0156*	*0.0093*	***0.0163***	*0.0128*	
	$\widehat{\mu}_1$	-0.0146	**-0.0046**	0.0054	-0.0128	**-0.0026**	0.0073	-
		0.0410	***0.0414***	*0.0420*	*0.0399*	***0.0415***	*0.0410*	
	$\widehat{\mu}_2$	-0.0033	**0.0067**	0.0167	-0.0062	**0.0039**	0.0138	-
		0.0648	***0.0649***	*0.0650*	*0.0646*	***0.0650***	*0.0650*	
	$\widehat{F}$	0.0965	**-0.0035**	-0.1035	0.0992	**-0.0008**	-0.1008	-
		0.0496	***0.0496***	*0.0496*	*0.0497*	***0.0497***	*0.0497*	
	$\widehat{SUNK}$	0.0035	**0.0035**	0.0035	0.0008	**0.0008**	0.0008	0.0002
		0.0496	***0.0496***	*0.0496*	*0.0497*	***0.0497***	*0.0497*	*0.0553*

Table 3: Data generated from equilibrium 3 in Pesendorfer and Schmidt-Dengler (2008).

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.[6] 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 very 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

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

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 extends the single agent model in the paper as described in the Appendix. 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 into variable profits, operating cost and switching costs. We assume the variable profit is determined by the players competing in a capacity constrained Cournot game. The operating cost is a fixed profit that incurs when $a_{it} > 0$. The switching costs capture 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. The operating and switching costs parameters enter the payoff function additively and are parameters to be estimated using the dynamic model. These operating cost is non-zero whenever $a_{it} > 0$. For the switching costs we normalize the payoff for choosing action 0 to be zero. There are a total of 25 switching cost parameters to be estimated.[7]

The payoff function in our empirical model satisfies Assumptions MN and MP in the Appendix. So 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 also estimate the sunk entry and investment values based on the estimates from Methods A and B, as well as nonparametrically without normalization (cf. Corollaries

[7] 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).

1 and 2, and see the discussion in the Appendix).

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). We use 500 bootstrap samples and report the standard errors in *italics*.

Results

We estimate the model twice. Once using the data from before and after the implementation of the 1990 CAAA. We assume for illustrational purposes 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 and there is no other source of unobserved heterogeneity.[8]

Table 4 and 5 compile the results from estimating switching costs using the data from the years 1980 to 1990 and 1991 to 1998 respectively. Tables 6 and 7 give the estimates for the discount factor and fixed operating cost using the data from the corresponding periods. Table 8 compares the estimates of the sunk entry costs and sunk investment costs.

The signs and relative magnitudes of individually estimated switching costs almost uniformly make sensible economic sense. E.g., by reading down the columns in Tables 4 and 5, 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 alone 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.[9] Therefore we do not reject the capacity constrained Cournot game specification based on comparing the switching costs estimates. Comparing Tables 4 and 5 shows the entry and switching costs increase after

[8] Recently Otsu, Pesendorfer and Takahashi (2015) propose several tests to detect differences in the probability distribution of data across markets. If a test rejects then there is evidence data across markets should not be pooled together, which can point to possible violation of single equilibrium assumption and/or misspecification in terms of omitting other unobserved heterogeneity. They actually suggest Ryan's data in general should not be pooled together across markets. In particular there is a strong evidence against pooling data between 1980 and 1990, while the data from 1991 to 1998 did not get rejected by some of their poolability tests.

[9] 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.

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. An increase in switching costs can be partly attributed to the new plants using newer (or better maintained) equipment that require more certification and testing than previously.

We 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.[10] 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 implementation of 1990 CAAA.

Finally Table 8 reports sunk costs using different estimation methods. The estimates from Methods A and B can be found by computing $-SC^{a' \to a} - SC^{a \to a'}$ using individual switching costs in Tables 4 and 5. The N-N approach estimates the same object without the assumption that the payoff is zero upon choosing action 0. The signs and magnitudes of the sunk cost estimates are plausible. We find the sunk investment costs between any two capacity levels increase as the gap between levels grow, while we find the costs to be of similar magnitude when compared within the same capacity difference bands. We also find the sunk costs to have increased after the implementation of 1990 CAAA.

[10]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 4: Results from estimating switching costs using data from the years 1980 to 1990.

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 5: Results from estimating switching costs using data from the years 1991 to 1998.

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 6: Results from estimating the discount factor and fixed operating cost using data from the years 1980 to 1990.

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 7: Results from estimating the discount factor and fixed operating cost using data from the years 1991 to 1998.

		Before			After		
a_{it}	a_{it-1}	Method A	Method B	N-N	Method A	Method B	N-N
1	0	3.30	2.78	2.78	6.96	5.65	5.66
		0.36	*0.27*	*0.27*	*1.53*	*0.73*	*0.70*
2	0	10.50	10.48	10.48	17.04	17.75	17.74
		0.94	*0.69*	*0.69*	*1.72*	*1.38*	*1.49*
3	0	23.27	23.28	23.28	35.49	36.10	36.10
		1.41	*1.34*	*1.34*	*2.44*	*2.28*	*2.18*
4	0	41.02	41.42	41.42	51.54	51.84	51.83
		2.00	*1.81*	*1.80*	*3.06*	*2.20*	*1.61*
5	0	52.88	54.38	54.25	64.02	64.24	64.22
		2.28	*1.91*	*2.00*	*4.51*	*6.71*	*6.34*
2	1	2.98	2.66	2.44	3.84	3.98	3.30
		1.22	*2.54*	*0.25*	*0.31*	*0.61*	*0.36*
3	2	2.10	2.53	2.56	1.54	3.18	3.22
		1.18	*2.30*	*0.26*	*0.30*	*0.73*	*0.33*
4	3	1.81	2.68	2.58	3.40	4.90	4.81
		1.52	*4.33*	*0.28*	*0.42*	*2.45*	*0.50*
5	4	0.30	5.24	2.87	7.85	7.34	7.30
		2.50	*4.75*	*0.33*	*1.74*	*4.58*	*2.14*
3	1	10.36	10.09	10.01	13.53	13.81	13.05
		1.22	*2.12*	*0.75*	*0.79*	*1.24*	*0.98*
4	2	9.59	10.33	10.29	14.69	16.01	16.07
		1.54	*3.31*	*0.77*	*0.81*	*2.13*	*1.25*
5	3	9.71	12.66	10.91	24.23	24.78	24.21
		1.45	*4.83*	*0.91*	*1.37*	*6.09*	*5.22*
4	1	22.66	22.99	22.76	34.92	34.16	34.02
		1.78	*3.29*	*1.37*	*1.45*	*1.93*	*1.42*
5	2	23.22	25.98	24.05	34.76	35.95	34.79
		1.83	*4.64*	*1.79*	*1.59*	*6.89*	*6.34*
5	1	39.88	41.20	40.21	43.73	42.79	41.67
		2.40	*4.68*	*2.60*	*2.08*	*6.82*	*6.40*

Table 8: Results from estimating the sunk entry and investment costs.

6 Concluding Remarks

We show the discount factor can be identified jointly with the payoff function under the linear-in-parameter specification. The key property we exploit is the conditional linearity of the choice-specific value functions for a given value of the discount factor. The discount factor can in fact be identified even if the payoff parameters cannot be identified. This has an important implication since many empirical problems have to normalize parts of the payoff parameters. Our result shows the discount factor can be identified independently of these normalization choices.

We also contribute to recent interest in the robust identification of combination of switching costs without any normalization as studied in Aguirregabiria and Suzuki (2014); also see Kalouptsidi, Scott and Souza-Rodrigues (2016a, 2016b). We provide closed-form identification results independently of the discount factor and other parts of the payoff function. We show some costs, such as sunk entry and investment costs, can be identified without any normalization and, for linear models, even when an incorrect normalization is used.

We have shown our parametric and nonparametric identification approaches can deliver substantially different (flavors of) results. But there is a considerable overlap in practice when it comes to estimating the switching costs as a payoff function can satisfy both P and N(i). However, there are notable implications for our nonparametric results that extend beyond the linear model. First, a researcher may want to use a nonlinear parametric specification on parts of the payoffs; for example to impose positivity. Second, our nonparametric identification result is valid pointwise for each observed state, therefore it is immediately applicable to models with continuous states; e.g. see Srisuma and Linton (2012). In these cases Assumption P has no implication on our nonparametric identification results.

Finally our main message is that one should generally attempt to identify and estimate the discount factor in dynamic decision problems and games. Clearly we do not expect the linear specification to be necessary for identification, but an analysis with nonlinear a parametric payoff function will be substantially more difficult. Similarly, outside of discrete choice models, e.g. for games with supermodular payoff functions (see Bajari, Benkard and Levin (2007) and Srisuma (2013)), joint identification and estimation of the discount factor and payoff parameters should also be possible. However, the practical implementation can be burdensome since there is no obvious way to reduce the parameter space even when the payoff functions take a linear-in-parameter structure.

Appendix

The Appendix has two parts. A.1 extends the results on identification of switching costs to dynamic games. A.2 provides a high level sufficient condition for the identification of the discount factor. Since the single agent decision problem is a special case of a game, we also present the results in A.2 in the context of a game.

A.1 Identification of the Switching Costs in Dynamic Games

We shall keep our description of the basic elements of the game very brief. The notation we use directly extends what we describe in Sections 2 and 3. Consider a game with I players, indexed by $i \in \mathcal{I} = \{1, \ldots, I\}$. The random variables in the game are the actions: $a_t \equiv (a_{it}, a_{-it}) \in A^I$, $A = \{0, 1, \ldots, J\}$; past actions $w_t \equiv (w_{it}, w_{-it}) \in A^I$; $s_{it} \equiv (w_t, x_t, \varepsilon_{it}) \in A^I \times X \times \mathbb{R}^{J+1}$, where $X = \{1, \ldots, K\}$, and $\varepsilon_{it} \equiv (\varepsilon_{it}(0), \ldots, \varepsilon_{it}(J)) \in \mathbb{R}^{J+1}$; and we let $s_t \equiv (w_t, x_t, \varepsilon_{1t}, \ldots, \varepsilon_{It})$.

In an equilibrium $a_{it} = \alpha_i(s_{it})$ for all i, such that

$$\alpha_i(s_i) = \max_{a_i \in A}\{E[u_i(a_{it}, a_{-it}, s_i) | s_{it} = s_i, a_{it} = a_i] + \beta E[V_i(s_{it+1}) | s_{it} = s_i, a_{it} = a_i]\}, \tag{18}$$

where u_i and V_i are player i's payoff and value function respectively; in particular

$$V_i(s_i) = \sum_{t=0}^{\infty} \beta^t E[u_i(a_{it}, a_{-it}, s_{it}) | s_{i0} = s_i].$$

Assumption MN updates Assumptions M and N for games.

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

$$u_i(a_i, a_{-i}, w, x, \varepsilon_i) = \pi_i(a_i, a_{-i}, w, x) + \varepsilon_i(a_i).$$

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

$$P(x', \varepsilon' | x, \varepsilon, w, a) = \prod_{i=1}^{I} Q_i(\varepsilon_i') G(x' | x, w, a),$$

where Q_i is the cumulative distribution function of ε_{it} and G denotes the transition law of x_{t+1} conditioning on x_t, a_t. Furthermore, ε_{it} has finite first moments, and a positive, continuous and bounded density on $\mathbb{R}^{J+1}$.

(iii) (Finite Observed State) $X = \{1, \ldots, K\}$.

(iv) (Decomposition of Profits): For all a, w, x:

$$\pi_i(a_i, a_{-i}, w, x, \varepsilon) = \mu_i(a_i, a_{-i}, x) + \phi_i(a_i - w_i, w_{-i}, x),$$

such that $\phi_i(0, w_{-i}, x) = 0$*.*

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

Beside from explicitly separating out past actions from other observed state variables, MN(i) to MN(iii) are standard in the dynamic discrete choice game literature; e.g. see Aguirregabiria and Mira (2007), Bajari et al. (2007), Pakes and Berry (2007), and Pesendorfer and Schmidt-Dengler (2008). MN(iv) extends N(i). It assumes that strategic interactions can affect payoffs in μ_i directly but not ϕ_i, while past actions enter ϕ_i but not μ_i. The exclusion restrictions we impose are quite natural for components of μ_i such as per-period variable profits and operation costs, while switching costs that occur for each player are determined by her own actions. It will be useful to sometimes represent the switching cost using a more intuitive notation (cf. equation (9)):

$$\phi_i(a_i - w_i, w_{-i}, x) = SC_i^{w_i \to a_i}(w_{-i}, x).$$

MN(v) is a direct extension of N(ii).

As with the single agent case, our identification study will be based on the choice-specific value function:

$$v_i(a_i, w, x) = E[\pi_i(a_i, a_{-it}, w_t, x) | w_t = w, x_t = x] + \beta E[V_i(s_{t+1}) | w_t = w, x_t = x, a_t = a],$$

which can be recover from:

$$\Pr[a_{it} = a_i | w_t = w, x_t = x] = \Pr[\Delta v_i(a_i, w, x) - \Delta v_i(a'_i, w, x) > \varepsilon_{it}(a'_i) - \varepsilon_{it}(a_i) \quad \text{for all} \quad a'_i \neq a_i],$$

where $\Delta v_i(a_i, w, x) \equiv v_i(a_i, w, x) - v_i(0, w, x)$. Let also, $\Delta\lambda_i(a_i, a_{-i}, x) \equiv \lambda_i(a_i, a_{-i}, x) - \lambda_i(0, a_{-i}, x)$ and $\Delta\phi_i(a_i, w, x) \equiv \phi_i(a_i - w_i, w_{-i}, x) - \phi_i(-w_i, w_{-i}, x)$. Lemma 4 is a generalization of Lemma 1.

Lemma 4: *Under Assumption MN, we have for all* $i, a_i > 0$ *and* w, x:

$$\Delta v_i(a_i, w, x) = E[\Delta\lambda_i(a_i, a_{-it}, x) | w_t = w, x_t = x] + \Delta\phi_i(a_i, w, x),$$

where,

$$\begin{aligned} \Delta\lambda_i(a_i, a_{-i}, x) &\equiv \pi_i(a_i, a_{-i}, x) - \pi_i(0, a_{-i}, x) + \beta(\widetilde{m}_i(a_i, a_{-i}, x) - \widetilde{m}_i(0, a_{-i}, x)), \\ \widetilde{m}_i(a_i, a_{-i}, x) &\equiv E[m_i(w_{t+1}, x_{t+1}) | a_{it} = a_i, a_{-it} = a_{-i}, x_t = x], \\ m_i(w, x) &\equiv E[V_i(s_{it}) | w_t = w, x_t = x]. \end{aligned}$$

PROOF: Follows immediately from applying the law of iterated expectations (cf. the proof of Lemma 1).■

Since we have finite actions and states, we can collect $\Delta v_i(a_i, w, x)$ across w for each (i, a_i, x) into a vector of size $(J+1)^I$. Using a matrix form, we have:

$$\mathbf{\Delta v}_i(a_i, x) = \mathbf{Z}_i(x)\,\mathbf{\Delta\lambda}_i(a_i, x) + \mathbf{Q}_i(a_i, x)\,\boldsymbol{\phi}_i(a_i, x), \tag{19}$$

where $\mathbf{\Delta v}_i(a_i, x) = (\Delta v_i(a_i, w, x))_{w \in A^I}$, $\mathbf{\Delta\lambda}_i(a_i, x) = (\Delta\lambda_i(a_i, a_{-i}, x))_{a_{-i} \in A^{I-1}}$, $\mathbf{Z}_i(x)$ represents the matrix of conditional probabilities for computing a conditional expectation of a_{-it} given $(w_t = w, x_t = x)$, $\mathbf{Q}_i(a_i, x)\,\boldsymbol{\phi}_i(a_i, x)$ represents $(\Delta\phi_i(a_i, w, x))_{a \in A^I}$ with $\boldsymbol{\phi}_i(a_i, x) = (\phi_i(a_i - w_i, w_{-i}, x))_{w_i \in A, w_{-i} \in A^{I-1}}$ and $\mathbf{Q}_i(a_i, x)$ is a matrix of indicators (consisting of 0's and 1's) that pick up switching costs as appropriate.

Theorem 3 generalizes the closed-form identification of switching costs in Theorem 1 for dynamic games.

THEOREM 3: *Assume that Assumption MN holds. Let* $\mathbf{D}$ *be an* $\ell_1 \times (J+1)^I$ *matrix with* $\rho(\mathbf{D}) = \ell_1$ *such that* $(J+1)^{I-1} < \ell_1 \leq (J+1)^I$. *Denote* $\mathbf{DZ}_i(x)$ *by* $\widetilde{\mathbf{Z}}$ *and* $\rho(\widetilde{\mathbf{Z}})$ *by* ℓ_2. *Suppose also* $\mathbf{DQ}_i(a_i, x)\,\boldsymbol{\phi}_i = \widetilde{\mathbf{Q}}\widetilde{\boldsymbol{\phi}} + \boldsymbol{\phi}_0$ *for some* ℓ_3-*dimensional vectors* $\widetilde{\boldsymbol{\phi}}$ *and* $\boldsymbol{\phi}_0$ *that consist of elements, possibly combinations, of* $\boldsymbol{\phi}_i$ *such that* $\ell_3 \leq \ell_1 - \ell_2$, *and* $\widetilde{\mathbf{Q}}$ *is an* $\ell_1 \times \ell_3$ *matrix with* $\rho(\widetilde{\mathbf{Q}}) = \ell_3$. *If* $\rho([\widetilde{\mathbf{Z}} : \widetilde{\mathbf{Q}}]) = \ell_2 + \ell_3$ *then,*

$$\widetilde{\boldsymbol{\phi}} = (\widetilde{\mathbf{Q}}^\top \widetilde{\mathbf{P}} \widetilde{\mathbf{Q}})^{-1} \widetilde{\mathbf{Q}}^\top \widetilde{\mathbf{P}}\,(\mathbf{D}\mathbf{\Delta v}_i(a_i, x) - \boldsymbol{\phi}_0). \tag{20}$$

where $\widetilde{\mathbf{P}} = I_{\ell_1} - \widetilde{\mathbf{Z}}(\widetilde{\mathbf{Z}}^\top \widetilde{\mathbf{Z}})^\dagger \widetilde{\mathbf{Z}}^\top$.

Before presenting the proof to Theorem 3 some explanations on the notations will be useful. The crucial interpretation of our result rests on the relation: $\mathbf{DQ}_i(a_i, x)\,\boldsymbol{\phi}_i = \widetilde{\mathbf{Q}}\widetilde{\boldsymbol{\phi}} + \boldsymbol{\phi}_0$. The goal of Theorem 3 is to identify components, or combinations, of $(\phi_i(a_i, w, x))_{w \in A^I}$ using choice-specific value functions in equation (19) for a given (i, a_i, x). We denote the object of interest by $\widetilde{\boldsymbol{\phi}}$. We use $\boldsymbol{\phi}_0$ to account for components of switching costs that can be identified outside the dynamic model from the data or by normalization. Therefore $(\mathbf{D}, \widetilde{\mathbf{Q}})$ are user-chosen matrices and are known. We can also treat $\widetilde{\mathbf{Z}}_i$ as known since $\mathbf{Z}_i(x)$ is a matrix of observed choice probabilities.

PROOF OF THEOREM 3.

Note that $\ell_3 \geq 1$ since $\ell_2 \leq \min\{\ell_1, \rho(\mathbf{Z}_i(x))\}$ and $\rho(\mathbf{Z}_i(x)) \leq (J+1)^{I-1}$. Multiply equation (19) by $\mathbf{D}$ yields,

$$\mathbf{D}\mathbf{\Delta v}_i(a_i, x) = \widetilde{\mathbf{Z}}\mathbf{\Delta\lambda}_i(a_i, x) + \widetilde{\mathbf{Q}}\widetilde{\boldsymbol{\phi}} + \boldsymbol{\phi}_0.$$

By assumption, $\widetilde{\mathbf{P}}\widetilde{\mathbf{Q}}$ has full column rank. The result then follows from projecting $\mathbf{\Delta v}_i(a_i, x)$ orthogonally onto the null space of $\widetilde{\mathbf{Z}}$ and solve out for $\widetilde{\boldsymbol{\phi}}_i$.■

One systematic approach to apply Theorem 3 in practice is to first write out the matrix equation (19). Then choose $\mathbf{D}$ so that $\mathbf{DQ}_i(a_i, x)\boldsymbol{\phi}_i$ contains the switching costs of interest, and define $\widetilde{\mathbf{Q}}\widetilde{\boldsymbol{\phi}} + \boldsymbol{\phi}_0$ appropriately. We now illustrate this identifying strategy with a two-player binary choice game for different types of switching costs.

For notational compactness we will suppress x_t and assume that $SC_i^{w\to a}(w_{-i})$ is the same for all w_{-i}. We use $\Delta\nu_i(w_i, w_{-i}) \equiv v_i(1, w_i, w_{-i}) - v_i(0, w_i, w_{-i})$, $p_{-i}(w) \equiv \Pr[a_{-it} = 1 | w_t = w]$, and $\Delta\lambda_i(a_{-i}) \equiv \Delta\lambda_i(1, a_{-i})$. Then equation (19) represents:

$$\begin{bmatrix} \Delta v_i(0,0) \\ \Delta v_i(0,1) \\ \Delta v_i(1,0) \\ \Delta v_i(1,1) \end{bmatrix} = \begin{bmatrix} 1 - p_{-i}(0,0) & p_{-i}(0,0) \\ 1 - p_{-i}(0,1) & p_{-i}(0,1) \\ 1 - p_{-i}(1,0) & p_{-i}(1,0) \\ 1 - p_{-i}(1,1) & p_{-i}(1,1) \end{bmatrix} \begin{bmatrix} \Delta\lambda_i(0) \\ \Delta\lambda_i(1) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} SC_i^{0\to 1} \\ -SC_i^{1\to 0} \end{bmatrix}. \qquad (21)$$

In particular we have

$$\mathbf{Q}_i(a_i, x)\boldsymbol{\phi}_i = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} SC_i^{0\to 1} \\ -SC_i^{1\to 0} \end{bmatrix}.$$

We consider three examples of potential objects of interest.

EXAMPLE 1: SUNK ENTRY COST

Suppose we want to identify $-SC_i^{0\to 1} - SC_i^{1\to 0}$ that represents the sunk entry cost in the context of an entry game. We can subtract $\Delta v_i(0,0)$ from the first equation in (21) off the remaining three equations. This yields

$$\begin{bmatrix} \Delta v_i(0,1) \\ \Delta v_i(1,0) \\ \Delta v_i(1,1) \end{bmatrix} = \begin{bmatrix} p_{-i}(0,0) - p_{-i}(0,1) & p_{-i}(0,1) - p_{-i}(0,0) \\ p_{-i}(0,0) - p_{-i}(1,0) & p_{-i}(1,0) - p_{-i}(0,0) \\ p_{-i}(0,0) - p_{-i}(1,1) & p_{-i}(1,1) - p_{-i}(0,0) \end{bmatrix} \begin{bmatrix} \Delta\lambda_i(0) \\ \Delta\lambda_i(1) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \left[-SC_i^{0\to 1} - SC_i^{1\to 0}\right].$$

In particular, in this case,

$$\widetilde{\mathbf{Z}} = \begin{bmatrix} p_{-i}(0,0) - p_{-i}(0,1) & p_{-i}(0,1) - p_{-i}(0,0) \\ p_{-i}(0,0) - p_{-i}(1,0) & p_{-i}(1,0) - p_{-i}(0,0) \\ p_{-i}(0,0) - p_{-i}(1,1) & p_{-i}(1,1) - p_{-i}(0,0) \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \widetilde{\mathbf{Q}} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \widetilde{\boldsymbol{\phi}} = -SC_i^{0\to1} - SC_i^{1\to0}, \text{ and } \boldsymbol{\phi}_0 = 0.$$

The sunk entry cost can then be identified by the expression in equation (20).

Example 2: Menu cost under symmetry

Suppose we want to identify $SC_i^{0\to1}$ under the assumption that $SC_i^{0\to1} = SC_i^{1\to0}$. Then equation (21) becomes

$$\begin{bmatrix} \Delta v_i(0,0) \\ \Delta v_i(0,1) \\ \Delta v_i(1,0) \\ \Delta v_i(1,1) \end{bmatrix} = \begin{bmatrix} 1 - p_{-i}(0,0) & p_{-i}(0,0) \\ 1 - p_{-i}(0,1) & p_{-i}(0,1) \\ 1 - p_{-i}(1,0) & p_{-i}(1,0) \\ 1 - p_{-i}(1,1) & p_{-i}(1,1) \end{bmatrix} \begin{bmatrix} \Delta\lambda_i(0) \\ \Delta\lambda_i(1) \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} SC_i^{0\to1} \end{bmatrix}.$$

In this case

$$\widetilde{\mathbf{Z}} = \begin{bmatrix} 1 - p_{-i}(0,0) & p_{-i}(0,0) \\ 1 - p_{-i}(0,1) & p_{-i}(0,1) \\ 1 - p_{-i}(1,0) & p_{-i}(1,0) \\ 1 - p_{-i}(1,1) & p_{-i}(1,1) \end{bmatrix}, \mathbf{D} = I_4, \widetilde{\mathbf{Q}} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \widetilde{\boldsymbol{\phi}} = SC_i^{0\to1}, \text{ and } \boldsymbol{\phi}_0 = 0.$$

Example 3: Switching Costs with Normalizations

Suppose we want to identify $SC_i^{0\to1}$ under the assumption that $SC_i^{0\to1} = c_0$. For example, we may be interested in identifying the entry cost under the assumption that the scrap value is c_0. Then equation (21) becomes

$$\begin{bmatrix} \Delta v_i(0,0) \\ \Delta v_i(0,1) \\ \Delta v_i(1,0) \\ \Delta v_i(1,1) \end{bmatrix} = \begin{bmatrix} 1 - p_{-i}(0,0) & p_{-i}(0,0) \\ 1 - p_{-i}(0,1) & p_{-i}(0,1) \\ 1 - p_{-i}(1,0) & p_{-i}(1,0) \\ 1 - p_{-i}(1,1) & p_{-i}(1,1) \end{bmatrix} \begin{bmatrix} \Delta\lambda_i(0) \\ \Delta\lambda_i(1) \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} SC_i^{0\to1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -c_0 \\ -c_0 \end{bmatrix}.$$

In this case

$$\widetilde{\mathbf{Z}} = \begin{bmatrix} 1 - p_{-i}(0,0) & p_{-i}(0,0) \\ 1 - p_{-i}(0,1) & p_{-i}(0,1) \\ 1 - p_{-i}(1,0) & p_{-i}(1,0) \\ 1 - p_{-i}(1,1) & p_{-i}(1,1) \end{bmatrix}, \mathbf{D} = I_4, \widetilde{\mathbf{Q}} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \widetilde{\boldsymbol{\phi}} = SC_i^{0\to1}, \text{ and } \boldsymbol{\phi}_0 = \begin{bmatrix} 0 \\ 0 \\ -c_0 \\ -c_0 \end{bmatrix}.$$

In order to obtain the sunk costs when the number of actions is larger than two one has to combine identifiable objects across actions, e.g. see Corollary 2. Identification of objects for each action can be obtained as the examples above have shown. We use Theorem 3 to estimate the games such as those in our simulation study and the empirical model of capacity game in Section 5 of our paper.

A.2 A Sufficient Condition for Identification of the Discount Factor

In this part of the appendix we attempt to give a more analytical approach that ensures identification of the discount factor and payoff parameters in a dynamic game context. We first introduce some additional notations.

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 generalize the setup of Section 4 to dynamic games. The following is a straightforward extension of Assumptions M and P.

Assumption MP:

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

$$u_i\left(a_i, a_{-i}, x, \varepsilon_i; \theta\right) = \pi_i\left(a_i, a_{-i}, x; \theta\right) + \varepsilon_i\left(a_i\right).$$

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

$$P\left(x', \varepsilon' | x, \varepsilon, w, a\right) = \prod_{i=1}^{I} Q_i\left(\varepsilon_i'\right) G\left(x' | x, w, a\right),$$

where Q_i is the cumulative distribution function of ε_{it} and G denotes the transition law of x_{t+1} conditioning on x_t, a_t. Furthermore, ε_{it} has finite first moments, and a positive, continuous and bounded density on $\mathbb{R}^{J+1}$.

(iii) (Finite Observed State) $X = \{1, \ldots, K\}$.

(iv) (Linear-in-Parameters): For all $a_i, a_{-i}, x, \varepsilon_i$*:*

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

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

Our analysis will be based on the parameterized choice-specific value function:

$$v_i(a_i, x; \beta, \theta) = E[\pi_i(a_i, a_{-it}, x; \theta) | x_t = x] + \beta E[V_i(s_{t+1}; \beta, \theta) | x_t = x, a_{it} = a_i], \text{ where}$$

$$V_i(s_i; \beta, \theta) = \sum_{t=0}^{\infty} \beta^t E[u_i(a_{it}, a_{-it}, s_{it}; \theta) | s_{i0} = s_i].$$

Let $\Delta v_i(a_i, x; \beta, \theta) \equiv v_i(a_i, x; \beta, \theta) - v_i(0, x; \beta, \theta)$. We can use Δv_i from all players to define the pseudo-model and the corresponding notion of identification and observationally equivalence as in Section 4. We will omit this discussion to avoid repetition.

Our starting point will be the following lemma that generalizes Lemma 2.

LEMMA 5: *Under Assumption MP, we have for all* $i, a_i > 0$, $\Delta \mathbf{v}_i^{a_i}(\beta, \theta) \equiv (\Delta v_i(a_i, x; \beta, \theta))_{x \in X}$ *can collected in the following vector form for all* $(\beta, \theta) \in \mathcal{B} \times \Theta$*:*

$$\begin{aligned} \Delta \mathbf{v}_i^{a_i}(\beta, \theta) \;=\; & \Delta \mathbf{R}_{i0}^{a_i} + \beta \Delta \mathbf{H}_i^{a_i} (\mathbf{I}_K - \beta \mathbf{L})^{-1} \mathbf{R}_{i0} \\ & + \left(\Delta \mathbf{R}_{i1}^{a_i} + \beta \Delta \mathbf{H}_i^{a_i} (\mathbf{I}_K - \beta \mathbf{L})^{-1} \mathbf{R}_{i1} \right) \theta \\ & + \beta \Delta \mathbf{H}_i^{a_i} (\mathbf{I}_K - \beta \mathbf{L})^{-1} \boldsymbol{\epsilon}_i, \end{aligned} \tag{22}$$

where the elements in the above display are collected and explained in Tables C and D.

Matrix	Dimension	Representing
$\Delta \mathbf{R}_{i1}^{a_i}$	K by p	$E[\pi_{i1}(a_i, a_{-it}, x_t) - \pi_{i1}(0, a_{-it}, x_t) \| x_t = \cdot]$
$\mathbf{R}_1$	K by p	$E[\pi_{i1}(a_t, x_t) \| x_t = \cdot]$
$\mathbf{L}$	K by K	$E[\psi(x_{t+1}) \| x_t = \cdot]$
$\mathbf{H}_i^{a_i}$	K by K	$E[\psi(x_{t+1}) \| x_t = \cdot, a_{it} = a_i]$
$\Delta \mathbf{H}_i^{a_i}$	K by K	$E[\psi(x_{t+1}) \| x_t = \cdot, a_{it} = a_i] - E[\psi(x_{t+1}) \| x_t = \cdot, a_{it} = 0]$

Table C. 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
$\boldsymbol{\epsilon}_i$	$E\left[\varepsilon_{it}\left(a_{it}\right)\right\| x_t=\cdot]$
$\Delta\mathbf{R}_{i0}^{a_i}$	$E\left[\pi_{i0}\left(a_i,a_{-it},x_t\right)-\pi_{i0}\left(0,a_{-it},x_t\right)\|x_t=\cdot\right]$
$\mathbf{R}_{i0}$	$E\left[\pi_{i0}\left(a_t,x_t\right)\|x_t=\cdot\right]$
$\left(\mathbf{I}_K-\beta_i\mathbf{L}\right)^{-1}\mathbf{R}\boldsymbol{\Pi}_{ij}$	$\sum_{t=0}^{\infty}\beta^t E[\pi_{ij}\left(a_t,x_t\right)\|x_0=\cdot]$
$\Delta\mathbf{H}_i^{a_i}\left(\mathbf{I}_K-\beta_i\mathbf{L}\right)^{-1}\mathbf{R}\boldsymbol{\Pi}_{ij}$	$\sum_{t=0}^{\infty}\beta^t\left(E[\pi_{ij}\left(a_t,x_t\right)\|a_{i0}=a_i,x_0=\cdot]-E[\pi_{ij}\left(a_t,x_t\right)\|a_{i0}=0,x_0=\cdot]\right)$
$\Delta\mathbf{H}_i^{a_i}\left(\mathbf{I}_K-\beta_i\mathbf{L}\right)^{-1}\boldsymbol{\epsilon}_i$	$\sum_{t=0}^{\infty}\beta^t\left(E\left[\varepsilon_t\left(a_t\right)\right\| a_{i0}=a_i,x_0=\cdot]-E\left[\varepsilon_t\left(a_t\right)\right\| a_{i0}=0,x_0=\cdot]\right)$

Table D. The K by 1 vectors represent (differences in) expected payoffs.

Our strategy to show identification is to re-write Lemma 5 to set up a map that defines the data generating parameter as its fixed-point. One desired relation is the following.

LEMMA 6: *Under Assumption MP, (β,θ) is observationally equivalent to (β_0,θ_0) if and only if (β,θ) satisfies*

$$\mathbf{c}_i^{a_i}-\mathbf{D}_i^{a_i}\left(\beta\right)\theta-\mathbf{E}_i\left(\beta\right)=\mathbf{F}_i^{a_i}\begin{pmatrix}\theta\\ \beta\end{pmatrix} \tag{23}$$

for all $i,a_i>0$, where

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

PROOF: Equation (23) is obtained by re-arranging equation (22), after applying the identity that $\left(\mathbf{I}_K-\beta\mathbf{L}\right)^{-1}=\mathbf{I}_K+\beta\mathbf{L}\left(\mathbf{I}_K-\beta\mathbf{L}\right)^{-1}$ and replace $\Delta\mathbf{v}_i^{a_i}\left(\beta,\theta\right)$ by $\Delta\mathbf{v}_i^{a_i}\left(\beta_0,\theta_0\right)$. Therefore, by construction, (β,θ) satisfies (22) 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 $K\geq p+1$ and Assumption MP holds. 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 K 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*

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

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

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 (23), 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

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

and

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

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

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

It then follows that

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

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 trade-off 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 (23) 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 (23) 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.

(iii) Rank Deficiency. We have seen that sometimes not all components of the payoff functions can be identified. For example in the entry/exit game generally the entry cost and scrap value cannot be jointly identified. Then one may consider normalizing, say, the scrap value in order to estimate all the other parameters in the model. If the normalized value is incorrect one may expect that identification of the parameters of interest will not be possible since the model is misspecified. Yet, our simulation study suggests the discount factor can be consistently estimated while all other profit parameters are biased when an incorrect scrap value is assumed. We can also relax condition (i) in Theorem 4 in this direction and allow $\mathbf{F}_i^{a_i}$ to be rank deficient. In particular, recall from (23) that $\mathbf{F}_i^{a_i} = [\Delta\mathbf{R}_{i1}^{a_i} : \Delta\mathbf{H}_i^{a_i}(\mathbf{R}_{i0}+\boldsymbol{\epsilon}_i)]$, we can allow $\Delta\mathbf{R}_{i1}^{a_i}$ to be rank deficient. In such case there exists a full rank matrix $\mathbf{W}$ such that $\Delta\mathbf{R}_{i1}^{a_i}\mathbf{W} = [\Delta\widetilde{\mathbf{R}}_{i1}^{a_i} : 0]$ where $\Delta\widetilde{\mathbf{R}}_{i1}^{a_i}$ has full column rank; e.g. this is a consequence of Theorem 6.2.4 in Mirsky (1955). Then $\mathbf{F}_i^{a_i}\begin{pmatrix}\theta \\ \beta\end{pmatrix}$ in (23) becomes $\left[\Delta\widetilde{\mathbf{R}}_{i1}^{a_i} : 0 : \Delta\mathbf{H}_i^{a_i}(\mathbf{R}_{i0}+\boldsymbol{\epsilon}_i)\right]\begin{pmatrix}\mathbf{W}^{-1}\theta \\ \beta\end{pmatrix}$. Therefore, by inspection, the proof of Theorem 4 can be readily adapted by reparameterizing θ to show the identification of the discount factor is possible.

References

[1] Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcome," *Handbook of Econometrics*, vol. 6, eds. J. Heckman and E. Leamer. North-Holland.

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

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

[4] Arcidiacono, P. and R.A. Miller (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, **79**, 1823-1867.

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

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

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

[8] Ching, A. and M. Osborne (2017): "Identification and Estimation of Forward-looking Behavior: The Case of Consumer Stockpiling," *Working Paper, University of Toronto.*

[9] Dubé, J.P., G. Hitsch, and P. Jindal (2014): "The Joint Identification of Utility and Discount Functions from Stated Choice Data: an Application to Durable Goods Adoption," *Quantitative Marketing and Economics*, **12**, 331–377.

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

[11] Fang, H. and Y. Wang (2015): "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions," *International Economic Review*, **56**, 565-596.

[12] Hong, H., and M. Shum (2009): "Pairwise-Difference Estimation of a Dynamic Optimization Model," *Review of Economic Studies*, **77**, 273-304.

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

[14] 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.

[15] Kalouptsidi, M. (2014): "Time to Build and Fluctuations in Bulk Shipping," *American Economic Review*, **104**, 564-608.

[16] Kalouptsidi, M., P. Scott and E. Souza-Rodrigues (2016a): "On the Non-Identification of Counterfactuals in Dynamic Discrete Games", *forthcoming in International Journal of Industrial Organization.*

[17] Kalouptsidi, M., P. Scott and E. Souza-Rodrigues (2016b): "Identification of Counterfactuals in Dynamic Discrete Choice Models", NBER Working Paper 21527.

[18] Kasahara, H. and K. Shimotsu (2008): "Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models," *Journal of Econometrics*, **146**, 92-106.

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

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

[21] McFadden, D.L. (1974). Conditional logit analysis of qualitative choice behavior. Frontier in econometrics, New York.

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

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

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

[25] Mirsky, L. (1955). An Introduction to Linear Algebra. Oxford University Press.

[26] Myśliwski, M., F. Sanches, D. Silva Junior and S. Srisuma (2017): "Dynamic Pricing and Consumer Loyalty in a Multiproduct Oligopoly," *Working Paper, University College London.*

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

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

[29] 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.

[30] Pesendorfer, M. (2010): "Estimation of (Dynamic) Games: A Discussion," in Advances in Economics and Econometrics, Tenth World Congress, Volume 3, Econometrics, Cambridge University Press, 2013.

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

[32] Rust, J., 1987, Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica, 55, 999 - 1033.

[33] Rust, J., 1994, Estimation of dynamic structural models: Problems and prospects part I: Discrete decision processes, in: C.A. Sims, (Eds), Proceedings of the 6th World Congress of the Econometric Society, Cambridge University Press, pp. 119-171.

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

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

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

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

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

[39] Wang, Y. (2014): "Dynamic Implications of Subjective Expectations: Evidence from Adult Smokers," *American Economic Journal: Applied Economics Economics*, **6**, 1-37.