Journal of Econometrics 236 (2023) 105461

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Dynamic discrete choice models with incomplete data: Sharp identification[☆]

Yuya Sasaki [a], Yuya Takahashi [b,*], Yi Xin [c], Yingyao Hu [d]

[a] Department of Economics, Vanderbilt University, 415 Calhoun Hall, Nashville, TN 37240, United States of America
[b] Department of Economics, the University of Washington, Savery Hall 329, Box 353330, Seattle, WA 98195, United States of America
[c] Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E California Blvd, MC 228-77, Pasadena, CA 91125, United States of America
[d] Department of Economics, Johns Hopkins University, Wyman Park Building 544E, 3400 N. Charles Street, Baltimore, MD 21218, United States of America

ARTICLE INFO

Article history:
Received 18 December 2021
Received in revised form 3 January 2023
Accepted 21 April 2023
Available online 21 June 2023

JEL classification:
C18

Keywords:
Dynamic discrete choice
Incomplete data
Industry dynamics
Partial identification
Sharp identification

ABSTRACT

In many empirical studies, those states that are relevant for forward-looking economic agents to make decisions may not be included in the data to which researchers have access. This problem often arises in the context of declining/booming industries. In this paper, we develop the sharp identified sets of structural parameters and counterfactuals for dynamic discrete choice models when empirical data do not cover realizations of relevant future states. Applying the proposed method to the annual *Toyo Keizai* database, we study the behaviors of Japanese firms on foreign direct investments in China without observing the future states after Chinese economy slows down.

1. Introduction

In many empirical studies, those states that are relevant for economic agents to make dynamic decisions may not be included in the data to which researchers have access.[1] For example, firms make investment decisions taking into account the declining/expanding phases of the industries (see for example Takahashi, 2015; Igami and Uetake, 2016; Igami, 2017), but the data may not cover the realizations of the industry dynamics.[2] In life-cycle models, individuals make saving and

☆ This paper was previously circulated as "Inference of Dynamic Discrete Choice Models under Incomplete Data Coverage of Relevant States". We are grateful to the editor (Elie Tamer), an anonymous associate editor, and anonymous referees. We benefited from discussions with Victor Aguirregabiria, Jeremy Fox, Hiro Kasahara, Katsumi Shimotsu, Elie Tamer, Elena Manresa, and comments and suggestions by seminar participants at University of British Columbia, University of Calgary, 2017 Asian Meeting of Econometric Society, the Structural Microeconometrics Workshop in 2015 MOVE-Barcelona GSE Summer Forum, and 2022 ASSA Annual Meeting.

* Corresponding author.
E-mail addresses: yuya.sasaki@vanderbilt.edu (Y. Sasaki), ytakahas@uw.edu (Y. Takahashi), yixin@caltech.edu (Y. Xin), yhu@jhu.edu (Y. Hu).

1 There are two possible reasons why this may occur: (1) the missing state is realized during the sample period but the econometrician does not observe it due to data limitations; (2) some states have never been visited within the sample period despite its recurrency if the sample period is short in monotone industries. Our approach is applicable to both cases.

2 For example, in booming industries, the markets may have experienced only the low demand states and the econometrician may not observe the high demand states in empirical data. Another concrete empirical example of missing relevant states can be found in Section 5 (Japanese firms' investment decisions in China).

https://doi.org/10.1016/j.jeconom.2023.04.005
0304-4076/

occupational choices taking into account their future income flows, but many panel survey datasets only cover a limited number of years sampled from the complete course of a lifetime.[3] Other issues in survey data, such as top-coding, may also prevent empirical researchers from accessing all states relevant to agents' decision-making problem. The incomplete coverage of relevant states in data induces asymmetry in information between economic agents and researchers. It is a ubiquitous source of lack of point identification of structural parameters, and poses serious empirical challenges for evaluating policy effects.

A commonly exercised solution to this issue is to use a parametric extrapolation of choice probabilities. With an extrapolation, the economist effectively "observes" decisions at all relevant states including those not covered in available data. While it is convenient, this approach may incur a large extrapolation bias as we demonstrate via simulations. In this paper, we provide a robust method that deals with incomplete data coverage of relevant states without relying on parametric extrapolation. Specifically, we characterize the sharp identified set of structural parameters for a class of dynamic discrete-choice models when the conditional choice probabilities (CCPs) are partially identified. At first glance, it may appear counter-intuitive that we can obtain informative bounds — an econometrician does not observe future states at all, and hence any astronomical payoffs in the unforeseen future could appear to make observational equivalence. However, we can exploit the dynamic structure. Namely, economic agents make the current decisions taking into account the future state transition probabilities and their future payoffs. Any astronomical payoffs in the future can thus translate into extreme choice probabilities by economic agents today, such as the near-zero or near-unit conditional probability of entry/exit. Current decisions that are observed by an econometrician, combined with the restrictions of a structural model, therefore, can serve as informative signals for the econometrician to construct informative bounds in the adverse circumstance of unforeseen future from the econometrician's viewpoint. As such, the problem that we face is certainly specific to dynamic models, but the informative solution is also owing to the dynamics of the model.

Our sharpness result is obtained by exploiting model restrictions in a similar spirit to Aguirregabiria and Mira (2002, 2007) and Kasahara and Shimotsu (2012). The intuition is as follows. For a given vector of state transition laws $\vec{g}$, the model imposes fixed-point restrictions that conditional choice probabilities $\vec{p}$ must satisfy. Such a set of CCPs is smaller than the set directly identified by observed data without structural restrictions. These fixed points yield the sharp identified set for $\vec{p}$. Evaluating the structural inversion at each point $\vec{p}$ in its sharp identified set in turn yields the sharp identified set of structural parameters.

We illustrate our methods using a dataset of Japanese firms' FDI decisions in China from 1990–2005. In the sample period, we do not observe states where China has moderate economic growth as a WTO member. However, the firms are likely to take into account the slowdown of the future economic growth rate when making entry/exit decisions. The monotonic trends featured in our application is related to recent empirical studies on industry dynamics. For example, Igami (2017) and Igami and Uetake (2016) study various aspects of the hard disk drive industry where product quality and efficiency of production keep improving. Takahashi (2015) studies firms' exit behavior in the movie theater industry where demand is declining in the long run. In all of these studies, the econometrician would need an extrapolation to compute *future* demand/payoff from the econometrician's viewpoint.

There are a number of related methodological papers. First, Norets and Tang (2014) analyze partially identified semi-parametric dynamic discrete choice models. The sources of partial identification are different between their setting and our setting. While the non-identification results from a relaxation of the distributional assumption in Norets and Tang (2014), the non-identification in our framework results from the inability to observe agents' choices in relevant states, which is a common issue in empirical data of booming and declining industries. Second, Arcidiacono and Miller (2020) consider (non-) identification of non-stationary dynamic discrete choice models in short panels where relevant states are not observed. Their work is motivated by a similar empirical issue to what motivates our study. While (Arcidiacono and Miller, 2020) provide exclusion restrictions and normalizations to overcome the under-identification, we propose a method of inference based on sharp partial identification without imposing such restrictions or normalizations. We will come back to this discussion with a concrete example later. Third, Kasahara and Shimotsu (2009), Arcidiacono and Miller (2011), Hu and Shum (2012), Sasaki (2015, Example 1), Hu and Sasaki (2018), Berry and Compiani (2020), Hu and Xin (2020), Otsu et al. (2020), Kalouptsidi et al. (2021), and Aguirregabiria et al. (2021), to list but a few, study identification and estimation of dynamic discrete choice models with unobserved states/choices. Their focuses are on different types of incomplete data coverage issues.

The rest of the paper is organized as follows. We describe the model and the incomplete data coverage issue in Section 2. The main theoretical partial identification results are derived in Section 3. We discuss the Monte Carlo simulation exercises and the empirical application in Sections 4 and 5, respectively. Section 6 concludes.

[3] For example, Khwaja (2010) estimates a dynamic life-cycle model of health insurance demand for individuals aged 22–80. The paper uses HRS data spanning 1991–1998, which contains individuals aged 51–61 in 1991–1992.

2. Model and incomplete data

We consider a single-agent dynamic decision problem in discrete time, $t = 1, \ldots, \infty$.[4] In each time period, an agent makes a binary choice[5] $a \in \{0, 1\}$ under states (x, ε), where x denotes a state variable that has a finite support $\{1, \ldots, \bar{x}\}$ and is observed by the econometrician, and ε denotes a vector of random payoff shocks that are not observed by the econometrician. The period payoff depends on the choice and states in the current period. Specifically, we assume additive separability of the deterministic payoff and the random shock:

$$\pi_{a,x} + \varepsilon_{a,x}. \tag{2.1}$$

For simplicity and following the literature, we assume that the private shock $\varepsilon_{a,x}$ independently follows the Type I Extreme Value (Gumbel) distribution:[6]

$$\varepsilon_{a,x} \overset{iid}{\sim} \text{Gumbel}(0, 1). \tag{2.2}$$

The state variable x evolves according to the first-order Markov process and the transition rule is denoted by

$$g_{x',a,x} \; = \; \Pr(X_{t+1} = x' \mid A_t = a, X_t = x),$$

where A_t and X_t denote the choice and the observable state, respectively, at period t. Note that this transition rule does not depend on t, and hence we assume time-homogeneous laws.[7] The observable state x may not yet be in the ergodic distribution at the beginning of the decision process, but the transition probability and conditional choice probabilities defined below do not depend on the calendar time.

Based on these primitives, an agent maximizes the sum of the discounted profits

$$\mathbb{E}\left[\sum_{t=1}^{\infty} \beta^{t-1}\left(\pi_{A_t,X_t} + \varepsilon_{A_t,X_t}\right)\right],$$

where $\beta < 1$ is the discount factor and the expectation is taken over the possible realizations of x and ε. We follow the convention in the literature to assume that β is known.[8] Let $d\,(a, x, \varepsilon)$ denote the optimal decision rule that equals to one if a is chosen when the state is (x, ε) and zero otherwise. By integrating out ε, we obtain the choice probability conditional on the observable state x, i.e., the conditional choice probability given by

$$p_{a,x} \; = \; \Pr(A_t = a \mid X_t = x) = \int d(a, x, \varepsilon) dF_{\varepsilon}.$$

The integrated value function V is obtained as the fixed point of the following equation:

$$V(x) \; = \; \sum_{a=0}^{1} p_{a,x} \left\{ \pi_{a,x} + \bar{\varepsilon} - \ln p_{a,x} + \beta \sum_{x'} g_{x',a,x} V(x') \right\},$$

where $\bar{\varepsilon} := \mathrm{E}[\varepsilon_{a,x}] \approx 0.577$ is the Euler constant under (2.2).

We consider a case where $g_{x',a,x}$ and $p_{a,x}$ are partially identified. (For the main result of deriving the sharpness, we further focus on the case where $g_{x',a,x}$ is point-identified but $p_{a,x}$ are partially identified as is the case with most applications.) With Δ^d denoting the d-dimensional simplex, let $\mathcal{G}_{a,x} \subseteq \Delta^{\bar{x}-1}$ and $\mathcal{P}_x \subseteq \Delta^1$ be the identified sets for the probability vectors $\vec{g}_{a,x} := \left(g_{1,a,x}, \ldots, g_{\bar{x},a,x}\right)$ and $\vec{p}_x := \left(p_{0,x}, p_{1,x}\right)$, respectively. They can be singletons as a special case, i.e., $\mathcal{G}_{a,x}$ and $\mathcal{P}_x$ are singletons if $\vec{g}_{a,x}$ and $\vec{p}_x$ are directly observed in data. On the other hand, they are the entire simplexes when the data do not cover the relevant states. We let the Cartesian products of the identified sets be denoted by $\mathcal{G} = \mathcal{G}_{0,1} \times \mathcal{G}_{1,1} \times \cdots \times \mathcal{G}_{0,\bar{x}} \times \mathcal{G}_{1,\bar{x}}$ and $\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_{\bar{x}}$.

Example 1 (*Dynamic Model of Entry and Exit*)**.** $X_t = (S_t, Z_t)$ consists of an endogenous state S_t and an exogenous state Z_t, where S_t is determined by the lagged action, i.e., $S_t = A_{t-1}$. Both A_t and S_t are supported on $\mathcal{A} = \mathcal{S} = \{0, 1\}$, and Z_t is supported on $\mathcal{Z} = \{1, \ldots, \bar{z}\}$, and thus $\bar{x} = |\mathcal{S}| \times |\mathcal{Z}| = 2 \cdot \bar{z}$. Specifically, $S_t = 1$ indicates that the firm is in the

[4] We focus on infinite horizon models throughout the paper. Our approach can also be applied to finite horizon models. Instead of characterizing the sharp identified set for the conditional choice probabilities through fixed point restrictions informed by the model, we can derive the restrictions that the CCPs need to satisfy using backward induction for finite horizon models.

[5] We focus on dynamic binary choice in the main text of this paper for ease of exposition, but the same principle extends to general multinomial models — see Online Appendix B for details.

[6] The assumption of this particular distribution is not crucial for our results, but we make this assumption following the common practice in the literature.

[7] This assumption of time homogeneous transition rules is not uncommon in the dynamic discrete choice literature and is often imposed in industrial organization and labor economics applications.

[8] See for example the identifiability discussions of β by Rust (1994) and Magnac and Thesmar (2002). A recent paper by Abbring and Daljord (2020) provides comprehensive identification results of the discount factor for dynamic discrete choice models.

market, and Z_t indicates the demand faced by the firm. If the industry is new in the sense that every market is in state $Z_t = 1$ at $t = 1$ and if the demand state increments by at most one at each time, then the markets have experienced only the low demand states, and an econometrician may not observe the high demand states $Z_t > T$ in empirical data available today at $t = T$. In this case, $\mathcal{P}_{(s,z)} = \{(1 - \mathrm{E}[A_t \mid S_t = s, Z_t = z], \mathrm{E}[A_t \mid S_t = s, Z_t = z])\}$ is a singleton for every $(s, z) \in \mathcal{S} \times \{1, \ldots, T\}$, but $\mathcal{P}_{(s,z)} = \Delta^1$ for every $(s, z) \in \mathcal{S} \times \{T + 1, \ldots, \bar{z}\}$. Likewise, $\mathcal{G}_{a,(s,z)}$ is a singleton if $z < T$, and is the simplex $\Delta^{2 \cdot \bar{z} - 1}$ otherwise. This yields set identification, as opposed to point identification, of $\mathcal{G}$ and $\mathcal{P}$. □

Remark 1. As emphasized earlier, we remark that we consider time-homogeneous g throughout, and that this time homogeneity of g is not incompatible with Example 1. To see this, consider a time-homogeneous transition rule $g(X_{t+1}|X_t)$ given by $g(1|1) > 0$, $g(2|1) > 0$, $g(3|1) = 0$, $g(1|2) = 0$, $g(2|2) > 0$, $g(3|2) > 0$, $g(1|3) = 0$, $g(2|3) > 0$, and $g(3|3) > 0$. Suppose that the initial marginal distribution at the genesis of an industry of X_1 is given by the mass $f_{X_1}(1) = 1$ and $f_{X_1}(2) = f_{X_1}(3) = 0$. Thus, the support of X_1 is the singleton $\{1\}$. The support of X_2 is $\{1, 2\}$ and the support of X_3 is $\{1, 2, 3\}$. If an econometrician collects data at the end of period $t = 2$ and has not observed $t = 3$, then the CCP and state transition rule given X_1 and X_2 are point identified but those given X_3 are unidentified. This simplified example illustrates that a time-homogeneous g is not incompatible with Example 1. □

Remark 2. We also emphasize that our problem in this paper concerns about a data generating process in the population in cross section, i.e., cross sectional sample size is infinite for the identification arguments. As illustrated in Remark 1, certain states (e.g., 3) may not be included in data if an econometrician collects data at the end of period $T = 2$ before a marginal distribution of X_t with a full support realizes. □

Remark 3. We emphasize that a state will enter in the calculations for rational forward-looking agents whenever that state is recurrent in the Markov chain. This feature is irrelevant to whether an econometrician observes those states in finite T before these recurrent states have been visited. Again, consider the simple example in Remark 1. State 3 is a recurrent state and hence it enters the calculation for rational forward-looking agents. This nature of the model is independent of the setting where an econometrician who has collected data at the end of period $t = 2$ does not observe state 3 in his/her data. □

Remark 4. We also emphasized that our discussion is not restricted to models with a macro-level exogenous state variable. For example, we can consider a quality-ladder model where it takes time for firms to accumulate the quality of their product. If firms need at least ten years to reach the highest quality level and the number of time periods in the data at hand is less than ten, then the researcher would not observe conditional choice probabilities when the quality of the product is at its maximum. □

Remark 5. We remark that in our paper the support of the unobserved state variable is assumed to be known by the econometricians. In Online Appendix D.1, we conduct simulations to examine how the estimated parameters would change (1) when alternative assumptions on the support of the state variable are imposed and (2) when the support is misspecified.

Under the Markov decision process, the Markov law of state–action transition can be written as

$$\Pr(A_{t+1} = a', X_{t+1} = x' \mid A_t = a, X_t = x) \;=\; p_{a',x'} \cdot g_{x',a,x}$$

Thus, we can write the Markov transition matrix for $\Pr(A_{t+1}, X_{t+1} \mid A_t, X_t)$ as a function of $(\vec{g}, \vec{p})$ by

$$M(\vec{g}, \vec{p}) \;=\; \begin{pmatrix} p_{0,1} \cdot g_{1,0,1} & p_{0,1} \cdot g_{1,1,1} & \cdots & p_{0,1} \cdot g_{1,0,\bar{x}} & p_{0,1} \cdot g_{1,1,\bar{x}} \\ p_{1,1} \cdot g_{1,0,1} & p_{1,1} \cdot g_{1,1,1} & \cdots & p_{1,1} \cdot g_{1,0,\bar{x}} & p_{1,1} \cdot g_{1,1,\bar{x}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{0,\bar{x}} \cdot g_{\bar{x},0,1} & p_{0,\bar{x}} \cdot g_{\bar{x},1,1} & \cdots & p_{0,\bar{x}} \cdot g_{\bar{x},0,\bar{x}} & p_{0,\bar{x}} \cdot g_{\bar{x},1,\bar{x}} \\ p_{1,\bar{x}} \cdot g_{\bar{x},0,1} & p_{1,\bar{x}} \cdot g_{\bar{x},1,1} & \cdots & p_{1,\bar{x}} \cdot g_{\bar{x},0,\bar{x}} & p_{1,\bar{x}} \cdot g_{\bar{x},1,\bar{x}} \end{pmatrix}$$

where $\vec{g} := (\vec{g}_{0,1}, \vec{g}_{1,1}, \ldots, \vec{g}_{0,\bar{x}}, \vec{g}_{1,\bar{x}})$ and $\vec{p} := (\vec{p}_1, \ldots, \vec{p}_{\bar{x}})$ for concise notations. The τ-th order transition matrix is $M(\vec{g}, \vec{p})^\tau$. Its element in row $a' + 2x' - 1$ and column $a + 2x - 1$ represents the τ-th order transition probability $\Pr(A_{t+\tau} = a', X_{t+\tau} = x' \mid A_t = a, X_t = x)$, and we denote it by

$$h^\tau_{a',x',a,x}(\vec{g}, \vec{p}) \;=\; M(\vec{g}, \vec{p})^\tau [a' + 2x' - 1 : a + 2x - 1]. \tag{2.3}$$

3. Partial identification and the sharpness

We are interested in partial identifying the structural parameters and counterfactual outcomes. By using the model restriction like Aguirregabiria and Mira (2002, 2007) and Kasahara and Shimotsu (2012), we derive the *sharp identified sets* for these parameters.

3.1. The identified sets

We summarize the payoff parameters by the $2\bar{x}$-dimensional vector $\pi := \left[\pi_{0,1}, \pi_{1,1}, \ldots, \pi_{0,\bar{x}}, \pi_{1,\bar{x}}\right]'$. Economic structures impose restrictions on π with primitive parameters, which we denote by $\theta \in \mathbb{R}^k$. Suppose that the following linear restriction equation holds for some $2\bar{x}$-by-k restriction matrix R, which is non-stochastic and known by the econometrician.

$$\pi = R\theta. \tag{3.1}$$

In particular, since the structural parameters $\pi_{a,x}$ are identified only up to unknown location, we normalize at least one of them, say $\pi_{0,0} \equiv 0$. This sort of normalizing restriction ought to be included as one of the restrictions in (3.1). In addition to the linear restriction (3.1), we maintain the traditional assumption that the true parameter θ_0 resides in a certain admissible set Θ of structural parameters.

Example 1 (*Dynamic Model of Entry and Exit, Continued*)**.** Consider Example 1 again. Let κ and ϕ denote the entry cost and the exit value, respectively. If π_z denotes the profit that the firm earns in the market with demand state $Z_t = z$, then $\pi_{a,x}$ is defined by $\theta = (\pi_1, \ldots, \pi_{\bar{z}}, \phi, \kappa)$ through

$$\pi_{a,(s,z)} = \begin{cases} 0 & \text{if } a = 0 \text{ and } s = 0 \\ \pi_z + \phi & \text{if } a = 0 \text{ and } s = 1 \\ -\kappa & \text{if } a = 1 \text{ and } s = 0 \\ \pi_z & \text{if } a = 1 \text{ and } s = 1 \end{cases}$$

for each $z \in \mathcal{Z}$. See Appendix A.1 in the appendix for how to construct R and Θ. □

To reflect the restriction (3.1) in our identifying formulas, we define the $\bar{x}$-by-k matrix $\tilde{H}(\vec{g}, \vec{p}, \beta)$ and the $\bar{x}$-dimensional vector $\tilde{Y}(\vec{g}, \vec{p}, \beta)$ by

$$\tilde{H}(\vec{g}, \vec{p}, \beta) = \begin{bmatrix} H(1; \vec{g}, \vec{p}, \beta)R \\ \vdots \\ H(\bar{x}; \vec{g}, \vec{p}, \beta)R \end{bmatrix} \quad \text{and} \quad \tilde{Y}(\vec{g}, \vec{p}, \beta) = \begin{bmatrix} Y(1; \vec{g}, \vec{p}, \beta) \\ \vdots \\ Y(\bar{x}; \vec{g}, \vec{p}, \beta) \end{bmatrix},$$

respectively, where $H(x; \vec{g}, \vec{p}, \beta)$ is the $2\bar{x}$-dimensional vector

$$H(x; \vec{g}, \vec{p}, \beta) := \begin{bmatrix} H_{0,1}(x; \vec{g}, \vec{p}, \beta) - \mathbb{1}\{x = 1\} \\ H_{1,1}(x; \vec{g}, \vec{p}, \beta) + \mathbb{1}\{x = 1\} \\ \vdots \\ H_{0,\bar{x}}(x; \vec{g}, \vec{p}, \beta) - \mathbb{1}\{x = \bar{x}\} \\ H_{1,\bar{x}}(x; \vec{g}, \vec{p}, \beta) + \mathbb{1}\{x = \bar{x}\} \end{bmatrix}'$$

and $Y(x; \vec{g}, \vec{p}, \beta)$ is the scalar

$$\begin{aligned} Y(x; \vec{g}, \vec{p}, \beta) := \sum_{x'=1}^{\bar{x}} \Big[& \big(H_{1,x'}(x; \vec{g}, \vec{p}, \beta) + \mathbb{1}\{x = x'\}\big) \cdot \ln p_{1,x'} \\ & + \big(H_{0,x'}(x; \vec{g}, \vec{p}, \beta) - \mathbb{1}\{x = x'\}\big) \cdot \ln p_{0,x'} \\ & - \big(H_{1,x'}(x; \vec{g}, \vec{p}, \beta) + H_{0,x'}(x; \vec{g}, \vec{p}, \beta)\big) \cdot \bar{\varepsilon} \Big] \end{aligned}$$

with $H_{a',x'}(x; \vec{g}, \vec{p}, \beta) := \sum_{\tau=1}^{\infty} \beta^\tau \left(h^\tau_{a',x',1,x}(\vec{g}, \vec{p}) - h^\tau_{a',x',0,x}(\vec{g}, \vec{p})\right)$ for each x, x', and a'. With these short-hand notations, given the vectors $(\vec{g}, \vec{p})$ of transition probabilities and CCPs, we state the restrictions of and solution to the structural parameters θ as follows.

Lemma 1. *(i) If $\vec{p}$ is generated from the model with structural parameters θ and transition probabilities $\vec{g}$, then*

$$\left[\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{H}(\vec{g}, \vec{p}, \beta)\right]\theta = \left[\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{Y}(\vec{g}, \vec{p}, \beta)\right] \tag{3.2}$$

holds. (ii) If, in addition, the rank condition

$$Rank\left(\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{H}(\vec{g}, \vec{p}, \beta)\right) = k \tag{3.3}$$

is satisfied, then the equality $\theta = \vartheta(\vec{g}, \vec{p})$ holds where

$$\vartheta(\vec{g}, \vec{p}) = \left[\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{H}(\vec{g}, \vec{p}, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{Y}(\vec{g}, \vec{p}, \beta)\right]. \tag{3.4}$$

Part (ii) of this lemma is already well known in the literature, e.g., Hotz et al. (1994) – also see Aguirregabiria and Mira (2002), Pesendorfer and Schmidt-Dengler (2008), Norets and Tang (2014), Sanches et al. (2016), Hu and Sasaki (2018), Buchholz et al. (2021), and Kalouptsidi et al. (2020b). The statement of part (i) on the other hand is new to our knowledge, although it follows on the way to proving part (ii). Part (i) paves the way for characterizing identified sets of structural parameters in the absence of point identification.[9] We state and prove part (ii) as well as part (i) for completeness and for convenience of readers. We also remark that the rank condition invoked for part (ii) is analogous to the rank condition required by Pesendorfer and Schmidt-Dengler (2008) – we refer readers to Pesendorfer and Schmidt-Dengler (2008) and Buchholz et al. (2021) for discussions and intuitions of this condition. With Lemma 1, we can narrow down the structural parameters θ by evaluating (3.2) at various points of $(\vec{g}, \vec{p})$ in a set $\mathcal{GP} \subset \mathcal{G} \times \mathcal{P}$ that is consistent with the observed data and relevant restrictions as formally stated in the following theorem.

Theorem 1. *(i) Suppose that the current-time payoff is given by (2.1) with (2.2), $\theta_0 \in \Theta$, and $\beta \in (0, 1)$, then*

$$\Theta_I \;=\; \left\{ \theta \in \Theta \mid \left[\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{H}(\vec{g}, \vec{p}, \beta)\right]\theta = \left[\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{Y}(\vec{g}, \vec{p}, \beta)\right] \text{ and } (\vec{g}, \vec{p}) \in \mathcal{GP} \right\}$$

is an identified set of the structural primitive parameters θ. (ii) If, in addition, $\mathcal{G} = \{\vec{g}_0\}$ is a singleton and the rank condition (3.3) is satisfied for $\vec{g} = \vec{g}_0$ and for all $\vec{p} \in \mathcal{P}$, then Θ_I is written as

$$\Theta_I \;=\; \left\{ \vartheta(\vec{g}_0, \vec{p}) \;\middle|\; \vec{p} \in \mathcal{P} \right\} \cap \Theta.$$

(iii) If, in addition, $\mathcal{GP} \subset \mathcal{G} \times \mathcal{P}$ is the sharp identified set for $(\vec{g}, \vec{p})$, then so is Θ_I for θ.

A proof is given in Appendix A.4 in the appendix. Note that the basic identification result of Theorem 1 does not use any dynamic model information to restrict the set $\mathcal{GP} \subset \mathcal{G} \times \mathcal{P}$. The next subsection proposes a way to construct the sharp identified set $\mathcal{GP}$ for $(\vec{g}, \vec{p})$ when $\mathcal{G} = \{\vec{g}_0\}$ is a singleton, and thus the sharp identified set Θ_I for θ by virtue of Theorem 1 (iii).

3.2. The sharp identified sets

In this section, we focus on the case where $\mathcal{G} = \{\vec{g}_0\}$ is a singleton, which holds under mild and standard assumptions in the literature of empirical industrial organization such as the time-homogeneous incremental/decremental state transition probabilities. For example, the standard incremental state-transition model, $P(X_{t+1} - X_t = 0|A_t = a, X_t = x) = \rho_0(a)$, $P(X_{t+1} - X_t = 1|A_t = a, X_t = x) = \rho_1(a)$, and $P(X_{t+1} - X_t = 2|A_t = a, X_t = x) = \rho_2(a)$, with t- and x-invariant $(\rho_0(a), \rho_1(a), \rho_2(a))$, which is commonly used in the literature of monotone industry as well as mileages run (regardless of complete or incomplete data), allows us to identify $(\rho_0(a), \rho_1(a), \rho_2(a))$ and thus $\vec{g}$ from just two time periods of data without having to observe future states yet to observe in currently available data (see Rust (1987)). Unlike an extrapolation of the period utility, this type of time-homogeneous incremental/decremental state transition (which effectively extrapolates $\vec{g}$ to non-visited states) is not so objectionable and has therefore been commonly assumed in many applications in the literature. Hence we proceed with this model setting in the current subsection.

Theorem 1 claims that the identified set Θ_I for the structural parameters θ is sharp provided that the identified set $\mathcal{GP} \subset \{\vec{g}_0\} \times \mathcal{P}$ for the state transition probabilities and the CCPs is sharp. We propose a way to construct the sharp identified set $\mathcal{GP}$ for $(\vec{g}, \vec{p})$ by using the structural restrictions in a similar manner to Aguirregabiria and Mira (2002, 2007) and Kasahara and Shimotsu (2012). Consequently, we also propose how to obtain the sharp identified set of the structural parameters as well. The model restrictions narrow the identified set for the CCPs, $\vec{p}$, because the CCPs are the structural consequences of endogenous behaviors prescribed by the model restrictions. In particular, we use the fact that the structure provides the following additional restriction.

Lemma 2 (*Restrictions*). *Suppose that the current-time payoff is given by (2.1) with (2.2), $\beta \in (0, 1)$, and the rank condition (3.3) is satisfied for $\vec{g} = \vec{g}_0$ and for all $\vec{p} \in \mathcal{P}$. Then, the true CCPs $\vec{p} \in \mathcal{P}$ satisfy the restriction*

$$p_{1,x} \;=\; \frac{\exp\left\{\Lambda_{1,x}\left(R\vartheta(\vec{g}_0, \vec{p}), \vec{g}_0, \vec{p}, \beta\right)\right\}}{1 + \exp\left\{\Lambda_{1,x}\left(R\vartheta(\vec{g}_0, \vec{p}), \vec{g}_0, \vec{p}, \beta\right)\right\}}$$

for each $x \in \{1, \ldots, \bar{x}\}$, where $\Lambda_{1,x}(\pi, \vec{g}, \vec{p}, \beta)$ is defined by

$$\begin{aligned}
&\Lambda_{1,x}(\pi, \vec{g}, \vec{p}, \beta) \;=\; \pi_{1,x} - \pi_{0,x} + \\
&\sum_{\tau=1}^{\infty}\sum_{x'=1}^{\bar{x}} \beta^{\tau} \cdot \left[h^{\tau}_{1,x',1,x}(\vec{g}, \vec{p}) \cdot \left(\pi_{1,x'} + \bar{\varepsilon} - \ln p_{1,x'}\right) + h^{\tau}_{0,x',1,x}(\vec{g}, \vec{p}) \cdot \left(\pi_{0,x'} + \bar{\varepsilon} - \ln p_{0,x'}\right)\right] - \\
&\sum_{\tau=1}^{\infty}\sum_{x'=1}^{\bar{x}} \beta^{\tau} \cdot \left[h^{\tau}_{1,x',0,x}(\vec{g}, \vec{p}) \cdot \left(\pi_{1,x'} + \bar{\varepsilon} - \ln p_{1,x'}\right) + h^{\tau}_{0,x',0,x}(\vec{g}, \vec{p}) \cdot \left(\pi_{0,x'} + \bar{\varepsilon} - \ln p_{0,x'}\right)\right]
\end{aligned}$$

[9] Note that Eq. (3.2) can be replaced with $\tilde{H}(\vec{g}, \vec{p}, \beta)\,\theta \;=\; \tilde{Y}(\vec{g}, \vec{p}, \beta)$. This equation and Eq. (3.2) contain the same information about θ. Relatedly, the identified set in part (i) of Theorem 1 can be equivalently expressed as $\Theta_I \;=\; \left\{\theta \in \Theta \mid \tilde{H}(\vec{g}, \vec{p}, \beta)\theta = \tilde{Y}(\vec{g}, \vec{p}, \beta) \text{ and } (\vec{g}, \vec{p}) \in \mathcal{GP}\right\}$.

A proof is given in Appendix A.5 in the appendix. This lemma implies that, given the true transition probabilities $\vec{g}_0$, the true CCPs $\vec{p}$ can be characterized as a fixed point of the self map $\Psi_{\vec{g}_0} : \mathcal{P} \to \mathcal{P}$ defined by

$$\Psi_{\vec{g}_0}(\vec{p}) \;=\; \left[\; \left[\; \begin{array}{c} \frac{1}{1+\exp\left\{\Lambda_{1,1}\left(R\vartheta(\vec{g}_0,\vec{p}),\vec{g}_0,\vec{p},\beta\right)\right\}} \\ \frac{\exp\left\{\Lambda_{1,1}\left(R\vartheta(\vec{g}_0,\vec{p}),\vec{g}_0,\vec{p},\beta\right)\right\}}{1+\exp\left\{\Lambda_{1,1}\left(R\vartheta(\vec{g}_0,\vec{p}),\vec{g}_0,\vec{p},\beta\right)\right\}} \end{array} \;\right]' \;\cdots\; \left[\; \begin{array}{c} \frac{1}{1+\exp\left\{\Lambda_{1,\bar{x}}\left(R\vartheta(\vec{g}_0,\vec{p}),\vec{g}_0,\vec{p},\beta\right)\right\}} \\ \frac{\exp\left\{\Lambda_{1,\bar{x}}\left(R\vartheta(\vec{g}_0,\vec{p}),\vec{g}_0,\vec{p},\beta\right)\right\}}{1+\exp\left\{\Lambda_{1,\bar{x}}\left(R\vartheta(\vec{g}_0,\vec{p}),\vec{g}_0,\vec{p},\beta\right)\right\}} \end{array} \;\right]' \;\right]'.$$

Aguirregabiria and Mira (2002, 2007) and Kasahara and Shimotsu (2012) exploit this additional model restriction as means of estimation and inference. We use a similar idea to shrink the identified set to the sharp one. Consider the set $\mathcal{P}(\vec{g}_0)$ defined by

$$\mathcal{P}(\vec{g}_0) \;:=\; \left\{\vec{p} \in \mathcal{P} \;\middle|\; \vec{p} = \Psi_{\vec{g}_0}(\vec{p}),\, \vartheta(\vec{g}_0,\vec{p}) \in \Theta \right\}.$$

Under the setting in which the true transition probabilities $\vec{g}_0$ are known, i.e., $\mathcal{G} = \{\vec{g}_0\}$, the set $\mathcal{P}$ of CCPs can be shrunk to the sharp set $\mathcal{P}(\vec{g}_0)$ as formally stated in the following theorem.

Theorem 2 (*Sharp Identified Set of $\vec{p}$*)**.** *Suppose that the current-time payoff is given by (2.1) with (2.2), $\beta \in (0,1)$, $\mathcal{G} = \{\vec{g}_0\}$, and the rank condition (3.3) is satisfied for $\vec{g} = \vec{g}_0$ and for all $\vec{p} \in \mathcal{P}$. Then,*

$$\mathcal{GP}^{\dagger} := \left(\{\vec{g}_0\} \times \mathcal{P}(\vec{g}_0)\right) \tag{3.5}$$

is the sharp identified set of $(\vec{g}, \vec{p})$.

A proof is given in Appendix A.6. Consequently, the identified set Θ_I (Theorem 1) constructed from this sharp identified set $\mathcal{GP} = \mathcal{GP}^{\dagger}$ is the sharp identified set of structural parameters θ.

Corollary 1 (*Sharp Identified Set of θ*)**.** *Suppose that the current-time payoff is given by (2.1) with (2.2), $\beta \in (0,1)$, $\mathcal{G} = \{\vec{g}_0\}$, and the rank condition (3.3) is satisfied for $\vec{g} = \vec{g}_0$ and for all $\vec{p} \in \mathcal{P}$. If $\theta_0 \in \Theta$, then the sharp identified set $\Theta_I^{\dagger}$ of the structural primitive parameters θ is given by*

$$\Theta_I^{\dagger} \;=\; \left\{\vartheta(\vec{g}_0,\vec{p}) \;\middle|\; (\vec{g}_0,\vec{p}) \in \mathcal{GP}^{\dagger}\right\}.$$

3.3. Identified sets for counterfactual outcomes

In structural econometric analysis, the objects of interest are not necessarily the structural parameters *per se*. Instead, researchers often use the identified structural parameters to make inference about counterfactual outcomes.[10] In this section, we remark that our partial identification result for the structural parameters from the previous subsection straightforwardly extends to partial identification of counterfactuals.

Suppose that a scalar-valued counterfactual policy outcome C is computed using the structural primitive parameters θ by

$$C = \Gamma(\theta, \vec{g}, \vec{p}).$$

We can obtain its bounds as a direct consequence of Theorem 1 as follows.

Corollary 2 (*Bounds of Counterfactual Outcomes*)**.** *Suppose that the current-time payoff is given by (2.1) with (2.2), $\beta \in (0,1)$, $\mathcal{G} = \{\vec{g}_0\}$, and the rank condition (3.3) is satisfied for $\vec{g} = \vec{g}_0$ and for all $\vec{p} \in \mathcal{P}$. The identified set $\mathcal{C}_I$ of the counterfactual outcome C is given by*

$$\left\{\Gamma\left(\vartheta(\vec{g},\vec{p}),\vec{g},\vec{p}\right) \;\middle|\; (\vec{g},\vec{p}) \in \mathcal{GP}\right\}.$$

If $\mathcal{GP}$ is the sharp identified set for $(\vec{g},\vec{p})$, then so is $\mathcal{C}_I$ for C.

A proof is given in Appendix A.7 in the appendix. By the last sentence of this corollary, the sharpness of this identified set also follows from Corollary 1 by using $\mathcal{GP} = \mathcal{GP}^{\dagger}$ defined in (3.5). If Γ is continuous and the counterfactual outcome is scalar-valued, then the identified set $\mathcal{C}_I$ is guaranteed to be an interval even if the counterfactual outcome map Γ is highly nonlinear — See Proposition 2 in Appendix A.8 in the appendix.

10 See for example Norets and Tang (2014) and Kalouptsidi et al. (2020a).

4. Simulation

4.1. Setup and baseline results

Let us revisit the dynamic model of entry and exit introduced in Example 1. For simplicity, suppose that there are $\bar{z} = 3$ exogenous states and an econometrician observes $T = 2$ time periods of dynamic decisions.[11] That is, a researcher does not observe CCPs when $(S_t, Z_t) = (0, 3)$ and $(S_t, Z_t) = (1, 3)$.[12,13] The transition law for the exogenous state variable Z_t is specified by the Markov matrix

$$\begin{pmatrix} 1-\lambda_1 & \lambda_1 & 0 \\ 0 & 1-\lambda_2 & \lambda_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

This matrix describes an increasing industry, where the state advances from 1 to 2 with probability λ_1, and advances from 2 to 3 with probability λ_2. Once the state with $Z_t = \bar{z}$ is reached, the industry will stay there with probability one.

We assume that the deterministic period payoff consists of two parts. The first part depends on the current state variable only. An example is the operating flow profit earned this period. The second part depends on the previous state variables and the firm's action. Specifically, if a firm was not active in the previous period but decides to be active, the firm incurs the entry cost of κ. Furthermore, if a firm was active in the previous period but decides to exit the market, the firm collects the exit value of ϕ. We set the exit value to $\phi = 0$ and assume that a researcher knows its value throughout this simulation exercise. We set the other structural payoff parameters as follows.

$$\kappa = 20 \qquad (\pi_1, \pi_2, \pi_3) = (2.5, 4.0, 6.0).$$

In terms of the state transition rules, we first consider the case where $\lambda_1 = \lambda_2 = 0.5$. That is, the probability that Z_t advances from 1 to 2 equals to the probability that it advances from 2 to 3. Thus, the econometrician can infer the latter probability from the data, even though he/she observes only $T = 2$ time periods. All results reported in Sections 4.1–4.3 are based on the specification where $\lambda_1 = \lambda_2$ and λ's are assumed known by the econometrician. In Section 4.4, we perform a Monte Carlo analysis for the case in which $\lambda_1 \neq \lambda_2$, so that λ_2 is only set-identified.[14] Throughout this simulation exercise, we assume that $\varepsilon_{a,(s,z)}$ follows the Gumbel distribution with the scaling parameter of 10. Finally, we impose the monotonicity restriction as described in Example 1.

We provide details of the criterion-based approach to estimating the sharp identified set in Appendix A.9.[15] Monte Carlo simulation results based on 2000 iterations are summarized in Table 1 for each of the sample sizes $N = 1000$, 3000, 5000 and 10000. Since the projected identified set is an interval (see Appendix A.8 in the appendix for details), we focus on the lower and upper bounds. The table lists the Monte Carlo means of the bounds for the payoff parameters, and their standard deviations in parentheses.

For each sample size, the true value of each parameter is located between the mean lower bound and the mean upper bound. Overall, our method gives reasonably tight bounds for the structural parameters with the sample size in typical empirical applications ($N = 3,000$ or $N = 5,000$). As the sample size increases, the lower bound (respectively, the upper bound) increases (respectively, decreases) to the direction of the true parameter value. Even at a very large sample size, the lower and upper bounds do not converge to the true parameter values as we would expect, since the model is not point identified.

We also produce the lower and upper bounds for the CCPs. We find that the lower and upper bounds coincide for the CCPs corresponding to $(S_t, Z_t) \in \{(0, 1), (0, 2), (1, 1), (1, 2)\}$. For the states that are not observed in the data, i.e., $(S_t, Z_t) \in \{(0, 3), (1, 3)\}$, our approach does a good job shrinking the bounds of the CCPs through the fixed point representation. More detailed discussions can be found in Online Appendix E.

4.2. Sharp identified set

In theory, the set of the maxima of the likelihood function should coincide with our identified set in a large sample — see Tamer (2010) and Chen et al. (2011). Fig. 1 in Appendix A.10 plots likelihood values over parameter values. The

11 Our approach requires that the econometrician specifies the support of the state variables ex-ante. To check how sensitive are the estimated parameters to the alternative assumptions on the support of the state variable, we perform a Monte Carlo analysis for the case in which $\bar{z} = 4$, and the econometrician only observes $T = 2$ time periods of dynamic decisions. We also compare the bound estimates for the structural parameters with the estimates when the support of the state variable is misspecified. The details of these Monte Carlo simulations are provided in Appendix D.1.

12 Given this data availability, if we normalize π_2, then π_1 is point-identified (see Arcidiacono and Miller, 2020). This result is useful when a researcher is not interested in the value of π_3 and only the relative value π_1/π_2 matters.

13 We also consider another set of Monte Carlo simulation exercises where the econometrician does not observe $Z_t = 2$. The bound estimates for the structural parameters under this specification are provided in Panel (A) of Table D.1 in Online Appendix D.1.

14 When $\lambda_1 \neq \lambda_2$ and the econometrician only observes $T = 2$ time periods of dynamic decisions, we cannot identify the probability that the state advances from 2 to 3 in the data (as state 3 has not yet realized), thus λ_2 is only partially identified.

15 In the implementation, we chose the tuning parameter $\varkappa = 0.00001$.

Table 1
Monte Carlo simulation results based on 2000 iterations. $\lambda_1 = \lambda_2 = 0.5$, and the value of λ's are assumed known in the simulation exercise. The displayed numbers for the lower and upper bounds are the Monte Carlo means. The numbers in parentheses indicate the standard deviations.

N		True	Lower bound		Upper bound	
1,000	κ	20.000	17.206	(1.420)	23.205	(1.828)
	π_1	2.500	−0.552	(1.279)	3.721	(0.589)
	π_2	4.000	2.678	(0.522)	6.009	(0.625)
	π_3	6.000	4.560	(0.634)	8.184	(1.207)
3,000	κ	20.000	17.997	(0.843)	22.186	(0.974)
	π_1	2.500	−0.004	(0.773)	3.629	(0.371)
	π_2	4.000	2.868	(0.316)	5.747	(0.378)
	π_3	6.000	4.706	(0.391)	7.665	(0.723)
5,000	κ	20.000	18.328	(0.643)	21.810	(0.750)
	π_1	2.500	0.196	(0.634)	3.570	(0.303)
	π_2	4.000	2.938	(0.262)	5.653	(0.301)
	π_3	6.000	4.787	(0.315)	7.447	(0.575)
10,000	κ	20.000	18.721	(0.495)	21.338	(0.552)
	π_1	2.500	0.473	(0.473)	3.485	(0.219)
	π_2	4.000	3.021	(0.199)	5.536	(0.218)
	π_3	6.000	4.924	(0.233)	7.167	(0.423)

four graphs display the profiled plots over κ, π_1, π_2, and π_3 from the top to the bottom. Each gray point corresponds to a point that is collected by the MCMC algorithm for the exercise in Section 4.1 with the tuning parameter $\varkappa = 1$ and sample size $N = 1,000,000$.[16] The vertical lines indicate the true parameter values. Among 100,000 points collected by the MCMC algorithm, the bottom one percentile in terms of our criterion function $Q(\vec{g}, \vec{p})$ (see the construction of the criterion function in Appendix A.9) is highlighted in black. These black points are roughly what we would collect by the MCMC algorithm with a much smaller value of $\varkappa$ as in Footnote 4.1. We can see from Fig. 1 that the region of the black dots exactly coincides with the region of maximum likelihood value.

We also look at the sharp identified set in two dimensions. The detailed discussions are provided in Online Appendix C.

4.3. Identified set and logit extrapolation

In empirical applications, it is often the case that part of relevant states is not observed in data. A common practice in the literature is to impose a parametric restriction (such as logit) on CCPs and interpolate/extrapolate for state variables that are not observed in data. This subsection investigates consequences of such a parametric restriction in our context, namely, when CCPs are extrapolated for states that have not been reached.[17]

As above, we assume $\bar{z} = 3$ and $T = 2$. To focus on the identification issue setting aside sampling variations, we continue to use $N = 1,000,000$. We use the following logit model for CCPs:

$$a_{it} = \mathbb{1}\{\alpha_0 + \alpha_1\sqrt{z_{it}} + \alpha_2 s_{it} + \varepsilon_{it}^1 > \varepsilon_{it}^0\}, \tag{4.1}$$

where $\left(\varepsilon_{it}^0, \varepsilon_{it}^1\right)$ follows the i.i.d. Type I Extreme Value distribution.[18] After estimating $(\alpha_0, \alpha_1, \alpha_2)$ by maximum likelihood, we compute the CCPs for all observed and unobserved states $(z, s) \in \{1, 2, 3\} \times \{0, 1\}$. For the sake of comparisons, we also estimate (4.1) using a linear term $\alpha_1 z_{it}$ instead of $\alpha_1\sqrt{z_{it}}$.

Table 2 shows simulation results for four different parameterizations (cases 1 through 4). Let us first focus on comparisons between our method and the model with $\alpha_1\sqrt{z_{it}}$ (second last column). Case 1 uses the same set of parameters as the base case, confirming our discussion above that the parameters are not point identified. In this case, the parameters obtained by the logit model do not converge to the true value, as it is misspecified. However, it performs reasonably well. This may be because (π_1, π_2, π_3) align in a somewhat linear fashion. We change the degree of non-linearity of the payoff function and investigate how our method and the logit model perform. In Case 2, π changes in a convex fashion. While our identified set contains the true values and gives sharp bounds, the logit model performs surprisingly well. On the

16 In this exercise, we use the large sample size so that we can focus on the identification issue while setting aside sampling variations. Setting the tuning parameter to a large value helps to pick up parameter values that lie outside of the identified set.

17 In Online Appendix D.2, we perform a Monte Carlo analysis when the econometricians do not observe realizations when $Z_t = 2$ and the CCPs for the missing state are interpolated. We consider two different logit models for interpolation; the estimation results are reported in Panel (B) in Table D.2.

18 In the generated dataset, z takes only two values; i.e., $z = 1$ or $z = 2$. We cannot use both of linear and quadratic terms for z since the coefficients for z and z^2 cannot be separately identified. We try using $\alpha_1 z$, $\alpha_1\sqrt{z}$, and $\alpha_1 z^2$ and find that $\alpha_1\sqrt{z}$ has the best performance in terms of the bias. We also try including an interaction term between z and s in the logit model and find that the results do not differ much.

Table 2
Monte Carlo simulation results to compare our bounds with point estimates using logit extrapolation. The results are based on 2000 iterations. To ignore the effect of sampling variation, we set $N = 1,000,000$.

		True values	Sharp identified set		Logit	Logit
			Lower bound	Upper bound	with $\alpha_1\sqrt{z_{it}}$	with $\alpha_1 z_{it}$
Case 1	κ	20.000	19.532	20.254	20.000	20.000
	π_1	2.500	1.193	3.259	3.173	3.752
	π_2	4.000	3.242	5.251	3.327	2.748
	π_3	6.000	5.222	6.507	6.422	6.791
Case 2	κ	20.000	19.592	20.270	19.999	19.999
	π_1	2.500	−1.144	2.739	2.574	3.691
	π_2	3.000	2.738	6.609	2.926	1.809
	π_3	9.000	6.635	9.236	9.049	9.827
Case 3	κ	20.000	19.711	20.321	19.999	19.999
	π_1	2.500	1.866	5.245	5.018	5.962
	π_2	8.000	5.247	8.609	5.482	4.539
	π_3	9.000	8.579	10.990	10.774	11.464
Case 4	κ	20.000	19.678	20.301	20.000	20.000
	π_1	0.000	−0.634	5.986	5.353	7.148
	π_2	12.000	6.059	12.578	6.648	4.853
	π_3	13.000	12.534	17.835	17.301	18.823

other hand, a different picture emerges in Case 3, when π changes in a concave fashion. Above all, the shape of the profit function estimated by logit exhibits strong convexity, which is opposite to the true shape. This bias becomes severe when the degree of concavity becomes higher (see Case 4). This is interesting given that we are using a concave function of z in the reduced-form CCP function in (4.1). That is, even if a researcher has knowledge about the shape of payoff function (e.g., concave in an observable variable), it would not help the researcher pick an appropriate functional form for the CCP estimation.

Table 2 also reports the logit extrapolation with a linear term $\alpha_1 z_{it}$ instead of $\alpha_1\sqrt{z_{it}}$ (last column). The linear model performs worse than the original logit model. In particular, the monotonicity of π is violated, even though the CCP is modeled to be monotonic in z. This illustrates the difficulty of imposing a meaningful restriction on primitives by imposing a parametric restriction on CCPs.

Finally, we consider a case in which the entry cost depends on z. In Example 1, a researcher may want to estimate $(\kappa_1, \kappa_2, \kappa_3)$ separately instead of a single value κ. If the major part of entry costs is the cost of land acquisition, it is natural that the cost of entry changes with demand or growth rates. In theory, the model is still point identified if CCPs are observed in all possible states. However, when CCPs are partially observed (*i.e.*, $T = 2$), extrapolation performs poorly. When z changes, so do π and entry costs, both of which change value functions. Therefore, it is difficult even for a flexible function of z in the CCP estimation to fully capture the effect of z on the value function.[19]

4.4. Unknown state transition rules

We have thus far focused on simulation exercises with $\lambda_1 = \lambda_2$, so that the state transition rules can be inferred from the data even if the econometrician observes $T = 2$ time periods of dynamic decisions. In this section, we consider a scenario in which $\lambda_1 \neq \lambda_2$. In such a case, the state transition rules are unknown to the econometrician and λ_2 is unidentified.[20]

In the current simulation exercise, we set $\lambda_1 = 0.6$ and $\lambda_2 = 0.7$ and consider two strategies to address the issue that λ_2 is only set-identified. First, we impose a parametric restriction on the state transition probabilities and make an extrapolation for state variables that are not observed in the data. Specifically, we use the following logit model for the transition rule of Z_t:

$$Z_{it+1} = Z_{it} + \mathbb{1}\{\gamma Z_{it} + \varepsilon^1_{it} > \varepsilon^0_{it}\}, \qquad \text{if } Z_{it} \leq 2 \tag{4.2}$$

where $(\varepsilon^0_{it}, \varepsilon^1_{it})$ follows the i.i.d. Type I Extreme Value distribution. Second, we estimate the bounds for λ_2 together with the payoff parameters. This corresponds to the case where we draw $\vec{g}$ along with $\vec{p}$, as outlined in Appendix A.9, since λ_2

19 In this case, the sharp identified set also gives wide bounds. The details of simulation exercise for this case is available from the authors upon request.

20 In Section 3, we characterize the sharp identified set of structural parameters for a class of dynamic discrete-choice models when the state transition rules ($\vec{g}$) are point identified. However, we still allow for set-identified $\vec{g}$ for the characterization of the identified set as claimed in Theorem 1(i). In this section and Section 5.4, we characterize the identified set of structural parameters.

Table 3
Monte Carlo simulation results based on 2000 iterations. $N = 1000$, $\lambda_1 = 0.6$, $\lambda_2 = 0.7$. The displayed numbers for the lower and upper bounds are the Monte Carlo means. The numbers in parentheses indicate the standard deviations. Panel (A) shows the results when we impose a parametric restriction on the state transition probabilities and make an extrapolation for state variables that are not observed in the data. Panel (B) shows the results when we estimate the bounds for λ_2 together with the payoff parameters (imposing the restriction that $1 \geq \lambda_2 > \lambda_1$).

(A) Extrapolate state transition rules					
	True	Lower bound		Upper bound	
κ	20.000	17.134	(1.436)	23.301	(1.878)
π_1	2.500	−2.153	(1.837)	3.837	(0.645)
π_2	4.000	2.799	(0.566)	6.238	(0.599)
π_3	6.000	4.911	(0.605)	7.367	(0.842)
(B) Estimate bounds for λ_2					
	True	Lower bound		Upper bound	
κ	20.000	17.083	(1.444)	23.086	(1.853)
π_1	2.500	−2.333	(1.835)	3.757	(0.627)
π_2	4.000	2.752	(0.570)	6.256	(0.591)
π_3	6.000	4.924	(0.590)	7.544	(0.902)
λ_2	0.700	0.595	(0.016)	0.969	(0.017)

is only set-identified. In this exercise, we impose the restriction that $1 \geq \lambda_2 > \lambda_1$. Imposing a lower bound of λ_2 helps to bound the identified set of π_3 from above, because any value of π_3 can be rationalized by a small value of λ_2.[21]

Monte Carlo simulation results based on 2000 iterations are summarized in Table 3 for sample size $N = 1000$. From this table we can see that, the bounds for payoff parameters and entry cost are similar to those in the base case where the state transition rules are assumed to be known by the econometrician. Extrapolating state transition rules (see results shown in Panel (A)) works reasonably well in our simulation exercises, as the extrapolated transition probabilities are close to their true values. When we jointly estimate bounds for λ_2 (see results shown in Panel (B)), the true value of λ_2 is located between [0.594, 0.973]. As we would expect, the estimated mean lower bound and the mean upper bound are close to the true values of λ_1 and 1, respectively.

5. Japanese FDI in China

In the last 30 years,[22] a large number of Japanese firms opened foreign affiliates in China to exploit low local wages or to sell their products in the growing local market. The high rate of growth in China attracted many investors. In addition, China's accession to the WTO in 2001 accelerated this trend. As the Chinese economy matures, economic growth will slow down, and the Chinese market will be less attractive compared to other growing markets. Dynamic investors will take this future into account, but we have not observed states where China has moderate economic growth as a WTO member. Therefore, FDI decisions by Japanese firms in China serve as a good illustrating example for our method.

5.1. Data

We create a dataset using the annual *Toyo Keizai* database, which contains information on all foreign affiliates of parent companies that are headquartered in Japan. For each foreign affiliate, we observe the location/country of the affiliate, the name of the parent company, the industry code, and the number of employees. We aggregate affiliate-level information to the level of parent companies. If a parent firm in Japan opens an affiliate in China for the first time, we say that the parent firm enters the Chinese market. If the firm closes all affiliates in China, then we say that the firm exits the Chinese market. To homogenize the industries and products, we focus on Japanese FDI in the machinery industries (machinery, electronics, automobiles, transportation, and precision machinery).[23] By connecting the annual database from 1990 to 2005, we define the years of entry and exit for each parent company. In addition, using the *World Development Indicators*, we collect the time series of China's GDP growth rates. Table 4 summarizes the number of incumbents, entry, and exit, as well as other macroeconomic variables.

To estimate the model, we need to identify the set of potential entrants. We define all firms that opened at least one foreign affiliate in machinery industries in some country outside of Japan during the sample period as potential entrants.

21 The rationale for assuming $\lambda_2 \geq \lambda_1$ is that at the initial development of an industry, it is usually harder for the market demand to increase. When the market enters the high growth stage, it is easier for the market demand to further increase, possibly due to network effects, marketing, and mature infrastructure development, etc.

22 This statement is as of 2015 when we had the first version of this paper presented in public.

23 Examples of precision machinery (SIC code is 3599) include watches and medical semiconductors, etc.

Table 4
Summary statistics.

Year	Incumbent	Entry	Exit	GDP growth	WTO member
1990	45	8	2	4.1	0
1991	51	28	2	3.8	0
1992	77	41	2	9.2	0
1993	116	79	1	14.2	0
1994	194	101	13	14.0	0
1995	282	126	11	13.1	0
1996	397	43	15	10.9	0
1997	425	34	12	10.0	0
1998	447	28	15	9.3	0
1999	460	23	19	7.8	0
2000	464	50	14	7.6	0
2001	500	80	23	8.3	1
2002	557	129	28	9.1	1
2003	658	99	29	10.0	1
2004	728	71	34	10.1	1
2005	765	62	33	11.3	1

As a result, we identified $N = 2,197$ potential entrants. That is, approximately 35% $(= 765/2197)$ of potential entrants were active in the Chinese market in 2005.

5.2. Model

We adopt the dynamic model of entry and exit described in Example 1 to model Japanese firms' FDI decisions in China.[24] We use s_{it} to denote the endogenous state variable that equals one if firm i operates in China in t, and zero otherwise. The exogenous state variable $z_t = (y_t, w_t)$ contains y_t that indicates the category of the GDP growth rate of China in t and w_t that indicates whether China is a member of WTO in t. Specifically, $y_t = 1, 2,$ and 3 indicate the GDP growth rate of $(-\infty, 5\%)$, $[5\%, 10\%)$, and $[10\%, +\infty)$, respectively. The binary indicator w_t takes the value of one if China is a member of WTO and zero otherwise.[25]

We continue to assume that the period payoff consists of two parts as in Example 1. The payoff that depends on the observable exogenous state variables is written as $\pi_z = \pi_{(y,w)}$. We set the exit value to $\phi = 0$ while we estimate the entry cost κ.

The exogenous state variable y_t is assumed to evolve according to the following Markov matrix

$$\begin{pmatrix} \lambda_y & 1-\lambda_y & 0 \\ \frac{1-\lambda_y}{2} & \lambda_y & \frac{1-\lambda_y}{2} \\ 0 & 1-\lambda_y & \lambda_y \end{pmatrix};$$

likewise, we assume that

$$w_{t+1} = \begin{cases} 1 & \text{w.p. } \lambda_w \\ 0 & \text{w.p. } 1-\lambda_w \end{cases} \qquad \text{if } w_t = 0,$$

and $w_{t+1} = 1$ with probability one if $w_t = 1$.[26] This implies that China's accession to the WTO is stochastic, but once it becomes a member, it will not withdraw forever. In this application, we separately estimate (λ_y, λ_w) by maximum likelihood and treat their estimate $(\lambda_y = 0.733, \lambda_w = 0.091)$ as known by the econometrician. In Section 5.4, we present the estimation results when the transition rule of y_t is not known for states where China has moderate economic growth as a WTO member. The set estimates for the payoff parameters do not change qualitatively when the state transition rules are assumed unknown by the econometrician.

We impose the following restrictions on the shape of $\pi_{(y,w)}$.

(I) $\pi_{(y,w)} > \pi_{(y',w)}$ for $y > y'$ and $w \in \{0, 1\}$;
(II) $\pi_{(y,1)} = \pi_{(y,0)} + \pi_{wto}$ for all y.

24 Online Appendix B extends this simple model to the one with a larger state space and multinomial choices.

25 The per-capita income, wage rate, and other variables related to investment climates in China would also affect investor's decisions. However, a preliminary regression analysis suggests that China's GDP growth rate and its WTO membership are major determinants of firms' entry and exit. Therefore, we focus on these two variables in this analysis.

26 We impose parametric assumptions on the transition rule of y_t mainly due to the limited time periods we observe in the data. Alternatively, we can nonparametrically estimate the transition rule of y_t. The bound estimates for the payoff parameters do not change qualitatively when the transition rule is estimated nonparametrically.

Table 5
Empirical results are displayed in panel (A). The numbers in the first two columns indicate the set estimates. The numbers in the last column indicate the point estimates under the logit extrapolation. Bootstrap credible regions and confidence intervals are displayed in panel (B).

(A) Bounds for the structural parameters

	Set estimates		Extrapolation
κ	[61.753	68.761]	64.499
$\pi_{(1,0)}$	[−7.715	−0.478]	−3.142
$\pi_{(2,0)}$	[−1.469	2.877]	0.209
$\pi_{(3,0)}$	[1.636	5.305]	3.487
π_{wto}	[0.515	2.226]	1.491

(B) Credible regions for the structural parameters

	Set estimates 95% CR		Extrapolation 95% CI	
κ	[60.939	69.151]	[63.220	66.038]
$\pi_{(1,0)}$	[−8.835	−0.381]	[−3.849	−2.557]
$\pi_{(2,0)}$	[−1.754	3.241]	[−0.070	0.424]
$\pi_{(3,0)}$	[1.439	5.664]	[3.035	4.093]
π_{wto}	[0.356	2.453]	[1.202	1.828]

That is, the period payoff increases with the GDP growth rate. In addition, for any GDP growth, the WTO membership increases (or decreases) the period payoff by the same magnitude, which is represented by a parameter π_{wto}. With these shape restrictions, the parameter vector to be estimated is $(\pi_{(1,0)}, \pi_{(2,0)}, \pi_{(3,0)}, \pi_{wto}, \kappa)$. The above shape restrictions, (I) and (II), fail to reduce dimensions sufficiently enough for a point identification, unlike common forms of parametric shape restrictions, e.g., $\pi_{(y,w)} = \alpha_1 y + \alpha_2 w$.[27]

For the sake of comparison, we also estimate the parameters by assuming that the CCP has the logit model:

$$a_{it} = \mathbb{1}\{\alpha_0 + \alpha_1 y_t + \alpha_2 w_t + \alpha_3 s_{it} + \varepsilon_{it}^1 > \varepsilon_{it}^0\},$$

where $\left(\varepsilon_{it}^0, \varepsilon_{it}^1\right)$ follows the i.i.d. Type I Extreme Value distribution. Then, for all states, we can compute

$$\Pr(a_{it} = 1|s_{it}, y_t, w_t) = \frac{\exp(\hat{\alpha}_0 + \hat{\alpha}_1\widetilde{y}_t + \hat{\alpha}_2 w_t + \hat{\alpha}_3 s_{it})}{1 + \exp(\hat{\alpha}_0 + \hat{\alpha}_1\widetilde{y}_t + \hat{\alpha}_2 w_t + \hat{\alpha}_3 s_{it})}$$

where

$$\widetilde{y}_t = \begin{cases} 2.5 & \text{if } y_t = 1 \\ 7.5 & \text{if } y_t = 2 \\ 12.5 & \text{if } y_t = 3 \end{cases}.$$

Note that the logit assumption may be considered as a more restrictive version of our shape restrictions, and this strong parametric shape restriction fully reduces dimensions so that a point identification is achieved.

5.3. Results

Results are summarized in Table 5. The first two columns in panel (A) show estimates of the bound for each of structural parameter. It should be emphasized that this is the marginal bound for each parameter. Therefore, the identified region is smaller than the naive Cartesian product of these five intervals. The last column in panel (A) reports the point estimates obtained with the logit model. For each parameter, the estimate obtained from the extrapolation of CCPs is contained in the set estimates obtained by our method. Indeed, the point estimate is included in the set estimates, as the value of $\widehat{Q}_n^*(\vec{g}, \vec{p})$ evaluated at the point estimate is as small as the one evaluated at other parameter vectors in the set estimates. Panel (B) shows the 95% credible regions and confidence intervals corresponding to the two estimates in panel (A).[28]

While the parameter estimate obtained from extrapolation does not lie outside of the set estimates, this result does not mean that extrapolation is innocuous. Note that all points in the set estimates are consistent with the data. With somewhat wide bounds of our set estimates, it is possible that the bias from extrapolation may be significant. If one had

27 Indeed, we observe $y = 1$, 2, and 3 under $w = 0$, as well as $y = 2$ and 3 under $w = 1$, and hence it may appear that a point identification is achieved under restriction (II). However, due to the dynamic nature of the model, $\pi_{(y,0)}$ could not be pinned down from the CCPs under $w = 0$ alone. As such, the parameters in this model are only partially identified.

28 The credible regions reported in Panel (B) of Table 5 ignores the sampling error from estimating (λ_y, λ_w). To account for this, we can estimate the bounds for (λ_y, λ_w) together with other payoff parameters. Overall, the set estimates and 95% credible regions for the structural parameters do not change qualitatively when λ's are jointly estimated. The results are available from the authors upon request.

to make a point decision out of an interval, the fact that the point estimates based on extrapolation lie approximately around the middle of the sets can be considered as a better outcome (cf. Song, 2014).[29]

Using the set estimates for the structural parameters, we conduct several counterfactual exercises, such as reducing entry costs or increasing exit values. In general, we find non-trivial bounds for the counterfactual outcomes and the counterfactual predictions obtained from an extrapolation can be very different from the truth. More detailed discussion on our counterfactual analysis are provided in Online Appendix E.

5.4. Unknown state transitions

In this section, we relax the assumption adopted in Section 5.2 that the state transition rules can be directly estimated from the data and treated as known by the econometrician.

We assume that the state transition rule of y_t depends on China's WTO membership status. Specifically, the exogenous state variable y_t is assumed to evolve according to the following Markov matrix if $w_t = 0$:

$$\begin{pmatrix} \lambda_y & 1-\lambda_y & 0 \\ \frac{1-\lambda_y}{2} & \lambda_y & \frac{1-\lambda_y}{2} \\ 0 & 1-\lambda_y & \lambda_y \end{pmatrix}.$$

When $w_t = 1$, we consider two Markov matrices for the transition of y_t.

$$\text{Specification 1:} \quad \begin{pmatrix} 1 & 0 & 0 \\ \frac{1-\tilde{\lambda}_y}{2} & \tilde{\lambda}_y & \frac{1-\tilde{\lambda}_y}{2} \\ 0 & 1-\lambda_y & \lambda_y \end{pmatrix}, \quad \text{Specification 2:} \quad \begin{pmatrix} 1 & 0 & 0 \\ \tilde{\tilde{\lambda}}_y & \tilde{\lambda}_y & 1-\tilde{\tilde{\lambda}}_y-\tilde{\lambda}_y \\ 0 & 1-\lambda_y & \lambda_y \end{pmatrix}.$$

The transition probability of w_t remains the same as in Section 5.2.

In both specifications, we assume that $y_t = 1$ is an absorbing state when China enters WTO. In Specification 1, the transition rule for state $y_t = 2$, governed by the new parameter $\tilde{\lambda}_y$, may differ from the transition rule when China is not a WTO member. Since in the data we do not observe states where China has a moderate economic growth as a WTO member, $\tilde{\lambda}_y$ cannot be directly recovered from data, and thus is set-identified. The transition rule for y_t when $w_t = 1$ in the second specification is slightly more flexible. We allow two free parameters $(\tilde{\lambda}_y, \tilde{\tilde{\lambda}}_y)$ to govern the transition probabilities for state $y_t = 2$. Both of these parameters are set-identified.

Similar to Section 5.2, we estimate (λ_y, λ_w) by maximum likelihood and treat their estimate as known by the econometrician. With the new specifications of the transition rules, our estimates $\lambda_y = 0.769$, and $\lambda_w = 0.091$. In Specification 1, we jointly estimate the bounds for $\tilde{\lambda}_y$ together with other payoff parameters, imposing the restriction that $1 \geq \tilde{\lambda}_y > \lambda_y$. In Specification 2, we estimate bounds for $(\tilde{\lambda}_y, \tilde{\tilde{\lambda}}_y)$, imposing the restriction that $1 \geq \tilde{\lambda}_y > \lambda_y > \tilde{\tilde{\lambda}}_y$ and $1 - \tilde{\tilde{\lambda}}_y - \tilde{\lambda}_y \geq 0$.[30]

The estimates of the bound for each structural parameter are summarized in Table 6. Overall, the set estimates of the structural payoff parameters in Table 6 do not differ much from the ones in Table 5 when the state transition rules are assumed known by the econometrician. When the state transition rules are not known, we obtain a slightly wider range for the estimate of π_{wto}. The bound estimates for the payoff parameters are similar across the two specifications. In the second specification where the transition rule for $y_t = 2$ is more flexibly specified, the bound for $\tilde{\tilde{\lambda}}_y$ is $[0, 0.247]$, which is wider than the bound in the first specification (i.e., $\frac{1-\tilde{\lambda}_y}{2} \in [0, 0.124]$) as expected.

6. Conclusions

For a class of dynamic discrete choice models, we provide a robust empirical method that deals with incomplete data coverage of relevant states without relying on parametric extrapolation. Exploiting the model restriction *à la* Aguirregabiria and Mira (2002, 2007) and Kasahara and Shimotsu (2012), we characterize the sharp identified set of structural parameters when the conditional choice probabilities are only partially identified.

Through simulation studies, we find that our method gives informative bounds for the structural parameters with the sample size in typical empirical applications. We also confirm that the set of the maxima of the likelihood function coincides with our sharp identified set in a large sample. Using our sharp set, we study the performance of logit extrapolations and find that some specifications work well while others do not.

Focusing on a problem that is relevant to common situations of industry dynamics, we present the sharp identification result. Estimation and statistical inference for partially identified parameters and identified sets are by the present

29 With this said, we remark that the conclusion of Song (2014) does not exactly apply to our setting though, as he considers the case of explicit interval estimators which are different from our estimator.

30 In both specifications 1 and 2, we assume that $\tilde{\lambda}_y > \lambda_y$ because it is potentially easier for a country to maintain a median growth rate than a high growth rate. In specification 2, assuming $1 - \tilde{\tilde{\lambda}}_y - \tilde{\lambda}_y \geq 0$ allows the possibility that the economy transits from a median growth rate state to a high growth rate state.

Table 6
Set estimates of structural parameters when state transition rules are not known. We collect 10,000 points for each specification using the MCMC algorithm provided in Appendix A.9.

	Specification 1		Specification 2	
	Lower bound	Upper bound	Lower bound	Upper bound
κ	61.880	68.532	61.968	68.627
$\pi_{(1,0)}$	−5.295	−0.335	−6.217	−0.273
$\pi_{(2,0)}$	−1.249	2.348	−1.376	2.521
$\pi_{(3,0)}$	1.453	4.341	1.165	4.797
π_{wto}	0.172	2.703	−0.445	3.287
$\tilde{\lambda}_y$	0.753	1.000	0.751	0.999
$\tilde{\tilde{\lambda}}_y$	N/A	N/A	0.000	0.247

day established by a rich literature. In particular, given the criterion-based implementation procedure outlined in Appendix A.9, methods of inference based on criterion are applicable (e.g., Chernozhukov et al., 2007).[31] Extending the proposed approach to dynamic discrete choice models with unobserved heterogeneity or dynamic discrete games are left for future exploration.

Appendix A

A.1. On construction of the restriction matrix

In this section, we provide an example of constructing the restriction matrix R and the parameter set Θ. Consider Example 1 on the dynamic model of entry/exit. Let $\pi = (\pi_{0,(0,1)}, \ldots, \pi_{0,(0,\bar{z})}, \pi_{0,(1,1)}, \ldots, \pi_{0,(1,\bar{z})}, \pi_{1,(0,1)}, \ldots, \pi_{1,(0,\bar{z})}, \pi_{1,(1,1)}, \ldots, \pi_{1,(1,\bar{z})})'$ denote the vector of static payoffs, and let $\theta = (\pi_1, \ldots, \pi_{\bar{z}}, \phi, \kappa)'$ denote the vector of primitive parameters. The aforementioned restriction $\pi = R\theta$ can be formed by

$$R = \begin{bmatrix} 0_{\bar{z}\times\bar{z}} & 0_{\bar{z}\times 1} & 0_{\bar{z}\times 1} \\ I_{\bar{z}\times\bar{z}} & 1_{\bar{z}\times 1} & 0_{\bar{z}\times 1} \\ 0_{\bar{z}\times\bar{z}} & 0_{\bar{z}\times 1} & -1_{\bar{z}\times 1} \\ I_{\bar{z}\times\bar{z}} & 0_{\bar{z}\times 1} & 0_{\bar{z}\times 1} \end{bmatrix}$$

where $0_{r\times c}$ denotes the $r\times c$ matrix of zeros, $1_{r\times c}$ denotes the $r\times c$ matrix of ones, and $I_{r\times c}$ denotes the $r\times c$ identity matrix where $r = c$. In addition, the restriction, $\pi_1 \leqslant \cdots \leqslant \pi_{\bar{z}}$, of non-decreasing per-period profit with respect to demand can be imposed by defining the compact parameter set by $\Theta = \left\{(\pi_1, \ldots, \pi_{\bar{z}}, \phi, \kappa)' \in I_1 \times \cdots \times I_{\bar{z}} \times I_\phi \times I_\kappa \mid \pi_1 \leqslant \cdots \leqslant \pi_{\bar{z}}\right\}$ where $I_1, \cdots, I_{\bar{z}}, I_\phi$, and I_κ are compact subsets of $\mathbb{R}$.

A.2. The closed-form inversion

We obtain the following auxiliary lemma in the same manner as (Hotz et al., 1994) and Aguirregabiria and Mira (2002) — also related is (Pesendorfer and Schmidt-Dengler, 2008), Sanches et al. (2016), and Buchholz et al. (2021).

Lemma 3. *Suppose that the current-time payoff is given by (2.1) with (2.2) and that $\beta \in (0, 1)$. For true $(\vec{g}, \vec{p})$, we obtain the restriction*

$$\begin{aligned} &\sum_{x'=1}^{\bar{x}} \left(H_{1,x'}(x; \vec{g}, \vec{p}, \beta) + \mathbb{1}\{x = x'\}\right) \cdot \pi_{1,x'} + \sum_{x'=1}^{\bar{x}} \left(H_{0,x'}(x; \vec{g}, \vec{p}, \beta) - \mathbb{1}\{x = x'\}\right) \cdot \pi_{0,x'} \\ &= \sum_{x'=1}^{\bar{x}} \Big[\left(H_{1,x'}(x; \vec{g}, \vec{p}, \beta) + \mathbb{1}\{x = x'\}\right) \cdot \ln p_{1,x'} + \left(H_{0,x'}(x; \vec{g}, \vec{p}, \beta) - \mathbb{1}\{x = x'\}\right) \cdot \ln p_{0,x'} \\ &\qquad - \left(H_{1,x'}(x; \vec{g}, \vec{p}, \beta) + H_{0,x'}(x; \vec{g}, \vec{p}, \beta)\right) \cdot \bar{\varepsilon}\Big]. \end{aligned}$$

for each $x \in \{0, \ldots, \bar{x}\}$, where $H_{a',x'}(x; \vec{g}, \vec{p}, \beta) := \sum_{\tau=1}^{\infty} \beta^\tau \left(h^\tau_{a',x',1,x}(\vec{g}, \vec{p}) - h^\tau_{a',x',0,x}(\vec{g}, \vec{p})\right)$.

[31] We also provide a non-comprehensive list of papers on statistical inference about partially identified parameters available to date for convenience of readers: Imbens and Manski (2004), Chernozhukov et al. (2007), Beresteanu and Molinari (2008), Rosen (2008), Andrews and Guggenberger (2009), Stoye (2009), Andrews and Soares (2010), Bugni (2010), Canay (2010), Romano and Shaikh (2010), Chen et al. (2011), Andrews and Barwick (2012), Kitagawa (2012), Moon and Schorfheide (2012), Andrews and Shi (2013), Kline and Tamer (2013), Armstrong (2014), Romano et al. (2014), Chen et al. (2015), Bugni et al. (2017), Kaido et al. (2016), and Liao and Simoni (2016).

Proof. For the current-time payoff defined by (2.1), the policy value function v can be written as

$$v(a,x) \;=\; \pi_{a,x} + \beta \sum_{x'=1}^{\bar{x}} g_{x',a,x} V(x').$$

From this equation, we can write

$$\begin{aligned}
&\mathrm{E}[\beta \cdot V(X_{t+1}) \mid A_t = 1, X_t = x] - \mathrm{E}[\beta \cdot V(X_{t+1}) \mid A_t = 0, X_t = x] \\
&= \beta \sum_{x'=1}^{\bar{x}} g_{x',1,x} V(x') - \beta \sum_{x'=1}^{\bar{x}} g_{x',0,x} V(x') \\
&= v(1,x) - v(0,x) - \pi_{1,x} + \pi_{0,x} \;=\; \ln p_{1,x} - \ln p_{0,x} - \pi_{1,x} + \pi_{0,x}
\end{aligned} \tag{A.1}$$

where the third equality follows from (2.2) and the inversion theorem of Hotz and Miller (1993). On the other hand, the conditional expectation of the value function can be computed under (2.2) as

$$\mathrm{E}\left[\,\beta \cdot V(X_{t+1}) | A_t, X_t\right] = \mathrm{E}\left[\sum_{s=t+1}^{\infty} \sum_{a'=0}^{1} \beta^{s-t} \cdot p_{a',X_s} \cdot \left(\pi_{a',X_s} + \bar{\varepsilon} - \ln p_{a',X_s}\right) \middle| A_t, X_t \right].$$

for any $s > t$. Using the notation (2.3) for the transition probability $\Pr(A_{t+\tau} = a', X_{t+\tau} = x' \mid A_t = a, X_t = x)$, we can thus write

$$\begin{aligned}
\mathrm{E}\left[\,\beta \cdot V(X_{t+1}) | A_t = a, X_t = x\right] &= \sum_{s=t+1}^{\infty} \sum_{a'=0}^{1} \sum_{x'=1}^{\bar{x}} \beta^{s-t} \cdot h^{s-t}_{a',x',a,x}(\vec{g},\vec{p}) \cdot \left(\pi_{a',x'} + \bar{\varepsilon} - \ln p_{a',x'}\right) \\
&= \sum_{s=t+1}^{\infty} \sum_{x'=1}^{\bar{x}} \beta^{s-t} \cdot h^{s-t}_{1,x',a,x}(\vec{g},\vec{p}) \cdot \left(\pi_{1,x'} + \bar{\varepsilon} - \ln p_{1,x'}\right) \\
&\quad + \sum_{s=t+1}^{\infty} \sum_{x'=1}^{\bar{x}} \beta^{s-t} \cdot h^{s-t}_{0,x',a,x}(\vec{g},\vec{p}) \cdot \left(\pi_{0,x'} + \bar{\varepsilon} - \ln p_{0,x'}\right)
\end{aligned}$$

Substituting this expression on the left-hand side of (A.1) yields

$$\begin{aligned}
&\sum_{x'=1}^{\bar{x}} H_{1,x'}(x;\vec{g},\vec{p},\beta) \cdot \left(\pi_{1,x'} + \bar{\varepsilon} - \ln p_{1,x'}\right) + \sum_{x'=1}^{\bar{x}} H_{0,x'}(x;\vec{g},\vec{p},\beta) \cdot \left(\pi_{0,x'} + \bar{\varepsilon} - \ln p_{0,x'}\right) \\
&\quad = \ln p_{1,x} - \ln p_{0,x} - \pi_{1,x} + \pi_{0,x}
\end{aligned}$$

where $H_{a',x'}(x;\vec{g},\vec{p},\beta) := \sum_{\tau=1}^{\infty} \beta^{\tau} \left(h^{\tau}_{a',x',1,x}(\vec{g},\vec{p}) - h^{\tau}_{a',x',0,x}(\vec{g},\vec{p}) \right)$ for a short-hand notation. We can rewrite this equation conveniently as

$$\begin{aligned}
&\sum_{x'=1}^{\bar{x}} \left(H_{1,x'}(x;\vec{g},\vec{p},\beta) + \mathbb{1}\{x = x'\}\right) \cdot \pi_{1,x'} + \sum_{x'=1}^{\bar{x}} \left(H_{0,x'}(x;\vec{g},\vec{p},\beta) - \mathbb{1}\{x = x'\}\right) \cdot \pi_{0,x'} \\
&= \sum_{x'=1}^{\bar{x}} \Big[\left(H_{1,x'}(x;\vec{g},\vec{p},\beta) + \mathbb{1}\{x = x'\}\right) \cdot \ln p_{1,x'} + \left(H_{0,x'}(x;\vec{g},\vec{p},\beta) - \mathbb{1}\{x = x'\}\right) \cdot \ln p_{0,x'} \\
&\qquad - \left(H_{1,x'}(x;\vec{g},\vec{p},\beta) + H_{0,x'}(x;\vec{g},\vec{p},\beta)\right) \cdot \bar{\varepsilon} \Big].
\end{aligned}$$

This proves the proposition. □

To better understand the intuition about the matrix $\tilde{H}(\vec{p},\vec{q},\beta)$, it is convenient to put aside the matrix R and to focus on $H(x;\vec{g},\vec{p},\beta)$ for now. In this case, a typical element $H_{a',x'}(x;\vec{g},\vec{p},\beta)$ of $\tilde{H}(\vec{g},\vec{p},\beta)$ without R, albeit the current state indicators $\mathbb{1}\{x = x'\}$, can be written as the future discounted cumulative probability difference $\sum_{\tau=1}^{\infty} \beta^{\tau} \Pr(a',x'|1,x)^{\tau} - \sum_{\tau=1}^{\infty} \beta^{\tau} \Pr(a',x'|0,x)^{\tau}$. Note that this object does '*not*' convey the typical intuition in terms of dynamic paths taken by agents, because the future action-state pair (a',x') is fixed throughout across the future time periods for each element $H_{a',x'}(x;\vec{g},\vec{p},\beta)$. However, when multiplied by $\pi = R\theta$, they together yield the future discounted cumulative values summed across all the action-state pairs conditionally on the current state $X_t = x$ and action $A_t = 1$ minus the future discounted cumulative values summed across all the action-state pairs conditionally on the current state $X_t = x$ and action $A_t = 0$.

Now, recall that a typical element $H_{a',x'}(x;\vec{g},\vec{p},\beta)$ of $\tilde{H}(\vec{g},\vec{p},\beta)$ without R also includes the current state indicators $\mathbb{1}\{x=s'\}$. Multiplied by π, these parts represent the difference of the deterministic part of the current values across the two actions.

A typical element $Y(x;\vec{g},\vec{p},\beta)$ of $\tilde{Y}(\vec{g},\vec{p},\beta)$ without R can be interpreted similarly to a typical element $H_{a',x'}(x;\vec{g},\vec{p},\beta)$ of $\tilde{H}(\vec{g},\vec{p},\beta)$ without R, except that it represents the future discounted cumulative value difference associated with the $(\bar{\varepsilon}-\ln p_{a,x})$ part, instead of that associated with the $\pi_{a,x}$ part.

A.3. Proof of Lemma 1

Proof. With the short-hand notations $H(x;\vec{g},\vec{p},\beta)$, π, and $Y(x;\vec{g},\vec{p},\beta)$, the restriction provided in Lemma 3 can be succinctly rewritten as

$$H(x;\vec{g},\vec{p},\beta)\,\pi \;=\; Y(x;\vec{g},\vec{p},\beta) \tag{A.2}$$

for each $x\in\{1,\ldots,\bar{x}\}$. Combining the linear restrictions (3.1) and (A.2) together, we can write the degenerated restriction as follows.

$$\tilde{H}(\vec{g},\vec{p},\beta)\,\theta \;=\; \tilde{Y}(\vec{g},\vec{p},\beta) \tag{A.3}$$

Thus, we can form the restriction of the form

$$\tilde{H}(\vec{g},\vec{p},\beta)'\tilde{H}(\vec{g},\vec{p},\beta)\,\theta \;=\; \tilde{H}(\vec{g},\vec{p},\beta)'\tilde{Y}(\vec{g},\vec{p},\beta),$$

which proves part (i) of the lemma. Under the rank condition (3.3), we can solve for θ as

$$\theta \;=\; \left[\tilde{H}(\vec{g},\vec{p},\beta)'\tilde{H}(\vec{g},\vec{p},\beta)\right]^{-1}\left[\tilde{H}(\vec{g},\vec{p},\beta)'\tilde{Y}(\vec{g},\vec{p},\beta)\right],$$

which proves part (ii) of the lemma. □

A.4. Proof of Theorem 1

Proof. Part (i) follows immediately from Lemma 1(i) and the assumption in the statement of the theorem that $\theta_0\in\Theta$.

Part (ii) follows from part (i), Lemma 1 (ii), and the additional assumptions that $\mathcal{G}=\{\vec{g}_0\}$ is a singleton and the rank condition (3.3) is satisfied for $\vec{g}=\vec{g}_0$ and for all $\vec{p}\in\mathcal{P}$.

Part (iii): Assume by way of contradiction that Θ_I is not sharp. In other words, assume that there exists $\theta_*\in\Theta_I$ such that $\theta_*=\theta_0$ cannot be true given the available information $(\mathcal{G},\mathcal{P},\beta)$. By the definition of Θ_I, the inclusion $\theta_*\in\Theta_I$ implies that there exists $(\vec{g}_*,\vec{p}_*)\in\mathcal{GP}$ such that

$$\theta_* \;=\; \left[\tilde{H}(\vec{g}_*,\vec{p}_*,\beta)'\tilde{H}(\vec{g}_*,\vec{p}_*,\beta)\right]^{-1}\left[\tilde{H}(\vec{g}_*,\vec{p}_*,\beta)'\tilde{Y}(\vec{g}_*,\vec{p}_*,\beta)\right].$$

Note also that

$$\theta_0 \;=\; \left[\tilde{H}(\vec{g}_0,\vec{p}_0,\beta)'\tilde{H}(\vec{g}_0,\vec{p}_0,\beta)\right]^{-1}\left[\tilde{H}(\vec{g}_0,\vec{p}_0,\beta)'\tilde{Y}(\vec{g}_0,\vec{p}_0,\beta)\right]$$

is true. Since $\theta_*=\theta_0$ cannot be true given the available information $(\mathcal{G},\mathcal{P},\beta)$, $(\vec{g}_*,\vec{p}_*)=(\vec{g}_0,\vec{p}_0)$ cannot be true given this information. It thus follows that $(\vec{g}_0,\vec{p}_0)$ is partially identified by the set $\mathcal{GP}\backslash\{(\vec{g}_*,\vec{p}_*)\}$, showing that $\mathcal{GP}$ is not a sharp identified set. The claimed statement follows by the contrapositive argument. □

A.5. Proof of Lemma 2

Proof. Note that the CCP of $a=1$ given state x under (2.1) and (2.2) is written as

$$p_{1,x} \;=\; \frac{\exp\left\{\pi_{1,x}-\pi_{0,x}+\mathrm{E}\left[\beta\cdot V(X_{t+1})\mid A_t=1,X_t=x\right]-\mathrm{E}\left[\beta\cdot V(X_{t+1})\mid A_t=0,X_t=x\right]\right\}}{1+\exp\left\{\pi_{1,x}-\pi_{0,x}+\mathrm{E}\left[\beta\cdot V(X_{t+1})\mid A_t=1,X_t=x\right]-\mathrm{E}\left[\beta\cdot V(X_{t+1})\mid A_t=0,X_t=x\right]\right\}}$$

In the proof of Lemma 3, the terms of the form $\mathrm{E}[\beta \cdot V(X_{t+1}) | A_t = a, X_t = x]$ is shown to be identified by

$$\mathrm{E}[\beta \cdot V(X_{t+1}) | A_t = a, X_t = x] = \sum_{s=t+1}^{\infty} \sum_{x'=1}^{\bar{x}} \beta^{s-t} \cdot h^{s-t}_{1,x',a,x}(\vec{g}_0, \vec{p}) \cdot \left(\pi_{1,x'} + \bar{\varepsilon} - \ln p_{1,x'}\right) + \sum_{s=t+1}^{\infty} \sum_{x'=1}^{\bar{x}} \beta^{s-t} \cdot h^{s-t}_{0,x',a,x}(\vec{g}_0, \vec{p}) \cdot \left(\pi_{0,x'} + \bar{\varepsilon} - \ln p_{0,x'}\right)$$

Hence, the above CCP $p_{1,x}$ may be compactly written as

$$p_{1,x} = \frac{\exp\left\{\Lambda_{1,x}\left(\pi, \vec{g}_0, \vec{p}, \beta\right)\right\}}{1 + \exp\left\{\Lambda_{1,x}\left(\pi, \vec{g}_0, \vec{p}, \beta\right)\right\}}$$

where $\Lambda_{1,x}(\pi, \vec{g}, \vec{p}, \beta)$ is defined by

$$\begin{aligned} \Lambda_{1,x}(\pi, \vec{g}, \vec{p}, \beta) &= \pi_{1,x} - \pi_{0,x} + \\ &\sum_{\tau=1}^{\infty} \sum_{x'=1}^{\bar{x}} \beta^{\tau} \cdot \left[h^{\tau}_{1,x',1,x}(\vec{g}, \vec{p}) \cdot \left(\pi_{1,x'} + \bar{\varepsilon} - \ln p_{1,x'}\right) + h^{\tau}_{0,x',1,x}(\vec{g}, \vec{p}) \cdot \left(\pi_{0,x'} + \bar{\varepsilon} - \ln p_{0,x'}\right)\right] - \\ &\sum_{\tau=1}^{\infty} \sum_{x'=1}^{\bar{x}} \beta^{\tau} \cdot \left[h^{\tau}_{1,x',0,x}(\vec{g}, \vec{p}) \cdot \left(\pi_{1,x'} + \bar{\varepsilon} - \ln p_{1,x'}\right) + h^{\tau}_{0,x',0,x}(\vec{g}, \vec{p}) \cdot \left(\pi_{0,x'} + \bar{\varepsilon} - \ln p_{0,x'}\right)\right] \end{aligned}$$

Since the above equality for $p_{1,x}$ has to be satisfied under the true payoff parameters $\pi = R\theta_0$, we obtain the restriction

$$p_{1,x} = \frac{\exp\left\{\Lambda_{1,x}\left(R\theta_0, \vec{g}_0, \vec{p}, \beta\right)\right\}}{1 + \exp\left\{\Lambda_{1,x}\left(R\theta_0, \vec{g}_0, \vec{p}, \beta\right)\right\}}.$$

Furthermore, because the true structural parameters θ_0 are written in terms of the true $(\vec{g}_0, \vec{p}_0)$ by

$$\theta_0 = \left[\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)'\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)\right]^{-1} \left[\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)'\tilde{Y}(\vec{g}_0, \vec{p}_0, \beta)\right],$$

it follows that the identified set $\mathcal{P}$ restricts to the set of $\vec{p}$ satisfying the equation

$$p_{1,x} = \frac{\exp\left\{\Lambda_{1,x}\left(R\left[\tilde{H}(\vec{g}_0, \vec{p}, \beta)'\tilde{H}(\vec{g}_0, \vec{p}, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}_0, \vec{p}, \beta)'\tilde{Y}(\vec{g}_0, \vec{p}, \beta)\right], \vec{g}_0, \vec{p}, \beta\right)\right\}}{1 + \exp\left\{\Lambda_{1,x}\left(R\left[\tilde{H}(\vec{g}_0, \vec{p}, \beta)'\tilde{H}(\vec{g}_0, \vec{p}, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}_0, \vec{p}, \beta)'\tilde{Y}(\vec{g}_0, \vec{p}, \beta)\right], \vec{g}_0, \vec{p}, \beta\right)\right\}}.$$

for each $x \in \{1, \ldots, \bar{x}\}$. □

A.6. Proof of Theorem 2

Proof. First, note that $\vec{g}_0 \in \mathcal{G}$ holds by the assumption that $\mathcal{G} = \{\vec{g}_0\}$. Since the true $\vec{p}_0$ must satisfy $\vec{p}_0 \in \mathcal{G}(\vec{g}_0)$ by Lemma 2, it follows that $(\vec{g}_0, \vec{p}_0) \in \{\vec{g}_0\} \times \mathcal{P}(\vec{g}_0)$. This containment shows that $\mathcal{GP}^{\dagger}$ is an identified set of $(\vec{g}, \vec{p})$.

In order to show sharpness, assume by way of contradiction that there exists $(\vec{g}_*, \vec{p}_*) \in \mathcal{GP}^{\dagger}$ such that $(\vec{g}_*, \vec{p}_*) = (\vec{g}_0, \vec{p}_0)$ cannot be true given the available information. This can be divided into two cases. The first case is where $\vec{g}_* = \vec{g}_0$ cannot be true, but this is a contradiction with the assumption that $\mathcal{G} = \{\vec{g}_0\}$. The second case is where $\vec{g}_* = \vec{g}_0$ can be true, but $\vec{p}_* = \vec{p}_0$ cannot be true whenever $\vec{g}_* = \vec{g}_0$ is true. Note that the true equilibrium CCP vector $\vec{p}_0$ has to be the fixed point of the self map $\Phi_{\vec{g},\theta} : \mathcal{P} \to \mathcal{P}$ defined by

$$\Phi_{\vec{g},\theta}(\vec{p}) = \left[\; \frac{1}{1+\exp\left\{\Lambda_{1,1}\left(R\theta,\vec{g},\vec{p},\beta\right)\right\}} \quad \frac{\exp\left\{\Lambda_{1,1}\left(R\theta,\vec{g},\vec{p},\beta\right)\right\}}{1+\exp\left\{\Lambda_{1,1}\left(R\theta,\vec{g},\vec{p},\beta\right)\right\}} \quad \cdots \quad \frac{1}{1+\exp\left\{\Lambda_{1,\bar{x}}\left(R\theta,\vec{g},\vec{p},\beta\right)\right\}} \quad \frac{\exp\left\{\Lambda_{1,\bar{x}}\left(R\theta,\vec{g},\vec{p},\beta\right)\right\}}{1+\exp\left\{\Lambda_{1,\bar{x}}\left(R\theta,\vec{g},\vec{p},\beta\right)\right\}} \;\right]'$$

for $\vec{g} = \vec{g}_0$ and $\theta = \theta_0$. If $\vec{p}_* = \vec{p}_0$ cannot be true when $\vec{g}_* = \vec{g}_0$ is true, then $\vec{p}_*$ cannot be a fixed point of $\Phi_{\vec{g},\theta}$ for $\vec{g} = \vec{g}_* = \vec{g}_0$ for any $\theta \in \Theta$. But this is a contradiction with Lemma 2 and our choice of $(\vec{g}_*, \vec{p}_*)$ as an element of $\mathcal{GP}^{\dagger}$, i.e., $\vec{p}_* = \Psi_{\vec{g}_*}(\vec{p}_*) = \Psi_{\vec{g}_0}(\vec{p}_*) = \Phi_{\vec{g}_*,\vartheta(\vec{p}_*,\vec{g}_*)}(\vec{p}_*) = \Phi_{\vec{g}_0,\vartheta(\vec{p}_*,\vec{g}_0)}(\vec{p}_*)$ must hold. Therefore, the second case is also ruled out. □

A.7. Proof of Corollary 2

Proof. This corollary is proved in a similar manner to Theorem 1. Let $(\vec{g}_0, \vec{p}_0)$ denote the true element in $\mathcal{GP}$, and let C_0 denote the true counterfactual outcome. Since these are the truths, the restrictions (3.1) and (A.2) must hold with $(\vec{g}, \vec{p}) = (\vec{g}_0, \vec{p}_0)$ and $\theta = \theta_0$. But then, $\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)\,\theta_0 = \tilde{Y}(\vec{g}_0, \vec{p}_0, \beta)$ holds, and it thus follows that

$$\begin{aligned}
C_0 &= \Gamma\left(\theta_0, \vec{g}_0, \vec{p}_0\right) \\
&= \Gamma\left(\left[\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)'\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)'\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)\right]\theta_0, \vec{g}_0, \vec{p}_0\right) \\
&= \Gamma\left(\left[\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)'\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)'\tilde{Y}(\vec{g}_0, \vec{p}_0, \beta)\right], \vec{g}_0, \vec{p}_0\right) \in \mathcal{C}_I
\end{aligned}$$

where the last inclusion is due to $(\vec{g}_0, \vec{p}_0) \in \mathcal{GP}$ and by the definition of $\mathcal{C}_I$. This proves that $\mathcal{C}_I$ is an identified set for C_0.

Now, assume by way of contradiction that $\mathcal{C}_I$ is not sharp. In other words, assume that there exists $C_* \in \mathcal{C}_I$ such that $C_* = C_0$ cannot be true given the available information $(\mathcal{G}, \mathcal{P}, \beta)$. By the definition of $\mathcal{C}_I$, the inclusion $C_* \in \mathcal{C}_I$ implies that there exists $(\vec{g}_*, \vec{p}_*) \in \mathcal{GP}$ such that

$$C_* = \Gamma\left(\left[\tilde{H}(\vec{g}_*, \vec{p}_*, \beta)'\tilde{H}(\vec{g}_*, \vec{p}_*, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}_*, \vec{p}_*, \beta)'\tilde{Y}(\vec{g}_*, \vec{p}_*, \beta)\right], \vec{g}_*, \vec{p}_*\right).$$

Note also that

$$C_0 = \Gamma\left(\left[\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)'\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}_0, \vec{p}_0, \beta)'\tilde{Y}(\vec{g}_0, \vec{p}_0, \beta)\right], \vec{g}_0, \vec{p}_0\right)$$

is true. Since $C_* = C_0$ cannot be true given the available information $(\mathcal{G}, \mathcal{P}, \beta)$ and since Γ is a well-defined function, $(\vec{g}_*, \vec{p}_*) = (\vec{g}_0, \vec{p}_0)$ cannot be true given this information. It thus follows that $(\vec{g}_0, \vec{p}_0)$ is partially identified by the set $\mathcal{GP}\backslash\{(\vec{g}_*, \vec{p}_*)\}$, showing that $\mathcal{GP}$ is not a sharp identified set. The claimed statement follows by the contrapositive argument. □

A.8. Connectedness of the identified sets

Proposition 1 (*Interval*)**.** *Suppose that the assumptions in Theorem 1 are satisfied. If $\mathcal{GP}$ is a connected set, then so is the identified set Θ_I. In particular, its projection Θ_I to each coordinate is given by an interval.*

Proof. The assumptions in Theorem 1, namely that $\beta \in (0, 1)$ is true and that the rank condition (3.3) is satisfied for all $\vec{g} \in \mathcal{G}$ and $\vec{p} \in \mathcal{P}$, guarantee that the map $(\vec{g}, \vec{p}) \overset{\phi}{\mapsto} \left[\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{H}(\vec{g}, \vec{p}, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}, \vec{p}, \beta)'\tilde{Y}(\vec{g}, \vec{p}, \beta)\right]$ is continuous on $\mathcal{GP}$. Since a continuous function maps a connected set to a connected set, the identified set $\Theta_I = \phi(\mathcal{GP})$ is connected. Note that the projection mapping ψ is also continuous, and hence the projection $\psi(\Theta_I)$ of the connected identified set Θ_I is also connected. If ψ maps to $\mathbb{R}$, then $\psi(\Theta_I)$ is an interval since any connected set in $\mathbb{R}$ is an interval. □

A similar result holds for the identified set for the counterfactual policy outcomes.

Proposition 2 (*Interval*)**.** *Suppose that the assumptions in Corollary 2 are satisfied. If $\mathcal{GP}$ is a connected set and the counterfactual mapping Γ is continuous, then the identified set $\mathcal{C}_I$ of the counterfactual outcome C is interval-valued.*

Proof. Under the stated assumptions, the map ϕ introduced in the proof of Proposition 1 is continuous. Since Γ is continuous and $\mathcal{GP}$ is a connected set by assumption, it follows that $\mathcal{C}_I = \left\{\Gamma(\phi(\vec{g}, \vec{p}), \vec{g}, \vec{p}) \mid (\vec{g}, \vec{p}) \in \mathcal{GP}\right\}$ is also connected. Since C is scalar-valued, the connected identified set $\mathcal{C}_I \in \mathbb{R}$ must be an interval. □

A.9. Implementation

A.9.1. The criterion

Theorem 2 provides the sharp identified set $\mathcal{PG}^\dagger$ for the CCPs and the transition probabilities. Corollary 1 provides the associated sharp identified set $\Theta_I^\dagger$ for the structural parameters. Because of the closed-form partial identification and closed-form restrictions, one could certainly proceed with a constructive analog method of estimating the identified sets in practice. In this section, we propose a criterion-based approach to estimating the sharp identified set, which is compatible with an existing practitioner-friendly method of inference.

Given a preliminary set $\mathcal{G} \times \mathcal{P}$ (i.e., the set directly identified by observed data without structural restrictions), recall the sharp identified set is defined by

$$\mathcal{GP}^\dagger := \bigcup_{\vec{g}\in\mathcal{G}} \left(\{\vec{g}\} \times \mathcal{P}(\vec{g})\right) \qquad \text{where} \qquad \mathcal{P}(\vec{g}) := \left\{\vec{p} \in \mathcal{P} \;\middle|\; \vec{p} = \Psi_{\vec{g}}(\vec{p}), \vartheta(\vec{g}, \vec{p}) \in \Theta \right\}.$$

Equivalently, the sharp identified set can be characterized as the set of zeros of the criterion function $Q : \mathcal{G} \times \mathcal{P} \to \mathbb{R}$ defined by

$$\begin{aligned} &Q(\vec{g}, \vec{p}) := d_1\left(\vec{g}, \mathcal{G}\right) + d_2\left(\vec{p}, \mathcal{P}\right) + \left\|\vec{p} - \Psi_{\vec{g}}(\vec{p})\right\|^2 + d_3\left(\vartheta(\vec{p}, \vec{g}), \Theta\right) \qquad \text{where} \\ &d_l\left(\vec{s}, \mathcal{S}\right) := \inf\left\{\rho_l(\vec{s}, \vec{s}') \mid \vec{s}' \in \mathcal{S}\right\} \qquad \text{for each } l \in \{1, 2, 3\} \end{aligned}$$

with the Euclidean norm $\|\cdot\|$ and suitable metrics ρ_1, ρ_2, and ρ_3. The first term in $Q(\vec{g}, \vec{p})$ ensures that $\vec{g}$ be contained in $\mathcal{G}$ because the union is taken for $\vec{g} \in \mathcal{G}$ in the definition of $\mathcal{GP}^\dagger$. Similarly, the second term ensures that $\vec{p}$ be contained in $\mathcal{P}$ because the definition of $\mathcal{P}(\vec{g})$ requires $\vec{p} \in \mathcal{P}$. The third term ensures that the fixed point restriction be satisfied, which is required in the above definition of $\mathcal{P}(\vec{g})$. The fourth term ensures that the identified set for the structural parameters is contained in an admissible parameter set, which is also required in the above definition of $\mathcal{P}(\vec{g})$. As such, each of these four terms is indispensable for characterization of the sharp identified set $\mathcal{GP}^\dagger$.

In case $\vec{g}_{a,x}$ and $\vec{p}_x$ are observed for some (a, x), we can write the first two terms of $Q(\vec{g}, \vec{p})$ simply as

$$\begin{aligned} d_1\left(\vec{g}, \widehat{\mathcal{G}}\right) &= \sum_{(a,x):\ \text{observed}} \left\|\widehat{g}^*_{a,x} - \widehat{g}^{**}_{a,x} \cdot \vec{g}_{a,x}\right\|^2 \qquad \text{and} \\ d_2\left(\vec{p}, \widehat{\mathcal{P}}\right) &= \sum_{x:\ \text{observed}} \left\|\widehat{p}^*_x - \widehat{p}^{**}_x \cdot \vec{p}_x\right\|^2, \end{aligned}$$

where $\widehat{g}^*_{a,x}/\widehat{g}^{**}_{a,x}$ and $\widehat{p}^*_x/\widehat{p}^{**}_x$ constitute sample-mean estimates for $\vec{g}_{a,x}$ and $\vec{p}_x$, respectively, i.e.,

$$\begin{aligned} \widehat{g}^*_{a,x} &= \left(\sum_{i=1}^{n}\sum_{t=1}^{T-1} \frac{\mathbb{1}\{(X_{i,t+1}, A_{i,t}, X_{i,t}) = (1, a, x)\}}{n(T-1)}, \ldots, \sum_{i=1}^{n}\sum_{t=1}^{T-1} \frac{\mathbb{1}\{(X_{i,t+1}, A_{i,t}, X_{i,t}) = (\bar{x}, a, x)\}}{n(T-1)}\right) \\ \widehat{g}^{**}_{a,x} &= \sum_{i=1}^{n}\sum_{t=1}^{T-1} \frac{\mathbb{1}\{(A_{i,t}, X_{i,t}) = (a, x)\}}{n(T-1)} \qquad \text{and} \end{aligned}$$

$$\begin{aligned} \widehat{p}^*_x &= \left(\sum_{i=1}^{n}\sum_{t=1}^{T} \frac{\mathbb{1}\{(A_{i,t}, X_{i,t}) = (0, x)\}}{nT}, \sum_{i=1}^{n}\sum_{t=1}^{T} \frac{\mathbb{1}\{(A_{i,t}, X_{i,t}) = (1, x)\}}{nT}\right) \\ \widehat{p}^{**}_x &= \sum_{i=1}^{n}\sum_{t=1}^{T} \frac{\mathbb{1}\{X_{i,t} = x\}}{nT}. \end{aligned}$$

Thus the sample criterion $\widehat{Q}_n$ can be given by

$$\begin{aligned} \widehat{Q}_n(\vec{g}, \vec{p}) := &\sum_{(a,x):\ \text{observed}} \left\|\widehat{g}^*_{a,x} - \widehat{g}^{**}_{a,x} \cdot \vec{g}_{a,x}\right\|^2 + \sum_{x:\ \text{observed}} \left\|\widehat{p}^*_x - \widehat{p}^{**}_x \cdot \vec{p}_x\right\|^2 \\ &+ \left\|\vec{p} - \Psi_{\vec{g}}(\vec{p})\right\|^2 + d\left(\vartheta(\vec{p}, \vec{g}), \Theta\right). \end{aligned}$$

Example 1 (*Dynamic Model of Entry and Exit, Continued*)**.** Consider Example 1 again. Recall that Z_t is observed up to $Z_t \leqslant T$. In this case, (a, s, z) is observed for all $(a, s, z) \in \mathcal{A} \times \mathcal{S} \times \{1, \ldots, T-1\}$, and (s, z) is observed for all $(s, z) \in \mathcal{S} \times \{1, \ldots, T\}$. Thus, the sample criterion $\widehat{Q}_n$ is

$$\begin{aligned} \widehat{Q}_n(\vec{g}, \vec{p}) := &\sum_{a=0}^{1}\sum_{s=0}^{1}\sum_{z=1}^{T-1} \left\|\widehat{g}^*_{a,(s,z)} - \widehat{g}^{**}_{a,(s,z)} \cdot \vec{g}_{a,(s,z)}\right\|^2 + \sum_{s=0}^{1}\sum_{z=1}^{T} \left\|\widehat{p}^*_{(s,z)} - \widehat{p}^{**}_{(s,z)} \cdot \vec{p}_{(s,z)}\right\|^2 \\ &+ \left\|\vec{p} - \Psi_{\vec{g}}(\vec{p})\right\|^2 + d\left(\vartheta(\vec{p}, \vec{g}), \Theta\right). \end{aligned}$$

To impose $\Theta = \left\{(\pi_1, \ldots, \pi_{\bar{z}}, \phi, \kappa)' \in I_1 \times \cdots \times I_{\bar{z}} \times I_\phi \times I_\kappa \mid \pi_1 \leqslant \cdots \leqslant \pi_{\bar{z}}\right\}$, the last term in the above sample criterion can be written as

$$d(\theta, \Theta) = \sum_{\zeta=1}^{\bar{z}-1} \left|\theta_\zeta - \theta_{\zeta+1}\right|_+^2$$

where $|\cdot|_+$ returns $\cdot$ if it is positive and zero otherwise. □

A.9.2. Computation

Kline and Tamer (2013) propose numerical procedures to compute the set of zeros of criterion functions. We adapt their suggestion to our framework as follows. Define the function

$$\tilde{f}_\varkappa(\vec{g}, \vec{p}) = \exp\left(\frac{-Q(\vec{g}, \vec{p})}{\varkappa}\right)$$

where a small number $\varkappa > 0$ is a tuning parameter. For this pseudo-density function, we implement the following MCMC algorithm — the slice sampling.

1. Let $(\vec{g}_1, \vec{p}_1) \in \arg\min_{(\vec{g}, \vec{p}) \in \mathcal{G} \times \mathcal{P}} Q(\vec{g}, \vec{p})$ be an initial point.
2. For $(\vec{g}_{m-1}, \vec{p}_{m-1})$, sample $u_m \in \left(0, \tilde{f}_\varkappa(\vec{g}_{m-1}, \vec{p}_{m-1})\right)$ uniformly.
3. Sample $(\vec{g}'_m, \vec{p}'_m) \in \mathcal{G} \times \mathcal{P}$ uniformly.
4. If $\tilde{f}_\varkappa(\vec{g}'_m, \vec{p}'_m) \geqslant u_m$, then accept $(\vec{g}'_m, \vec{p}'_m)$ as $(\vec{g}_m, \vec{p}_m)$, increment m, and move to Step 2.
5. If $\tilde{f}_\varkappa(\vec{g}'_m, \vec{p}'_m) < u_m$, then reject $(\vec{g}'_m, \vec{p}'_m)$ and move to Step 2 without incrementing m.
6. Repeat steps 2–5 to obtain M points $\left\{(\vec{g}_m, \vec{p}_m)\right\}_{m=1}^M$.

With our model with the fixed point restriction, the first step may be established using the iterative algorithm of Aguirregabiria and Mira (2002, 2007) and Kasahara and Shimotsu (2012). The set $\left\{(\vec{g}_m, \vec{p}_m)\right\}_{m=1}^M$ of M points obtained through this procedure approximates the sharp identified set $\mathcal{GP}^\dagger$. We remark that it is desired to set M to a large value when the state space is large and especially when the number of states non-visited in data is large.

Once the sharp identified set $\mathcal{GP}^\dagger$ of the CCPs and the transition probabilities is numerically approximated by a sample $\left\{(\vec{g}_m, \vec{p}_m)\right\}_{m=1}^M$, one can substitute these M points in the formula (3.4) to approximate the identified set $\Theta_I^\dagger$ of the structural parameters. Specifically, $\Theta_I^\dagger$ is approximated by the following set of M points.

$$\left\{\vartheta(\vec{g}_m, \vec{p}_m)\right\}_{m=1}^M = \left\{\left[\tilde{H}(\vec{g}_m, \vec{p}_m, \beta)'\tilde{H}(\vec{g}_m, \vec{p}_m, \beta)\right]^{-1}\left[\tilde{H}(\vec{g}_m, \vec{p}_m, \beta)'\tilde{Y}(\vec{g}_m, \vec{p}_m, \beta)\right]\right\}_{m=1}^M .$$

With this numerical method to approximate the identified sets, we can directly apply the Bayesian bootstrap method proposed by Kline and Tamer (2013).

A.9.3. On the method of inference

Formal large sample properties of the method described above are presented in Section 5 of Kline and Tamer (2013). We chose to employ this particular method of inference for partially identified parameters over numerous alternative methods that are available in the literature for the following three reasons. First, our identified set is not characterized by moment inequalities, and hence we could not take advantage of the rich set of inference methods for moment inequality models. Second, among the methods of inference based on criterion functions, we wanted to circumvent a grid search given generally large dimensions of parameter vectors in structural dynamic discrete choice models. Hence, we eliminated those methods of inference based on test inversion from our candidate list. The MCMC algorithm of Kline and Tamer (2013) allows to draw a credible region by sampled points. Third, as mentioned above, we can feed a fixed-point value obtained from the iterative algorithm of Aguirregabiria and Mira (2002, 2007) and Kasahara and Shimotsu (2012) as the initial value for the MCMC algorithm of Kline and Tamer (2013), which is an advantage specific to the model of our interest.

A.10. Figures

See Fig. 1.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2023.04.005.

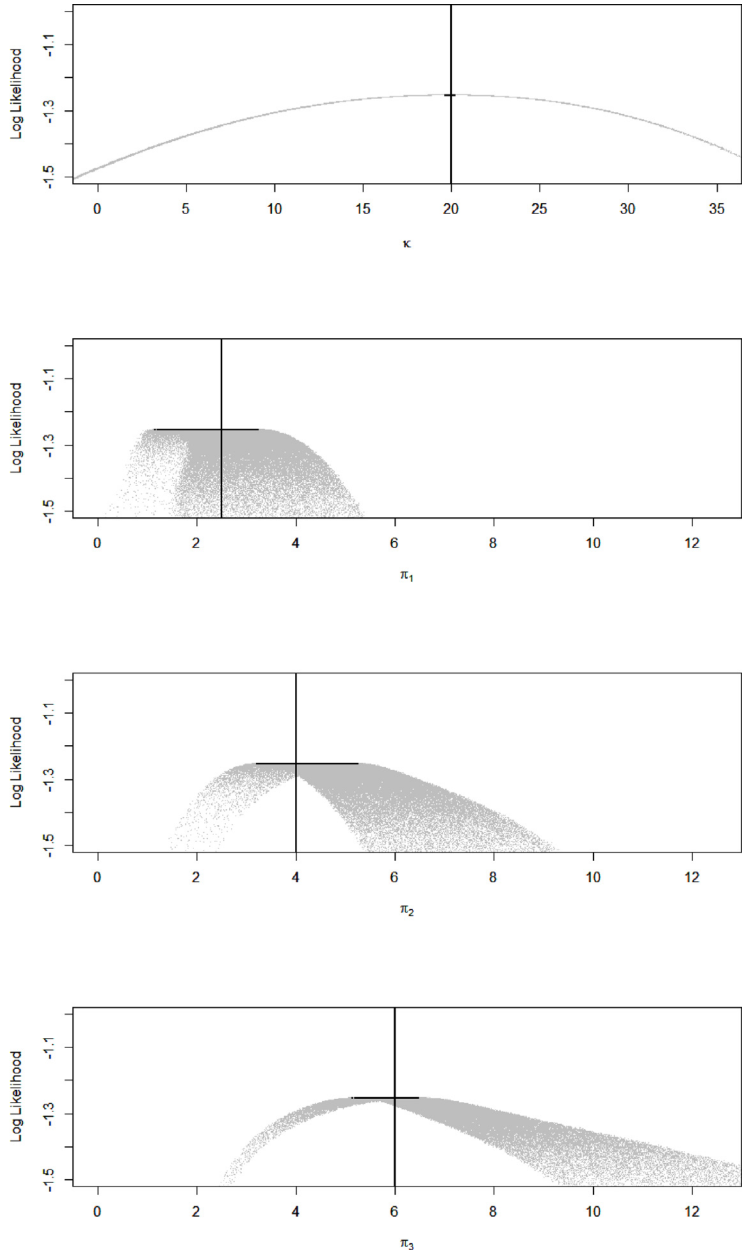

Fig. 1. Plots of likelihood values over parameter values with the monotonicity restriction and the sample size of $N = 1,000,000$. The four graphs display the profiled plots over κ, π_1, π_2, and π_3 from the top to the bottom. The vertical lines indicate the true parameter values. The black dots indicate the bottom one percentile in terms of our objective Q. That these black dots coincide with the region of maximum likelihood value evidence the sharpness of our identified set — see Tamer (2010) and Chen et al. (2011). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

Abbring, Jaap H., Daljord, Øystein, 2020. Identifying the discount factor in dynamic discrete choice models. Quant. Econ. 11 (2), 471–501.

Aguirregabiria, Victor, Gu, Jiaying, Luo, Yao, 2021. Sufficient statistics for unobserved heterogeneity in structural dynamic logit models. J. Econometrics 223 (2), 280–311.

Aguirregabiria, V., Mira, P., 2002. Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models. Econometrica 70 (4), 1519–1543.

Aguirregabiria, V., Mira, P., 2007. Sequential estimation of dynamic discrete games. Econometrica 75 (1), 1–53.

Andrews, D.W., Barwick, P.J., 2012. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. Econometrica 80 (6), 2805–2826.

Andrews, D.W., Guggenberger, P., 2009. Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities. Econom. Theory 25 (3), 669–709.
Andrews, D.W., Shi, X., 2013. Inference based on conditional moment inequalities. Econometrica 81 (2), 609–666.
Andrews, D.W., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78 (1), 119–157.
Arcidiacono, P., Miller, R.A., 2011. Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. Econometrica 79 (6), 1823–1867.
Arcidiacono, P., Miller, R.A., 2020. Identifying dynamic discrete choice models off short panels. J. Econometrics 215 (2), 473–485.
Armstrong, T.B., 2014. Weighted KS statistics for inference on conditional moment inequalities. J. Econometrics 181 (2), 92–116.
Beresteanu, A., Molinari, F., 2008. Asymptotic properties for a class of partially identified models. Econometrica 76 (4), 763–814.
Berry, Steven T., Compiani, Giovanni, 2020. An Instrumental Variable Approach to Dynamic Models. No. w27756, National Bureau of Economic Research.
Buchholz, N., Shum, M., Xu, H., 2021. Semiparametric estimation of dynamic discrete choice models. J. Econometrics 223 (2), 312–327.
Bugni, F.A., 2010. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. Econometrica 78 (2), 735–753.
Bugni, F.A., Canay, I.A., Shi, X., 2017. Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models. Working Paper.
Canay, I.A., 2010. EL inference for partially identified models: Large deviations optimality and bootstrap validity. J. Econometrics 156 (2), 408–425.
Chen, X., Christensen, T., Tamer, E., 2015. MCMC Confidence Sets for Identified Sets. Working Paper.
Chen, X., Tamer, E., Torgovitsky, A., 2011. Sensitivity Analysis in Semiparametric Likelihood Models. Working Paper.
Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75 (5), 1243–1284.
Hotz, V.J., Miller, R.A., 1993. Conditional choice probabilities and the estimation of dynamic models. Rev. Econom. Stud. 60 (3), 497–529.
Hotz, V.J., Miller, R.A., Sanders, S., Smith, J., 1994. A simulation estimator for dynamic models of discrete choice. Rev. Econom. Stud. 61 (2), 265–589.
Hu, Y., Sasaki, Y., 2018. Closed-form identification of dynamic discrete choice models with proxies for unobserved state variables. Econom. Theory 34 (1), 166–185.
Hu, Y., Shum, M., 2012. Nonparametric identification of dynamic models with unobserved state variables. J. Econometrics 171 (1), 32–44.
Hu, Y., Xin, Y., 2020. Identification and estimation of dynamic structural models with unobserved choices. Available at SSRN: https://ssrn.com/abstract=3634910.
Igami, M., 2017. Estimating the innovator's dilemma: Structural analysis of creative destruction in the hard disk drive industry, 1981–1998. J. Political Economy 125 (3), 798–847.
Igami, M., Uetake, K., 2016. Mergers, Innovation, and Entry-Exit Dynamics: Consolidation of the Hard Disk Drive Industry, 1996-2015. Working Paper, Yale University..
Imbens, G.W., Manski, C.F., 2004. Confidence intervals for partially identified parameters. Econometrica 72 (6), 1845–1857.
Kaido, H., Molinari, F., Stoye, J., 2016. Inference for Projections of Identified Sets. Manuscript, Cornell University.
Kalouptsidi, Myrto, Kitamura, Yuichi, Lima, Lucas, Souza-Rodrigues, Eduardo A., 2020a. Partial Identification and Inference for Dynamic Models and Counterfactuals. No. w26761, National Bureau of Economic Research.
Kalouptsidi, Myrto, Scott, Paul T., Souza-Rodrigues, Eduardo, 2020b. Identification of counterfactuals in dynamic discrete choice models. Quant. Econ. 12 (2), 351–403.
Kalouptsidi, Myrto, Scott, Paul T., Souza-Rodrigues, Eduardo, 2021. Linear IV regression estimators for structural dynamic discrete choice models. J. Econometrics 222 (1), 778–804.
Kasahara, H., Shimotsu, K., 2009. Nonparametric identification of finite mixture models of dynamic discrete choices. Econometrica 77 (1), 135–175.
Kasahara, H., Shimotsu, K., 2012. Sequential estimation of structural models with a fixed point constraint. Econometrica 80 (5), 2303–2319.
Khwaja, A., 2010. Estimating willingness to pay for medicare using a dynamic life-cycle model of demand for health insurance. J. Econometrics 156 (1), 130–147.
Kitagawa, T., 2012. Estimation and Inference for Set-Identified Parameters using Posterior Lower Probability. Manuscript, UCL.
Kline, B., Tamer, E., 2013. Default Bayesian Inference in a Class of Partially Identified Models. Working Paper.
Liao, Y., Simoni, A., 2016. Bayesian Inference for Partially Identified Convex Models: Is It Valid for Frequentist Inference? Working Paper.
Magnac, T., Thesmar, D., 2002. Identifying dynamic discrete decision processes. Econometrica 70 (2), 801–816.
Moon, H.R., Schorfheide, F., 2012. Bayesian and frequentist inference in partially identified models. Econometrica 80 (2), 755–782.
Norets, A., Tang, X., 2014. Semiparametric inference in dynamic binary choice models. Rev. Econom. Stud. 81 (3), 1229–1262.
Otsu, T., Pesendorfer, M., Sasaki, Y., Takahashi, Y., 2020. Estimation of (Static Or Dynamic) Games under Equilibrium Multiplicity. Centre for Economic Policy Research, 2020.
Pesendorfer, M., Schmidt-Dengler, P., 2008. Asymptotic least squares estimators for dynamic games. Rev. Econom. Stud. 75 (3), 901–928.
Romano, J.P., Shaikh, A.M., 2010. Inference for the identified set in partially identified econometric models. Econometrica 78 (1), 169–211.
Romano, J.P., Shaikh, A.M., Wolf, M., 2014. A practical two-step method for testing moment inequalities. Econometrica 82 (5), 1979–2002.
Rosen, A.M., 2008. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. J. Econometrics 146 (1), 107–117.
Rust, John, 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica 99, 9–1033.
Rust, J., 1994. Structural estimation of Markov decision processes. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North Holland, Amsterdam, pp. 3081–3143.
Sanches, F., Silva, Jr., D., Srisuma, T., 2016. Ordinary least squares estimation of a dynamic game model. Internat. Econom. Rev. 57 (2), 623–634.
Sasaki, Y., 2015. Heterogeneity and selection in dynamic panel data. J. Econometrics 188 (1), 236–249.
Song, K., 2014. Point decisions for interval-identified parameters. Econom. Theory 30 (2), 334–356.
Stoye, J., 2009. More on confidence intervals for partially identified parameters. Econometrica 77 (4), 1299–1315.
Takahashi, Y., 2015. Estimating a war of attrition: The case of the U.S. movie theater industry. Am. Econ. Rev. 105 (7), 2204–2241.
Tamer, E., 2010. Partial identification in econometrics. Annu. Rev. Econ. 2, 167–195.