Essays in Industrial Organization and Econometrics

by

Jason R. Blevins

Department of Economics
Duke University

Date: _____

Approved:

_____
Han Hong, Co-Advisor

_____
Shakeeb Khan, Co-Advisor

_____
Paul B. Ellickson

_____
Andrew Sweeting

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2010

ABSTRACT
(Economics)

Essays in Industrial Organization and Econometrics

by

Jason R. Blevins

Department of Economics
Duke University

Date: _____

Approved:

_____
Han Hong, Co-Advisor

_____
Shakeeb Khan, Co-Advisor

_____
Paul B. Ellickson

_____
Andrew Sweeting

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2010

Abstract

This dissertation consists of three chapters relating to identification and inference in dynamic microeconometric models including dynamic discrete games with many players, dynamic games with discrete and continuous choices, and semiparametric binary choice and duration panel data models.

The first chapter provides a framework for estimating large-scale dynamic discrete choice models (both single- and multi-agent models) in continuous time. The advantage of working in continuous time is that state changes occur sequentially, rather than simultaneously, avoiding a substantial curse of dimensionality that arises in multi-agent settings. Eliminating this computational bottleneck is the key to providing a seamless link between estimating the model and performing post-estimation counterfactuals. While recently developed two-step estimation techniques have made it possible to estimate large-scale problems, solving for equilibria remains computationally challenging. In many cases, the models that applied researchers estimate do not match the models that are then used to perform counterfactuals. By modeling decisions in continuous time, we are able to take advantage of the recent advances in estimation while preserving a tight link between estimation and policy experiments. We also consider estimation in situations with imperfectly sampled data, such as when we do not observe the decision not to move, or when data is aggregated over time, such as when only discrete-time data are available at regularly spaced intervals. We illustrate the power of our framework using several large-scale Monte Carlo experiments.

The second chapter considers semiparametric panel data binary choice and duration models with fixed effects. Such models are point identified when at least one regressor has full support on the real line. It is common in practice, however, to have only discrete or continuous, but possibly bounded, regressors. We focus on identification, estimation, and inference for the identified set in such cases, when the parameters of interest may only be partially identified. We develop a set of general results for criterion-function-based estimation and inference in partially identified models which can be applied to both regular and irregular models. We apply our general results first to a fixed effects binary choice panel data model where we obtain a sharp characterization of the identified set and propose a consistent set estimator, establishing its rate of convergence under different conditions. Rates arbitrarily close to $n^{-1/3}$ are possible when a continuous, but possibly bounded, regressor is present. When all regressors are discrete the estimates converge arbitrarily fast to the identified set. We also propose a subsampling-based procedure for constructing confidence regions in the models we consider. Finally, we carry out a series of Monte Carlo experiments to illustrate and evaluate the proposed procedures. We also consider extensions to other fixed effects panel data models such as binary choice models with lagged dependent variables and duration models.

The third chapter considers nonparametric identification of dynamic games of incomplete information in which players make both discrete and continuous choices. Such models are commonly used in applied work in industrial organization where, for example, firms make discrete entry and exit decisions followed by continuous investment decisions. We first review existing identification results for single agent dynamic discrete choice models before turning to single-agent models with an additional continuous choice variable and finally to multi-agent models with both discrete and continuous choices. We provide conditions for nonparametric identification of the utility function in both cases.

Dedication

To my lovely wife, Kelly, without whom this dissertation would not have been possible.

Contents

List of Tables

List of Figures

Acknowledgements

I am deeply indebted to the members of my dissertation committee, Han Hong, Shakeeb Khan, Paul Ellickson, and Andrew Sweeting, for their invaluable guidance, encouragement, and support throughout my work on this project and many others. I am also grateful to Peter Arcidiacono, Patrick Bayer, Arie Beresteanu, Joseph Hotz, Ahmed Khwaja, Curtis Taylor, and Chris Timmins for their comments, insights, and support throughout my years at Duke. Chapter 1 is based on joint work with Peter Arcidiacono, Patrick Bayer, and Paul Ellickson.

I am also grateful to Duke University for financial support through the Summer Research Fellowship in 2007 and 2008 and to Han Hong and the Department of Economics at Stanford University for hosting me in July of 2008, during which much of Chapter 3 was written.

1

Estimation of Dynamic Discrete Choice Models in Continuous Time

1.1 Introduction

Empirical models of single-agent dynamic discrete choice (DDC) problems have a rich history in structural applied microeconometrics, starting with the pioneering work of Gotz and McCall (1980), Miller (1984), Wolpin (1984), Pakes (1986), and Rust (1987). These methods have been applied to a wide range of economic problems including investment under uncertainty, savings and retirement, human capital accumulation, fertility decisions, labor market participation, resource extraction, and political decisions, among many others. Because dynamic decision problems are naturally high-dimensional, the empirical DDC literature has been accompanied from the outset by a parallel methodological literature aimed at reducing the computational burden of both estimating and computing these models.

The computational challenges raised by the dimensionality of these problems are even greater in the context of multi-agent strategic games, where the simultaneous actions of competing players introduces a further curse of dimensionality in computing

expectations over rivals' actions. Although a recent series of papers (Aguirregabiria and Mira, 2007; Bajari, Benkard, and Levin, 2007; Pesendorfer and Schmidt-Dengler, 2007; Pakes, Ostrovsky, and Berry, 2007) have shown how to extend two-step *estimation* techniques, originally developed by Hotz and Miller (1993) and Hotz, Miller, Sanders, and Smith (1994) in the context of single-agent dynamics, to more complex multi-agent settings, the computation of these models remains formidable, despite a growing number of methods for solving for equilibria (Pakes and McGuire, 1994, 2001; Doraszelski and Satterthwaite, 2010).[1]

A curse of dimensionality naturally arises in simultaneous move games because, in order to solve for their optimal policies, players must form expectations over all combinations of actions that each of their rivals can take. The burden of computing these expectations grows exponentially in the number of players and so, in many applications, the model that researchers can estimate (using two-step procedures) is far richer than what can be used for counterfactual policy simulations, leading some to suggest alternatives to the Markov perfect equilibrium concept in which firms condition on long run averages (regarding rivals' states) instead of current information (Weintraub, Benkard, and Van Roy, 2008). The goal of this chapter is to exploit the sequential structure of continuous time games to break the computational curse, create a tight link between estimation and counterfactuals, and open the door to more complex and realistic models of strategic interaction.

We are not the first to address these computational challenges. Pakes and McGuire (2001) extend their seminal approach to solving dynamic games (Pakes and McGuire, 1994) by replacing explicit integration with simulation and utilizing an adaptive algorithm that targets only the recurrent class of states. Their computational approach is able to

[1] Two-step estimation of dynamic discrete games was originally proposed by Rust (1994a). Rust recommended substituting non-parametric estimates of rivals' reaction functions into each player's dynamic optimization problem, turning a complex equilibrium solution into a collection of simpler games against nature.

alleviate the curse of dimensionality that arises when calculating expectations over future states as well as the increasing size of the state space itself, but does rely on the recurrent class being small. In theoretical work that is closest to ours, Doraszelski and Judd (2008) exploit the structure of continuous time to break the curse of dimensionality associated with the calculation of expectations over rival actions. Players in their model make continuous decisions that control the hazard rate of state changes (e.g., choose investment which results stochastically in a discrete productivity gain). Because state changes occur only one agent at a time, the dimension of expectations over rival actions grows linearly in the number of players, rather than exponentially, resulting in computation times that are orders of magnitude faster than those of discrete time.

Building on the theoretical insights of Doraszelski and Judd, we seek to connect the computational advantages of continuous time with the empirical tractability of discrete choice models. To do so, we recast the dynamic decision problem as one in which competing Poisson processes stochastically control when players are able to move, with players then facing a standard discrete choice problem when given the opportunity to make a decision. This structure results in a simple, yet flexible mathematical structure that is computationally light enough to make even full solution (i.e. nested fixed point) estimation feasible for very large problems. The model also inherits many features of the standard discrete choice framework and, as a result, many of the insights and tools commonly used in discrete time settings, such as two-step CCP (conditional choice probability) estimation, are directly applicable within our continuous time approach, further relaxing the computational burden of estimation. Having estimated the model, it is straightforward to re-solve the model in continuous time to perform counterfactuals or simulate data. The continuous time framework thus offers a seamless link between estimation and computation, allowing the same underlying model to be used throughout. We demonstrate the power of our approach using several large scale Monte Carlo

experiments, many of which would be infeasible using previous methods.

Our framework easily accommodates a number of more complex sampling schemes, including some that are especially challenging in a discrete time setting. We show how to handle situations in which certain observations are missing (e.g. passive actions, such as the decision not to invest) or where the data are only sampled at discrete intervals (e.g. quarterly or yearly). Both extensions are likely to be empirically relevant given the limitations of publicly available datasets (most of which are collected at regular intervals, rather than in real time). The mathematical structure of continuous time makes time aggregation simple. It is straightforward, for example, to calculate the likelihood of transitioning from any initial state to any final state over any discrete period of time in a manner that explicitly accounts for the complex combinations of possible actions and state changes that might have led to that final state. Time aggregation is much more difficult in discrete time and, as a result, researchers generally adopt the convention that players move at the same periodicity with which the data is observed, even when players clearly move far more frequently. The potential advantages of modeling decisions using a continuous time framework extend beyond computation, highlighting aspects of strategic interaction that are muted by discrete time (e.g. first-mover advantage) and mitigating unnatural implications that can arise from simultaneity (e.g. ex post regret). The empirical relevance of these issues will depend on the institutional setting. In this way, a continuous time approach provides researchers with the option of modeling moves as sequential rather than simultaneous when this accords more closely with the actual economic setting, even if the data is only observed at regular intervals.

The chapter is structured as follows. Section 1.2 reviews some basic properties of continuous time Markov jump processes. Section 1.3 introduces our model in a simple single-agent context in order to build intuition. Section 1.4 extends the model to the multi-agent setting. Concrete and canonical examples are provided in both cases. Section 1.5

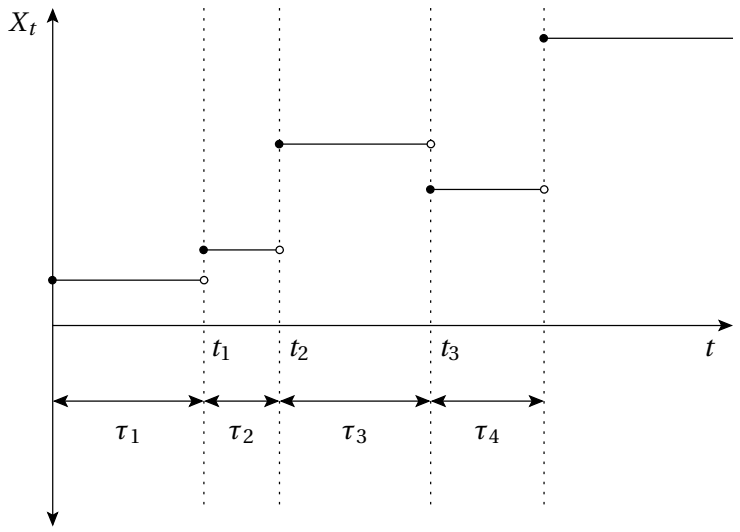

FIGURE 1.1: Markov jump process: a representative sample path with jumps at times t_n and inter-arrival times τ_n.

develops our estimators, including both full-solution and two-step approaches, and discusses issues associated with partial observation and time aggregation. Section 1.6 contains the results of several Monte Carlo studies relating to full-solution and two-step estimation in both settings, including time aggregation, unobserved passive moves, and comparisons of computational times. Section 1.7 concludes.

1.2 Background

The models we describe below are based on Markov jump processes and we briefly review their properties here. A Markov jump process is a stochastic process X_t indexed by $t \in [0,\infty)$ taking values in some discrete state space $\mathscr{X}$. If we begin observing this process at some arbitrary time t and state X_t, it will remain in this state for a duration of random length τ before transitioning to some other state $X_{t+\tau}$. The trajectory of such a process is a piecewise-constant, right-continuous function of time. This is illustrated in Figure 1.1, where a single realization x_t is plotted along with corresponding jump times t_n and inter-arrival times τ_n, with n denoting the n-th jump.

Jumps occur according to a Poisson process and the length of time between jumps is therefore exponentially distributed. The probability density function (pdf) of the exponential distribution with rate parameter $\lambda > 0$ is

$$f(x;\lambda) = \begin{cases} \lambda \, \mathrm{e}^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

and the cumulative distribution function (cdf) is

$$F(x;\lambda) = \begin{cases} 1 - \mathrm{e}^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

The mean is $1/\lambda$, the inverse of the rate parameter or frequency, and the variance is $1/\lambda^2$.

We consider stationary processes with finite state spaces $\mathscr{X} = \{1, \ldots, K\}$. Before proceeding, we first review some fundamental properties of Markov jump processes, presented without proof. For details see, for example, Karlin and Taylor (1975, section 4.8).

A finite Markov jump process can be summarized by it's *intensity matrix*

$$Q = \begin{bmatrix} -q_{11} & q_{12} & \cdots & q_{1K} \\ q_{21} & -q_{22} & \cdots & q_{2K} \\ \vdots & \vdots & \vdots & \vdots \\ q_{K1} & q_{K2} & \cdots & -q_{KK} \end{bmatrix}$$

where for $i \neq j$

$$q_{ij} = \lim_{h \to 0} \frac{\Pr\left(X_{t+h} = j | X_t = i\right)}{h}$$

represents the probability per unit of time that the system transitions from i to j and

$$q_{ii} = \sum_{j \neq i} q_{ij}$$

denotes the rate at which the system transitions out of state i. Thus, transitions out of i follow an exponential distribution with rate parameter q_{ii} and, conditional on leaving state i, the system transitions to $j \neq i$ with probability

$$p_{ij} = \frac{q_{ij}}{\sum_{k \neq i} q_{ik}}. \tag{1.1}$$

Finally, let $P_{ij}(t) = \Pr\left(X_{t+s} = j \mid X_s = i\right)$ denote the probability that the system has transitioned to state j after a period of length t given that it was initially in state i. Let $P(t) = (P_{ij}(t))$ denote the corresponding matrix of these probabilities, the *transition matrix*. $P(t)$ can be found as the unique solution to the system of ordinary differential equations

$$P'(t) = P(t)Q,$$

$$P(0) = I.$$

frequently referred to as the forward equations. It follows that

$$P(t) = \mathrm{e}^{tQ} = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!}. \tag{1.2}$$

This quantity is the matrix exponential, the matrix analog of the scalar exponential e^x.[2]

Finally, we review some properties of the exponential distribution which will be required for constructing the value function later. In particular, we note that if there are n competing Poisson processes (or exponential distributions) with rates λ_i for $i = 1, \ldots, n$, then distribution of the minimum wait time is exponential with rate $\sum_{i=1}^{n} \lambda_i$ and, furthermore, conditional on an arrival, the probability that it is due to process i is $\lambda_i / \sum_{j=1}^{n} \lambda_j$. These properties are well known, but we present the following proposition for completeness.

[2] In practice, since we cannot calculate the infinite sum (1.2) directly, we compute e^{tQ} using known algorithms implemented in the Fortran package Expokit (Sidje, 1998).

Proposition 1.1. *Suppose* $\tau_i \sim \text{Expo}(\lambda_i)$*, for* $i = 1, \ldots, n$*, are independent and define* $\tau \equiv \min_i \tau_i$ *and* $\iota \equiv \arg\min_i \tau_i$*. Then*

$$\tau \sim \text{Expo}(\lambda_1 + \cdots + \lambda_n)$$

and

$$\Pr(\iota = i) = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}.$$

Proof. See Appendix A. ■

This proposition allows us to treat the n competing Poisson processes $(\tau_1, \ldots, \tau_n)$ as a single joint process (τ, ι) where the joint distribution is given above.

1.3 Single-Agent Dynamic Discrete Choice Models

In this section, we introduce a dynamic discrete choice model of single-agent decision-making in continuous time. The single-agent model provides a simple setting in which to describe the main features of our continuous time framework; we show how these extend directly to multi-agent settings in the following section. We begin this section by laying out the notation and structure of the model in a general setting. We then introduce an example—the classic Rust bus engine replacement problem—to fix ideas. The example also serves as the basis for the Monte Carlo analysis presented later in the paper.

Consider a single agent decision problem in which time is a continuous variable $t \in [0, \infty)$. The state of the model at any time t can be summarized by an element x of some finite state space $\mathscr{X}$. Two competing Poisson processes drive the dynamics of the model. First, a continuous-time Markov jump process on $\mathscr{X}$ with intensity matrix Q_0 represents moves by nature—state changes that aren't a direct result of actions by the agent. At each time t, if a jump occurs next, the state jumps immediately to the new value. The agent may not influence this process. Second, a Poisson arrival process with

rate λ governs when the agent can move.[3] When the agent has an opportunity to move, the agent chooses an action a from the discrete choice set $\mathscr{A} = \{1, \ldots, J\}$, conditional on the current state $k \in \mathscr{X}$. The set $\mathscr{A}$ contains all possible actions the agent can take when given the opportunity to move.

The agent is forward looking and discounts future payoffs at a rate ρ. While the model is in state k, the agent receives flow utility u_k. Thus, if the model remains in state k over some interval $[0, \tau)$, the present discounted value of the payoff obtained over this period from the perspective of time 0 is $\int_0^\tau \mathrm{e}^{-\rho t} u_k \, dt$.

Upon receiving a move arrival when the current state is k, the agent chooses an action $j \in \mathscr{A}$. The agent then receives an instantaneous payoff $\psi_{jk} + \varepsilon_j$ associated with making choice j in state k, where ε_j is a choice-specific payoff shock that is iid over time and across choices. Let σ_{jk} denote the probability that the agent optimally chooses choice j in state k. Let v_{jk} denote the continuation value received by the agent after making choice j in state k. In most cases, v_{jk} will consist of a particular element of the value function, for example, if the state is unchanged after the action then we might have $v_{jk} = V_k$, where V_k denotes the value function at state k (defined below). On the other hand, if there is a terminal action after which the agent is no longer active, then we might have $v_{jk} = 0$.[4]

We can now write the Bellman equation, a recursive expression for the value function V_k which gives the present discounted value of all future payoffs obtained from starting in some state k and behaving optimally in future periods. Without loss of generality, we use time 0 as the initial time. Let τ denote the time until the next event: either an exogenous

[3] Note that it is completely straightforward to include state-specific move arrival rates λ_k. However, since this would significantly increase the data requirements for estimation, we assume throughout that $\lambda_k = \lambda$ for all k.

[4] There might also be uncertainty about the resulting state. In such cases we let ϕ_{jkl} denote the probability with which the model transitions to state l after the agent takes action j in state k, where for each j and k we have $\sum_{l=1}^K \phi_{jkl} = 1$. In many cases, such as an exit decision, these probabilities will be degenerate. In this notation, for example, one might express the future value term as $v_{jk} = \sum_{l=1}^K \phi_{jkl} V_l$. Since there are many possible scenarios, we use the notation v_{jk} for generality.

state change or a move opportunity for the agent. In state k we have

$$V_k = \mathrm{E}\left[\int_0^\tau \mathrm{e}^{-\rho t}\, u_k\, dt + \mathrm{e}^{-\rho\tau}\,\frac{1}{\lambda + q_{kk}}\left(\sum_{l\neq k} q_{kl} V_l + \lambda \max_j \{\psi_{jk} + \varepsilon_j + v_{jk}\}\right)\right]. \tag{1.3}$$

Here we have used Proposition 1.1 and the law of iterated expectations to evaluate the expectation over the joint distribution of (τ, ι) by first conditioning on τ.

The value function, as expressed recursively in (1.3), is the expectation of two terms. The first term represents the flow utility obtained in state k from the initial time until the next event (a move or jump), at time τ. The second term represents the discounted expected future value obtained from the time of the event onward, where $\lambda/(\lambda + q_{kk})$ is the probability that the event is a move opportunity and $q_{kl}/(\lambda + q_{kk})$ is the probability that the event is a jump to some state $l \neq k$. The expectation is taken with respect to both τ and ε. It is important to note that although we have used the time of the next *event*, τ, as the point of reference in (1.3), another intuitive possibility is the time of the next *move* by the agent in question, τ_i (here, τ_1). We use the next event form in practice as it is computationally simpler.[5]

A policy rule is a function $\delta : \mathscr{X} \times \mathscr{E} \to \mathscr{A}$ which assigns to each state k and vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_J)$ an action from $\mathscr{A}$. The optimal policy rule satisfies the following inequality condition:

$$\delta(k, \varepsilon) = j \iff \psi_{jk} + \varepsilon_j + v_{jk} \geq \psi_{lk} + \varepsilon_l + v_{lk} \quad \forall\, l \in \mathscr{A}.$$

That is, when given the opportunity to choose an action, δ assigns the action that maximizes the agent's expected future discounted payoff. Thus, under the optimal policy rule, the conditional choice probabilities σ_{jk} satisfy

$$\sigma_{jk} = \Pr[\delta(k, \varepsilon) = j \mid k].$$

[5] The next move form introduces matrix exponentials which must be calculated to evaluate the first term.

Note that the move arrival rate, λ, and the choice probabilities of the agent, σ_{jk}, also imply a Markov jump process on $\mathscr{X}$ with intensity matrix Q_1, where Q_1 is a function of both λ and σ_{jk} for all j and k. In particular, the hazard rate of action j in state k is simply $\lambda\sigma_{jk}$, the product of the move arrival rate and the choice probability. The choice probability σ_{jk} is thus the proportion of moves in state k, which occur at rate λ, that result in action j. Summing the intensity matrices Q_0 and Q_1 yields the intensity matrix of the combined (compound) process. This simple and intuitive structure is especially important in extending the model to include multiple agents, and in estimation with discrete time data.

Example: A Single Agent Renewal Model

Our first example is a simple single-agent renewal model, based on the bus engine (capital) replacement problem analyzed by Rust (1987). The state space represents accumulated mileage and is indexed by the finite set $\mathscr{X} = \{1, \ldots, K\}$. The agent has a binary choice set $\mathscr{A} = \{0, 1\}$, which represents the choice over whether or not to replace the engine, thereby reseting the mileage to its baseline level. The agent faces a cost minimization problem where the flow cost incurred in mileage state k is $u_k = -\beta k$ where $\beta > 0$. The action $j = 0$ represents continuation, where the state remains unchanged, and the choice $j = 1$ causes the state to reset to $k = 1$.

A representative sample path generated by this model is shown in Figure 1.2. Inter-arrival times are indicated by τ_{in}, where i denotes the identity of the player (with $i = 0$ denoting nature) and n denotes the event number. The agent's decisions (a_{t_n}) are indicated at each decision time. For example, at time t_1, the agent chooses to continue without replacement ($a_{t_1} = 0$), while at time t_4, the agent chooses to replace ($a_{t_4} = 1$), resetting the mileage.

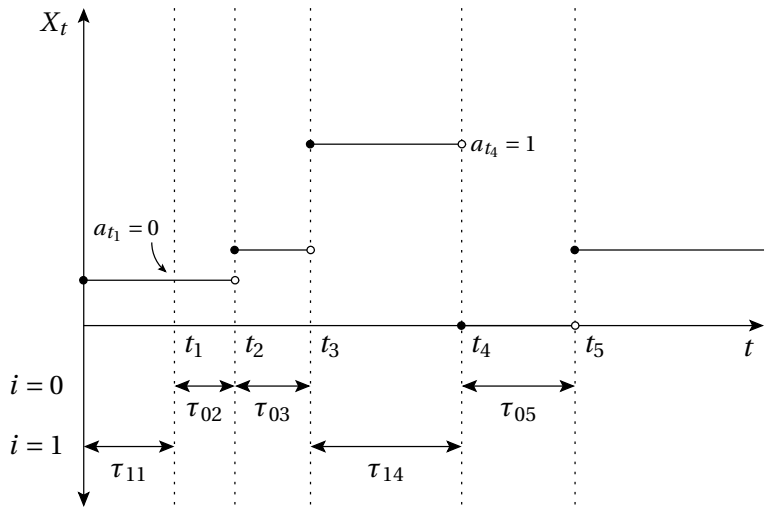

FIGURE 1.2: Single agent model: a representative sample path where t_n, τ_{in}, and a_{in} denote, respectively, the time, inter-arrival time, and action corresponding to n-th event. Moves by the agent are denoted by $i = 1$ while $i = 0$ denotes a state change (a move by nature).

The $K \times K$ intensity matrix for the jump process on $\mathscr{X}$ is

$$Q_0 = \begin{bmatrix} -q_1 - q_2 & q_1 & q_2 & 0 & \dots & 0 \\ 0 & -q_1 - q_2 & q_1 & q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & -q_1 - q_2 & q_1 & q_2 \\ 0 & 0 & \dots & 0 & -q_1 - q_2 & q_1 + q_2 \\ 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}.$$

Thus, the state can only move forward until it reaches the final state K, at which point it remains there until it is reset to state 1 by the agent. For any state $1 \le k < K-1$ the state may jump forward either one level or two (and only one at state $K-1$). Conditional on jumping, the probabilities of moving forward one level or two are $q_1/(q_1+q_2)$ and $q_2/(q_1+q_2)$ respectively.

In the notation of the general model above, the continuation values are

$$v_{jk} = \begin{cases} V_k, & \text{if } j = 0, \\ V_1, & \text{if } j = 1. \end{cases}$$

That is, when the model is in state k and the agent chooses to continue, $j = 0$, the state is unchanged and the continuation value is simply V_k. On the other hand, when the agent chooses to reset the state, $j = 1$, the continuation value is V_1, the present discounted value of being in state 1. Although no cost is incurred from continuation, the agent incurs a one-time cost of c when choosing to reset the state to the initial value:

$$\psi_{jk} = \begin{cases} 0, & \text{if } j = 0, \\ -c, & \text{if } j = 1. \end{cases}$$

The value function for this model can thus be represented recursively as

$$V_k = \mathrm{E}\Bigg[\int_0^\tau \mathrm{e}^{-\rho t}\, u_k\, dt + \mathrm{e}^{-\rho\tau}\Big(\frac{q_1}{\lambda + q_1 + q_2} V_{k+1} \frac{q_2}{\lambda + q_1 + q_2} V_{k+2} + \frac{\lambda}{\lambda + q_1 + q_2} \max\{\varepsilon_0 + V_k, -c + \varepsilon_1 + V_1\}\Big)\Bigg]$$

for $k \leq K - 2$. It is similar for $K - 1 \leq k \leq K$, with the appropriate adjustments being made at the boundary of the state space.

If we assume that the ε_j are iid with $\varepsilon_j \sim \mathrm{TIEV}(0, 1)$ then we can simplify this expression further using the closed form representation of the expected future value (cf. McFadden, 1984) and the law of iterated expectations (replacing $\mathrm{E}_{\tau,\varepsilon}$ with $\mathrm{E}_\tau \mathrm{E}_{\varepsilon|\tau}$) to obtain:

$$\mathrm{E}\left[\max\{V_k + \varepsilon_0, V_1 - c + \varepsilon_1\}\right] = \ln\left[\exp(V_k) + \exp(V_1 - c)\right],$$

and thus,

$$V_k = \mathrm{E}\Bigg[\int_0^\tau \mathrm{e}^{-\rho t}\, u_k\, dt + \mathrm{e}^{-\rho\tau}\Big(\frac{q_1}{\lambda + q_1 + q_2} V_{k+1} \frac{q_2}{\lambda + q_1 + q_2} V_{k+2} + \frac{\lambda}{\lambda + q_1 + q_2} \ln\big(\exp(V_k) + \exp(V_1 - c)\big)\Big)\Bigg]. \tag{1.4}$$

The value function summarizes the present discounted value of all future cost flows from the perspective of an arbitrary point in time, without loss of generality taken to be time 0, and at an arbitrary state $k \in \mathscr{X}$. Here, τ represents the length of time until the arrival of the next event. At each point in time, the agent makes a decision based on an expected future utility comparison, with the expectation taken with respect to the next event time τ, and ε. Inside the expectation, the first term provides the expected flow utility accumulated over the time interval $[0, \tau)$. Since the agent does not move during this time, the state evolves undeterred according the Markov jump process defined by the intensity matrix Q_0, resulting in a cost flow u_k at each instant. The second term is the present discounted value of future utility from time τ onward, after the next event occurs. At the arrival time τ, the state jumps to $k + l$, $l \in \{1, 2\}$ with probability $q_l/(\lambda + q_1 + q_2)$, while with probability $\lambda/(\lambda + q_1 + q_2)$, the agent gets to move and makes an expected future utility maximizing choice of $j \in \{0, 1\}$. The agent may choose $j = 0$ and simply continue accumulating the flow cost until the next arrival, or choose $j = 1$ and reset the state to 1 by paying a cost c. The type I extreme value assumption also yields closed forms for the associated CCPs:

$$\sigma_{jk} = \begin{cases} \frac{\exp(V_k - V_1 + c)}{\exp(V_k - V_1 + c) + 1}, & \text{if } j = 0, \\ \frac{1}{\exp(V_k - V_1 + c) + 1}, & \text{if } j = 1. \end{cases} \tag{1.5}$$

We return to this example in Section 1.6 below, where we conduct Monte Carlo experiments for various parameterizations of this single-agent renewal problem.

1.4 Multi-Agent Dynamic Discrete Games

Extending the single-agent model of Section 1.3 to the case of dynamic discrete games with many players is simply a matter of modifying the intensity matrix governing the market-wide state vector to incorporate players' beliefs regarding the actions of their rivals. Following Harsanyi (1973), we treat the dynamic discrete game as a collection

of single-agent games against nature, in which moves by rival agents are distributed in accordance with players' beliefs. As is standard in the literature, we focus on Markov strategies, eliminating the need to keep track of the full history of play. We begin this section by describing the general structure of the model followed by an example—the Ericson-Pakes quality ladder model—to fix ideas.

Suppose there are $\bar{N}$ players indexed by $i = 1, \ldots, \bar{N}$. The state space $\mathscr{X}$ is now a set of vectors of length $\bar{N}$, with the i-th component corresponding to the state of player i. Player i's discount rate is ρ_i. We shall simplify the notation later by assuming symmetry and anonymity, but for generality we index all other quantities by i, including the flow utility in state k, u_{ik}, the choice probabilities, σ_{ijk}, instantaneous payoffs, ψ_{ijk}, and transition probabilities resulting from the action of a player, ϕ_{ijkm}.

Although it is still sufficient to have only a single state jump process on $\mathscr{X}$ (with some intensity matrix Q_0) to capture moves by nature, there are now $\bar{N}$ competing Poisson processes with rates λ_i generating move arrivals for each of the $\bar{N}$ players.[6] The next event in the model is determined by the earliest arrival of one of these $\bar{N}+1$ processes.

By assuming that the iid shocks to the instantaneous payoffs are private information of the individual players, we can re-interpret the multi-agent model as a collection of games against nature, and incorporate the uncertainty about the moves of rival firms into the intensity matrix. This allows us to construct the value function for the multi-agent model in much the same way as in the single-agent case. Let τ denote the time of the next event, a state jump or a move opportunity for any player, which is the minimum of a collection of competing Poisson processes with rates given by the intensity matrix Q_0 and the move arrival rates λ_i for $i = 1, \ldots, \bar{N}$.

A representative sample path from a two-player game is shown in Figure 1.3. Moves by nature are indicated by $i = 0$. The moves and inter-arrival times at the n-th event are

[6] As in the single player model, we assume here that the move arrival rate is constant across states. However, with a large enough dataset, the inclusion of state specific arrival rates would be straightforward.

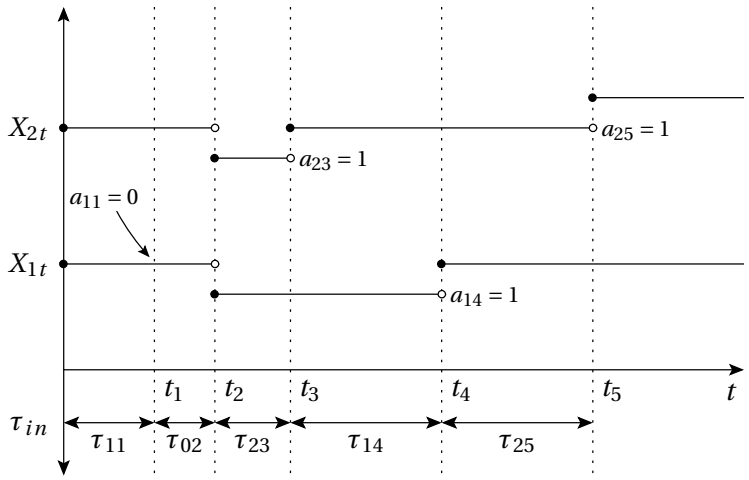

FIGURE 1.3: Multi-agent model: a representative sample path for two players ($i = 1,2$) and nature ($i = 0$). Event times are denoted by t_n, inter-arrival times are denoted τ_{in}, and actions are denoted a_{in}. Here, at t_1, player 1 chooses $a_{11} = 0$ which has no effect on the state and at t_2 an exogenous state change decreases both players' states. The final three events are moves by players 1 and 2 where action 1 is chosen by both, increasing the player-specific state variables in each case.

denoted by a_{it} and τ_{in} respectively. Here, for example, player 1 moves at time t_1 and chooses action 0 which does not change the state. The move by nature (an exogenous state change) at t_2 decreases both players' states. Player 2 then moves at t_3, choosing action 1 which increases player 2's state. Similar moves by players 1 and 2 follow at times t_4 and t_5.

Returning to the model, note that in the interval between the previous event time and τ, no other events may take place since, by definition, τ is the time of the *next* event. In some state k, the probability that the event is a move by player i is proportional to λ_i and the probability that the state jumps from k to $l \neq k$ is proportional to q_{kl}. The denominator of these probabilities is the sum of all of the rates involved,[7] so that the

[7] For simplicity, we assume here that all players are active in all states. In this case, the rate at which moves occur is the sum $\sum_i \lambda_i$, which is independent of the state. In general, these move arrival rates may be state-specific, as illustrated in the quality ladder model of the following section, where inactive firms cannot move (i.e., $\lambda_{ik} = 0$ in states k where firm i is inactive).

probability that the next event in state k is a move opportunity for player i is

$$\frac{\lambda_i}{\sum_{l=1}^{\bar{N}} \lambda_l + q_{kk}},$$

where $q_{kk} = \sum_{l \neq k} q_{kl}$, and the probability that the state jumps from k to m is

$$\frac{q_{km}}{\sum_{l=1}^{\bar{N}} \lambda_l + q_{kk}}.$$

As before, let σ_{ijk} denote the probability that action j is chosen optimally by player i in state k. These choice probabilities are determined endogenously in the model. The continuation values are denoted v_{ijk}, and ϕ_{ijkl} denotes the probability that immediately after player i takes action j, the state jumps to another state l.

Given the above notation, the value function for player i in state k is

$$V_{ik} = \mathrm{E}\Big[\int_0^\tau \mathrm{e}^{-\rho_i t}\, u_{ik}\, dt + \mathrm{e}^{-\rho_i \tau} \frac{1}{\sum_{i=1}^{\bar{N}} \lambda_i + q_{kk}} \Big(\sum_{l \neq k} q_{kl} V_{il} + \sum_{l \neq i} \lambda_l \sum_{j=1}^{J} \sigma_{ljk} \sum_{m=1}^{K} \phi_{ljkm} V_{im} + \lambda_i \max_j \{\psi_{ijk} + \varepsilon_{ij} + v_{ijk}\} \Big) \Big]. \tag{1.6}$$

This expression is complicated only for the sake of generality. In many applications, it will be the case that the ϕ_{ljkm} terms are degenerate, with deterministic state transitions following moves. Further simplifications are also possible under symmetry.

A policy rule is then a function $\delta_i : \mathscr{X} \times \mathscr{E}_i \to \mathscr{A}_i$ which maps each state k and vector $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iJ})$ to an action from $\mathscr{A}_i$. Given a set of beliefs σ_{ljk} for each rival $l \neq i$ regarding the probability that player l chooses j in state k (which enter Q_{-i}), the optimal policy rule satisfies the following condition:

$$\delta_i(k, \varepsilon_i) = j \iff \psi_{ijk} + \varepsilon_{ij} + v_{ijk} \geq \psi_{ilk} + \varepsilon_{il} + v_{ilk} \quad \forall\, l \in \mathscr{A}_i. \tag{1.7}$$

That is, when given the opportunity to choose an action, δ_i assigns the action that maximizes the agent's expected future discounted payoff given the specified beliefs.

Then, under a given policy rule, the conditional choice probabilities of player i, σ_{ijk}, satisfy

(1.8) $\sigma_{ijk} = \Pr[\delta_i(k, \varepsilon_i) = j \mid k]$.

A *Markov perfect equilibrium* is a collection of policy rules $(\delta_1, \ldots, \delta_{\bar{N}})$ and a set of beliefs $\{\sigma_{ijk} : i = 1, \ldots, \bar{N};\ j = 1, \ldots, J;\ k = 1, \ldots, K\}$ such that both (1.7) and (1.8) hold for all i.

Example: A Quality Ladder Model

To illustrate the application to dynamic games we consider a discrete control version of the quality ladder model proposed by Ericson and Pakes (1995). This model is widely used in industrial organization and has been studied extensively by Pakes and McGuire (1994, 2001), Doraszelski and Satterthwaite (2010), Doraszelski and Pakes (2007), and several others. The model consists of at most $\bar{N}$ firms who compete in a single product market. The products are differentiated in that the product of firm i has some quality level $\omega_i \in \Omega$, where $\Omega = \{1, 2, \ldots, \bar{\omega}, \bar{\omega} + 1\}$ is the finite set of possible quality levels, with $\bar{\omega} + 1$ denoting the "quality" of inactive firms. Firms with $\omega_i < \bar{\omega} + 1$ are incumbents. In contrast to Pakes and McGuire (1994), all controls here are discrete: given a move arrival, firms choose whether or not to move up the quality ladder, not how much to spend to increase their chances of doing so.

We consider the particular example of price competition with a single differentiated product where firms make entry, exit, and investment decisions, however, the quality ladder framework is quite general and can be easily adapted to other settings. For example, Doraszelski and Markovich (2007) use this framework in a model of advertising where, as above, firms compete in a differentiated product market by setting prices, but where the state ω_i is the share of consumers who are aware of firm i's product. Gowrisankaran (1999a) develops a model of endogenous horizontal mergers where ω_i is a capacity level

and the product market stage game is Cournot with a given demand curve and cost functions that enforce capacity constraints depending on each firm's ω_i.

State Space Representation

We make the usual assumption that firms are symmetric and anonymous. That is, the primitives of the model are the same for each firm and only the distribution of firms across states, not the identities of those firms, is payoff-relevant. We also assume players share the same discount rate, $\rho_i = \rho$ for all i, and move arrival rate, $\lambda_i = \lambda$, for all i. By imposing symmetry and anonymity, the size of the state space can be reduced from the total number of distinct market structures, $(\bar{\omega}+1)^{\bar{N}}$, to the number of possible distributions of $\bar{N}$ firms across $\bar{\omega}+1$ states. The set of payoff-relevant states is thus the set of ordered tuples of length $\bar{\omega}+1$ whose elements sum to $\bar{N}$:

$$\mathscr{S} = \{(s_1, \ldots, s_{\bar{\omega}+1}) : \sum_j s_j = \bar{N}, s_j \in \mathbb{Z}^*\},$$

where $\mathbb{Z}^*$ is the set of nonnegative integers. In this notation, each vector $\omega = (\omega_1, \ldots, \omega_{\bar{N}})$ in $\Omega^{\bar{N}}$ maps to an element $s = (s_1, \ldots, s_{\bar{\omega}+1}) \in \mathscr{S}$ with $s_j = \sum_{i=1}^{\bar{N}} 1\{\omega_i = j\}$ for each j.

In practice we map the multidimensional space $\mathscr{S}$ to an equivalent one-dimensional state space $\mathscr{X} = \{1, \ldots, |\mathscr{S}|\}$.[8] Payoff relevant market configurations from the perspective of firm i are then uniquely described by two integers (x, ω_i), where $x \in \mathscr{X}$ denotes the market structure and ω_i is firm i's own quality level.

Product Market Competition

Again, we follow Pakes and McGuire (1994) in assuming a continuum of consumers with measure $M > 0$ and that consumer j's utility from choosing the good produced by firm i is $g(\omega_i) - p_i + \varepsilon_i$, where ε_i is iid across firms and consumers and follows a type I extreme

[8] In particular, we use the "probability density space" encoding algorithm described in Gowrisankaran (1999b), to map market structure tuples $s \in \mathscr{S}$ to integers $x \in \mathscr{X}$.

value distribution. Pakes and McGuire (1994) use the g function to enforce an upper bound on profits. As in Pakes, Gowrisankaran, and McGuire (1993), for some constant ω^* we specify the function

$$g(\omega_i) = \begin{cases} \omega_i & \text{if } \omega_i \leq \omega^*, \\ \omega_i - \ln(2 - \exp(\omega^* - \omega_i)) & \text{if } \omega_i > \omega^*. \end{cases}$$

Let $\varsigma_i(\omega, p)$ denote firm i's market share given the state ω and prices p. From McFadden (1974), we know that the share of consumers purchasing good i is

$$\varsigma_i(\omega, p) = \frac{\exp(g(\omega_i) - p_i)}{1 + \sum_{j=1}^{\bar{N}} \exp(g(\omega_j) - p_j)}.$$

In a market of size M, firm i's demand is $q_i(\omega, p) = M\varsigma_i$.

All firms have the same constant marginal cost $c \geq 0$. Taking the prices of other firms, p_{-i}, as given, the profit maximization problem of firm i is

$$\max_{p_i \geq 0} q_i(p, \omega)(p_i - c).$$

Caplin and Nalebuff (1991) show that (in this single-product firm setting) there is a unique Bertrand-Nash equilibrium, which is given by the solution to the first order conditions of the firm's problem:

$$\frac{\partial q_i}{\partial p_i}(p, \omega)(p_i - c) + q_i(p, \omega) = 0.$$

Given the functional forms above, the first order conditions become

$$-(p_j - c)(1 - \varsigma_j) + 1 = 0.$$

We solve this nonlinear system of equations numerically using the Newton-Raphson method to obtain the equilibrium prices and the implied profits

$$\pi(\omega_i, \omega_{-i}) = q_i(p, \omega)(p_i - c)$$

earned by each firm i in each state (ω_i, ω_{-i}).

Incumbent Firms

We consider a simple model in which incumbent firms have three choices upon receiving a move arrival. Firms may continue without investing at no cost, they may invest an amount κ in order to increase the quality of their product from ω_i to $\omega_i' = \min\{\omega_i + 1, \bar{\omega}\}$, or they may exit the market and receive some scrap value η. We denote these choices, respectively, by the choice set $\mathscr{A}_i = \{0, 1, 2\}$. When an incumbent firm exits the market, ω_i jumps deterministically to $\bar{\omega} + 1$. Associated with each choice j is a private shock ε_{ijt}. These shocks are iid over firms, choices, and time and follow a type I extreme value distribution. Given the future value associated with each choice, the resulting choice probabilities are defined by a logit system.

Due to the complexity of the state space, we now introduce some simplifying notation. For any market-wide state $k \in \mathscr{X}$, let $\omega_k = (\omega_{1k}, \ldots, \omega_{\bar{N}k})$ denote its counterpart in $\Omega^{\bar{N}}$. In the general notation introduced above, the instantaneous payoff ψ_{ijk} to firm i from choosing choice j in state k is

$$\psi_{ijk} = \begin{cases} 0 & \text{if } j = 0, \\ -\kappa & \text{if } j = 1, \\ \eta & \text{if } j = 2. \end{cases}$$

Similarly, the continuation values are

$$v_{ijk} = \begin{cases} V_{ijk} & \text{if } j = 0, \\ V_{ijk'} & \text{if } j = 1, \\ 0 & \text{if } j = 2, \end{cases}$$

where state k' is the element of $\mathscr{X}$ such that $\omega_{k'i} = \min\{\omega_{ki} + 1, \bar{\omega}\}$ and $\omega_{k'j} = \omega_{kj}$ for all $j \neq i$. Note that we are considering only incumbent firms with $\omega_{ki} < \bar{\omega} + 1$.

The value function for an incumbent firm in state k is thus

$$V_{ik} = \mathrm{E}\Bigg[\int_0^{\tau} \mathrm{e}^{-\rho_i t}\pi_{ik}\,dt + \mathrm{e}^{-\rho_i \tau}\frac{1}{N_k\lambda + q_{kk}}\Big(\sum_{l\neq k} q_{kl}V_{il} + \sum_{l\neq i}\lambda\sum_{j=1}^{J}\sigma_{ljk}\sum_{m=1}^{K}\phi_{ljkm}V_{im} + \lambda\max\{V_{ik}+\varepsilon_{i0}, V_{ik'}-\kappa+\varepsilon_{i1}, \eta+\varepsilon_{i2}\}\Big)\Bigg]$$

where π represents the flow profit accruing from product market competition, N_k denotes the number of active incumbents and potential entrants in state k, and the expectation is with respect to τ and ε_{ij} for all i and j. Conditional upon moving while in state k, incumbent firms face the following maximization problem:

$$\max\{V_{ik}+\varepsilon_{i0}, -\kappa + V_{ik'}+\varepsilon_{i1}, \eta+\varepsilon_{i2}\}.$$

The resulting choice probabilities are

$$\sigma_{i0k} = \frac{\exp(V_{ik})}{\exp(V_{ik}) + \exp(-\kappa + V_{ik'}) + \exp(\eta)},$$

$$\sigma_{i1k} = \frac{\exp(-\kappa + V_{ik'})}{\exp(V_{ik}) + \exp(-\kappa + V_{ik'}) + \exp(\eta)},$$

$$\sigma_{i2k} = 1 - \sigma_{i0k} - \sigma_{i1k},$$

where, as before, k' denotes the resulting state after investment.

Potential Entrants

Whenever the number of incumbent is smaller than $\bar{N}$, a single potential entrant receives the opportunity to enter at rate λ. Potential entrants are short-lived and do not consider the option value of delaying entry. The potential entrant is counted in N_k, the total number of active firms in state k, and thus the rate at which *incumbents* receive the opportunity to move is $(N_k - 1)\lambda$ but the rate at which *any* type of move opportunity occurs is $N_k\lambda$. If firm i is a potential entrant with the opportunity to move it has two choices: it can choose to enter ($a_i = 1$), paying a setup cost η^e and entering the market

immediately in a predetermined entry state $\omega^e \in \Omega$ or it can choose not to enter ($a_i = 0$) at no cost. Associated with each choice j is a stochastic private payoff shock ε^e_{ijt}. These shocks are iid across firms, choices, and time and are distributed according to the type I extreme value distribution.

In the general notation of Section 1.4, for entrants ($j = 1$) in state k, the instantaneous payoff is $\psi_{i1k} = -\eta^e$ and the continuation value is $v_{i1k} = V_{ik'}$ where k' is the element of $\mathscr{X}$ with $\omega_{k'i} = \omega^e$ and $\omega_{k'j} = \omega_{kj}$ for all $j \neq i$. For firms that choose not to enter ($j = 0$) in state k, we have $\psi_{i0k} = V_{i0k} = 0$. Thus, conditional upon moving in state k, a potential entrant faces the problem

$$\max\{\varepsilon^e_{i0}, -\eta^e + V_{ik'} + \varepsilon^e_{i1}\}$$

yielding the conditional entry-choice probabilities

$$\sigma_{i1k} = \frac{\exp(V_{ik'} - \eta^e)}{1 + \exp(V_{ik'} - \eta^e)}.$$

State Transitions

In addition to state transitions that result directly from entry, exit, or investment decisions, the overall state of the market follows a jump process where at some rate γ, the quality of each firm i jumps from ω_i to $\omega'_i = \max\{\omega_i - 1, 1\}$. This process represents an industry-wide (negative) demand shock, interpreted as an improvement in the outside alternative.

Being a discrete-time model, Pakes and McGuire (1994) assume that each period this industry-wide quality deprecation happens with some probability δ, implying that the quality of all firms falls on average every $1/\delta$ periods. Our assumption of a rate γ is also a statement about this frequency in that $1/\gamma$ is the average length of time until the outside good improves.

We construct the corresponding intensity matrix Q_0 as follows. We map each market structure s to an integer k and map the resulting structure after deprecation s' to an

integer k'. The (k,k) element of Q_0 for each eligible state k is $-\gamma$ while the corresponding (k,k') element is γ.

1.5 Estimation

Methods that solve for the value function directly and use it to obtain the implied choice probabilities for estimation are referred to as full-solution methods. The nested-fixed point (NFXP) algorithm of Rust (1987), which uses value function iteration inside of an optimization routine that maximizes the likelihood, is the classic example of a full-solution method. Su and Judd (2008) provide an alternative MPEC (mathematical program with equilibrium constraints) approach which solves the constrained optimization problem directly, bypassing the repeated solution of the dynamic programming problem.

CCP-based estimation methods, on the other hand, are two-step methods pioneered by Hotz and Miller (1993) and Hotz et al. (1994) and later extended by Aguirregabiria and Mira (2007), Bajari et al. (2007), Pesendorfer and Schmidt-Dengler (2007), Pakes et al. (2007), and Arcidiacono and Miller (2008). The CCPs are estimated in a first step and used to approximate the value function in a closed-form inversion or simulation step. The approximate value function is then used in the likelihood function to estimate the structural parameters of the model using a maximum pseudo-likelihood procedure (or similarly "plugged-in" to a corresponding GMM criterion function).

Full-solution methods have the advantage that the exact CCPs are known once the value function is found—they do not have to be estimated—and thus the model can be estimated using full-information maximum likelihood. These methods can become quite computationally expensive for complex models with many players or a large state space. Many candidate parameter vectors must be evaluated during estimation and, if the value function is costly to compute, even if solving the model once might be feasible, doing so many times may not be. In the presence of multiple equilibria, they also require

researchers to make an assumption on the equilibrium selection mechanism and solve for all the equilibria (cf. Bajari, Hong, and Ryan, 2007).[9] The Su and Judd (2008) MPEC approach provides one solution to the issue of multiplicity. CCP methods provide another attractive alternative, allowing the value function to be computed very quickly and the pseudo-likelihood function to condition upon the equilibrium that is played in the data.

Our model has the advantage of being estimable via either approach. As in Doraszelski and Judd (2008), the use of continuous time breaks one primary curse of dimensionality in that only a single player moves at any particular instant. An attractive and novel feature of our framework is that it is easily estimable using standard CCP methods. This greatly reduces the computational costs of estimation relative to full solution methods. Having estimated a large problem with CCP methods, it is then straightforward to use the model for post-estimation exercises, since the computational burden of computing the equilibrium a few times for these purposes is not as great as nesting several such solutions into an estimation routine. In this way, our framework preserves a tight link between the estimated model and that used for post-estimation analysis, something which has proven infeasible for many empirical applications that have been modeled in discrete time.

This section is organized as follows. We begin by discussing estimation via full-solution methods with continuous time data in Section 1.5.1 before turning to cases with imperfectly sampled data. We consider the case when some moves may be unobserved in Section 1.5.2, and in Section 1.5.3 we consider the case where the data is only sampled at discrete intervals. We consider two-step CCP-based estimation in Section 1.5.4 and close with a discussion of extensions to models which admit unobserved heterogeneity in Section 1.5.5.

[9] When performing full-solution estimation in this paper, we assume that the equilibrium selection rule assigns probability one to the equilibrium obtained by our numerical fixed point routine. The computational advantages of continuous time, however, make it easier to explore more complex specifications with non-degenerate weightings.

1.5.1 Full-Solution Estimation

We begin our discussion of estimation by considering the application of a full solution method in a setting where continuous time data is available. In particular, suppose we observe all continuous-time events in a single market over the time interval $[0, T]$. Let N denote the total number of observed events during the observation window, that is, the number of actions (moves) plus the number of exogenous state changes (jumps). The n-th event is characterized by five pieces: the elapsed time since the last event, τ_n, the index of the player associated with the event, i_n (where $i_n = 0$ indicates a move by nature), the observed choice, a_n, and the states immediately before and after the event, x_n and x'_n.

Let $\ell_n(\theta)$ denote the likelihood of the n-th event where θ is the vector of parameters of interest which includes the move-arrival rates, the parameters of the process generating exogenous state changes, and the parameters of the payoff functions. The likelihood function depends on the aggregate and individual intensity matrices, which are in turn functions of θ. We let $q(x, x'; \theta)$ denote the absolute value of the (x, x') element of the aggregate intensity matrix $Q \equiv Q_0 + \sum_{i=1}^{\tilde{N}} Q_i$. Taking the absolute value serves to ensure that all rates are positive, since the diagonal elements, the rates of leaving a particular state, are negative. Define $q_0(x, x'; \theta)$ and $q_{\tilde{N}}(x, x'; \theta)$ similarly for the matrices Q_0 and $\sum_{i=1}^{\tilde{N}} Q_i$.

The likelihood of a single event consists of several parts, the first of which is the pdf of the inter-event interval, $g(\tau_n; q(x_n, x_n; \theta))$, where g is the pdf of the exponential distribution. The rate parameter in this case is $q(x_n, x_n; \theta)$, the negative of the diagonal element of the aggregate intensity matrix Q, which is the overall rate at which the process leaves state x_n. The remaining components of the likelihood for jumps (indicated by $i_n = 0$) are the probability that the event is indeed a jump, $q_0(x_n, x_n; \theta) / q(x_n, x_n; \theta)$, and the state transition probability, $p(x_n, x'_n; \theta) \equiv q_0(x_n, x'_n; \theta) / q_0(x_n, x_n; \theta)$. On the other

hand, for moves (indicated by $i_n > 0$), the remaining components are the probability that the event is a move, $q_{\bar{N}}(x_n, x_n; \theta)/q(x_n, x_n; \theta)$, the choice probability, $\sigma(i_n, a_n, x_n; \theta)$, and the probability of the resulting state transition, $\phi(i_n, a_n, x_n, x_n'; \theta)$.[10]

Combining these components, the likelihood of a single event is

$$\ell_n(\theta) = g(\tau_n; q(x_n, x_n; \theta)) \left[\frac{q_0(x_n, x_n; \theta)}{q(x_n, x_n; \theta)} \cdot p(x_n, x_n'; \theta) \right]^{1\{i_n = 0\}}$$
$$\times \left[\frac{q_{\bar{N}}(x_n, x_n; \theta)}{q(x_n, x_n; \theta)} \cdot \sigma(i_n, a_n, x_n; \theta) \cdot \phi(i_n, a_n, x_n, x_n'; \theta) \right]^{1\{i_n > 0\}}.$$

When the last observation occurs before the end of the observation window (i.e., when $t_N < T$), we must also account for the fact that *no* event was observed over the interval $[t_N, T]$. The likelihood of this is the probability of observing no events over an interval of length $T - t_N$ while the state was x_N, given by $1 - G(T - t_N; q(x_N, x_N; \theta))$, where G is the cdf of the exponential distribution. Thus, after simplifying the observation likelihood, the log likelihood for the time period $[0, T]$ is

$$\ln L_T(\theta) = \sum_{n=1}^{N} \ln g(\tau_n; q(x_n, x_n; \theta)) + \sum_{n=1}^{N} 1\{i_n = 0\} \ln q_0(x_n, x_n'; \theta)$$
$$+ \sum_{n=1}^{N} 1\{i_n > 0\} \left[\ln q_{\bar{N}}(x_n, x_n; \theta) + \ln \sigma(i_n, a_n, x_n; \theta) + \ln \phi(i_n, a_n, x_n, x_n'; \theta) \right]$$
$$- \sum_{n=1}^{N} \ln q(x_n, x_n; \theta) + \ln \left[1 - G(T - t_N, q(x_N, x_N; \theta)) \right].$$

We have replaced the transition probabilities by the relevant elements of the intensity matrices, canceled the $q_0(x_n, x_n; \theta)$ terms for move observations, and collected the common $q(x_n, x_n; \theta)$ terms. Note that it is straightforward to generalize this to the case where multiple markets are observed over potentially market-specific time horizons.

[10] In many cases, the state will change deterministically with the choice.

1.5.2 *Partially Observed Moves*

We continue using the same notation as in the previous sections, but now we suppose that the choice set is $\mathscr{A} = \{0, \ldots, J-1\}$ and that only actions a_n for which $a_n > 0$ are observed by the econometrician, where, without loss of generality, $a_n = 0$ denotes the unobserved action.[11] This assumption is likely to be the relevant one in most empirical settings with continuous-time data, as we generally expect the arrival of the right to move to be latent.[12] This complicates the estimation as now we only observe the truncated joint distribution of move arrival times and actions. Estimating λ using only the observed move times for observations with $a_n > 0$ would introduce a downward bias, corresponding to a longer average waiting time, because there could have been many unobserved moves in any interval between observed moves. Thus, in this setting τ_n is now the interval since the last *observed* event. For simplicity, we will consider only estimation of the single agent model of Section 1.3.

Over an interval where the state variable is constant at x_n, the choice probabilities for each action, $\sigma(i_n, a_n, x_n; \theta)$, are also constant. On this interval, conditional on receiving a move arrival, the move will be observed by the researcher with probability $1 - \sigma(i_n, 0, x_n; \theta)$.

For a given state x_n we can derive the likelihood of the waiting times between observed moves by starting with the underlying Poisson process generating the move arrivals. Let $N(t)$ denote the total cumulative number of move arrivals at time t and let $N_a(t)$ denote the number of move arrivals for which the agent chose action a. We will write $N_+(t)$ to denote $\sum_{a>0} N_a(t)$. We also define the waiting time before receiving a move arrival with corresponding action a, $W_a(t)$, defined as the smallest value of $\tau \geq 0$ such that

[11] We also assume that information about the unobserved action is not revealed through changes in the state, that is, the state remains constant following the choice $a_n = 0$. Stated formally, for all x_n and $i_n > 0$, $\phi(i_n, 0, x_n, x_n; \theta) = 1$.

[12] Estimation with time-aggregated data, discussed in Section 1.5.3, naturally accounts for this.

$N_a(t+\tau) - N_a(t) \geq 1$. Let $W_+(t)$ and $W(t)$ be defined similarly.

By the properties of Poisson processes we know that $W(t)$, the waiting time until the next move arrival (both observed and unobserved), is independent of t and has an exponential distribution with parameter λ. We have a similar result for $W_+(t)$. Because the probability of truncation (the probability of choosing $a = 0$) depends on x, so will the distribution of $W_+(t)$. We will derive the distribution for intervals where the state is constant, which will be sufficient for the purposes of the likelihood function.

Proposition 1.2. *Let the state of the model be x and let $\sigma(i,a,x)$ denote the choice probability of player i for action a in state x. Then $W_+(t)$ has an exponential distribution with rate parameter $(1-\sigma(i,0,x))\,\lambda$.*

Proof. See Appendix A. ■

The primary difference here is that we will use $\tilde{\lambda}_{x_n} = [1-\sigma(i_n,0,x_n;\theta)]\,\lambda$, the rate of observed moves in state x_n, to construct an intensity matrix for the *observed* processes, which will be used for estimation. We let $\tilde{q}(x,x';\theta)$ denote the absolute value of the (x,x') element of the resulting intensity matrix $\tilde{Q}$. Similarly, the choice probability $\sigma(i_n,a_n,x_n;\theta)$ is replaced by the choice probability conditional on having observed the choice,

$$\tilde{\sigma}(i_n,a_n,x_n;\theta) = \frac{\sigma(i_n,a_n,x_n;\theta)}{1-\sigma(i_n,0,x_n;\theta)}.$$

As before, we first write the likelihood for a single observation $(\tau_n, i_n, a_n, x_n, x_n')$:

$$\ell_n(\theta) = g(\tau_n;\tilde{q}(x_n,x_n;\theta)) \left[\frac{\tilde{q}_0(x_n,x_n;\theta)}{\tilde{q}(x_n,x_n;\theta)} \cdot p(x_n,x_n';\theta)\right]^{1\{i_n=0\}}$$
$$\times \left[\frac{\tilde{q}_{\bar{N}}(x_n,x_n;\theta)}{\tilde{q}(x_n,x_n;\theta)} \cdot \tilde{\sigma}(i_n,a_n,x_n;\theta) \cdot \phi(i_n,a_n,x_n,x_n';\theta)\right]^{1\{i_n>0\}}.$$

Estimation can now proceed as before by constructing and maximizing the log-likelihood function of the full sample.

1.5.3 Time Aggregaton

Having considered settings with partially observed moves, now suppose we only observe the process at N discrete points in time $\{t_1, t_2, \ldots, t_N\}$. Let $\{x_1, x_2, \ldots, x_N\}$ denote the corresponding states. Through the aggregate intensity matrix $Q \equiv Q_0 + \sum_i Q_i$, these discrete-time observations provide information about the underlying state jump process as well as the rate of move arrivals and the conditional choice probabilities. We use these observations to estimate the structural parameters θ, which appear in Q both directly and indirectly through the conditional choice probabilities σ_{ijk}. In this section, we describe a full-solution approach in which the value function is solved for each value of θ in order to obtain the implied CCPs which, in turn, are used to construct Q.

Let $P(t)$ denote the transition probability function from (1.2) corresponding to the aggregate intensity matrix Q. These probabilities summarize the relevant information about a pair of observations (t_{n-1}, x_{n-1}) and (t_n, x_n). That is, $P_{x_{n-1},x_n}(t_n - t_{n-1})$ is the probability of the process moving from x_{n-1} to x_n after an interval of length $t_n - t_{n-1}$. This includes cases where $x_n = x_{n-1}$ since the transition probabilities account for there having been no jump or any of an infinite number of combinations of jumps to intermediate states before coming back to the initial state. The likelihood for a sample $\{(t_n, x_n)\}_{n=1}^N$ is thus

$$\ln L_N(\theta) = \sum_{n=1}^{N} \ln P_{x_{n-1},x_n}(t_n - t_{n-1}).$$

In this way, dealing with time aggregation is remarkably simple in continuous time. Because transition probabilities for an interval of any length can easily be calculated from the underlying intensity matrix, it is straightforward to characterize the likelihood of transitioning from any observed initial state to any observed final state over any given interval. This approach naturally accounts for all of the myriad ways that such a transition could have occurred, something which would be much more computationally demanding

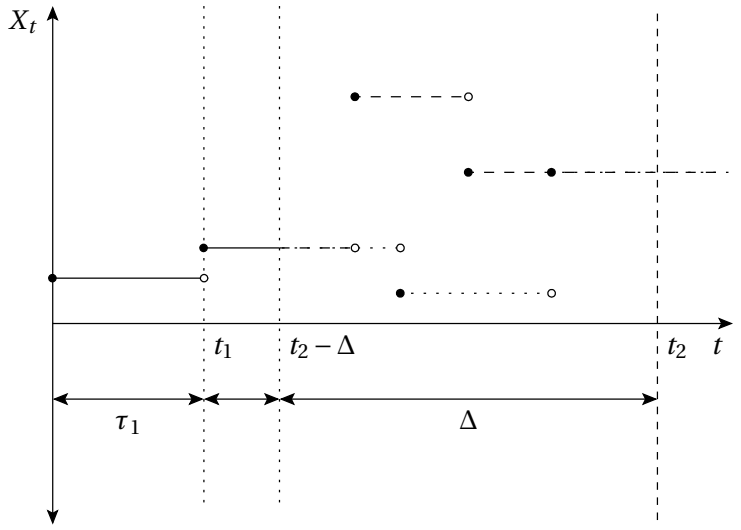

FIGURE 1.4: Time aggregation: two distinct paths which begin in the same state at $t_2 - \Delta$ and end in the same state at t_2 but differ over intermediate interval of length Δ.

in discrete time.[13]

To illustrate the issues involved, Figure 1.4 displays two distinct paths which coincide both before and after an interval of length Δ, but which take different intermediate steps. Consider the possible paths of the process between times $t_2 - \Delta$ and t_2. The dashed path first moves to a higher state before arriving at the resulting state x_{t_2}, while the dashed and dotted path first moves to a lower state and arrives in x_{t_2} at a later time (but before t_2). There are an infinite number of such paths since time is continuous, but the dynamics of the process over the interval are summarized by the transition matrix $P(\Delta)$.

For example, consider the single agent renewal model of Section 1.3 with $K = 5$ states. The intensity matrix Q_0 gives the rates at which the state changes due to nature. Suppose that the state increases one level at rate q_1 and two levels at rate q_2. Then, Q_0 for this

[13] In a discrete-time, sequential-move model with a prespecified number of sub-periods, time aggregation would require enumerating and integrating over all possible paths. Conveniently, in continuous time this information is summarized by the matrix exponential from Section 1.2.

model is

$$Q_0 = \begin{bmatrix} -q_1 - q_2 & q_1 & q_2 & 0 & 0 \\ 0 & -q_1 - q_2 & q_1 & q_2 & 0 \\ 0 & 0 & -q_1 - q_2 & q_1 & q_2 \\ 0 & 0 & 0 & -q_1 - q_2 & q_1 + q_2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Let σ_k denote the conditional choice probability of choosing to renew—moving the state back to 1 deterministically—in state k. Note that σ_k is determined endogenously and depends on the parameters θ through the value function as in (1.5). Under our assumptions, σ_k will have a logit form. If λ is the rate at which moves arrive, then Q_1 is

$$Q_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \lambda\sigma_2 & -\lambda\sigma_2 & 0 & 0 & 0 \\ \lambda\sigma_3 & 0 & -\lambda\sigma_3 & 0 & 0 \\ \lambda\sigma_4 & 0 & 0 & -\lambda\sigma_4 & 0 \\ \lambda\sigma_5 & 0 & 0 & 0 & -\lambda\sigma_5 \end{bmatrix}.$$

The first row contains only zeros because the model remains at state 1 regardless of which action is taken. The remaining diagonal elements are $-\lambda + \lambda(1-\sigma_k)$ where λ is the rate at which the model potentially *leaves* state k and $\lambda(1-\sigma_k)$ is the rate at which the state potentially remains unchanged yielding a net exit rate of $-\lambda\sigma_k$. The aggregate intensity matrix in this case is $Q = Q_0 + Q_1$, where the corresponding probability function $P(t)$ is used for estimation.

Embeddability and Identification

There are two fundamental issues concerning estimation using time aggregated data, which arise in both the continuous-time and discrete-time settings. The first is the *embeddability* problem: could the observed discrete-time transition matrix $P(\Delta)$, associated with a time interval of length Δ, have in fact been generated by the proposed data generating process (some continuous-time Markov structure with intensity matrix Q or some discrete-time chain over fixed time periods of length δ). If so, then we can address the

second question regarding *identification*: whether there is a unique parametrization of the underlying model which, when sampled at intervals of length Δ, gives rise to the transition matrix $P(\Delta)$.

In the continuous-time framework, let $\mathscr{Q}$ be the set of all valid intensity matrices and $\mathscr{P}_0 \equiv \{\exp(Q) : Q \in \mathscr{Q}\}$ be the set of all possible generated transition matrices. The question of embeddability asks whether $P(\Delta) \in \mathscr{P}_0$.[14] Singer and Spilerman (1976) provide several necessary conditions for embeddability involving testable conditions on the determinant and eigenvalues of $P(\Delta)$. Essentially, this is simply a model specification issue: was the data actually generated by a continuous time process? In this paper, we assume that the model is correctly specified and therefore, such a Q matrix exists.

The problem of identification is about whether there is a unique matrix $Q \in \mathscr{Q}$ such that $P(\Delta) = \exp(Q)$. Singer and Spilerman (1976) provide several sufficient conditions that must be verified to guarantee that Q is unique, for example if the eigenvalues of $P(\Delta)$ are distinct, real, and positive, if $\min_i\{P_{ii}(\Delta)\} > 1/2$, or if $\det P(\Delta) > e^{-\pi}$.

Although these two issues are typically discussed in the context of continuous-time models, both apply to discrete-time models as well. Consider a discrete-time model with a fixed move interval of length δ which may be different from the fixed interval Δ at which observations are sampled. In practice, researchers typically assume (implicitly) that $\delta = \Delta$, where $\Delta = 1$ is normalized to be some specific unit of time (e.g., one quarter).[15] This assumption is convenient but effectively assumes away the embeddability and identification problems.

In this case, the issue of embeddability is then whether the observed discrete time transition matrix $P(\Delta)$ could have in fact been generated by (another) discrete time

[14] This problem was first proposed by Elfving (1937). Kingman (1962) derived the set of embeddable processes with $K = 2$ and Johansen (1974) gives an explicit description of the set for $K = 3$.

[15] Our continuous-time model is more flexible than the discrete-time model in the sense that we estimate the rate of move arrivals, rather than fixing it at unity. The closest analog to discrete time is found by setting $\lambda \equiv 1$.

transition matrix over intervals of length δ. Stated somewhat differently, embeddability is satisfied if there exists a matrix P_0 such that $P_0^{\Delta/\delta} = P(\Delta)$. In general, the root P_0 need not exist. The identification question is then, if the root P_0 exists, is it unique? In general there may be multiple such matrices (Singer and Spilerman, 1976, p. 49). These issues become trivial under the aforementioned assumption that $\delta = \Delta$.

1.5.4 CCP-Based Estimation

We introduce CCP estimation in terms of the single-agent model for simplicity. Application to the multi-agent model follows directly and is discussed at the end of this section. CCP estimation relies on finding a mapping from the CCPs σ_{jk} to the value function V_k. When separated at the time of the next *event*, the value function as expressed in (1.3) contains both terms involving V_k directly, as well as the familiar "social surplus" term which is typically used to obtain the inverse mapping. These extra terms preclude the use of the usual inverse CCP mapping. However, when the value function is separated instead at the time of the player's next *move*, application of the inverse mapping is straightforward.

The derivation is very similar to the next-event representation of Section 1.3, but we now need to consider that between any two moves by the agent, any number of other state jumps could have occurred. For example, if the model is initially in state k and no move arrival occurs on the interval $[0, \tau_1)$ while the state follows the dynamics of the underlying Markov jump process, we know that the probability of being in any state l at time $t \in [0, \tau_1)$ is $P_{kl}(t)$, where $P(t)$ are the jump probabilities associated with the intensity matrix Q_0. The total payoff obtained over $[0, \tau_1)$, discounted to the beginning of the interval, is therefore $\int_0^{\tau_1} \mathrm{e}^{-\rho t} \sum_{l=1}^{K} P_{kl}(t) u_l \, dt$.

The next-move representation of the value function in state k, is

$$V_k = \mathrm{E}\left[\int_0^{\tau_1} \mathrm{e}^{-\rho t} \sum_{l=1}^{K} P_{kl}(t) u_l \, dt + \mathrm{e}^{-\rho \tau_1} \sum_{l=1}^{K} P_{kl}(\tau_1) \max_j \{\psi_{jl} + \varepsilon_j + \upsilon_{jl}\} \right]. \tag{1.9}$$

Note that this is simply an alternate representation of the value function in (1.3), ex-

pressed in terms of the next *move* time instead of the next *event* time. These representations are equivalent.[16]

The first term above represents the flow utility obtained from the initial time until the first move arrival at time τ_1. The second term represents the expected instantaneous and future utility obtained from making a choice at time τ_1. The resulting state l at time τ_1 is stochastic, as is the optimal choice j and, possibly even the continuation value v_{jl}. The expectation operator is needed because τ_1 is also random and unknown a priori.

If $\varepsilon_j \sim \text{TIEV}(0,1)$, then the CCPs admit the following closed form:

$$\sigma_{jk} = \frac{\exp(\psi_{jk} + v_{jk})}{\sum_{m=1}^{J} \exp(\psi_{mk} + v_{mk})}. \tag{1.10}$$

Suppose we wish to express this probability with respect to another state, say state 1, then we can write

$$\sigma_{jk} = \frac{\exp(\psi_{jk} + v_{jk} - \psi_{j1} - v_{j1})}{\sum_{m=1}^{J} \exp(\psi_{mk} + v_{mk} - \psi_{m1} - v_{m1})}. \tag{1.11}$$

The ψ_{jk}'s typically have closed forms in terms of the parameters. Thus, if we know differences in the continuation values $v_{jk} - v_{j1}$, we effectively know the CCPs and can estimate the model. In what follows, we show how to obtain these differences using first stage estimates of the CCPs and a closed form inverse relationship with the value function.

First, note that from (1.10) we can write

$$\ln\left[\sum_{m=1}^{J} \exp(\psi_{mk} + v_{mk})\right] = -\ln\sigma_{jk} + \psi_{jk} + v_{jk}. \tag{1.12}$$

The left side of this expression is precisely the closed form for the ex-ante future value term in the value function.

[16] In Section 1.2 we noted that the next-move representation introduces matrix exponential calculations, increasing the computational burden of evaluating the functional mapping. However, in the context of CCP estimation this can be completed in a preliminary step and only needs to be carried out for the states actually observed in the data and the neighboring states required for calculating expectations.

Single-Agent Example

In many model specifications we can then obtain an expression for the differences in (1.11) by choosing an appropriate normalizing state.[17] We use the example model of Section 1.3 to illustrate this point. In terms of this model, we can write (1.12), for $j = 1$ as

$$\ln\left[\exp(V_k) + \exp(V_1 - c)\right] = -\ln\sigma_{1k} + V_1 - c. \tag{1.13}$$

Note that the left-hand side of the above equation is exactly the expression in the value function as expressed in (1.4). Substituting (1.13) into (1.4) gives the following expression for the value function for each state k:

$$\begin{aligned} V_k &= \mathrm{E}\left[\int_0^{\tau_1} \mathrm{e}^{-\rho t} \sum_{l=1}^{K} P_{kl}(t)\, u_l\, dt + \mathrm{e}^{-\rho\tau_1} \sum_{l=1}^{K} P_{kl}(\tau_1)\,(-\ln\sigma_{1l} + V_1 - c)\right] \\ &= \mathrm{E}\left[\int_0^{\tau_1} \mathrm{e}^{-\rho t} \sum_{l=1}^{K} P_{kl}(t)\, u_l\, dt - \mathrm{e}^{-\rho\tau_1} \sum_{l=1}^{K} P_{kl}(\tau_1) \ln\sigma_{1l} + \mathrm{e}^{-\rho\tau_1}(V_1 - c)\right] \end{aligned}$$

where in the second equality we have used the fact that $V_1 - c$ does not depend on l and that the probabilities $P_{kl}(t)$ must sum to one over $l = 1, \ldots, K$. Evaluating the above expression at $k = 1$ and differencing gives

$$V_k - V_1 = \mathrm{E}\left[\int_0^{\tau_1} \mathrm{e}^{-\rho t} \sum_{l=1}^{K} \left[P_{kl}(t) - P_{1l}(t)\right] u_l\, dt - \mathrm{e}^{-\rho\tau_1} \sum_{l=1}^{K} \left[P_{kl}(\tau_1) - P_{1l}(\tau_1)\right] \ln\sigma_{1l}\right].$$

This expression gives differences in the value function in terms of the conditional choice probability σ_{1l}. With first-stage estimates of σ_{1l} for each l we can use this expression to "invert" the estimated CCPs to obtain an approximation of $V_k - V_1$ which can then be used, along with (1.5), to approximate $\sigma(a_t, x_t; \theta)$ in the likelihood. The result is a pseudo-likelihood function which can be maximized to obtain an estimate of θ.

[17] See Arcidiacono and Miller (2008) for a general discussion.

In dynamic games, in the interval between an arbitrary time $t < \tau_i$ and τ_i, any combination of state jumps and moves by other players may take place. Q_0 describes the dynamics of state jumps, and we can construct similar intensity matrices Q_i that describe the dynamics of events caused by the actions of rival players. In any state k, player i moves at a rate λ_i which is constant across k.

Thus, the rate at which the model leaves state k due to player i is λ_i. The rate at which the model enters another state $l \neq k$, the (k,l) element of Q_i, is given by the sum

$$\lambda_i \sum_{j=1}^{J} \sigma_{ijk} \phi_{ijkl},$$

which accounts for uncertainty both over the choice and the resulting state. Intuitively, this is the probability of moving to state l expressed as a proportion of λ_i, the rate at which the model leaves state k. Note that we must also allow for the state to remain at k, in which case the diagonal (k,k) element of Q_i is

$$-\lambda_i + \lambda_i \sum_{j=1}^{J} \sigma_{ijk} \phi_{ijkk}.$$

From the perspective of player i, the dynamics of the model follow an intensity matrix $Q_{-i} \equiv Q_0 + \sum_{j \neq i} Q_j$ which captures all events caused by nature and player i's rivals. With this intensity matrix in hand, the flow utility portion of the value function can be expressed exactly as before with $P^{-i}(t)$ being constructed using the intensity matrix Q_{-i}: $\int_0^{\tau_i} \mathrm{e}^{-\rho t} \sum_{l=1}^{K} P_{kl}^{-i}(t) u_{il} \, dt$. The value function for player i is then

$$V_{ik} = \mathrm{E}\left[\int_0^{\tau_i} \mathrm{e}^{-\rho_i t} \sum_{l=1}^{K} P_{kl}^{-i}(t) u_{il} \, dt + \mathrm{e}^{-\rho \tau} \sum_{l=1}^{K} P_{kl}^{-i}(\tau_i) \max_j \{\psi_{ijl} + \varepsilon_{ij} + v_{ijl}\} \right]. \tag{1.14}$$

CCP estimation of the quality ladder model, for example, can now proceed as in the single agent case by recognizing that exiting is a terminal state. Hence, at the time of

the next move, the continuation value can be expressed simply as the negative of the log probability of exiting.

Computational Issues

There are several computational issues to consider when evaluating the next-move based value function in both the single- and multi-agent cases. In practice, for CCP estimation, we are actually interested in approximating the difference $V_k - V_1$. For simplicity, we will discuss methods for approximating V_k. Approximating the difference is straightforward using the same procedures.

Consider the single-agent version in (1.9). The expectation of the $\max\{\cdot\}$ term is the ex-ante expected future value of choosing optimally in state l. We can isolate this term using the law of iterated expectations, replacing $\mathrm{E}_{\tau_1,\varepsilon}$ with $\mathrm{E}_{\tau_1}\mathrm{E}_{\varepsilon|\tau_1}$. If we then make the standard assumption that the ε_j are iid and distributed according to the type I extreme value distribution, we can simplify this expression using the known closed form for the maximum of J values $\{\delta_1+\varepsilon_1,\ldots,\delta_J+\varepsilon_J\}$:

$$\mathrm{E}\left[\max\{\delta_1+\varepsilon_1,\ldots,\delta_J+\varepsilon_J\}\right] = \ln\left[\exp(\delta_1)+\cdots+\exp(\delta_J)\right].$$

See, for example, McFadden (1984) for details.

Now, we must still choose how to evaluate both the flow utility term as well as the expectation over τ_1. We describe two Monte Carlo integration methods for doing so. The first involves simulating from the distribution of τ_1 and using a closed form for the flow utility term. The second involves averaging the flow utility and discounted future value over many simulated paths of the combined jump process, starting from the current time and ending at the next move by the player in question. The first method involves a lower-dimensional integral but requires many matrix exponential calculations. The second approach involves approximating a more complex integral, but avoids potentially costly matrix calculations.

In the first approach, we simply approximate the expectation over τ_1 using Monte Carlo integration by drawing R values of τ_1, $\{\tau_1^s\}_{s=1}^R$, and forming the following approximation:

$$V_k \approx \frac{1}{R}\sum_{s=1}^{R}\left[\int_0^{\tau_1^s} \mathrm{e}^{-\rho t}\sum_{l=1}^{K} P_{kl}(t)\,u_l\,dt + \mathrm{e}^{-\rho\tau_1^s}\sum_{l=1}^{K} P_{kl}(\tau_1^s)\,\mathrm{E}\max_j\{\psi_{jk}+\varepsilon_j+v_{jk}\}\right]. \tag{1.15}$$

In matrix notation, the flow utility term has a relatively simple closed form which allows (1.15) to be calculated directly. To see this, let $b_i(\tau_1) = \int_0^{\tau_1} \mathrm{e}^{-\rho s}\sum_j P_{ij}(s)\,u(x_j)\,ds$, $B(\tau_1) = (b_1(\tau_1), b_2(\tau_1), \ldots, b_K(\tau_1))^\top$ and $U = (u(x_1), \ldots, u(x_n))^\top$. Define $C \equiv -(\rho I - Q)$ for simplicity. Then we can write the first term inside the expectation in matrix notation as

$$\begin{aligned} B(\tau_1) &= \int_0^{\tau_1} \mathrm{e}^{-\rho s I}\,\mathrm{e}^{sQ}\,U\,ds = \left[\int_0^{\tau_1} \mathrm{e}^{-s(\rho I - Q)}\,ds\right] U \\ &= \left[\int_0^{\tau_1} C^{-1} C\,\mathrm{e}^{sC}\,ds\right] U = C^{-1}\left[\int_0^{\tau_1} C\,\mathrm{e}^{sC}\,ds\right] U = C^{-1}\left[\mathrm{e}^{\tau_1 C} - I\right] U. \end{aligned}$$

Finally, substituting for C we have

$$B(\tau_1) = -(\rho I - Q)^{-1}\left[\mathrm{e}^{-\tau_1(\rho I - Q)} - I\right] U.$$

The alternate approach is able to avoid the computation of matrix exponentials altogether. We can approximate V_k using a forward simulation procedure where we simulate R paths of the joint jump process governed by the aggregate intensity matrix $Q = Q_0 + \sum_i Q_i$. Each path begins at the current time, in state k, and ends when the player in question moves next (a simulated realization of τ_1). The flow utility obtained over each path is accumulated and the discounted future value term at the final state is calculated (via CCP inversion when working in differences). Averaging both the flow utility and future value terms over the R simulated paths and discounting appropriately provides an approximation to V_k.

In either case, it is important to note that the value function only needs to be approximated at states that are relevant for estimation. We can focus only on those states that

are actually observed in the sample and any related states which are used in the choice probability calculations that appear in the log likelihood function. That is, we only need to know the value function at each observed state k and each additional state that might arise as a result of some action at state k (e.g., exit or investment). As a result, even when the state space is very large the number of components of the value function that need to be calculated is simply a function of the observations in the sample. This can result in considerable computational savings.

1.5.5 Unobserved Heterogeneity

Incorporating permanent unobserved heterogeneity into the models above follows the same method commonly used in the dynamic discrete choice literature. Namely, we can use finite mixture distributions to allow for permanent unobserved characteristics.[18]

Consider, for example, the bus engine problem but where certain buses now have higher replacement costs or mileage transitions. The type-specific likelihood for a particular bus is composed of the type-specific probabilities of the mileage and engine transitions over the course of the sample period. The log likelihood for a particular bus with the unobserved state integrated out is then the log of the sum of the type-specific likelihoods weighted by the population probabilities of being each of the different types. For the nested fixed point algorithm, estimation is then straightforward. With CCP estimation, the techniques developed by Arcidiacono and Miller (2008) apply to the continuous time setting as well.

1.6 Monte Carlo Experiments

1.6.1 Single Agent Dynamic Discrete Choice

Here, we generate data according to the simple single player binary choice model of Section 1.3. The primitives of the model are the payoff (mileage cost) parameter β, the

[18] See Keane and Wolpin (1997), Eckstein and Wolpin (1999), Arcidiacono (2005) and several others.

intensity matrix (mileage transition) parameters q_1 and q_2, the reset (engine replacement) cost c, the discount rate ρ, and the move arrival rate λ. We fix $\rho = 0.05$ and focus on estimating $\theta = (\lambda, q_1, q_2, \beta, c)$.

In the first set of experiments, we use a full solution approach to estimate the model. Namely, the value function is obtained through value function iteration for each value of θ while the log likelihood function is maximized in an outer loop. We estimate the model under several different scenarios including full continuous-time data, continuous-time data when the decision not to replace the engine is not observed, and discrete time data of varying resolution. In each experiment we fixed the discount rate, $\rho = 0.05$, the number of states, $K = 10$, and the number of draws used for Monte Carlo integration, $R = 250$. Additional details regarding data generation and estimation can be found in the appendix. The means and standard deviations of the parameter estimates are reported in Table 1.1. All are centered around their true values and estimated quite precisely. The loss in precision from moving away from continuous time data is initially greatest for the move arrival rate, λ, yet all estimates of this parameter are still very precise. The replacement cost, c, also loses precision with more coarsely sampled data, but the increases are not large until we move to seeing only one in four events on average in the sampling period.

Results for two-step estimation using conditional choice probabilities are displayed in Table 1.2. We use a simple bin estimator to obtain the CCPs in a first stage. Details about how these first-stage estimates were obtained in the time aggregation and partial move cases can be found in Appendix B. Using CCPs increases the standard deviations slightly, reflecting noise from the first stage. However, the estimates are still very good, particularly when the average number of state changes per sampling interval is small.

Finally, we also estimated the model with continuous-time data while allowing for buses to be of two distinct types, where the type is not observed by the econometrician. In this specification, the type affected both the mileage transition probabilities and

payoff parameters. In particular, with probability π, the bus is of the first type and with probability $1-\pi$, the bus is of the second type. For buses of type $m = 1,2$, the mileage jumps forward one unit at rate q_1 and two units at rate q_{2m}, the cost of mileage is β, and the cost of replacement is c_m. Again, estimation proceeded quickly with little difficulty in separating the unobserved heterogeneity from the other model parameters. The results are reported in Table 1.3.

1.6.2 A Dynamic Discrete Game

Our second set of Monte Carlo experiments corresponds to the quality ladder model described in Section 1.4. We estimate models ranging from 10 to 20 firms with 7 possible quality levels. The size of the state space for our largest problem is over four and a half million. In all experiments, as before, we fixed $\rho = 0.05$ and used $R = 250$ draws for Monte Carlo integration. Further details can again be found in Appendix B.

Table 1.4 summarizes the results for full-solution estimation, where we obtain the value function using value function iteration for each trial value of θ. Table 1.5 presents the analogous results obtained using CCP estimation, where we assume the true CCPs are available. In all cases, both full-solution methods and CCP estimation perform extremely well and there is virtually no change in the standard deviations across the different state space sizes.

We then compare the computational time required for both full-solution and CCP estimation in Table 1.6. We first report the number of players $\bar{N}$, the market size M, and the total number of states K. For each model, computational times are reported for only one replication. Since we consider many models, the overall trends are clear despite the fact that we do not report averages.[19]

The first timing column reports the time required to obtain the value function V for

[19] All reported times are for estimation on a desktop PC with a quad-core 64-bit AMD Phenom II X4 920 processor. Our programs are written in Fortran and take advantage of parallel processing in obviously parallel segments of code. Again, we use L-BFGS-B to maximize the log-likelihood function in each case.

each model specification. This step is necessary to either generate a dataset or to simulate the model (e.g. to perform counterfactuals). In particular, we use datasets consisting of $M = 200$ markets with $T = 100$ continuous time events observed in each. Next we report the time required to estimate the first stage parameters λ and γ. This step is independent of the method used to obtain the value function. Next, we report the total time required to estimate the second stage parameters κ, η, and η^e via full solution estimation. For each new trial value of θ, we use the value function at the previous θ as the starting value for the value function iteration. Finally, we report the setup time required to perform the initial forward simulation procedure described in Section 1.5.4 (with $R = 250$), the time required to estimate the second-stage parameters, and the sum of these two times (the total time).[20]

Even with over four and a half million states, full solution estimation took under five hours. Conditional on already having the CCPs from a first stage, two-step estimation times were incredibly fast, with the longest taking less than two minutes. To put these numbers in perspective, Doraszelski and Judd (2008) note that it would take about a year to just solve for the equilibrium of a 14 player game (with 9 levels of quality) using the basic Pakes-McGuire algorithm.[21] Our continuous-time approach takes about 20 minutes to solve the game and under two hours to estimate the parameters using a full solution (NFXP) approach. CCP estimation requires less than a minute. These computational times suggest that very large classes of problems can be easily estimated in a continuous-time framework. Furthermore, the computational time required to calculate the fixed point once in continuous time is small even for very large problems. This implies that simulating counterfactuals from large-scale models will not be an issue.

[20] This table does not address the time required to estimate the first-stage CCPs, which can vary significantly depending on which method is used. Parametric methods can clearly be quite fast while fully nonparametric methods can be computationally intensive.

[21] Similar computational times are also reported in Doraszelski and Pakes (2007).

1.7 Conclusion

While recently developed two-step estimation methods have made it possible to estimate large-scale dynamic games, performing simulations for counterfactual work or generating data remains severely limited by the curse of dimensionality that arises from simultaneous moves. We recast the standard discrete-time, simultaneous-move game as a sequential-move game in continuous time. This breaks the curse of dimensionality, greatly expanding the breadth and applicability of these structural methods and making even full-solution estimation feasible for very large games.

Furthermore, by building on an underlying discrete-choice random utility model, our model preserves many of the desirable features of discrete-time models. In particular, we show that the insights from two-step estimation methods can be applied directly in our framework, resulting in another order of magnitude computational gain during estimation. We also show how to extend the model to accommodate incomplete sampling schemes, including missing actions and time-aggregated data. Both are likely to be relevant for real-world datasets.

Our framework suggests a number of areas for future research. First, we currently do not allow players to influence the arrival rate of move opportunities. It is reasonably straightforward to endogenize these rates (over a finite set) using the methods described in Puterman (2005). This would, however, increase the data requirements substantially. Second, we have focused exclusively on models with discrete state spaces. Future work is needed to extend these models to continuous state spaces, which will be key to allowing for asymmetric information in continuous time games. If players only observe their rivals' actions with a lag, then the time since player i's last move, which is continuous, becomes a relevant state variable for player i's value function. We believe this is a particularly promising area for future research.

Table 1.1: Single player Monte Carlo results: NFXP estimation

Sampling	n	q_1	q_2	λ	β	c
Population	∞	0.150	0.050	0.200	1.000	1.250
Continuous Time	10,000	0.150	0.050	0.200	1.009	1.254
		(0.002)	(0.001)	(0.003)	(0.068)	(0.054)
Passive Moves	7,176	0.150	0.050	0.204	1.010	1.271
		(0.002)	(0.001)	(0.020)	(0.127)	(0.126)
$\Delta = 0.625$	40,000	0.137	0.053	0.189	1.107	1.305
		(0.003)	(0.002)	(0.019)	(0.213)	(0.238)
$\Delta = 1.25$	20,000	0.145	0.051	0.191	1.074	1.191
		(0.003)	(0.002)	(0.024)	(0.210)	(0.297)
$\Delta = 2.5$	10,000	0.147	0.051	0.198	1.014	1.167
		(0.004)	(0.002)	(0.027)	(0.334)	(0.408)
$\Delta = 5.0$	5,000	0.151	0.050	0.195	1.088	1.233
		(0.007)	(0.003)	(0.019)	(0.249)	(0.402)
$\Delta = 10.0$	2,500	0.158	0.048	0.200	1.010	1.108
		(0.019)	(0.007)	(0.022)	(0.397)	(0.618)

We simulate 100 datasets over an observation window of length $T = 25,000$ and report the means and standard deviations of the parameter estimates. Here, Δ denotes the discrete-time observation interval and n denotes the average number discrete-time observations or continuous-time events.

Table 1.2: Single player Monte Carlo results: CCP estimation

Sampling	n	q_1	q_2	λ	β	c
Population	∞	0.150	0.050	0.200	1.000	1.250
Continuous Time	10,000	0.150	0.050	0.200	1.015	1.256
		(0.002)	(0.001)	(0.003)	(0.064)	(0.053)
Passive Moves	7,176	0.150	0.050	0.187	0.830	1.157
		(0.002)	(0.001)	(0.012)	(0.148)	(0.094)
$\Delta = 0.625$	40,000	0.137	0.053	0.196	1.114	1.367
		(0.003)	(0.002)	(0.041)	(0.267)	(0.272)
$\Delta = 1.25$	20,000	0.145	0.051	0.211	1.066	1.370
		(0.003)	(0.002)	(0.053)	(0.301)	(0.325)
$\Delta = 2.5$	10,000	0.147	0.051	0.219	1.094	1.377
		(0.004)	(0.002)	(0.103)	(0.333)	(0.421)
$\Delta = 5.0$	5,000	0.151	0.050	0.222	1.092	1.350
		(0.007)	(0.003)	(0.089)	(0.373)	(0.499)
$\Delta = 10.0$	2,500	0.154	0.049	0.241	1.159	1.356
		(0.018)	(0.008)	(0.157)	(0.516)	(0.733)

We simulate 100 datasets over an observation window of length $T = 25,000$ and report the means and standard deviations of the parameter estimates. Here, Δ denotes the discrete-time observation interval and n denotes the average number discrete-time observations or continuous-time events. The CCPs were estimated in a first step using a bin estimator for continuous-time data and via logistic regression on x and $\ln x$ for estimation with time aggregation.

Table 1.3: Single player Monte Carlo results with unobserved heterogeneity

M	n	q_1	q_{21}	q_{22}	π	λ	β	c_1	c_2
∞	∞	0.150	0.050	0.030	0.700	0.200	1.000	1.000	2.000
25	100	0.150	0.051	0.031	0.677	0.201	1.040	0.986	2.003
		(0.006)	(0.004)	(0.005)	(0.115)	(0.005)	(0.303)	(0.111)	(0.255)
50	100	0.151	0.050	0.030	0.693	0.201	1.045	0.995	2.001
		(0.004)	(0.003)	(0.004)	(0.070)	(0.004)	(0.188)	(0.067)	(0.141)
100	100	0.151	0.051	0.030	0.689	0.201	1.023	0.994	1.994
		(0.003)	(0.002)	(0.002)	(0.058)	(0.003)	(0.137)	(0.049)	(0.107)
25	200	0.150	0.050	0.030	0.685	0.200	1.025	1.004	2.002
		(0.003)	(0.003)	(0.003)	(0.092)	(0.004)	(0.176)	(0.061)	(0.118)
50	200	0.151	0.050	0.030	0.694	0.201	1.033	1.009	2.008
		(0.003)	(0.002)	(0.002)	(0.073)	(0.003)	(0.136)	(0.041)	(0.102)
100	200	0.151	0.050	0.030	0.701	0.201	1.014	1.002	1.995
		(0.002)	(0.001)	(0.002)	(0.047)	(0.002)	(0.096)	(0.029)	(0.062)

We simulate 100 datasets containing M markets each with n observed continuous-time events and report the means and standard deviations of the parameter estimates.

Table 1.4: Quality ladder Monte Carlo results: NFXP estimation

$\bar{N}$	K	M	n_{avg}	ω_{avg}	λ	γ	κ	η	η^e
	Population				1.800	0.200	0.800	4.000	5.000
10	80,080	5.0	6.62	3.79	1.820	0.201	0.798	3.986	4.967
					(0.005)	(0.001)	(0.026)	(0.204)	(0.171)
11	136,136	7.0	7.79	3.62	1.819	0.201	0.791	3.990	4.952
					(0.005)	(0.002)	(0.031)	(0.198)	(0.174)
12	222,768	8.0	8.29	3.47	1.821	0.201	0.798	4.010	5.007
					(0.006)	(0.001)	(0.024)	(0.192)	(0.163)
13	352,716	9.0	8.81	3.35	1.821	0.200	0.801	4.043	5.044
					(0.006)	(0.001)	(0.031)	(0.184)	(0.157)
14	542,640	10.0	9.32	3.22	1.821	0.200	0.801	4.043	5.044
					(0.006)	(0.001)	(0.031)	(0.184)	(0.157)
15	813,960	11.0	9.86	3.13	1.822	0.201	0.811	4.012	5.047
					(0.005)	(0.002)	(0.073)	(0.211)	(0.257)
16	1,193,808	13.0	10.89	3.00	1.822	0.200	0.837	3.967	5.152
					(0.005)	(0.001)	(0.103)	(0.223)	(0.402)
17	1,716,099	15.0	11.88	2.90	1.820	0.201	0.836	3.984	5.181
					(0.006)	(0.002)	(0.107)	(0.201)	(0.363)
18	2,422,728	17.0	12.90	2.81	1.821	0.200	0.808	3.999	5.030
					(0.006)	(0.002)	(0.060)	(0.195)	(0.268)
19	3,364,900	19.0	13.91	2.72	1.820	0.201	0.809	3.987	5.139
					(0.006)	(0.002)	(0.078)	(0.190)	(0.272)
20	4,604,600	21.0	14.92	2.64	1.820	0.200	0.801	4.009	5.129
					(0.006)	(0.0020	(0.084)	(0.194)	(0.308)

The mean and standard deviation of the parameter estimates for 25 samples are shown for different choices of $\bar{N}$, the total number of players, and M, the market size, with $\bar{\omega}$ fixed at 7. K denotes the total number of distinct states, n_{avg} denotes the average number of active players, and ω_{avg} denotes the average quality level. Samples consisted of 1000 markets each with 100 observed events.

Table 1.5: Quality ladder Monte Carlo results: CCP estimation

$\bar{N}$	K	M	n_{avg}	ω_{avg}	λ	γ	κ	η	η^e
	Population				1.800	0.200	0.800	4.000	5.000
10	80,080	5.0	6.62	3.79	1.822	0.202	0.777	4.075	5.072
					(0.005)	(0.001)	(0.012)	(0.236)	(0.235)
11	136,136	7.0	7.79	3.62	1.821	0.202	0.774	4.099	5.080
					(0.005)	(0.002)	(0.014)	(0.253)	(0.247)
12	222,768	8.0	8.29	3.47	1.823	0.202	0.775	4.088	5.086
					(0.006)	(0.001)	(0.013)	(0.236)	(0.228)
13	352,716	9.0	8.81	3.35	1.823	0.202	0.779	4.076	5.071
					(0.006)	(0.001)	(0.010)	(0.226)	(0.227)
14	542,640	10.0	9.32	3.22	1.823	0.202	0.780	4.076	5.069
					(0.005)	(0.001)	(0.015)	(0.238)	(0.228)
15	813,960	11.0	9.86	3.13	1.824	0.202	0.782	4.068	5.060
					(0.005)	(0.002)	(0.014)	(0.218)	(0.210)
16	1,193,808	13.0	10.89	3.00	1.824	0.202	0.787	4.067	5.060
					(0.005)	(0.001)	(0.016)	(0.212)	(0.199)
17	1,716,099	15.0	11.88	2.90	1.822	0.202	0.782	4.066	5.064
					(0.006)	(0.002)	(0.013)	(0.209)	(0.201)
18	2,422,728	17.0	12.90	2.81	1.823	0.202	0.784	4.070	5.064
					(0.006)	(0.002)	(0.014)	(0.210)	(0.189)
19	3,364,900	19.0	13.91	2.72	1.822	0.202	0.782	4.055	5.047
					(0.006)	(0.002)	(0.016)	(0.210)	(0.185)
20	4,604,600	21.0	14.92	2.64	1.822	0.201	0.783	4.061	5.050
					(0.006)	(0.002)	(0.017)	(0.204)	(0.189)

The mean and standard deviation of the parameter estimates for 25 samples are shown for different choices of $\bar{N}$, the total number of players, and M, the market size, with $\bar{\omega}$ fixed at 7. K denotes the total number of distinct states, n_{avg} denotes the average number of active players, and ω_{avg} denotes the average quality level. Samples consisted of 1000 markets each with 100 observed events. The true CCPs were used in estimation.

Table 1.6: Computational time: NFXP vs CCP

					NFXP		CCP	
$\bar{N}$	M	K	Solve V	First Stage	Estimate	Setup	Estimate	Total
5	0.5	2,310	9.05	0.42	51.61	0.51	0.30	0.81
6	1.0	5,544	15.59	0.30	107.41	1.54	0.32	1.86
7	2.0	12,012	29.26	0.28	172.40	2.17	0.47	2.64
8	3.0	24,024	58.27	0.30	256.19	4.02	0.62	4.64
9	4.0	45,045	107.40	0.26	375.26	5.58	1.01	6.59
10	5.0	80,080	185.78	0.35	535.83	7.13	1.42	8.55
11	7.0	136,136	325.28	0.34	639.98	11.20	1.80	13.00
12	8.0	222,768	518.57	0.33	1,069.52	13.47	3.21	16.69
13	9.0	352,716	821.83	0.34	1,411.32	14.96	3.63	18.59
14	10.0	542,640	1,228.98	0.39	2,436.61	17.21	4.10	21.31
15	11.0	813,960	1,719.72	0.38	3,413.42	19.67	7.15	26.82
16	13.0	1,193,808	2,499.98	0.44	4,765.67	23.85	7.06	30.91
17	15.0	1,716,099	3,642.02	0.43	13,513.96	27.28	8.38	35.66
18	17.0	2,422,728	5,109.30	0.41	10,807.54	30.93	10.54	41.47
19	19.0	3,364,900	6,929.01	0.43	13,737.87	35.67	15.37	51.04
20	21.0	4,604,600	9,377.57	0.41	16,069.89	37.26	15.60	52.86

Times are reported in seconds for estimation on a desktop PC with a quad-core AMD Phenom II X4 920 processor using code written in Fortran using OpenMP for parallel processing in obviously parallel segments of code. Times are reported for only one replication of each specification. $\bar{N}$ denotes the total possible number of players, M denotes the market size, and K denotes the total number of distinct states. We have fixed the number of possible quality levels at $\bar{\omega} = 7$. Obtaining the value function v once is required for generating data. Obtaining the first stage estimates is a common step for both NFXP and CCP estimation.

2

Partial Identification and Inference in Binary Choice and Duration Panel Data Models

2.1 Introduction

Many economic variables of interest are qualitative in nature and therefore discrete response models have become a standard tool in applied econometrics and their properties have been studied thoroughly in the econometrics literature (McFadden, 1974; Maddala, 1983; Amemiya, 1985). Semiparametric methods such as maximum score have emerged to estimate such models without tenuous parametric assumptions, however, these methods typically assume the existence of an exogenous explanatory variable with full support (Manski, 1975, 1985; Horowitz, 1992). Similar rank conditions have been successful in estimating more general regression models but the known conditions for point identification still include a full support condition (Han, 1987; Abrevaya, 2000). In practice, however, it is not uncommon to encounter datasets with genuinely discrete or bounded variables. In general, without a regressor with full support on the real line, under semiparametric assumptions the models we consider are only partially identified (Horowitz, 1998).

This chapter develops estimators for a general class of partially identified models with limited support regressors and provides conditions for consistency, obtaining rates of convergence, and constructing confidence regions. While the previous literature has focused on partially identified *regular* models which give rise to set estimators that are essentially $\sqrt{n}$-consistent,[1] this chapter provides conditions under which *irregular* rates of convergence may also arise. Our analysis is motivated by several semiparametric fixed effects panel data models including binary choice and duration models. We apply our general results to several models and show that depending on the assumptions made on the support of the regressors, the set estimators may achieve nearly cube-root convergence or they may converge arbitrarily fast.

In a broad sense, this chapter concerns econometric models characterized by a finite vector of parameters θ which lie in some parameter space Θ. Our particular focus is on semiparametric models which also have unknown infinite-dimensional components, such as the distribution of disturbances, which are not specified a priori and are not of interest themselves. However, to address the concepts of partial identification it suffices to consider a standard parametric model. Suppose that the data generating process, the distribution of observables, is induced by a true parameter $\theta_0 \in \Theta$ which is unknown by the researcher and is the primary object of interest. The model is *point identified* if θ_0 is the only element of Θ such that the model would be consistent with the population distribution P_{θ_0}, assuming for a moment that it were perfectly observable. On the other hand, the model is *partially identified* if there are multiple elements $\theta \in \Theta$ that are observationally equivalent to θ_0, that is, such that $P_\theta = P_{\theta_0}$. The set of all such θ is the *identified set* and is denoted Θ_I. See Manski (2003) and Tamer (2009) for surveys of partial identification in econometric models.

This chapter contributes to both the emerging literature on partial identification and

[1] That is, they can achieve rates arbitrarily close to $1/\sqrt{n}$ as in Chernozhukov, Hong, and Tamer (2007).

the broad literature on nonlinear panel data models. First, it presents general inference results for two new classes of models: models with continuous but potentially bounded regressors which may have non-standard rates of convergence and models with discrete regressors which are characterized by a discontinuity in the population objective function at the boundary of the identified set. Our results parallel those of Chernozhukov et al. (2007) in that we propose criterion-function-based set estimators, derive their rates of convergence, and propose a subsampling-based (Politis, Romano, and Wolf, 1999) procedure for obtaining confidence regions. We obtain these results under new conditions which are applicable to the specific cases we consider: binary choice panel data models and panel data duration models with discrete or continuous (but potentially bounded) regressors. Thus, this chapter also contributes to the subset of the partial identification literature which is concerned with semiparametric estimation of models with limited support regressors, as well as to the nonlinear semiparametric panel data literature. We provide sharp characterizations of the identified sets of the fixed effects models we consider which are then used to motivate estimators. The consistency and rates of convergence of these estimators are established, as is the validity of subsampling for constructing confidence regions in these models.

This chapter is organized as follows. First, Section 2.2 provides a brief review of the related literature. Then, in Section 2.3, we formally describe the specific models and assumptions that motivate our analysis. Subsequent sections first introduce general definitions or theorems and then apply them to the panel data binary choice models we consider. In particular, Section 2.4 focuses on identification, Section 2.5 discusses consistent estimation and rates of convergence, and Section 2.6 proposes a subsampling-based algorithm for performing inference in a class of discrete models. We discuss extensions to a class of panel data duration models in Section 2.7. Several Monte Carlo experiments are described in Section 2.8 and Section 2.9 concludes.

2.2 Related Literature

This chapter is related to several topics in the econometrics literature. First, it contributes to a series of papers on criterion-function-based estimation and inference in partially identified models beginning with Manski and Tamer (2002), who consider regression models with interval data. They derive the sharp identified set in a semiparametric binary response model with an interval-valued regressor under a conditional quantile restriction and propose a set estimator which is defined as an appropriately-chosen contour set of a modified maximum score objective function. This estimator is shown to be consistent. In addition to nonparametric estimation, they also consider modified minimum distance and maximum likelihood estimation of parametric models. Chernozhukov et al. (2007) develop a general framework for criterion-function-based estimation of partially identified models, obtain rates of convergence, and construct confidence regions using subsampling. They apply their general results to models characterized by moment equalities and inequalities. Romano and Shaikh (2008, 2009) further explore subsampling-based inference in partially identified models. Bugni (2008), on the other hand, introduces a bootstrap procedure for performing inference. He also works within the criterion function framework and considers models characterized by a finite number of moment equalities and inequalities.

A second, fundamentally different method for constructing confidence regions in partially identified models is based on set expansion. Expanding the identified set requires a better understanding of its boundary, which is easy to characterize, for instance, when the identified set is an interval on the real line. See Horowitz and Manski (2000) and Imbens and Manski (2004) for examples of the use of set expansion. Beresteanu and Molinari (2008) extend this method to more general settings and develop inference procedures based on the theory of random sets for partially identified models where the identified set can be expressed as the Aumann expectation of a set valued random

variable.

There is also a distinction made in the literature, pointed out by Imbens and Manski (2004), between two possible objects of interest: the identified set itself, which is the focus of the present chapter, and individual points within the identified set, including the true parameter θ_0. Stoye (2009) observes that the conditions of Imbens and Manski (2004) implicitly assume the existence of a superefficient estimator of the width of the identified interval. He revisits the problem under assumptions that both weaken and remove this condition. Note that although some of the estimators proposed in this chapter are superefficient, this arises due to the inherent properties of the model, not as a result of an implicit assumption.

There are numerous other areas where partially identified econometric models have arisen including, games with multiple equilibria (Tamer, 2003; Andrews, Berry, and Jia, 2004; Pakes, Porter, Ho, and Ishii, 2006; Aradillas-Lopez and Tamer, 2008; Ciliberto and Tamer, 2009; Beresteanu, Molchanov, and Molinari, 2009), and models characterized by conditional moment inequalities (Khan and Tamer, 2009; Kim, 2009; Andrews and Shi, 2009).

Of particular relevance to the present chapter is a growing literature on semiparametric binary response models with limited support regressors, typically involving either discrete or interval-valued regressors. In terms of cross-sectional models, Bierens and Hartog (1988) show that there are infinitely many single-index representations of the mean regression of a dependent variable when all covariates are discrete. Horowitz (1998) discusses the non-identification of single-index and binary response models with only discrete regressors. Generic non-identification results such as these serve to motivate our analysis.

Manski and Tamer (2002) and Magnac and Maurin (2008) consider partial identification and estimation of binary choice models with an interval-valued regressor. This is a

related, but different source of partial identification than those that we consider. Honoré and Lleras-Muney (2006) estimate partially identified competing risk models with interval outcome data and discrete explanatory variables. Komarova (2008) considers partial identification in static binary response models with discrete regressors. Despite using a different methodology, part of the present chapter is similar to her work in that we consider a fixed effects panel extension of the static binary choice model with discrete regressors. However, our analysis differs substantially in that we also consider models with continuous regressors and analyze other unrelated models. Even similarities in the binary choice case are limited since, for example, sharpness of the identified set does not follow directly from the cross-sectional case since we must account for the distribution of the fixed effect in the panel case.

Previous papers have considered partial identification in panel data models, with different points of departure and quantities of interest. They highlight the importance of studying the identifying power of various assumptions and provide practitioners with methods to assess the robustness of their results. In particular, Honoré and Tamer (2006) analyze dynamic random effects panel data models and discus how to calculate the identified set using minimum distance, maximum likelihood, and linear programming methods. Chernozhukov, Fernández-Val, Hahn, and Newey (2009) derive bounds on marginal effects in nonlinear panel models with discrete regressors. Rosen (2009) considers partial identification in fixed effects panel data models under conditional quantile restrictions.

This chapter is also related to the point-identified fixed effects panel data literature, especially the semiparametric analysis of Manski (1987) for the basic fixed effects model and Honoré and Kyriazidou (2000) for dynamic models with lagged dependent variables. Our characterizations of the identified sets in the models we consider are based in part on known necessary conditions for point identification established in these papers, however,

establishing sharpness in partially identified models requires additional work.

2.3 Models and Assumptions

We consider panel data models where observations are available at times $t = 0, \ldots, T-1$ for each individual. An individual in the model is described completely by a random vector $(y_0, x_0, u_0, \ldots, y_{T-1}, x_{T-1}, u_{T-1}, c)$, where y_t is a binary response variable in period t, x_t is a vector of k observed explanatory variables, u_t is an unobserved disturbance in period t, and c is a time invariant individual-specific unobserved effect. Let $y \equiv (y_0, \ldots, y_{T-1})$ and define x and u similarly. Let F denote the joint distribution of (y, x, u, c) and let P denote the underlying probability measure generating F. In this case, F_{yx} is the joint distribution of the observed variables. Our first objective is to combine our knowledge of F_{yx} and a set of weak semiparametric assumptions on F to determine the identified set of parameters of interest. We let θ denote the finite vector of parameters of interest and we will denote the set of possible values of θ by Θ. We assume F is induced by some true unknown parameter θ_0.

In the models we consider, the distribution of the available regressors may not be rich enough to point identify θ_0 without additional assumptions. Therefore, we focus instead on the identified set Θ_I which contains θ_0 itself, as well as all other parameter vectors which cannot be distinguished from θ_0. We address these issues in depth in Section 2.4.

Our goal is to combine data and prior knowledge about the joint distribution F to learn about θ. First, note that we can always write F as the product of conditional distributions $F = F_{y|xcu} F_{u|xc} F_{c|x} F_x$. In principle, F_x is observable and therefore any restrictions on it should be determined by the data. Much of the literature assumes assumes the presence of at least one component of x, say x_1, which has full support conditional on the remaining components $x_2, \ldots, x_k$. Instead, we consider what can be learned about θ without this assumption in order to develop methods which are

appropriate to datasets with only discrete regressors, regressors with compact support, or which otherwise fail to satisfy a full support condition. The present chapter focuses on models for which $F_{y|xcu}$ will be fully specified. For example, in panel data discrete choice models, $F_{y|xcu}$ is determined by an underlying latent variable model. Following the fixed effects literature, $F_{c|x}$ will not be restricted in any way. We will, however, restrict $F_{u|xc}$ with a standard stationarity assumption used in the literature.

2.3.1 Basic Fixed Effects Panel Data Model

We begin with the fundamental restriction on $F_{y|xcu}$ which defines the basic linear-index fixed effects binary response model.

Model 2.1 (Fixed Effects Model)**.** *For all t,*

$$y_t = 1\{x_t'\beta + c + u_t \geq 0\} \tag{2.1}$$

where x_t is a random variable with support $\mathscr{X} \subseteq \mathbb{R}^k$, c is a real-valued random variable, and $\theta = \beta$ is the parameter of interest, a member of some parameter space $\Theta \subseteq \mathbb{R}^k$. In addition, for all x and c, $F_{u_t|xc}$ satisfies the following:

a. $F_{u_t|xc} = F_{u_0|xc}$ for all t.

b. The support of u_t is $\mathbb{R}$.

Here, $1\{\cdot\}$ denotes the indicator function, equal to one when the event $\{\cdot\}$ is true and zero otherwise. Condition a above is a substantive restriction, necessary for the estimation methods we introduce below. It requires u_t to be is stationary conditional on the identity of the panel member—that is, conditional on (x, c). Note, however, that it does not restrict the form of serial dependence of u_t in any way. Condition b is a regularity condition which serves to ensure that for any c, the event $y_1 \neq y_0$ occurs with positive probability. Otherwise, the model provides no information about θ.

2.3.2 *Limited Support Regressors*

Now, turning to F_x, we begin by reviewing existing conditions for point identification. In the cross-sectional model with a conditional median restriction, analogous to the fixed effects model above, Manski (1985) showed that a full rank, full support condition on x was sufficient to point identify β up to scale. That is, he assumes that x is not contained in a proper linear subspace of $\mathbb{R}^k$ and that the first component of x has positive density everywhere on $\mathbb{R}$ for almost every value of the remaining components. The same conditions were invoked by Han (1987) for the maximum rank correlation estimator and Horowitz (1992) for the smoothed maximum score estimator. The panel version of this assumption (for $T = 2$) was used by Manski (1987) to establish point identification of β up to scale in a semiparametric fixed effects panel data model of the kind considered in the present chapter.

Thus, modulo assumptions on the disturbances, point identification of β hinges on the assumptions one is willing to make on the underlying data generating process. The validity of a full support assumption depends critically on the particular explanatory variables available for and relevant to a particular application. It is therefore up to the researcher to determine whether it holds. Many common variables such as age, number of children, years of education, and gender are inherently discrete and so in many cases the decision will be clear. Similarly, many variables such as income have only partial support on the real line (e.g., $\mathbb{R}^+ \subset \mathbb{R}$). The estimators proposed in this chapter do not distinguish between the point identified and partially identified cases. They exploit additional information available from regressors with full support if available, but do not require it.

We consider two alternatives to the full support condition. The first applies when x_t is a discrete random variable with finite support. The second applies when at least one component of $x_t - x_{t-1}$ is continuous but may fail to have full support on $\mathbb{R}$.

Assumption 2.1 (Discrete Regressors)**.** x_t is a discrete random vector with finite support $\mathscr{X} \subset \mathbb{R}$. That is, $|\mathscr{X}| < \infty$, where $|\mathscr{X}|$ denotes the cardinality of the set $\mathscr{X}$.

This assumption applies to models which include only genuinely discrete explanatory variables, including indicator variables.

Assumption 2.2 (Continuous Regressor)**.** The first component of the vector $x_1 - x_0$ has positive density everywhere on a set $\mathscr{W}_1 \subseteq \mathbb{R}$ for almost every value of the remaining components.

Note that this assumption does not rule out the possibility that $\mathscr{W}_1 = \mathbb{R}$ but it also includes cases where the support $x_1 - x_0$ is bounded in some sense. Therefore, this condition includes variables with one-sided support such as income, which is non-negative. As we discuss in detail below, the implications of these two assumptions for estimation are very different.

2.3.3 Lagged Dependent Variable Model

We also consider a lagged dependent variable model, an extension to the basic fixed effect model which allows for state dependence. Since we do not observe y_t in periods prior to the sample, the model is left unspecified in the first period.

Model 2.2 (Lagged Dependent Variable Model)**.** *The choice probabilities in the first period are* $P(y_0 = 0 \mid x, c) = p_0(x, c)$*, where* p_0 *is unknown and* $0 < p_0(x, c) < 1$ *for all* x *and* c*. In subsequent periods* $t = 1, \ldots, T$*,*

$$y_t = 1\{x_t'\beta + \gamma y_{t-1} + c + u_t \geq 0\} \tag{2.2}$$

where x_t *is a random vector with support* $\mathscr{X}$*,* c *is a real-valued random variable, and* $\theta = (\beta, \gamma)$ *are the parameters of interest which lie in some parameter space* $\Theta \subseteq \mathbb{R}^{k+1}$*. In addition, the unobservables* u_t *are serially independent, identically distributed with cdf* $F_{u_t|xc} = F_{u_0|xc}$ *for all* t*, and have full support on* $\mathbb{R}$*.*

Note that in this model, as opposed to the basic fixed effects model, we do not allow serial correlation in the disturbances. The full support assumption on u_t is a regularity condition which guarantees that certain events used for estimation occur with positive probability.

2.3.4 Panel Data Duration Models

We also consider estimation of fixed effects panel data versions of a general class of transformation models.

Model 2.3 (Panel Data Transformation Model)**.** *For all t,*

$$\Lambda(y_t) = x_t'\beta + c + u_t \tag{2.3}$$

where Λ is a strictly monotonic function, x_t is a random vector with support $\mathscr{X}$, c is a real-valued random variable, and $\theta = \beta$ is the parameter of interest which lies in some parameter space $\Theta \subseteq \mathbb{R}^k$. The disturbances u_t are serially independent with identical distribution $F_{u_0|xc}$ and independent of x.

Here, t denotes a single spell. The covariates x_t remain constant within a spell, but vary may across spells. Again, c is a time-invariant individual-specific unobserved variable.

This model is quite general and contains many common duration models in their panel data forms with individual specific time invariant unobserved heterogeneity. For example, the generalized accelerated failure time (GAFT) model of Ridder (1990) is of this form. The mixed proportional hazards model arises when u_t has the minus extreme value distribution with $F_{u_0|xc}(u) = 1 - \exp(-\mathrm{e}^u)$ and Λ is the log integrated baseline hazard function

2.4 Identification

We begin our identification analysis by developing a broad definition of the identified set in a generic regression model which can later be applied to the specific models we consider. Let F_{yx} denote the joint distribution of (y, x), the observable variables, and v, a vector of unobservables. In Model 2.1, for example, we have $v = (c, u)$. Let θ be a vector of parameters of interest and let Θ be the parameter space, the set of all feasible values of θ. Assume that we observe the marginal distributions $F_{y|x}$ and F_x, but not F_v. The unknown primitives of the model are thus θ and F_v. Let $\pi(\cdot \mid \theta, F_v, x)$ denote the distribution of $y \mid x$ implied by the model under θ and F_v. The set of primitives that are observationally equivalent to $F_{y|x}$ is thus

$$\Psi(F_{yx}) = \{(\theta, F_v) : \pi(y \mid \theta, F_v, x) = F_{y|x}(y \mid x)\ F_x - \text{a.s.}, y - \text{a.e.}\}.$$

Definition. The *identified set* for θ given $F_{y|x}$ is

$$\Theta_I(F_{yx}) = \left\{\theta \in \Theta : \exists F_v \text{ such that } (\theta, F_v) \in \Psi(F_{yx})\right\}. \tag{2.4}$$

This set is *sharp* by definition in the sense that each $\theta \in \Theta_I(F_{yx})$ is consistent with F_{yx} and cannot be rejected given the maintained assumptions of the model. Henceforth, we simply write Θ_I, with the dependence on F_{yx} understood.

We also assume throughout that the model is correctly specified: $\Theta_I \neq \emptyset$. See Komarova (2008) for a discussion of misspecification in terms of the closely-related static binary choice model.

Note that we do not rule out cases where point identification obtains. If the model is actually point identified, then our estimates will converge to a point. In practice, models with richer regressor support will have a smaller identified set. Our Θ_I characterizes this set, but when $x_1 - x_0$ has richer support, Θ_I naturally becomes smaller. For example, in the fixed effects model with discrete regressors, considered below, the number of

equality constraints defining Θ_I increases with the cardinality of the support of $x_1 - x_0$. Intuitively speaking, when a component is continuous, but perhaps bounded, the number of equalities becomes infinite. When a component has full support, Θ_I collapses to a singleton. This may happen in other situations as well.

2.4.1 Fixed Effects Model

In Model 2.1, the primitives of the model are β, $F_{u_0|xc}$, and $F_{c|x}$. We now provide a characterization of Θ_I in terms of observables and show that it is equivalent to the identified set defined above. Since c is unobserved, in order to estimate β we must find implications of the model that are independent of c.

Our identification analysis follows that of Manski (1987). Although our characterization of the identified set is based on a previously known necessary condition for point identification, our characterization of the identified set and the conclusion that it is sharp in this setting are new. The following theorem provides a tractable representation of the identified set, Θ_I, in terms of observables: $P(y_0 = 1 \mid x)$, $P(y_1 = 1 \mid x)$, and F_x.

Theorem 2.1. *In Model 2.1,*

$$\Theta_I = \left\{\theta \in \Theta : \operatorname{sgn}\left(P(y_1 = 1 \mid x) - P(y_0 = 1 \mid x)\right) = \operatorname{sgn}\left((x_1 - x_0)'\beta\right)\ F_x - a.s.\right\}. \tag{2.5}$$

Proof. See Appendix C.4. ■

Henceforth, in discussions of Model 2.1, we use (2.5) to characterize the identified set rather than the general definition given in (2.4).

2.4.2 Lagged Dependent Variable Model

In this section we turn to the identification of Model 2.2. Our analysis follows along the lines of Chamberlain (1985) and Honoré and Kyriazidou (2000) and we focus on the case where $T = 4$. Again, although we build on a previously established necessary condition

for identification in point identified models, our characterization of the identified set in the present partially identified model is new.

The identification of the model is based on comparing observations for which we observe the same outcome in periods 0 and 3 but different outcomes in periods 1 and 2. Consider the following events for given values of $d_0, d_3 \in \{0,1\}$:

$$A = \{y_0 = d_0, y_1 = 0, y_2 = 1, y_3 = d_3\},$$

$$B = \{y_0 = d_0, y_1 = 1, y_2 = 0, y_3 = d_3\}.$$

Letting G denote $F_{u_0|xc}$ for simplicity, the corresponding choice probabilities are:

$$\begin{aligned}
P(A \mid x, c, x_2 = x_3) = {} & p_0(x,c)^{1-d_0}(1-p_0(x,c))^{d_0} G(-x_1'\beta - \gamma d_0 - c) \\
& \times \left[1 - G(-x_2'\beta - c)\right] G(-x_2'\beta - \gamma - c)^{1-d_3} \\
& \times \left[1 - G(-x_2'\beta - \gamma - c)\right]^{d_3}, \\
P(B \mid x, c, x_2 = x_3) = {} & p_0(x,c)^{1-d_0}(1-p_0(x,c))^{d_0} \left[1 - G(-x_1'\beta - \gamma d_0 - c)\right] \\
& \times G(-x_2'\beta - \gamma - c) G(-x_2'\beta - c)^{1-d_3} \\
& \times \left[1 - G(-x_2'\beta - c)\right]^{d_3}.
\end{aligned}$$

Note that the latter probability is nonzero since u_t has full support on $\mathbb{R}$ for all t and since $p_0(x,c) > 0$. Dividing, we have

$$\begin{aligned}
\frac{P(A \mid x, c, x_2 = x_3)}{P(B \mid x, c, x_2 = x_3)} = {} & \frac{G(-x_1'\beta - \gamma d_0 - c)}{G(-x_2'\beta - \gamma - c)} \times \frac{1 - G(-x_2'\beta - c)}{1 - G(-x_1'\beta - \gamma d_0 - c)} \\
& \times \left[\frac{G(-x_2'\beta - \gamma - c)}{G(-x_2'\beta - c)}\right]^{1-d_3} \times \left[\frac{1 - G(-x_2'\beta - \gamma - c)}{1 - G(-x_2'\beta - c)}\right]^{d_3}.
\end{aligned}$$

When $d_3 = 0$,

$$\frac{P(A \mid x, c, x_2 = x_3)}{P(B \mid x, c, x_2 = x_3)} = \frac{G(-x_1'\beta - \gamma d_0 - c)}{G(-x_2'\beta - \gamma d_3 - c)} \times \frac{1 - G(-x_2'\beta - \gamma d_3 - c)}{1 - G(-x_1'\beta - \gamma d_0 - c)},$$

and when $d_3 = 1$,

$$\frac{P(A \mid x, c, x_2 = x_3)}{P(B \mid x, c, x_2 = x_3)} = \frac{G(-x_1'\beta - \gamma d_0 - c)}{G(-x_2'\beta - \gamma d_3 - c)} \times \frac{1 - G(-x_2'\beta - \gamma d_3 - c)}{1 - G(-x_1'\beta - \gamma d_0 - c)}.$$

We have used the fact that when $d_3 = 0$, $\gamma d_3 = 0$, and when $d_3 = 1$, $\gamma d_3 = \gamma$. In both cases, by the monotonicity of G,

$$P(A \mid x, c, x_2 = x_3) \geq P(B \mid x, c, x_2 = x_3) \iff -x_1'\beta - \gamma d_0 - c \geq -x_2'\beta - \gamma d_3 - c,$$

or equivalently, since this event is independent of c,

$$\operatorname{sgn}\left(P(A \mid x, x_2 = x_3) - P(B \mid x, x_2 = x_3)\right) = \operatorname{sgn}\left((x_2 - x_1)'\beta + \gamma(d_3 - d_0)\right).$$

This condition provides the foundation for our characterization of the identified set and the results of the derivation above are formalized in the following theorem.

Theorem 2.2. *In Model 2.2,*

$$\begin{aligned}(2.6)\quad \Theta_I \subseteq \tilde{\Theta}_I = \big\{\theta \in \Theta : \operatorname{sgn}&\left(P(A \mid x, x_2 = x_3) - P(B \mid x, x_2 = x_3)\right) \\ &= \operatorname{sgn}\left((x_1 - x_2)'\beta + \gamma(d_3 - d_0)\right)\ F_x - a.s.\ \forall d_0, d_3 \in \{0,1\}\big\}.\end{aligned}$$

2.5 Consistent Estimation

In the remainder of the chapter we focus on criterion-function-based estimation and inference. In this section, we first propose consistent estimators for the identified set in a class of models that satisfy a set of general conditions. We also provide rates of convergence for models with objective functions that are either step functions in the limit (e.g., Model 2.1 with discrete regressors) or that are bounded by a polynomial in $d(\theta, \Theta_I)$ on regions away from the identified set (e.g., Model 2.1 with a continuous regressor). In both cases our conditions are new. In the latter case we provide new conditions which allow analysis of irregular models with non-standard rates of convergence. We then verify the conditions of the general theorems for the specific models we consider.

First, we assume that an iid sample is available for use in estimation.

Assumption 2.3 (Sampling)**.** We observe a iid sample $\{(x_{i,0}, \ldots, x_{i,T-1}, y_{i,0}, \ldots, y_{i,T-1})\}_{i=1}^n$ drawn from the joint distribution F_{yx}.

Furthermore, we assume the existence of a population criterion function Q and a finite sample objective function Q_n. These functions must satisfy certain conditions which are stated formally below. A requirement of the population criterion function Q is that the set of parameters at which it attains its maximum must equal the identified set. The analogy principle then suggests estimating Θ_I using the set of maximizers of the sample objective function Q_n. However, in general, taking only the set of maximizers may result in an inconsistent estimator. Instead, we define the estimator $\hat{\Theta}_n(\tau_n)$ to be a contour set of Q_n for some non-negative sequence τ_n:

$$\hat{\Theta}_n(\tau_n) \equiv \left\{\theta \in \Theta : Q_n(\theta) \geq \sup_{\Theta} Q_n - \tau_n\right\}. \tag{2.7}$$

The "slackness" sequence τ_n was introduced by Manski and Tamer (2002) and has been used by Chernozhukov et al. (2007), Bugni (2008), Kim (2009), and others. Below, we determine the properties of the sequence τ_n such that $\hat{\Theta}_n$ is a consistent estimator of Θ_I.

To discuss consistency and convergence, we must be precise about which metric space we are working in. We define convergence in terms of the Hausdorff distance, a generalization of Euclidean distance for sets, on the space of all subsets of Θ. Let (Θ, d) be a metric space where d is the standard Euclidean distance. For a pair of subsets $A, B \subset \Theta$, the *Hausdorff distance* between A and B is

$$d_H(A,B) = \max\left\{\sup_{\theta\in B} d(\theta, A), \sup_{\theta\in A} d(\theta, B)\right\}, \tag{2.8}$$

where, in a slight abuse of notation, $d(\theta, A) \equiv \inf_{\theta'\in A} d(\theta,\theta')$ is the distance between a point θ and a set A. This is illustrated in Figure 2.1.

2.5.1 Consistency in General Models

This section develops generic consistency results and rates of convergence. In the following sections, the conditions of these theorems will be verified in the context of the

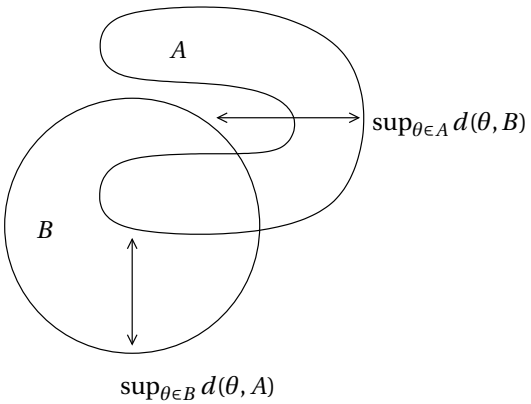

FIGURE 2.1: Hausdorff Distance

specific models discussed above. We first assume the existence of a population objective function $Q(\theta)$ that fully and exactly characterizes the identified set Θ_I. Note that once Θ_I is found, constructing Q is straightforward. Using the analogy principle, we then use the finite sample objective function $Q_n(\theta)$ to obtain a set estimator $\hat{\Theta}_n$. Finally, we shall prove that the sequence of set estimates $\hat{\Theta}_n$ converges in probability to the identified set Θ_I in the Hausdorff metric and obtain rates of convergence under different assumptions on the curvature of the objective function.

Assumption 2.4. Suppose the following conditions are satisfied:

a. Θ is a nonempty subset of $\mathbb{R}^k$ and is compact with respect to the Euclidean metric.

b. There exists a function $Q : \Theta \to \mathbb{R}$ such that $\arg\max_\Theta Q = \Theta_I$.

c. Q has a well-separated maximum in that for all $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that $\sup_{\Theta\setminus\Theta_I^\varepsilon} Q \leq \sup_\Theta Q - \delta_\varepsilon$.

d. There exists a function $Q_n : \Theta \times \mathscr{X}^T \times \mathscr{Y}^T \to \mathbb{R}$, denoted $Q_n(\theta)$, which converges uniformly in probability to Q at the $1/b_n$ rate. That is, $\sup_\Theta |Q_n - Q| = O_p(1/b_n)$ for some sequence $b_n \to \infty$.

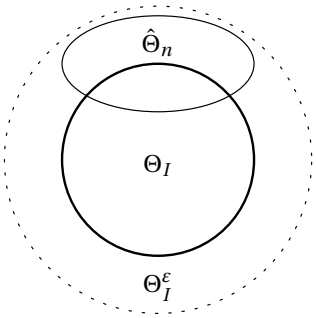

FIGURE 2.2: One-sided consistency without slackness

Part c is a regularity condition which rules out pathological cases that can arise without a continuity assumption. It is satisfied in the models we consider, for example, when Q is continuous or when Q is a step function.

Theorem 2.3 (Consistency in General Models)**.** *Suppose Assumption 2.4 holds.*

1. *If* $\tau_n \xrightarrow{p} 0$*, then* $\sup_{\theta \in \hat{\Theta}_n} d(\theta, \Theta_I) \xrightarrow{p} 0$.

2. *If* $\tau_n \xrightarrow{p} 0$ *and* $\tau_n b_n \xrightarrow{p} \infty$*, then* $\lim_{n\to\infty} P(\Theta_I \subseteq \hat{\Theta}_n) = 1$.

If both conditions hold then $d_H(\hat{\Theta}_n, \Theta_I) \xrightarrow{p} 0$.

Proof. See Appendix C.2. ■

Note that the first conclusion of Theorem 2.3 actually holds in general without slackness (i.e., with $\tau_n = 0$). This is formalized in the following corollary.

Corollary (One-Sided Consistency Without Slackness)**.** *Suppose Assumption 2.4 holds. If* $\tau_n = 0$*, then* $\sup_{\theta \in \hat{\Theta}_n} d(\theta, \Theta_I) \xrightarrow{p} 0$.

This corollary guarantees that asymptotically, without slackness, $\hat{\Theta}_n$ is close to Θ_I. The converse need not be true in general since $\sup_{\theta \in \Theta_I} d(\theta, \hat{\Theta}_n)$ may be large, as illustrated by Figure 2.2.

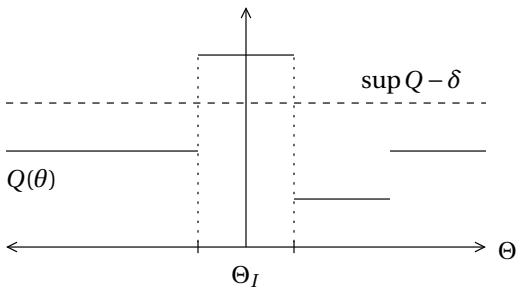

FIGURE 2.3: Infinite curvature of $Q(\theta)$

2.5.2 Rates of Convergence in General Models

The rate of convergence of the Hausdorff distance $d_H(\hat{\Theta}_n, \Theta_I)$ is the slowest rate at which the component distances $\sup_{\theta \in \Theta_I} d(\theta, \hat{\Theta}_n)$ and $\sup_{\theta \in \hat{\Theta}_n} d(\theta, \Theta_I)$ converge to zero. The second part of Theorem 2.3 establishes that with only Assumption 2.4, the first distance converges arbitrarily fast to zero in probability (because with probability approaching one, $\Theta_I \subseteq \hat{\Theta}_n$). The rate of convergence of the second component depends on the shape of the objective function. In the specific models we consider this shape depends in turn on the support of x_t. In this section, however, we prove general results by making assumptions about Q and Q_n. In later sections we provide conditions on the support of x_t that imply the required properties of these functions.

In particular, we show that when Q has a discrete jump at the boundary of Θ_I, then $\hat{\Theta}_n$ converges arbitrarily fast in probability to Θ_I. That is, for *any* sequence r_n, including powers of n and exponential forms, $r_n d_H(\hat{\Theta}_n, \Theta_I) \xrightarrow{p} 0$. This result also implies that $\hat{\Theta}_n = \Theta_I$ with probability approaching one.

On the other hand, when $Q_n(\theta)$ is stochastically bounded from above by a polynomial in $d(\theta, \Theta_I)$, we show that the rate of convergence of $\sup_{\theta \in \hat{\Theta}_n} d(\theta, \Theta_I)$ depends on both the curvature of the bounding polynomial and the rate at which τ_n converges to zero.

We begin with models that satisfy the following assumption, where Q exhibits a discrete jump at Θ_I:

Assumption 2.5 (Existence of a Constant Majorant)**.** There exists a positive constant δ

$$Q(\theta) \leq \sup_{\Theta} Q - \delta$$

for all $\theta \in \Theta \setminus \Theta_I$.

When the above condition holds, $\hat{\Theta}_n$ converges arbitrarily fast to Θ_I. This result is due to the discrete jump in Q at the boundary of Θ_I. As we will see later, this can happen when the regressors in a binary response model have discrete support. We present the theorem first, followed by a discussion of the intuition.

Theorem 2.4. *Suppose Assumptions 2.4 and 2.5 hold. If* $\tau_n \xrightarrow{p} 0$ *and* $\tau_n b_n \xrightarrow{p} \infty$*, then for any sequence* r_n*,* $r_n d_H(\hat{\Theta}_n, \Theta_I) \xrightarrow{p} 0$.

Proof. See Appendix C.2 ■

Figure 2.4 illustrates the notion that, due to the discrete nature of $Q_n(\theta)$, there are only a finite (though potentially very large) number of possible estimates $\hat{\Theta}_n$. For the realization of Q_n in the figure, the contour sets determine a partition of Θ into four disjoint sets: $\Theta = \Theta_1 \cup \Theta_2 \cup \Theta_3 \cup \Theta_4$. In the present framework, where $\hat{\Theta}_n$ is defined by a threshold τ_n, so that it includes all values of θ for which $Q_n(\theta) \geq \sup Q_n - \tau_n$, there are four possible estimates: Θ_2, $\Theta_2 \cup \Theta_3$, $\Theta_2 \cup \Theta_3 \cup \Theta_1$, and $\Theta_2 \cup \Theta_3 \cup \Theta_1 \cup \Theta_4$. In higher dimensions, and for large sample sizes, the combinatorics of the problem dictate that the number of possibilities becomes large very quickly. On the other hand, as $n \to \infty$, the contour sets of Q_n approach those of Q, and the set of possible estimates contains a set equal to Θ_I with probability approaching one. Intuitively, as we obtain more data, we are able to detect which values of θ belong to Θ_I with increasing accuracy since there is a discrete jump in $Q(\theta)$ for all θ not in Θ_I. Furthermore, since τ_n converges to zero in probability slower than Q_n converges uniformly to Q, $\hat{\Theta}_n$ converges to Θ_I. This is illustrated in Figure 2.5.

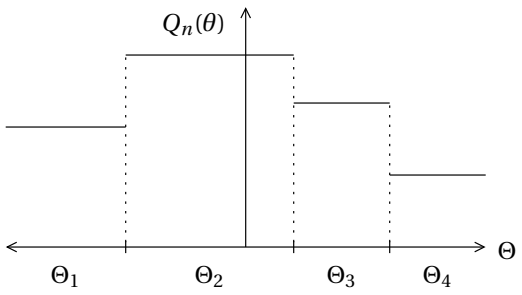

FIGURE 2.4: A realization of Q_n and the partition of Θ it generates

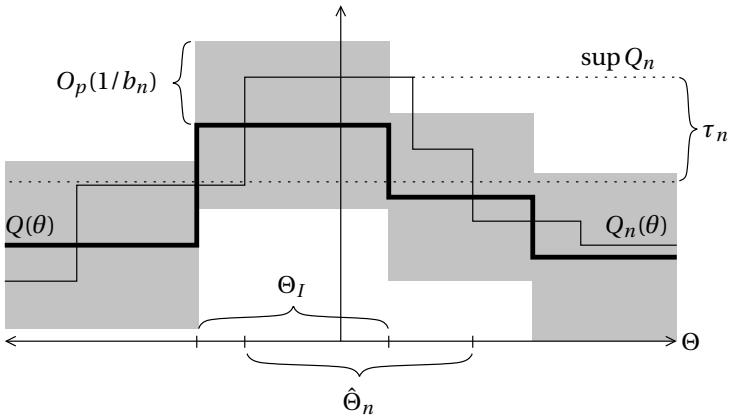

FIGURE 2.5: Convergence of Q_n to Q with b_n and τ_n

We now consider models for which Q and Q_n may be smooth, but which satisfy a curvature condition such that, outside of a shrinking neighborhood of Θ_I, Q_n is bounded in probability by a polynomial in the distance from the identified set. This condition is analogous to conditions used to obtain rates of convergence in point identified models.

Assumption 2.6 (Existence of a Polynomial Majorant)**.** There exist positive constants $(\delta, \kappa, \gamma_1, \gamma_2)$ with $\gamma_1 \geq \gamma_2$ such that for any $\varepsilon \in (0,1)$ there are $(\kappa_\varepsilon, n_\varepsilon)$ such that for all $n \geq n_\varepsilon$,

$$Q_n(\theta) \leq \sup_\Theta Q_n - \kappa \cdot (d(\theta, \Theta_I) \wedge \delta)^{\gamma_1}$$

uniformly on $\{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\varepsilon / b_n)^{1/\gamma_2}\}$ with probability at least $1 - \varepsilon$.

Theorem 2.5. *Suppose Assumptions 2.4 and 2.6 hold. If* $\tau_n \xrightarrow{p} 0$ *and* $\tau_n b_n \xrightarrow{p} \infty$*, then* $d_H(\hat{\Theta}_n, \Theta_I) = O_p(\tau_n^{1/\gamma_2})$.

Proof. See Appendix C.2 ■

2.5.3 Fixed Effects Binary Choice Model

In this section we focus on consistent estimation of Model 2.1. We first propose population and finite sample criterion functions and show that the population criterion function characterizes the identified set exactly. Then, we verify the conditions of Theorem 2.3, making use of empirical process techniques, to show that the estimator is consistent. Finally, we obtain the rate of convergence in two cases: models with only discrete regressors, under Assumption 2.1, and models with a continuous regressors, under assumption Assumption 2.2. In these cases we verify, respectively, the assumptions for Theorem 2.4 and Theorem 2.5.

Objective Function

The population objective function we propose for use in estimating Model 2.1 is the maximum score objective function of Manski (1987), a panel data analog of the cross-sectional maximum score objective function of Manski (1975, 1985):

$$Q(\theta) = \mathrm{E}\left[(y_1 - y_0)\,\mathrm{sgn}\left((x_1 - x_0)\beta\right)\right].$$

The corresponding finite sample analog objective function is

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_{i1} - y_{i0})\,\mathrm{sgn}\left((x_{i1} - x_{i0})\beta\right).$$

Note that although essentially the same objective function is used for maximum score estimation in the point identified case, the set estimators proposed here are fundamentally different since they are defined as contour sets of this function. Also, note that $Q(\theta)$ and

$Q_n(\theta)$ effectively condition on the event $y_1 \neq y_0$. This does not result in a loss of efficiency since, as established by Theorem 2.1, the event $y_1 = y_0$ is not informative about θ.

Lemma 2.1 below establishes the equivalence between the identified set Θ_I and the set of maximizers of the population objective function.

Lemma 2.1. *Under the maintained assumptions of Model 2.1,*

$$\arg\max_{\theta \in \Theta} Q(\theta) = \Theta_I.$$

Proof. See Appendix C.4. ■

Consistency

We verify each of the conditions of Assumption 2.4 in order to use the general consistency result of Theorem 2.3. In doing so, we will make use of empirical process concepts such as the subgraph of a function, Vapnik-Chervonenkis (VC) classes of sets, and Euclidean classes of functions. We refer the reader to Section 2 of Pakes and Pollard (1989) for definitions and further details. Essentially, we construct a class of functions $\mathscr{F}$, indexed by Θ, such that $Q(\theta) = Pf_\theta$ and $Q_n(\theta) = P_n f_\theta$ for $f_\theta \in \mathscr{F}$. We begin by defining $\mathscr{F}$ and establishing that it is Euclidean.

Lemma 2.2. *Let* $f(z, w, \theta) = z \cdot (2 \cdot 1\{w'\beta \geq 0\} - 1)$. *Then, the class* $\mathscr{F} = \{f(\cdot, \cdot, \theta) : \theta \in \Theta\}$ *is Euclidean for the constant envelope* $F = 1$.

Proof. See Appendix C.4. ■

Now that we have established that the objective function is generated by an underlying Euclidean class of functions, we can use tools from empirical process theory to establish the uniform convergence required for consistency. In particular, we make use of a result from Kim and Pollard (1990) to establish uniform convergence of Q_n to Q at the rate $1/b_n$ with $b_n = n^{-1/2}$.

Lemma 2.3 (Uniform Convergence of Q_n to Q)**.** *Under Assumption 2.3,*

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| = O_p(n^{-1/2}).$$

Proof of Lemma 2.3. $\mathscr{F}$ is Euclidean, so it is also manageable in the sense of Pollard (1989) (cf. Pakes and Pollard, 1989, p. 1033). Since $\int F^2 \, dP = 1 < \infty$, the result follows from Corollary 3.2 of Kim and Pollard (1990). ■

Finally, combining the above results, we can apply Theorem 2.3 to establish consistency of $\hat{\Theta}_n$ for Model 2.1.

Theorem 2.6. *Suppose Assumption 2.3 holds in Model 2.1. If* $\tau_n \xrightarrow{p} 0$*, and* $\tau_n n^{1/2} \xrightarrow{p} \infty$*, then* $d_H(\hat{\Theta}_n, \Theta_I) \xrightarrow{p} 0$.

Proof. See Appendix C.4. ■

Rates of Convergence

The rate of convergence of $\hat{\Theta}_n$ to Θ_I in Model 2.1 depends on the support of x_t. We obtain the rate under both Assumption 2.1 and Assumption 2.2. We show that when the support of x_t is finite, $\hat{\Theta}_n$ converges arbitrarily fast in probability to Θ_I. On the other hand, when at least one component of $x_2 - x_1$ is continuous, the estimator can achieve rates arbitrarily close to $n^{-1/3}$. The rate depends on τ_n and, although the exact rate $n^{-1/3}$ is not achievable, in practice, one can achieve convergence close to $n^{-1/3}$ by choosing, for example, $\tau_n \propto \sqrt{\ln n / n}$.

Discrete Regressors Here, we verify Assumption 2.5, the constant majorant condition, in the context of Model 2.1. We can then apply Theorem 2.4 to show that in this case, $\hat{\Theta}_n$ converges arbitrarily fast to Θ_I.

When the support of (x_0, x_1) is a finite set, henceforth $\mathscr{X}$, the objective function $Q(\theta)$ can be rewritten as follows:

$$\begin{aligned} Q(\theta) &= \mathrm{E}_x \mathrm{E}_{y|x} \left[(y_1 - y_0) \operatorname{sgn}\left((x_1 - x_0)'\beta \right) \right] \\ &= \sum_{x \in \mathscr{X}} P(x) \left[P(y_1 = 1 \mid x) - P(y_0 = 1 \mid x) \right] \operatorname{sgn}\left((x_1 - x_0)'\beta \right). \end{aligned}$$

Therefore, $Q(\theta)$ is a step function and there exists a real number $\delta > 0$ such that for all $\theta \in \Theta \setminus \Theta_I$, $Q(\theta) \leq \sup_\Theta Q - \delta$. In particular, δ is bounded below by the smallest nonzero value of $P(x) \left[P(y_1 = 0 \mid x) - P(y_0 = 1 \mid x) \right]$ for any $x \in \mathscr{X}$. Thus, applying Theorem 2.4, we have the following result.

Theorem 2.7. *Suppose that Assumption 2.1 holds in Model 2.1. For any sequence τ_n such that $\tau_n \xrightarrow{\mathrm{p}} 0$ and $n^{1/2}\tau_n \xrightarrow{\mathrm{p}} \infty$, then $\hat{\Theta}_n$ converges to Θ_I arbitrarily fast in probability in the Hausdorff metric. That is, for any sequence r_n, $r_n d_H(\hat{\Theta}_n, \Theta_I) \xrightarrow{\mathrm{p}} 0$.*

Continuous Regressors The properties of the maximum score objective function in the continuous covariate case have been studied carefully by Kim and Pollard (1990), Abrevaya and Huang (2005), and others. We follow Abrevaya and Huang (2005) in restricting the coefficient on one component of x, henceforth x_d, to be either 1 or -1 and consider β to be a vector in $\mathbb{R}^{k-1}$. Let $\tilde{x}$ denote the remaining components of x.[2]

Kim and Pollard's heuristic for cube root convergence translates almost directly to the set identified case. Let $\Gamma(\theta) \equiv Q(\theta) - Q(\theta_0)$ and $\Gamma_n(\theta) \equiv Q_n(\theta) - Q_n(\theta_0)$. We can decompose $\Gamma_n(\theta)$ into two components, a trend and a stochastic component: $\Gamma_n(\theta) = \Gamma(\theta) + [\Gamma_n(\theta) - \Gamma(\theta)]$. The limiting objective function is approximately quadratic near the identified set: $\Gamma(\theta) = O(d^2(\theta, \Theta_I))$. The variance of the empirical process component is $O_p(d(\theta, \Theta_I)/n)$. When the trend overtakes the noise, Γ_n very likely to be below the

[2] Alternatively, Kim and Pollard (1990) work with parameters in unit sphere $\mathbb{S}^{k-1} \equiv \{x \in \mathbb{R}^k : \|x\| = 1\}$ in $\mathbb{R}^k$ and assume that the angular component of x has continuous, bounded density with respect to the surface measure on $\mathbb{S}^{k-1}$.

maximum. Thus, the maximum is likely to occur when the standard deviation of the random component is of the same magnitude or larger than the trend. That is, when $\sqrt{d(\theta,\Theta_I)/n} > d^2(\theta,\Theta_I)$, or, $d(\theta,\Theta_I) < n^{-1/3}$. Therefore, $\hat{\Theta}_n$ the set of near maximizers of Γ_n, should be within an $n^{-1/3}$ neighborhood of Θ_I. In the set identified case, this is only one component of the distance. The other component, however, was shown to converge arbitrarily fast and therefore does not hinder the rate of convergence.

In terms of Theorem 2.5, the above argument corresponds to the case where $\gamma_1 = 2$ and $\gamma_2 = 3/2$. Since τ_n can be chosen arbitrarily close to $n^{-1/2}$, the rate of convergence can be made arbitrarily close to $(n^{-1/2})^{1/\gamma_2} = n^{-1/3}$. The following theorem formalizes this result. We also need several assumptions on the distribution of x, which are intentionally close to those made by Abrevaya and Huang (2005) in analyzing the cross-sectional model in the point identified case.

Let $w \equiv x_1 - x_0$ and $v \equiv u_1 - u_0$. Let F and f denote cdf and density of v and let G and g denote the cdf and density of w. Finally, let w_1 denote the first component of w and let $\tilde{w}$ denote the remaining $k-1$ components.

Theorem 2.8. *Suppose that Assumptions 2.2 and 2.3 hold in Model 2.1. In addition, suppose the following:*

a. The components of $\tilde{w}$ and $\tilde{w}\tilde{w}'$ have finite first absolute moments.

b. The function $g'(w_1 \mid \tilde{w})$ exists and, for some $M > 0$, $\left|g'(w_1 \mid \tilde{w})\right| < M$ and $\left|g(w_1 \mid \tilde{w})\right| < M$ for all w_1 and almost every $\tilde{w}$.

c. For all v in a neighborhood of 0, all w_1 in a neighborhood around $-\tilde{w}'\beta_0$, almost every $\tilde{w}$, and some $M > 0$, the function $f(v \mid \tilde{w}, w_1)$ exists and $f(v \mid \tilde{w}, w_1) < M$.

d. For all v in a neighborhood of 0, all w_1 in a neighborhood of $-\tilde{w}'\beta_0$, almost every $\tilde{w}$, and some $M > 0$, the function $\partial F(v \mid \tilde{w}, w_1)/\partial w_1$ exists and $|\partial F(v \mid \tilde{w}, w_1)/\partial w_1| < M$.

e. Θ_I *is contained in the interior of* Θ.

f. The matrix $V(\theta) \equiv \mathrm{E}\left[2f(0 \mid \tilde{w}, -\tilde{w}'\beta)g(-\tilde{w}'\beta \mid \tilde{w})\tilde{w}\tilde{w}'\right]$ *is positive semidefinite for all* $\theta \in \mathrm{bd}(\Theta_I)$.

Then for any sequence τ_n *such that* $\tau_n \xrightarrow{\mathrm{p}} 0$ *and* $n^{1/2}\tau_n \xrightarrow{\mathrm{p}} \infty$, $d_H(\hat{\Theta}_n, \Theta_I) = O_p(\tau_n^{2/3})$.

Proof. See Appendix C.4 ■

2.5.4 Lagged Dependent Variable Model

In this section, we propose a consistent estimator for Model 2.2. The proofs of the results in this section largely parallel those for the fixed effects model and therefore all proofs are reserved for Section C.5. For simplicity we only consider the lagged dependent variable model under Assumption 2.1 (discrete regressors). An extension to Assumption 2.2 (a continuous regressor) would involve the use of a kernel as in Honoré and Kyriazidou (2000), along with the associated assumptions. The kernel is used to condition on the event $x_3 = x_2$ and $x_3 - x_2$ is assumed to support in a neighborhood of zero. In the case of discrete regressors, this conditioning is accomplished with a simple indicator function.

We use the population objective function

$$Q(\theta) = \mathrm{E}\left[1\{x_2 = x_3\} \cdot (y_2 - y_1) \cdot \mathrm{sgn}((x_2 - x_1)'\beta + \gamma(y_3 - y_0))\right].$$

This function was used by Honoré and Kyriazidou (2000) for estimation in point identified models. The finite sample objective function is

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} 1\{x_{i2} = x_{i3}\} \cdot (y_{i2} - y_{i1}) \cdot \mathrm{sgn}((x_{i2} - x_{i1})'\beta + \gamma(y_{i3} - y_{i0})).$$

The set of maximizers of Q is indeed a sharp characterization of the identified set, as established by the following Lemma.

Lemma 2.4 (Objective Function Representation of $\tilde{\Theta}_I$)**.** *Under the maintained assumptions of Model 2.2,* $\arg\max_{\theta \in \Theta} Q(\theta) = \tilde{\Theta}_I$ *(as defined in Theorem 2.2).*

Proof. See Appendix C.5. ■

Next, we verify each of the conditions of Assumption 2.4 in order to use Theorem 2.3 to establish consistency of the estimator $\hat{\Theta}_n$ for $\tilde{\Theta}_I$. As in the fixed effects model, we begin by establishing that the objective function belongs to a Euclidean class of functions indexed by θ so that we can leverage results from empirical process theory.

Lemma 2.5 (Euclidean Property)**.** *The class of functions $\mathscr{F} = \{f_\theta : \theta \in \Theta\}$, where $f_\theta(x,y) = 1\{x_2 = x_3\}(y_2 - y_1)\left[2 \cdot 1\{(x_2 - x_1)'\beta + \gamma(y_3 - y_0) \geq 0\} - 1\right]$, is Euclidean for the constant envelope $F = 1$.*

Proof. See Appendix C.5. ■

As before, the Euclidean property allows us to immediately establish uniform convergence and the P-Donsker property which we will in turn use to show consistency and, later, the conditions required by our inference procedure.

Theorem 2.9. *Suppose Assumption 2.3 holds in Model 2.2. If $\tau_n \xrightarrow{p} 0$ and $\tau_n n^{1/2} \xrightarrow{p} \infty$, then $d_H(\hat{\Theta}_n, \tilde{\Theta}_I) \xrightarrow{p} 0$.*

Proof. See Appendix C.5. ■

Additionally, when Assumption 2.1 is satisfied, Q is again a step function. The argument is analogous to that for the basic fixed effects model and is reserved for the proof. Thus, applying Theorem 2.4, we again find that $\hat{\Theta}_n$ converges arbitrarily fast to $\tilde{\Theta}_I$ in probability.

Theorem 2.10. *Suppose that Assumptions 2.1 and 2.3 hold in Model 2.2. If $\tau_n \xrightarrow{p} 0$ and $n^{1/2}\tau_n \xrightarrow{p} \infty$, then for any sequence r_n, $r_n d_H(\hat{\Theta}_n, \tilde{\Theta}_I) \xrightarrow{p} 0$.*

Proof. See Appendix C.5. ■

2.6 Confidence Regions

Confidence regions for Θ_I can be formed using contour sets of Q_n in much the same way as we defined the estimator $\hat{\Theta}_n$ in (2.7). Let $C_n(\kappa_n)$ denote the set

$$C_n(\kappa_n) = \{\theta \in \Theta : b_n Q_n(\theta) \geq \sup_{\Theta} b_n Q_n - \kappa_n\}. \tag{2.9}$$

Inference is based on the statistic

$$\mathcal{Q}_n \equiv \sup_{\Theta} b_n Q_n - \inf_{\Theta_I} b_n Q_n$$

and the following equivalence:

$$\{\Theta_I \subseteq C_n(\kappa_n)\} \iff \{\mathcal{Q}_n \leq \kappa_n\}.$$

The sets $C_n(\kappa_n)$ defined in (2.9) have the same form as (2.7), except that the objective function is now normalized by b_n, the rate of uniform convergence. We apply this normalization in order to use subsampling to approximate quantiles of $\mathcal{Q}_n$. As a result, the sequence κ_n is analogous to $b_n\tau_n$. Thus, while in Theorem 2.3 we required $\tau_n \xrightarrow{p} 0$ and $b_n\tau_n \to \infty$ for consistent estimation using (2.7), we could obtain consistent estimates with (2.9) if $\kappa_n \xrightarrow{p} \infty$ and $\kappa_n / b_n \xrightarrow{p} 0$. That is, κ_n approaches infinity at a rate slower than that of b_n.

For smooth models, where $\hat{\Theta}_n$ converges at a polynomial rate and where the limiting distribution of $\mathcal{Q}_n$ is continuous, Chernozhukov et al. (2007) provide methods of constructing confidence regions which cover Θ_I asymptotically with probability $1-\alpha$ using subsampling. Their results are not applicable to the models we consider with discrete regressors due to the discrete nature of $\mathcal{Q}_n$. Instead, in the following sections, we provide conditions under which one can obtain *conservative* asymptotic confidence regions with coverage probability *at least* $1-\alpha$.

Our confidence regions are based on estimates of quantiles of $\mathcal{Q}$. To understand why the confidence regions we propose are conservative, consider the cdf and quantile functions of a generic discrete random variable X depicted in Figure 2.6. There, for example, the 0.50 and 0.75 quantiles are equal. If we use the x_2, the 0.50 quantile in an attempt to form a 50% confidence region, the coverage will actually be over 75%.

2.6.1 Confidence Regions in General Discrete Models

For now, we assume the availability of a consistent estimate $\hat{c}_n$ of the corresponding $1-\alpha$ quantile of $\mathcal{Q}$, the limiting distribution of $\mathcal{Q}_n$. In the following section, we describe an algorithm to construct such a sequence. Large sample inference with discrete regressors is based on the following lemma.[3] We require only that $\mathcal{Q}_n$ has a nondegenerate limiting distribution.

Assumption 2.7 (Convergence of $\mathcal{Q}_n$)**.** Suppose that $P\{\mathcal{Q}_n \leq c\} \to P\{\mathcal{Q} \leq c\}$ for each $c \in \mathbb{R}$, where $\mathcal{Q}$ has a nondegenerate distribution function on $\mathbb{R}$.

Lemma 2.6. *Suppose Assumption 2.7 holds. Then, for any sequence $\hat{c}_n$ such that $\hat{c}_n \xrightarrow{p} c(1-\alpha) \equiv \inf\{c : P\{\mathcal{Q} \leq c\} \geq 1-\alpha\}$ for some $\alpha \in (0,1)$,*

$$P\{\Theta_I \subseteq C_n(\hat{c}_n)\} \geq (1-\alpha) + o_p(1).$$

Proof. See Appendix C.3. ■

An appropriate sequence $\hat{c}_n$, and corresponding conservative confidence regions $C_n(\hat{c}_n)$ with asymptotic coverage probability of at least $1-\alpha$, can be constructed using the following algorithm.

---

[3] Lemma 2.6 is the discrete-distribution analog of Lemma 3.1 of Chernozhukov et al. (2007). The fundamental difference is that here, the distribution of $\mathcal{Q}$ may not be continuous. As a result, our confidence regions are conservative since we cannot place an upper bound on the coverage probability.

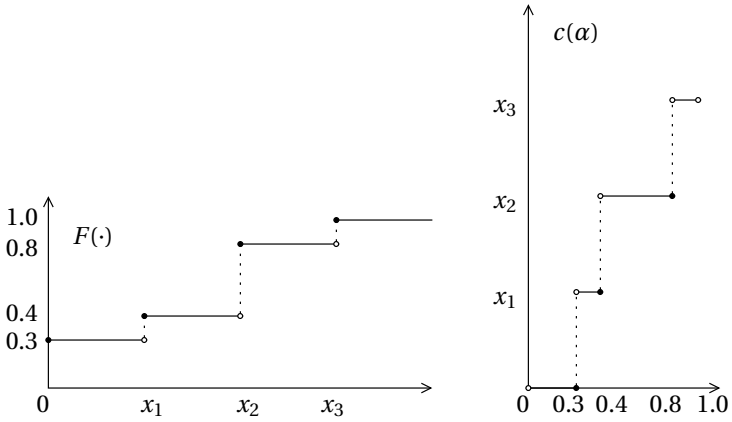

FIGURE 2.6: Cumulative distribution and quantile functions for a discrete distribution

Algorithm 2.1. 1. Choose a subsample size $m < n$ such that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$. Let M_n denote the number of subsets of size m and let κ_n be any sequence such that such that $C_n(\kappa_n)$ is a consistent estimator of Θ_I (e.g., $\kappa_n \propto \sqrt{\ln n}$).

2. Compute $\hat{c}_n$ as the $1-\alpha$ quantile of the values $\{\hat{\mathcal{Q}}_{n,m,j}\}_{j=1}^{M_n}$ where

$$\hat{\mathcal{Q}}_{n,m,j} \equiv \sup_{\theta \in \Theta} b_m Q_{n,m,j}(\theta) - \inf_{\theta \in C_n(\kappa_n)} b_m Q_{n,m,j}(\theta)$$

and $Q_{n,m,j}$ denotes the sample objective function constructed using the j-th subsample of size m.

3. Report $C_n(\kappa_n)$ as a consistent estimate of Θ_I and $C_n(\hat{c}_n)$ as a conservative confidence region.

The following theorem addresses the validity of this algorithm for obtaining the desired sequence $\hat{c}_n$. Let $a_n \downarrow a$ denote a sequence which eventually equals a, or in other words, a sequence which converges arbitrarily fast to a.[4]

Assumption 2.8 (Approximability of $\mathcal{Q}_n$)**.** Let Θ_n be a sequence of subsets of Θ such that $d_H(\Theta_n, \Theta_I) \downarrow 0$ in probability and let $\mathcal{Q}'_n = \sup_\Theta b_n Q_n - \inf_{\Theta_n} b_n Q_n$. Then $P(\mathcal{Q}'_n \leq c) \to$

[4] See Appendix C.1.1 for a precise definition of $a_n \downarrow a$, both deterministically and in probability.

$P(\mathcal{Q} \leq c)$ for each $c \in \mathbb{R}$.

Theorem 2.11. *Suppose that Assumptions 2.3, 2.4, 2.5, 2.7 and 2.8 hold and that $m \to \infty$, and $m/n \to 0$ as $n \to \infty$. Let $1-\alpha$ denote the desired coverage level, where the distribution of $\mathcal{Q}$ is continuous at $c(1-\alpha)$. Then, $\hat{c}_n \xrightarrow{p} c(1-\alpha)$.*

Proof. See Appendix C.3 ■

2.6.2 Fixed Effects Binary Choice Model

In this section we verify the conditions required for constructing confidence regions in the context of Model 2.1 under Assumption 2.1 (discrete regressors). The following lemma verifies both the convergence of $\mathcal{Q}_n$ required by Assumption 2.7 and the approximability of $\mathcal{Q}_n$ based on a sequence of estimates $\hat{\Theta}_n$, required by Assumption 2.8. Thus, this result establishes the validity of Algorithm 2.1 for constructing conservative confidence regions.

Lemma 2.7. *In Model 2.1 under Assumptions 2.1, and 2.3, both Assumptions 2.7 and 2.8 are satisfied.*

Proof. See Appendix C.4 ■

2.6.3 Lagged Dependent Variable Model

For the case of Model 2.2 with discrete regressors, the arguments to establish the validity of the subsampling procedure of Algorithm 2.1 are identical to those of the previous section for Model 2.1. This follows since both objective functions are of the same form in the underlying functions f_θ and both functions satisfy Assumption 2.5. That is, in both cases, for the appropriate class of functions $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$, we have $Q(\theta) = Pf_\theta$ and $Q_n(\theta) = P_n f_\theta$. Since both classes of functions $\mathcal{F}$ are Euclidean, it follows that Lemma 2.7 also applies to Model 2.2 under Assumption 2.1.

2.7 Panel Data Duration Models

This section considers Model 2.3 (defined on page 60), the fixed effects panel data duration model. Identification of this model and similar ones has been considered by a number of authors under a wide variety of conditions. For example, Ridder (1990) considers the nonparametric identification of the generalized accelerated failure time (GAFT) model, which contains both the mixed proportional hazards (MPH) model and the accelerated failure time (AFT). He shows that GAFT models are nonparametrically identified (up to an obvious normalization) with continuous duration data (and continuous covariates). Furthermore, it is identified even with discrete duration data with an additional parametric assumption on the regression function. We consider a similar model, when the observed durations are continuous but the covariates are discrete. Han (1987), Chen (2002), Abrevaya (2000) and others have considered point identification and estimation of various components of generalized regression models, which contain models of this type, but such studies are based on a full-support condition which we relax. Honoré and Lleras-Muney (2006) consider partial identification of a related competing risks model.

In many ways, this model is very similar to Model 2.1 and so many of the results will be familiar. When the disturbances are independent, we can carry out a similar ranking procedure relating the ordering of y_1 and y_0 to that of $x_1'\beta$ and $x_0'\beta$:

$$
\begin{aligned}
& P(y_1 \geq y_0 \mid x, c) \geq P(y_0 \geq y_1 \mid x, c) \\
& \quad \iff P(x_1'\beta + u_1 \geq x_0'\beta + u_0 \mid x, c) \geq P(x_0'\beta + u_0 \geq x_1'\beta + u_1 \mid x, c) \\
& \quad \iff P(u_0 - u_1 \leq (x_1 - x_0)'\beta \mid x, c) \geq P(u_1 - u_0 \leq (x_0 - x_1)'\beta \mid x, c) \\
& \quad \iff P(u_0 - u_1 \leq (x_1 - x_0)'\beta \mid x, c) \geq P(u_0 - u_1 \leq (x_0 - x_1)'\beta \mid x, c) \\
& \quad \iff (x_1 - x_0)'\beta \geq 0
\end{aligned}
$$

Note that we are able to exchange u_1 and u_0 due to the independence assumption.

Here we consider estimating the set suggested by the rank condition above:

$$\tilde{\Theta}_I = \left\{ \operatorname{sgn}\left(P(y_1 \geq y_0 \mid x, c) - P(y_1 \geq y_0 \mid x, c)\right) = \operatorname{sgn}\left((x_1 - x_0)'\beta\right) \right\}.$$

This set is guaranteed to contain Θ_I and establishing its relative sharpness is left for future research. The intuition underlying this set is that, due to the structure of the model, whenever $x_1'\beta \geq x_0'\beta$ it is likely also the case that $y_1 \geq y_0$.

Consider the following population objective function and sample analog:

$$Q(\theta) = \mathrm{E}\left[\operatorname{sgn}(y_1 - y_0) \cdot \operatorname{sgn}\left((x_1 - x_0)'\beta\right)\right]$$

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \operatorname{sgn}(y_{i1} - y_{i0}) \cdot \operatorname{sgn}\left((x_{i1} - x_{i0})'\beta\right)$$

Due to the similarity of the objective functions, it follows from the proof of Lemma 2.1 for Model 2.1 that Q is maximized exactly on $\tilde{\Theta}_I$.

As before, we can write $Q(\theta) = P f_\theta$ and $Q_n(\theta) = P_n f_\theta$ where

$$f_\theta(x, y) = 1\{y_1 > y_0\}\, 1\{x_1'\beta \geq x_0'\beta\} - 1\{y_1 < y_0\}\, 1\{x_1'\beta < x_0'\beta\}.$$

It should also be apparent from the arguments underlying Lemma 2.2 and Lemma 2.5 that the class $\mathscr{F} = \{f_\theta : \theta \in \Theta\}$ is Euclidean for the constant envelope $F = 1$. Therefore, the conditions of Theorem 2.3 are satisfied with $b_n = n^{1/2}$.

2.7.1 Bounding the Transformation Function

In this model, in addition to β, one might be interested in estimating the transformation function Λ. This section discusses estimating bounds for $\Lambda(\bar{y})$ at particular values of $\bar{y}$. Since we can only identify $\Lambda(\bar{y})$ up to differences with respect to $\Lambda(\bar{y}_0)$ at some value $\bar{y}_0$, we normalize $\Lambda(\bar{y}_0) = 0$.

Suppose first that θ_0 is known. Then, again following the maximum score principle, we could estimate bounds for $\Lambda(\bar{y})$ by collecting all values of λ which maximize

$$\frac{1}{n}\sum_{i=1}^{n} \left(1\{y_{i1} > \bar{y}\} - 1\{y_{i0} > \bar{y}_0\}\right) 1\{(x_{i0} - x_{i1})'\beta_0 \leq \lambda\}.$$

Estimating the set above is infeasible because θ_0 is unknown. However, given an estimated set $\hat{\Theta}_n$, the above method can be applied for each $\theta \in \hat{\Theta}_n$. This suggests using the function

$$\Gamma_n(\bar{y}, \lambda, \theta) = \frac{1}{n} \sum_{i=1}^{n} \left(1\{y_{i1} > \bar{y}\} - 1\{y_{i0} > \bar{y}_0\} \right) 1\{(x_{i0} - x_{i1})'\beta \leq \lambda\}$$

and forming a set estimate $\hat{\Lambda}_n(\bar{y})$ of $\Lambda(\bar{y})$ which consists of all values of λ which maximize Γ_n for some $\theta \in \hat{\Theta}_n$. That is,

$$\hat{\Lambda}_n(\bar{y}) = \left\{ \lambda : \lambda \in \arg\max \Gamma_n(\bar{y}, \lambda, \hat{\theta}) \text{ for some } \hat{\theta} \in \hat{\Theta}_n \right\}.$$

Establishing the asymptotic properties of two-stage estimators such as $\hat{\Lambda}_n(\bar{y})$, which depend on first-stage set estimators such as $\hat{\Theta}_n$, is a promising area for future work which we intend to pursue.

2.8 Monte Carlo Experiments

In this section we summarize the results of a series of Monte Carlo experiments intended to shed light on the finite sample properties of the proposed estimators defined in Section 2.5 and the inference procedures defined in Section 2.6.[5] First, we consider the estimator for Model 2.1 by replicating the following model:

$$y_{it} = 1\{x_{i1t} + \beta x_{i2t} + c_i + u_{it} \geq 0\}$$

where x_{i1t} and x_{i2t} are uniformly distributed for each t with $x_{i1t} \in \{-2, -1, 0, 1, 2\}$ and $x_{i2t} \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. The individual effect is generated as $c_i = (x_{i11} + x_{i12} + x_{i21} + x_{i22})/4$ and the disturbances are iid standard Normal draws. The population parameter used in the experiments is $\theta_0 = \beta_0 = -0.15$ which yields the identified set $\Theta_I = [-0.163, -0.148]$.

[5] Fortran 95 source code to reproduce all figures and tables in this section is available from the author's website at `http://jblevins.org/research/panel`.

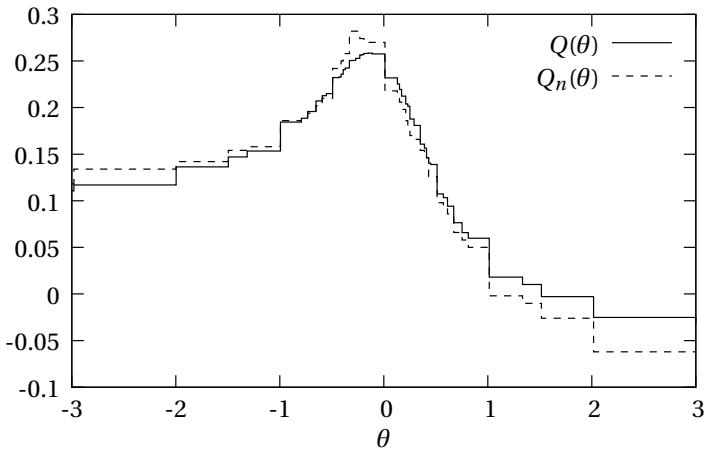

FIGURE 2.7: $Q(\theta)$ and one realization of $Q_n(\theta)$ for $n = 500$

Figure 2.7 displays one realization of $Q_n(\theta)$ for this model, with $n = 500$, along with the population objective function $Q(\theta)$. We compare the estimates for several sample sizes in Table 2.1, which lists the mean estimated set over 1000 replications for each sample size with $\kappa_n = C\sqrt{\ln n}$ (recall that $\tau_n = \kappa_n/\sqrt{n}$). We choose $C \in \{0.20, 0.10, 0.05, 0.01\}$. These values were chosen to be roughly around the same magnitude as $\mathscr{Q}_n$. For each sample size, the standard deviation of the endpoints of the estimated sets and the coverage frequency are also reported. By definition of consistency, the coverage probability should asymptotically approach one. Note that only observations for which $y_0 \neq y_1$ are used in estimation. The effective sample size for this specification is about $0.307n$.

As seen in Table 2.1, smaller constants C used to construct κ_n produce smaller estimated sets, but only at the expense of lower empirical coverage for small sample values of n. One interesting point to note about the estimates in the first panel of Table 2.1, with $C = 0.20$, is that the upper bound of the estimated interval plateaus at -0.003 for the small sample sizes shown. This corresponds to the large jump in the objective function at $\beta = -0.003$ that can be seen in Figure 2.7 and is even larger in the sample analog objective function. Since the sequence $\kappa_n = 0.20\sqrt{\ln n}$ is large relative to the other panels, the cutoff value does not rise above this jump as quickly.

Table 2.1: Estimates

κ_n	n	Mean $\hat{\Theta}_n$	St. Dev.	Coverage
$0.20\sqrt{\ln n}$	250	[-0.450, 0.077]	[0.149, 0.105]	0.96
	500	[-0.401, 0.044]	[0.096, 0.076]	0.98
	1000	[-0.365, 0.019]	[0.069, 0.051]	0.99
	2000	[-0.328, -0.001]	[0.045, 0.016]	0.99
	4000	[-0.305, -0.003]	[0.037, 0.003]	1.00
	8000	[-0.277, -0.003]	[0.032, 0.005]	1.00
	16000	[-0.256, -0.003]	[0.019, 0.000]	1.00
	32000	[-0.246, -0.003]	[0.010, 0.000]	1.00
	64000	[-0.239, -0.003]	[0.015, 0.000]	1.00
$0.10\sqrt{\ln n}$	250	[-0.318, -0.003]	[0.124, 0.103]	0.75
	500	[-0.301, -0.007]	[0.085, 0.070]	0.85
	1000	[-0.293, -0.005]	[0.064, 0.044]	0.94
	2000	[-0.268, -0.011]	[0.047, 0.034]	0.96
	4000	[-0.252, -0.012]	[0.036, 0.033]	0.99
	8000	[-0.236, -0.011]	[0.026, 0.031]	0.99
	16000	[-0.226, -0.012]	[0.024, 0.032]	0.99
	32000	[-0.216, -0.019]	[0.022, 0.040]	1.00
	64000	[-0.204, -0.026]	[0.015, 0.046]	1.00
$0.05\sqrt{\ln n}$	250	[-0.242, -0.079]	[0.114, 0.121]	0.39
	500	[-0.258, -0.041]	[0.083, 0.085]	0.66
	1000	[-0.247, -0.033]	[0.061, 0.065]	0.76
	2000	[-0.231, -0.044]	[0.048, 0.064]	0.81
	4000	[-0.215, -0.050]	[0.038, 0.065]	0.84
	8000	[-0.203, -0.055]	[0.031, 0.064]	0.88
	16000	[-0.196, -0.064]	[0.027, 0.062]	0.95
	32000	[-0.194, -0.078]	[0.021, 0.060]	0.97
	64000	[-0.188, -0.096]	[0.017, 0.051]	0.99
$0.01\sqrt{\ln n}$	250	[-0.242, -0.079]	[0.114, 0.121]	0.39
	500	[-0.210, -0.096]	[0.080, 0.097]	0.34
	1000	[-0.192, -0.109]	[0.058, 0.083]	0.29
	2000	[-0.178, -0.122]	[0.047, 0.068]	0.31
	4000	[-0.171, -0.126]	[0.039, 0.061]	0.30
	8000	[-0.173, -0.120]	[0.031, 0.056]	0.46
	16000	[-0.166, -0.129]	[0.027, 0.042]	0.50
	32000	[-0.168, -0.137]	[0.021, 0.027]	0.62
	64000	[-0.166, -0.138]	[0.018, 0.017]	0.77

The true parameter is $\theta_0 = -0.150$ and the corresponding identified set is $\Theta_I = [-0.163, -0.148]$.

Tables 2.2, 2.3, and 2.4 list, for $m = n^{2/5}$, $m = n^{3/5}$, and $m = n^{4/5}$ respectively, the empirical coverage frequencies of 1000 confidence regions for $1-\alpha \in \{0.75, 0.90, 0.95, 0.99\}$. For each of the 1000 datasets used for estimation and for each value of $1-\alpha$, a confidence region was constructed using Algorithm 2.1 of Section 2.6. These regions are based on the estimated sets from the same 1000 datasets as before. Increasing the subsample size from $n^{2/5}$ to $n^{3/5}$ seems to increase the speed of convergence of the lower quantiles. The results for the upper quantiles are largely the same for $n^{2/5}$, $n^{3/5}$, and $n^{4/5}$. Note that when the level of τ_n used for estimation is large, the finite sample confidence regions tend to have too little coverage, although it seems that larger subsample sizes are able to mitigate this to some extent.

Finally, in Tables 2.5 and 2.6, we present similar estimates and confidence regions with $\kappa_n = 0$ (for $m = n^{3/5}$ only). The estimates obtained with $\kappa_n = 0$ are tight, but have poor coverage in finite samples, as do the corresponding confidence regions.

2.9 Conclusion

We have developed new conditions for establishing both regular and irregular rates of convergence for set estimators in partially identified econometric models and proposed a method for performing inference in models whose estimators exhibit arbitrarily fast convergence. We have applied these general results to a standard binary choice panel data models with fixed effects. First we characterize the sharp identified set and we propose a consistent estimator which converges arbitrarily fast with fully discrete regressors and can achieve rates arbitrarily close to $n^{-1/3}$ when a continuous regressor is present. The validity of a subsampling-based inference procedure is established in the discrete regressor case. We also consider extensions to a lagged dependent variable and panel data duration models. Finally, a series of Monte Carlo experiments illustrates the estimation and inference procedures, which perform as expected.

Table 2.2: Confidence regions ($m = n^{2/5}$)

			Empirical Coverage			
κ_n	m	n	0.750	0.900	0.950	0.990
$0.20\sqrt{\ln n}$	$n^{2/5}$	250	0.700	0.941	0.973	0.992
		500	0.672	0.948	0.984	0.993
		1000	0.621	0.960	0.987	0.995
		2000	0.660	0.979	0.992	0.995
		4000	0.824	0.990	0.993	0.997
		8000	0.866	0.991	0.992	0.996
		16000	0.968	0.991	0.992	0.998
		32000	0.994	0.998	0.998	1.000
		64000	0.998	0.998	0.998	1.000
$0.10\sqrt{\ln n}$	$n^{2/5}$	250	0.464	0.662	0.793	0.933
		500	0.438	0.719	0.863	0.967
		1000	0.391	0.819	0.935	0.980
		2000	0.425	0.899	0.972	0.991
		4000	0.508	0.944	0.985	0.991
		8000	0.653	0.978	0.990	0.993
		16000	0.834	0.986	0.992	0.996
		32000	0.926	0.997	0.998	0.999
		64000	0.984	0.998	0.999	1.000
$0.05\sqrt{\ln n}$	$n^{2/5}$	250	0.399	0.464	0.525	0.573
		500	0.384	0.556	0.689	0.857
		1000	0.321	0.594	0.787	0.927
		2000	0.338	0.674	0.857	0.942
		4000	0.362	0.739	0.888	0.958
		8000	0.446	0.811	0.927	0.973
		16000	0.558	0.903	0.970	0.983
		32000	0.728	0.966	0.984	0.988
		64000	0.837	0.978	0.993	0.996
$0.01\sqrt{\ln n}$	$n^{2/5}$	250	0.404	0.465	0.535	0.574
		500	0.347	0.404	0.468	0.530
		1000	0.295	0.336	0.398	0.481
		2000	0.311	0.344	0.382	0.446
		4000	0.308	0.342	0.366	0.414
		8000	0.355	0.461	0.541	0.636
		16000	0.420	0.494	0.550	0.615
		32000	0.515	0.604	0.661	0.688
		64000	0.640	0.764	0.797	0.808

Table 2.3: Confidence regions ($m = n^{3/5}$)

			Empirical Coverage			
κ_n	m	n	0.750	0.900	0.950	0.990
$0.20\sqrt{\ln n}$	$n^{3/5}$	250	0.943	0.979	0.987	0.992
		500	0.934	0.978	0.994	0.993
		1000	0.910	0.976	0.984	0.996
		2000	0.936	0.989	0.991	0.997
		4000	0.978	0.989	0.990	0.995
		8000	0.985	0.991	0.994	0.995
		16000	0.986	0.994	0.997	0.997
		32000	0.997	1.000	1.000	1.000
		64000	1.000	1.000	1.000	1.000
$0.10\sqrt{\ln n}$	$n^{3/5}$	250	0.709	0.884	0.949	0.960
		500	0.776	0.923	0.952	0.970
		1000	0.838	0.931	0.965	0.983
		2000	0.903	0.961	0.978	0.990
		4000	0.922	0.972	0.982	0.987
		8000	0.945	0.979	0.993	0.992
		16000	0.965	0.989	0.995	0.994
		32000	0.984	0.996	0.996	1.000
		64000	0.994	0.999	1.000	1.000
$0.05\sqrt{\ln n}$	$n^{3/5}$	250	0.484	0.574	0.574	0.573
		500	0.608	0.775	0.877	0.918
		1000	0.721	0.853	0.897	0.936
		2000	0.808	0.912	0.931	0.946
		4000	0.859	0.926	0.939	0.957
		8000	0.900	0.918	0.926	0.945
		16000	0.898	0.954	0.965	0.974
		32000	0.939	0.975	0.980	0.982
		64000	0.969	0.983	0.992	0.995
$0.01\sqrt{\ln n}$	$n^{3/5}$	250	0.480	0.572	0.574	0.574
		500	0.434	0.519	0.535	0.538
		1000	0.378	0.468	0.497	0.498
		2000	0.387	0.421	0.453	0.465
		4000	0.367	0.412	0.432	0.447
		8000	0.554	0.629	0.649	0.667
		16000	0.583	0.609	0.612	0.617
		32000	0.665	0.671	0.676	0.688
		64000	0.798	0.800	0.804	0.807

Table 2.4: Confidence regions ($m = n^{4/5}$)

			Empirical Coverage			
κ_n	m	n	0.750	0.900	0.950	0.990
$0.20\sqrt{\ln n}$	$n^{4/5}$	250	0.952	0.977	0.989	0.990
		500	0.961	0.987	0.992	0.997
		1000	0.966	0.989	0.996	0.997
		2000	0.979	0.994	0.995	0.999
		4000	0.989	0.996	0.999	1.000
		8000	0.987	0.997	0.999	1.000
		16000	0.991	0.996	0.998	1.000
		32000	0.997	1.000	1.000	0.999
		64000	1.000	1.000	1.000	1.000
$0.10\sqrt{\ln n}$	$n^{4/5}$	250	0.807	0.876	0.893	0.904
		500	0.829	0.918	0.946	0.951
		1000	0.888	0.958	0.970	0.982
		2000	0.949	0.981	0.985	0.988
		4000	0.960	0.991	0.995	0.990
		8000	0.968	0.992	0.995	0.996
		16000	0.977	0.994	0.998	0.995
		32000	0.993	0.997	0.997	1.000
		64000	0.998	1.000	1.000	1.000
$0.05\sqrt{\ln n}$	$n^{4/5}$	250	0.544	0.551	0.556	0.556
		500	0.743	0.822	0.838	0.856
		1000	0.775	0.861	0.892	0.920
		2000	0.868	0.917	0.931	0.941
		4000	0.890	0.905	0.916	0.934
		8000	0.889	0.918	0.934	0.955
		16000	0.939	0.961	0.965	0.968
		32000	0.965	0.980	0.979	0.982
		64000	0.974	0.991	0.995	0.995
$0.01\sqrt{\ln n}$	$n^{4/5}$	250	0.545	0.553	0.555	0.555
		500	0.480	0.499	0.504	0.505
		1000	0.437	0.458	0.464	0.464
		2000	0.429	0.436	0.444	0.446
		4000	0.396	0.410	0.413	0.423
		8000	0.575	0.606	0.616	0.634
		16000	0.589	0.601	0.606	0.616
		32000	0.666	0.679	0.679	0.685
		64000	0.798	0.801	0.806	0.806

Table 2.5: Estimates without slackness

n	Mean $\hat{\Theta}_n$	St. Dev.	Coverage
125	[-0.281, -0.019]	[0.181, 0.156]	0.53
250	[-0.242, -0.079]	[0.114, 0.121]	0.39
500	[-0.210, -0.096]	[0.080, 0.097]	0.34
1000	[-0.192, -0.109]	[0.058, 0.083]	0.29
2000	[-0.178, -0.122]	[0.047, 0.068]	0.31
4000	[-0.171, -0.126]	[0.039, 0.061]	0.30
8000	[-0.166, -0.132]	[0.031, 0.047]	0.35
16000	[-0.162, -0.135]	[0.027, 0.037]	0.41
32000	[-0.163, -0.143]	[0.022, 0.020]	0.50
64000	[-0.161, -0.143]	[0.017, 0.014]	0.61

These estimates were obtained with $\kappa_n = 0$ and $m = n^{3/5}$.

Table 2.6: Confidence regions without slackness

	Empirical Coverage			
n	0.750	0.900	0.950	0.990
125	0.603	0.651	0.652	0.651
250	0.480	0.571	0.574	0.573
500	0.429	0.517	0.534	0.536
1000	0.377	0.461	0.495	0.501
2000	0.381	0.424	0.458	0.465
4000	0.372	0.416	0.430	0.447
8000	0.399	0.421	0.433	0.440
16000	0.442	0.457	0.459	0.461
32000	0.514	0.517	0.518	0.521
64000	0.622	0.622	0.622	0.622

These estimates were obtained with $\kappa_n = 0$ and $m = n^{3/5}$.

3

Nonparametric Identification of Dynamic Games with Discrete and Continuous Choices

3.1 Introduction

In this chapter we present new nonparametric identification results for both single-agent models and games in which players make both discrete and continuous choices. Such models are routinely used in industrial organization where firms in dynamic oligopoly models typically make discrete entry and exit decisions and continuous investment, pricing, or quantity choices. For example, in the theoretical framework of Ericson and Pakes (1995), each period incumbent firms first decide whether to continue in the industry or exit, and conditional upon continuing, they make a continuous investment decision.

Previous work regarding identification of dynamic structural models has focused primarily on discrete choice models. Identification of single-agent dynamic discrete choice models has been studied by Rust (1994b) and Magnac and Thesmar (2002). Several authors have also considered nonparametric identification of dynamic discrete games. Pesendorfer and Schmidt-Dengler (2007) provide a rank condition for identification in models with discrete state spaces while Bajari, Chernozhukov, Hong, and Nekipelov

(2007) show that models with continuous state spaces are identified under an exclusion restriction. Other identification results have been established in related models. Jofre-Bonet and Pesendorfer (2003) consider nonparametric identification of the cost distribution in a dynamic auction game with continuous choices. Heckman and Navarro (2005) consider semiparametric identification of dynamic discrete choice models and dynamic treatment effect models.

Our contribution relative to the existing literature is to establish conditions for the nonparametric identification of both single- and multi-agent models in which agents also make continuous choices in addition to the usual discrete choices. Given that identification of discrete-choice games has been established, it may not seem surprising at first that models with an additional continuous choice are also identified, since observing a continuous choice should provide more information than a observing a discrete one. However, in the continuous choice framework, for each state, the unknown primitives are infinite-dimensional functionals rather than finite-dimensional vectors. Thus, although more information is available, the objects of interest are much more complex.

We build on the insights of Hotz and Miller (1993), who develop a method for estimating single-agent dynamic discrete choice models which is based on a mapping from (observable) conditional choice probabilities to differences in the choice-specific value function (with respect to a normalizing choice). Bajari et al. (2007) use this idea in a preliminary step in establishing nonparametric identification of dynamic discrete games of incomplete information. They then show that the choice-specific value function can be recovered in levels by establishing that the functional operator in the recursive definition of the value function for the normalized choice is a contraction, and therefore has a unique fixed point. The utility function is then identified trivially by definition of the value function.

We follow a similar approach but we must account for an additional layer of com-

plication due to the introduction of a second, continuous choice. We define a *discrete*-choice-specific value function and show that it can be recovered, in differences, from the conditional choice probabilities as in Hotz and Miller (1993). We show that the relevant functional operators are also contractions in the models we consider, and can thus be used to identify the discrete-choice-specific value function in levels. The utility function can then be identified up to a normalization, using the first order condition implied by agents' optimal choice of the continuous variable. We will explicitly only consider models in which all components of the state vector are either serially independent or fully observable to the researcher. However, in light of recent work by Hu and Shum (2008a,b), our results can also be applied to models with serially correlated unobserved state variables.

This chapter proceeds as follows. Section 3.2 introduces a general discrete-time dynamic decision process and some fundamental assumptions. We then turn to nonparametric identification of the structural primitives in specific models. We approach the main result for dynamic games in three steps, each of which adds one level of complexity. Section 3.3 begins with a discussion of identification of single-agent discrete choice models in order to build intuition. Then, we consider single-agent models with the addition of a continuous choice in Section 3.4. Finally, we extend these results to multi-agent dynamic games in Section 3.5. Section 3.6 concludes.

3.2 Framework and Basic Assumptions

We consider a general class of discrete-time dynamic models with N players, indexed by $i = 1,\ldots,N$, over an infinite time horizon $t = 1,2,\ldots,\infty$. The state of the market at time t can be summarized by a state vector $s_t \in \mathscr{S}$ which is common knowledge to all players and evolves according to a first order Markov process. At the beginning of the period, players observe vectors of private choice-specific shocks $\varepsilon_{it} \in \mathscr{E}_i \subseteq \mathbb{R}^{K+1}$

and simultaneously make discrete choices $d_{it} \in \mathscr{D}_i = \{0, 1, \ldots, K\}$. Next, players observe private shocks $\eta_{it} \in \mathscr{H}_i \subseteq \mathbb{R}$ and simultaneously make continuous choices $c_{it} \in \mathscr{C}_i \subseteq \mathbb{R}$. Let $a_{it} = (d_{it}, c_{it})$ and $\nu_{it} = (\varepsilon_{it}, \eta_{it})$ denote, respectively, the vectors of choices and private shocks, and let a_t denote the vector consisting of the actions of all players at time t. For simplicity, we assume that all players have the same choice sets and that the support of each of the player-specific shocks is identical across players. We occasionally omit the time subscript on variables when the context is clear.

Upon making the choices a_t, each player i receives a payoff $U_i(a_t, s_t, \nu_{it})$ associated with making choice a_{it} in state s_t given that player i's rivals make choices a_{-it}, where we define $a_{-it} \equiv (a_{1t}, \ldots, a_{i-1,t}, a_{i+1,t}, \ldots, a_{Nt})$. Players are forward looking and discount future payoffs and share a common discount factor $\beta \in [0, 1)$. Players choose actions a_{it} in order to maximize their expected discounted future utility. When the market is state s_t this can be written as

$$\mathrm{E}\left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} U_i(a_\tau, s_\tau, \nu_{i\tau}) \,\middle|\, s_t\right],$$

where the expectation is taken over the infinite sequence of actions, states, and private shocks. Note that we have implicitly assumed that the utility functions and state transitions probabilities are time invariant. Combined with the Markov assumption, this implies that agents' optimal decision rules are stationary.

Before proceeding we make several standard assumptions to make the model more tractable (cf. Rust, 1994b; Aguirregabiria and Mira, 2009). Rust (1994b) imposes a conditional independence assumption which requires ν_t to be conditionally independent of ν_{t-1} given the state s_t. However, we make a stronger assumption that the private shocks are independent and identically distributed (iid), an assumption which is commonly used in practice. For example, many applications assume the discrete-choice-specific shocks are iid and have a type I extreme value distribution.

Assumption 3.1 (Private Information)**.** The private shocks $\nu_{it} = (\varepsilon_{it}, \eta_{it})$ are independent across i and t and have a known distribution G. Furthermore, the choice-specific components ε_{it} are independent with full support on $\mathbb{R}$ and finite first moments.

Additive separability assumptions are standard in discrete choice analysis, both in static and dynamic models. Given the potential addition of a continuous choice, we make the following modified additive separability condition.

Assumption 3.2 (Additive Separability)**.** The utility function is additively separable in ε_i:

$$U_i(a, s, \nu_i) = u_i(a, s, \eta_i) + \varepsilon_{id_i}.$$

Additionally, if a continuous choice is made, then u_i is further separable as

$$u_i(a, s, \eta_i) = u_{1i}(d, c, s) + u_{2i}(d, c, s)\eta_i$$

where $u_{2i}(d, c, s)$ is a known function. Otherwise, $u_i(a, s, \eta_i)$ is independent of η_i.

For the discrete choice, this assumption is equivalent to the usual additive separability assumption on the discrete-choice-specific shocks. In models where a continuous choice is made, we make the following additional assumption on the utility function.

Assumption 3.3 (Monotone Choice)**.** For all i, the utility function u_i is twice differentiable and $\partial^2 u_i / \partial c_i \partial \eta_i > 0$.

The monotone choice assumption is standard and will be used for identification in models with continuous choices. This assumption implies that agents' continuous choice policy rules are monotonic in η_i (Bajari et al., 2007). Note that we can always use $\tilde{\eta}_i \equiv -\eta_i$ in cases where the policy function is decreasing in η_i.

In the following sections, we analyze several common special cases of the general model described above. In each case, we first describe the model and then provide conditions for nonparametric identification accompanied by constructive proofs. Since

our approach extends previous results for discrete-choice models, we briefly review existing results in the context of a single-agent discrete choice model in Section 3.3. Next, we consider identification in a similar single agent model with the addition of a continuous choice variable in Section 3.4. Finally, we establish the main identification result for dynamic games with both discrete and continuous choices in Section 3.5.

3.3 Single Agent Dynamic Discrete Choice

Single agent dynamic discrete choice models are an important special case of the more general multi-agent model discussed above. These models have a long history in applied microeconomics, beginning with the pioneering work of Gotz and McCall (1980), Miller (1984), Wolpin (1984), Pakes (1986), and Rust (1987). See Eckstein and Wolpin (1989) for a survey of the early literature. Rust (1994b) also provides a survey, discusses identification, and develops a general framework for estimating such models. Hotz and Miller (1993) develop two-step methods for estimating these models which are based on first-step estimates of the conditional choice probabilities. Aguirregabiria and Mira (2002) extend this approach to develop a class of nested pseudo-likelihood estimators. Again, our focus is on nonparametric identification and in this section we review certain results of Hotz and Miller (1993) and Bajari et al. (2007) which we build upon in later sections.

Since there is only a single player ($N = 1$) we omit the i subscript from states and payoffs in this section. Furthermore, since there is only a discrete choice we have $\nu = \varepsilon$. Assumption 3.2 simplifies to the usual additive separability condition. Note that Assumption 3.3 has no meaning here since there is no continuous choice.

The value function for this model can be expressed recursively as follows:

$$V(s,\varepsilon) = \max_{d \in \mathscr{D}} \left[u(d,s) + \varepsilon_d + \beta \iint V(s',\varepsilon')\, G_\varepsilon(d\varepsilon' \mid s')\, P(ds' \mid s,d) \right].$$

Under Assumption 3.2, following Rust (1994b), we also define the choice-specific value

function

$$v(d,s) \equiv u(d,s) + \beta \iint V(s',\varepsilon')\, G_\varepsilon(d\varepsilon' \mid s')\, P(ds' \mid d,s) \tag{3.1}$$

which gives the expected discounted utility in the current period and all future periods resulting from choosing d when the current state is s, excluding the iid shock ε_d. Assumption 3.2 allows us to express this problem in a more compact form resembling a static discrete choice problem with the choice-specific value function playing the role of the period utility function. Let $\sigma(s,\varepsilon)$ denote the agent's optimal choice of d in state (s,ε). Then,

$$\sigma(s,\varepsilon) = \arg\max_{d\in\mathscr{D}} \left[v(d,s) + \varepsilon_d \right].$$

In this setting, we define the ex-ante value function

$$\bar{V}(s) \equiv \mathrm{E}\left[V(s,\varepsilon) \mid s \right] = \mathrm{E}\left[\max_{d\in\mathscr{D}} \left(v(d,s) + \varepsilon_d \right) \,\middle|\, s \right]. \tag{3.2}$$

This is social surplus function of McFadden (1981). Using this notation we can rewrite the choice-specific value function from (3.1) as

$$v(d,s) = u(d,s) + \beta \int \bar{V}(s')\, P(ds' \mid s,d). \tag{3.3}$$

Assuming β is known, the structural primitive of interest here is the utility function $u(d,s)$.

Example 3.1. A classic dynamic discrete choice model is the bus engine replacement model of Rust (1987). The state variable s_t is the accumulated mileage of the bus. Each period, a manager must decide whether or not to overhaul the engine. The manager pays a cost $u(1,s_t)$ upon replacement, and pays a cost $u(0,s_t)$ when operating the bus when the mileage is s_t. Associated with each choice d is a random component ε_{dt} which represents unobserved choice-specific costs or benefits incurred in period t. The manager chooses d in order to minimize his expected discounted costs.

The identification argument proceeds as follows. Hotz and Miller (1993) show that there is a one-to-one mapping between the conditional choice probabilities $\Pr(d \mid s)$, which are observable, and differences in the choice-specific value function, $\Delta(d,s) \equiv v(d,s) - v(0,s)$. Without loss of generality, we work in differences with respect to the choice $d = 0$. Bajari et al. (2007) show that this mapping can be used to recover $v(0,s)$ itself, through the use of a contraction mapping in the choice-specific value function. Then, the entire choice-specific value function can be recovered in levels which in turn allows one to recover the utility function, the primary structural primitive of interest, up to a standard normalization. We state the result, due to Bajari et al. (2007), before proceeding with the argument, which is instructive for the following sections.

Theorem 3.1. *Suppose Assumptions 3.1 and 3.2 are satisfied in the single-agent dynamic discrete choice model. Then the payoff function $u(d,s)$ is nonparametrically identified up to the normalization $u(0,s) = 0$ for all s.*

Proof of Theorem 3.1. Under Assumption 3.1, Hotz and Miller (1993) show that there is a one-to-one mapping Ψ from the conditional choice probabilities to differences in the choice-specific value function (3.1):

$$(v(1,s) - v(0,s), \ldots, v(K,s) - v(0,s)) = \Psi\left(\Pr(d = 1 \mid s), \ldots, \Pr(d = K \mid s)\right)$$

(see also Rust, 1994b, Lemma 3.1). This mapping depends on the distribution G and, given the choice probabilities, it is sufficient to identify the differences $\Delta(k,s) \equiv v(k,s) - v(0,s)$ for each $k = 1, \ldots, K$ and $s \in \mathscr{S}$.

For any function $f : \mathscr{D} \times \mathscr{S} \to \mathbb{R}$, define

$$\tilde{H}\left(f(0,s), f(1,s), \ldots, f(K,s)\right) \equiv \int_{\mathscr{E}} \max_{d \in \mathscr{D}} [f(d,s) + \varepsilon_d] \, G_\varepsilon(d\varepsilon).$$

When f is the utility function in a static discrete choice model, $\tilde{H}$ is McFadden's social surplus function. $\tilde{H}$ has the following additivity property (Rust, 1994b, Theorem 3.1):

$$\tilde{H}\left(f(0,s) + \alpha, f(1,s) + \alpha, \ldots, f(K,s) + \alpha\right) = \tilde{H}\left(f(0,s), f(1,s), \ldots, f(K,s)\right) + \alpha.$$

In particular, when we take $f = v$, we define

$$H(\Delta(1,s),\ldots,\Delta(K,s)) \equiv \tilde{H}(0,\Delta(1,s),\ldots,\Delta(K,s)) \tag{3.4}$$

and note that

$$\begin{aligned} H(\Delta(1,s),\ldots,\Delta(K,s)) + v(0,s) &= \tilde{H}(0,\Delta(1,s),\ldots,\Delta(K,s)) + v(0,s) \\ &= \tilde{H}(0, v(1,s) - v(0,s),\ldots, v(K,s) - v(0,s)) + v(0,s) \\ &= \tilde{H}(v(0,s), v(1,s),\ldots, v(K,s)). \end{aligned}$$

We can now write the functional mapping for the choice-specific value function in (3.3) as a function of $\Delta(d,s)$, $u(d,s)$, and $v(0,s)$:

$$v(d,s) = u(d,s) + \beta \int \left[H\left(\Delta(1,s'),\ldots,\Delta(K,s')\right) + v(0,s') \right] p(s' \mid s,d)\, ds$$

where H is defined above. Again, this function is specific to the distribution G. Note that $p(s' \mid s,d)$ is identified since it is observable. For $d = 0$, under the normalization $u(0,s) = 0$, the only remaining unknown is the functional $v(0,s)$. Thus, $v(0,s)$ is identified as the unique[1] fixed point to this functional equation. Given $v(0,s)$, the remainder of the choice-specific value functions are identified in levels since $v(d,s) = \Delta(d,s) + v(0,s)$.

It remains to identify the utility function $u(d,s)$ for $d = 1,\ldots,K$. We can express $u(d,s)$ in terms of the choice-specific value function and other identified quantities as

$$u(d,s) = v(d,s) - \beta \int \left[H\left(\Delta(1,s'),\ldots,\Delta(K,s')\right) + v(0,s') \right] P(ds' \mid s,d)$$

for any d and s. Thus, $u(d,s)$ is identified up to the normalization $u(0,s) = 0$ for all s. ■

Note that this identification result is not in conflict with the non-identification result of Rust (1994b) since the utility normalization rules out alternative specifications of the form $\tilde{u}(d,s) = u(d,s) + f(s) - \beta \mathrm{E}[f(s') \mid d,s]$.

[1] See Rust (1994b) for a proof that this functional operator is a contraction.

3.4 Single-Agent Dynamic Discrete-Continuous Choice

In this section we consider the single-agent dynamic discrete choice model from the previous section with the addition of a continuous choice. The choice sets are $\mathscr{D}$ for the discrete choice and $\mathscr{C}$ for the continuous choice. The associated random shocks in each period are ε_t and η_t. We assume that the discrete choice is made at the beginning of the period, prior to making the continuous choice and prior to learning the value of η_t.

The value function from the perspective of the beginning of the period is thus

$$V(s,\varepsilon) = \max_d \mathrm{E}_\eta \left\{ \sup_c \left[u(d,c,s,\eta) + \varepsilon_d + \beta \,\mathrm{E}\left[V(s',\varepsilon') \mid d,c,s \right] \right] \middle| \, d,s,\varepsilon \right\},$$

and the corresponding *discrete*-choice-specific value function can be written

$$v(d,s) = \mathrm{E}_\eta \left\{ \sup_c \left[u(d,c,s,\eta) + \beta \,\mathrm{E}\left[\bar{V}(s') \mid d,c,s \right] \right] \middle| \, d,s \right\}, \tag{3.5}$$

where we have made use of the ex-ante value function as defined in (3.2) and the fact that ε_d is independent of η.

Example 3.2. Timmins (2002) considers the problem of a municipal water utility administrator who chooses the price of water each period. The price may either be zero, or it may be some positive value. Consider a simplified version of the model in which $d_t \in D = \{0,1\}$ represents the decision of whether or not to set the price at zero, and, conditional on not choosing zero, $c_t \in C = \mathbb{R}^+$ represents the choice of the (positive) price. Associated with each discrete choice j is a random shock ε_{ij}. The continuous-choice-specific shock η_{it} represents unobservables affecting the cost of extracting water in period t.

The following theorem establishes that, as with the discrete choice model, the utility primitives of this model are nonparametrically identified up to an obvious normalization.

Theorem 3.2. *Suppose Assumptions 3.1–3.3 are satisfied. Then the payoff function u_1 is nonparametrically identified up to the normalization $u(0,c,s) = 0$ for all c and s.*

Proof of Theorem 3.2. The steps in the proof correspond conceptually to those of the proof of Theorem 3.1. We first use a conditional choice probability inversion to identify differences in the discrete-choice-specific value function. This is a different function from the previous model in that it represents the value before the continuous-choice-specific shock is known. Then, we show that the discrete-choice-specific value function is identified in levels by showing that it is the unique fixed point to a similar functional equation, appropriately modified to account for the continuous choice. Finally, we show that the utility function is identified up to a normalization using the first order condition and the fact that the continuous choice is monotonic in the corresponding shock.

Note that as before, we can now write the discrete choice probabilities in terms of the discrete-choice-specific value function as

$$\Pr(d = k \mid s) = \int 1\{k = \arg\max_{d} (v(d, s) + \varepsilon_d)\} G_\varepsilon(d\varepsilon).$$

Given Assumption 3.1, we can use the mapping Ψ from choice probabilities to $\Delta(d, s)$ to identify differences in the discrete-choice-specific value function for all d and s.

We show that $v(0, s)$ can be identified as before, through the use of a similar contraction mapping. First, note that by definition of the ex-ante value function $\bar{V}(s)$ and the function H, defined as in (3.4), we have

$$\begin{aligned}
\mathrm{E}_{s'|d,c,s} \bar{V}(s') &= \mathrm{E}_{s',\varepsilon'|d,c,s} \left[\max_{d' \in \mathscr{D}} \left(v(d', s') + \varepsilon'_{d'} \right) \right] \\
&= \mathrm{E}_{s'|d,c,s} \left[H\left(\Delta(1, s'), \ldots, \Delta(K, s') \right) + v(0, s') \right].
\end{aligned}$$

Using this identity in the expression above for $v(d, s)$ and evaluating it at $d = 0$ yields a functional equation for $v(0, \cdot)$:

$$\begin{aligned}
v(0, s) &= \mathrm{E}_\eta \left\{ \sup_c \left[u(0, c, s, \eta) + \beta \, \mathrm{E}_{s'|d=0,c,s} \bar{V}(s') \right] \right\}. \\
&= \mathrm{E}_\eta \left\{ \sup_c \left[u(0, c, s, \eta) + \beta \, \mathrm{E}_{s'|d=0,c,s} \left[H\left(\Delta(1, s'), \ldots, \Delta(K, s') \right) + v(0, s') \right] \right] \right\}.
\end{aligned}$$

Under the normalization $u(0,c,s) = 0$ for all s, everything in this expression is known except for $v(0,s)$. Lemma D.1 (see Appendix D) establishes that this mapping is a contraction and therefore $v(0,s)$ is identified since it is the unique fixed point. We can then use the identity $v(d,s) = \Delta(d,s) + v(0,s)$ to recover $v(d,s)$ for all $d = 1,\ldots,K$.

It remains to identify $u(d,c,s,\eta)$. In the case of single agent models with only discrete choices, we were able to obtain u directly from the choice-specific value function. With the addition of the continuous choice, however, we cannot simply identify u from $v(d,s)$ as before, due to the presence of the additional private shock η and the $\sup_c$ operator in (3.5). The monotone choice assumption allows us to overcome the first problem. It guarantees a one-to-one relationship between c and η so that given values of d, s, and c we can infer the value of η. We address the second problem by working with the first-order condition.

First, by the monotone choice assumption (Assumption 3.3) the policy function $c = \sigma_c(d,s,\eta)$ provides a one-to-one relationship between the private shock η and the continuous choice c (conditional on d and s). Let $\eta(d,s,c) \equiv \sigma_c^{-1}(d,s,c)$ denote the inverse mapping. The distribution $F_{c|d,s}$ is identified since c, d, and s are all observable and the distribution G of η is known. So, we have

$$\begin{aligned} F_{c|ds}(c \mid d,s) &= \Pr\left(\sigma_c(d,s,\eta) \leq c \mid d,s\right) \\ &= \Pr\left(\eta \leq \sigma_c^{-1}(d,s,c) \,\middle|\, d,s\right) \\ &= G_\eta\left(\sigma_c^{-1}(d,s,c)\right). \end{aligned}$$

Thus, if we observe the continuous choice c made when the discrete choice is d and the state is s, the value of η must have been

$$\eta = \sigma_c^{-1}(d,s,c) = G_\eta^{-1} \circ F_{c|ds}(c \mid d,s).$$

From now on we focus on identifying $u_1(d,c,s)$. The optimal choice of c, given by the

policy rule $\sigma_c(d,s,\eta)$, satisfies

$$\sigma_c(d,s,\eta) = \arg\sup_c \left[u(d,c,s,\eta) + \phi(d,c,s) \right],$$

where

$$\phi(d,c,s) \equiv \beta \, \mathrm{E}\left[H(\Delta(1,s'),\ldots,\Delta(K,s')) + v(0,s') \mid d,c,s \right]$$

is an identified function. Therefore, c satisfies the corresponding first-order condition

$$\frac{\partial}{\partial c} u(d,c,s,\eta) + \frac{\partial}{\partial c}\phi(d,c,s) = 0.$$

Applying Assumption 3.2 (additive separability) we have the equivalent condition

$$\frac{\partial}{\partial c} u_1(d,c,s) + \frac{\partial}{\partial c} u_2(d,c,s)\eta + \frac{\partial}{\partial c}\phi(d,c,s) = 0.$$

Rearranging and using the fact that $\eta = \sigma^{-1}(d,s,c)$, we have

$$\frac{\partial}{\partial c} u_1(d,c,s) = -\frac{\partial}{\partial c} u_2(d,c,s)\, \sigma_c^{-1}(d,s,c) - \frac{\partial}{\partial c}\phi(d,c,s).$$

$u_1(d,c,s)$ is now identified (up to a normalizing constant) for each d and s since

$$u_1(d,c,s) = -\int_{\underline{c}}^{c} \left[\frac{\partial}{\partial c} u_2(d,\tilde{c},s)\, \sigma_c^{-1}(d,s,\tilde{c}) + \frac{\partial}{\partial c}\phi(d,\tilde{c},s) \right] d\tilde{c} + u_1(d,\underline{c},s)$$

where $\underline{c} = \inf \mathscr{C}$. ■

3.5 Dynamic Games with Discrete and Continuous Choices

In this section we consider dynamic games with discrete and continuous choices involving $N > 1$ players. These models are becoming increasingly important in empirical work in applied microeconomics, especially in industrial organization, and many methods have been developed to estimate them. Aguirregabiria and Mira (2007) propose pseudo

maximum likelihood estimators for dynamic discrete games and Pakes et al. (2007) consider two-step method of moments based estimators. Pesendorfer and Schmidt-Dengler (2007) discuss a general class of asymptotic least squares estimators for such games. Bajari et al. (2007) develop simulation-based methods for estimating dynamic games with both discrete and continuous choices based on revealed preference conditions. See Aguirregabiria and Mira (2009) for a survey of this literature. We focus on nonparametric identification of these models.

In models with multiple players, each player's optimal decision depends on the expectations that player holds about the actions of the other players and so we require some sort of equilibrium concept. We assume that players use strategies that are consistent with a Markov perfect equilibrium (MPE). A Markov strategy for player i in this model is a vector-valued mapping $\sigma_i = (\sigma_{di}, \sigma_{ci})$ where $\sigma_{di} : \mathscr{S} \times \mathscr{E}_i \to \mathscr{D}_i$ denotes the discrete choice of player i in each state and $\sigma_{ci} : \mathscr{S} \times \mathscr{D} \times \mathscr{H}_i \to \mathscr{C}_i$ denotes the continuous choice of player i in each state, conditional on the discrete choices d. When behaving optimally, the present discounted value of player i's payoffs in state (s, ε_i) at the beginning of the period is

$$V_i(s, \varepsilon_i \mid \sigma_{-i}) = \max_{d_i \in \mathscr{D}_i} \mathrm{E}\left[W_i(d_i, \sigma_{d,-i}, s, \eta_i \mid \sigma_{c,-i}) + \varepsilon_{id_i} \mid d_i, s, \varepsilon_i \right], \tag{3.6}$$

where

$$W_i(d_i, d_{-i}, s, \eta_i \mid \sigma_{c,-i}) = \sup_{c_i \in \mathscr{C}_i} \mathrm{E}\left\{ u_i(d, c_i, \sigma_{c,-i}, s, \eta_i) + \beta \mathrm{E}\left[V_i(s', \varepsilon_i' \mid \sigma_{-i}) \mid d, c, s \right] \middle| d, c_i, s, \eta_i \right\} \tag{3.7}$$

The value function V_i gives the present discounted value of player i's payoffs when facing the discrete decision in state s after the discrete-choice-specific shocks ε_i are known. The expectation is with respect to ε_{-i}, the discrete-choice-specific shocks of player i's rivals, and η_i, player i's own continuous-choice-specific shock. W_i is the value function when

facing the continuous choice in state s after η_i is revealed, given that discrete choices d_i and d_{-i} were made. These functions provide a succinct recursive representation of the dynamic decision of players in this model. The expectation is with respect to η_{-i}, the continuous-choice-specific shocks of player i's rivals.

A *Markov perfect equilibrium* in this model is a collection of policy rules $(\sigma_1, \ldots, \sigma_N)$ such that for all $i = 1, \ldots, N$ and $s \in \mathscr{S}$, σ_{di} assigns the optimal discrete choice in (3.6) given beliefs σ_{-i} and σ_{ci} assigns the optimal continuous choice in (3.7) given beliefs σ_{-i}.

The primitives of the model are the discount factor β, the distribution of private shocks G, the utility functions $U_1, \ldots, U_N$, and the state transition kernel $P(ds' \mid s, a)$. The state transition kernel is observable, we assume β and G are known, and we focus on identifying the utility functions.

The following example outlines a model of the type we consider.

Example 3.3. Consider a dynamic game in which each period firms first choose whether or not to remain in the market ($d_{it} \in \mathscr{D}_i = \{0, 1\}$) and then choose quantities for competing in a product market ($c_{it} \in \mathscr{C}_i = \mathbb{R}+$). Suppose that there is learning by doing in the sense that firms' marginal costs are decreasing in their past cumulative production. The choice of quantity now has dynamic implications since more production today results in lower marginal costs in the future. The state vector is $s_{it} = (x_t, C_{it})$ where x_t is a market-wide state variable and C_{it} is the past cumulative production of firm i which evolves as $C_{i,t+1} = C_{it} + c_{it}$. Suppose for simplicity that C_{i1}, firm i's past cumulative production in the initial period, is known. Let $p(d_t, c_t, s_t)$ denote the inverse demand function. Marginal costs are a function of cumulative production so the utility (profit) function (conditional on continuing) is

$$U_i(d_{it} = 1, d_{-i,t}, c_t, s_t, \eta_{it}, \varepsilon_{it}) = c_{it}\left[p(d_t, c_t, s_t) - \left(\mu - \theta C_{it} + \eta_i\right)\right] + \varepsilon_{it1}$$

where μ is the baseline marginal cost, θ is the amount by which an additional unit of past cumulative production decreases the marginal cost, and η_i is a private marginal cost

shock. ε_{it1} is the private choice-specific shock for continuing. Upon exit ($d_{it} = 0$), firms receive a scrap payment ϕ and a random shock ε_{it0}. In this example,

$$u_{i1}(d, c, s) = c_{it}\left(p(d_t, c_t, s_t) - \mu + \theta c_{it}\right),$$

$$u_{i2}(d, c, s) = c_{it},$$

and so the assumption that u_{i2} is known simply amounts to assuming that the continuous-choice specific shock is a linear marginal cost shock. Furthermore, the monotone choice assumption holds since $\partial^2 u_i / \partial c_i \partial \eta_i = 1 > 0$.

With multiple players, we require an exclusion restriction to identify the utility function. Namely, we assume that the state vector can be written as $s = (s_1, \ldots, s_N)$ where s_i are the relevant state variables for player i.

Assumption 3.4 (Exclusion Restriction)**.** The utility function for player i satisfies

$$u_i(a, s, \eta_i) = u_i(a, s_i, \eta_i)$$

where s_i has at least one continuous component for each i.

Thus, some components of s_i may be common to all players but there must be at least one continuous component that is specific to player i. A similar exclusion restriction was used for identification purposes in dynamic discrete choice games by Bajari et al. (2007). Our condition is slightly different in that we require that there be at least one *continuous* component of s_i for each player. Note that the utility function in Example 3.3 satisfies this assumption.

Theorem 3.3. *Suppose Assumptions 3.1–3.4 are satisfied. Furthermore, suppose that both u_i and $\partial u_i / \partial c_i$ are continuous for all i and that the conditional expectation operator $\mathrm{E}_{c_{-i}|d,s}$ is one-to-one. Then the payoff function u_i is nonparametrically identified up to the normalization $u_i(0, d_{-i}, c_i, c_{-i}, s_i) = 0$.*

Proof of Theorem 3.3. For simplicity, we drop the policy rule notation (σ_{-i}) and simply treat d_{-i} and c_{-i} as random variables distributed according to the appropriate equilibrium beliefs. We can define the discrete-choice-specific value function as

$$v_i(d_i, s) \equiv \mathrm{E}_{\eta_i|d_i,s} \mathrm{E}_{d_{-i}|d_i,s} \sup_{c_i} \left\{ \mathrm{E}_{c_{-i}|d,c_i,s} \left[u_i(d_i, d_{-i}, c_i, c_{-i}, s, \eta_i) + \beta \mathrm{E}_{s'|d,c,s} \bar{V}_i(s') \right] \right\}$$

where $\bar{V}_i(s) = \mathrm{E}\left[V_i(s, \varepsilon_i) \mid s\right]$ is the ex-ante value function for player i. If we define H as in (3.4), then

$$\bar{V}_i(s) = \mathrm{E}\left[\max_{d_i} \left\{ v_i(d_i, s) + \varepsilon_{id_i} \right\} \middle| s \right] = H(\Delta_i(1, s), \ldots, \Delta_i(K, s)) + v_i(0, s).$$

Then, for $d_i = 0$, after normalizing $u_i(0, d_{-i}, c_i, c_{-i}, s, \eta_i) = 0$, we have

$$\begin{aligned} v_i(0, s) &= \mathrm{E}_{\eta_i|d_i,s} \mathrm{E}_{d_{-i}|d_i,s} \sup_{c_i} \left\{ \beta \mathrm{E}_{c_{-i}|d,c_i,s} \left[\mathrm{E}_{s'|d,c,s} \bar{V}_i(s') \right] \right\} \\ &= \mathrm{E}_{\eta_i|d_i,s} \mathrm{E}_{d_{-i}|d_i,s} \sup_{c_i} \left\{ \beta \mathrm{E}_{c_{-i}|d,c_i,s} \mathrm{E}_{s'|d,c,s} \left[H(\Delta_i(1, s'), \ldots, \Delta_i(K, s')) + v_i(0, s') \right] \right\}. \end{aligned}$$

Note that all quantities except the function $v_i(0, s)$ are known, including β, $\Delta_i(d, s)$, the distributions of η_i, $d_{-i} \mid d_i, s$, and $c_{-i} \mid d, c_i, s$, and the transition density of $s' \mid d, c, s$. As established by Lemma D.2 (see Appendix D), this defines a contraction mapping which identifies $v_i(0, s)$ and in turn, knowing $v_i(0, s)$ allows us to identify $v_i(d_i, s) = \Delta_i(d_i, s) + v_i(0, s)$ for all $d_i > 0$.

We now turn to identifying the utility function u. For any d, s, and η_i, the optimal choice of c_i maximizes

$$\mathrm{E}_{c_{-i}|d,c_i,s} \left[u_{i1}(d_i, d_{-i}, c_i, c_{-i}, s) + u_{i2}(d_i, d_{-i}, c_i, c_{-i}, s) \eta_i \right] + \phi(d, c_i, s)$$

where $\phi(d, c_i, s)$ is the known function

$$\phi(d, c_i, s) \equiv \beta \mathrm{E}_{c_{-i}|d,c_i,s} \mathrm{E}_{s'|d,c,s} \left[H(\Delta_i(1, s'), \ldots, \Delta_i(K, s')) + v_i(0, s') \right].$$

Therefore, c_i must satisfy the first order condition

$$\frac{\partial}{\partial c_i}\int \left[u_{i1}(d_i, d_{-i}, c_i, c_{-i}, s) + u_{i2}(d_i, d_{-i}, c_i, c_{-i}, s)\eta_i\right] p(c_{-i} \mid d, s)\, dc_{-i} + \frac{\partial}{\partial c_i}\phi(d, c_i, s) = 0$$

Here, we have used the fact that $p(c_{-i} \mid d, c_i, s)$ does not depend on c_i conditional on s and d since the relevant shocks are iid. Under the maintained assumptions, we can interchange the order of integration and differentiation:

$$\int \left[\frac{\partial}{\partial c_i}u_{i1}(d_i, d_{-i}, c_i, c_{-i}, s) + \frac{\partial}{\partial c_i}u_{i2}(d_i, d_{-i}, c_i, c_{-i}, s)\eta_i\right] p(c_{-i} \mid d, s)\, dc_{-i} + \frac{\partial}{\partial c_i}\phi(d, c_i, s) = 0$$

The partial derivatives of ϕ and u_{i2} are identified since both are known functions. Furthermore, due to the monotone choice assumption we can replace η_i with $\sigma_{ci}^{-1}(d, s, c_i)$, also a known quantity.

$\mathrm{E}_{c_{-i}|d,s}$ is one-to-one by assumption so we can apply the inverse operator to the right-hand side above to recover u_i, however we need to use the exclusion restriction $u_i(d_i, d_{-i}, c_i, c_{-i}, s) = u_i(d_i, d_{-i}, c_i, c_{-i}, s_i)$ if we hope to recover a *function* of d_{-i} and c_{-i}. The conditional expectation operator maps functions of $(d_i, d_{-i}, c_i, c_{-i}, s_i)$ (such as the utility function) to functions of (d_i, d_{-i}, c_i, s). The inverse operator reverses this mapping so that for fixed values of (d, c_i, s_i) it maps functions of s_{-i} to functions of c_{-i}.

Now for each (d, c_i, s_i) we can apply the inverse operator $\mathrm{E}^{-1}_{c_{-i}|d,s}$ to $-\frac{\partial}{\partial c_i}\phi(d, c_i, s)$ to recover

$$\frac{\partial}{\partial c_i}u_{i1}(d_i, d_{-i}, c_i, c_{-i}, s_i) + \frac{\partial}{\partial c_i}u_{i2}(d_i, d_{-i}, c_i, c_{-i}, s_i)\,\sigma_{ci}^{-1}(d, s_i, c_i).$$

Integrating this with respect to c_i and using the fact that we know u_{i2} and $\sigma_{ci}^{-1}(d, s_i, c_i)$ identifies u_{i1} up to a constant normalization for each c_i. ■

Intuitively, the assumption that the conditional expectation operator $E_{c_{-i}|d,s}$ is one-to-one requires there to be sufficient variation in the conditional distribution of c_{-i} for different values of d and s. Assumptions of this type have been used by Hu and Shum (2008a,b) in identifying the Markov kernel in dynamic discrete choice models and dynamic games with unobserved state variables. They are also key conditions for identification in many nonparametric econometric models such as instrumental regression models (Newey and Powell, 2003; Darolles, Florens, and Renault, 2007; Blundell, Chen, and Kristensen, 2007) and classical measurement error models (Chen and Hu, 2006; Hu and Schennach, 2008).

3.6 Conclusion

We have established conditions for nonparametric identification of both single agent models and dynamic games of incomplete information in which agents make both discrete and continuous choices. Our nonparametric identification results can serve as a point of reference for practitioners estimating parametric models. In the absence of conditions for parametric identification, our conditions, though likely stronger than necessary for highly parametrized models, can serve as a benchmark. Furthermore, our proofs are constructive and suggest the possibility of nonparametric or semiparametric estimators for such models.

Appendix A

Proofs for Chapter 1

Proof of Proposition 1.1. The result follows directly from the joint distribution function:

$$\begin{aligned}\Pr(\tau \le t) = \Pr\left(\min_i \tau_i \le t\right) &= 1 - \Pr(\tau_1 > t, \ldots, \tau_n > t) \\ &= 1 - \prod_{i=1}^{n} \Pr(\tau_i > t) = 1 - \prod_{i=1}^{n} \mathrm{e}^{-\lambda_i t} = 1 - \mathrm{e}^{-(\sum_{i=1}^{n} \lambda_i) t}.\end{aligned}$$

Therefore, τ has an exponential distribution with rate parameter $\sum_{i=1}^{n} \lambda_i$.

Furthermore,

$$\begin{aligned}\Pr(\tau_i \le \tau_j \ \forall j) &= \mathrm{E}_{\tau_i}\left[\Pr(\tau_j \ge \tau_i \ \forall j \ne i) \,\middle|\, \tau_i\right] \\ &= \int_0^{\infty} \left[\mathrm{e}^{-\sum_{j \ne i} \lambda_j}\right] \lambda_i \mathrm{e}^{-\lambda_i \tau_i} \, d\tau_i \\ &= \int_0^{\infty} \lambda_i \mathrm{e}^{-(\sum_{j=1}^{n} \lambda_j) \tau_i} \, d\tau_i \\ &= -\frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \left[\mathrm{e}^{-(\sum_{j=1}^{n} \lambda_j) \tau_i}\right]_{\tau_i = 0}^{\infty} \\ &= \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}.\end{aligned}$$

■

Proof of Proposition 1.2. We have

$$\begin{aligned}
\Pr(W_+(t) \geq \tau) &= \Pr\left[N_+(t+\tau) - N_+(t) = 0\right] \\
&= \sum_{k=0}^{\infty} \Pr\left[N(t+\tau) - N(t) = k, N_0(t+\tau) - N_0(t) = k\right] \\
&= \sum_{k=0}^{\infty} \Pr\left[N(t+\tau) - N(t) = k\right] \sigma(0,x)^k \\
&= \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-\lambda\tau}(\lambda\tau)^k}{k!} \sigma(0,x)^k \\
&= \mathrm{e}^{-\lambda\tau} \sum_{k=0}^{\infty} \frac{(\sigma(0,x)\lambda\tau)^k}{k!} \\
&= \mathrm{e}^{-\lambda\tau}\, \mathrm{e}^{\sigma(0,x)\lambda\tau} \\
&= \mathrm{e}^{-(1-\sigma(0,x))\lambda\tau},
\end{aligned}$$

and therefore the cdf of $W_+(t)$ is

$$\Pr(W_+(t) \leq \tau) = 1 - \mathrm{e}^{-(1-\sigma(0,x))\lambda\tau}.$$

For a given x, this is precisely the cdf of the exponential distribution with parameter $(1-\sigma(0,x))\lambda$. ■

Appendix B

Details of the Monte Carlo Experiments for Chapter 1

Single Agent Model

To generate data for the single agent model we first choose values for θ and then use numerical fixed point methods to determine the value function over the state space $\mathscr{X}$ to within a tolerance of $\epsilon = 10^{-6}$ in the relative sup norm. To evaluate the expectation over τ in (1.4), we use Monte Carlo integration as described in Section 1.5.4, drawing R arrival intervals according to the appropriate exponential distribution and approximating the integral using the sample average. We set R to 250. We then use the resulting value function to generate data for various values of T.

In the first set of experiments, we estimate the model using full solution methods. The value functions are obtained through value function iteration for each value of θ while maximizing the likelihood function using the L-BFGS-B algorithm (Byrd, Lu, and Nocedal, 1995; Zhu, Byrd, Lu, and Nocedal, 1997).[1] We generate 100 data sets over the interval

[1] While there are more efficient methods to evaluate the expectation over τ and taking analytic derivatives would clearly speed up estimation, the computational times are so fast that these steps were not needed.

$[0, T]$ with $T = 25,000$ for an average of $10,000$ events and then estimate the model under several sampling regimes: true continuous time data, continuous time data when passive actions ($a = 0$, the choice not to renew) are unobserved, and discrete time data observed at intervals $\Delta \in \{0.625, 1.25, 2.5, 5.0, 10.0\}$.[2]

We also carry out the same experiments using CCP-based estimation in the single agent model. Again, for $T = 25,000$, we estimate the model with full continuous-time data, a continuous-time dataset with missing passive actions, and several discrete-time datasets of varying granularity. For the full continuous-time dataset, we can nonparametrically estimate the CCPs using a simple bin estimator. When accounting for passive moves, we approximate the CCPs by dividing the number of times each particular observed choice was made in each state by the implied expected number of move arrivals in that state. Finally, when estimating the model with discrete-time data, we first jointly estimate the first-stage parameters (λ, q_1, and q_2) and the parameters of a logistic regression model for the probability of renewal with parameters α. The regressors in our logit model are a constant, the state x, and $\ln x$. Then, we invert the predicted CCPs obtained using the estimated parameters $\hat{\alpha}$ to obtain the value function which we use to estimate the remaining second stage parameters.

Quality Ladder Model

For the multi-agent quality ladder model, we obtain estimates of $\theta = (\lambda, \gamma, \kappa, \eta, \eta^e)$ for each of 25 simulated datasets and report the means and standard deviations (in parenthesis). In all experiments, we hold $\bar{\omega}$ fixed at $\bar{\omega} = 7$, set $\omega^e = \lfloor \frac{\bar{\omega}}{2} \rfloor$, and vary the maximum number of players, $\bar{N}$, and the market size, M.[3]

[2] One could view 25,000 as the number of months in the data with $\{0.625, 1.25, 2.5, 5.0, 10.0\}$ indicating the number of months (or fraction of months) between samples. While 25,000 implies having over 2,000 years of data, this is following *one* time series. An almost equivalent structure would follow 1000 decision-makers over two years.

[3] For a fixed value of $\bar{N}$, the size of the state space of our model with $\bar{\omega} = 7$ is roughly comparable to that of Doraszelski and Judd (2008) with $\bar{\omega} = 9$ since their model does not include entry and exit.

We also increase the market size M so that the average number of active players (n_{avg}) grows with the total number of possible players ($\bar{N}$). The average quality level of active firms is denoted ω_{avg}. We also report K, the number of states from the perspective of player i—the number of distinct (ω, ω_i) combinations. In these experiments, we used samples containing $T = 100$ continuous time events in each of $M = 1000$ markets. We fixed $\rho = 0.05$ and use $R = 250$ draws for Monte Carlo integration.

For the CCP estimation, we use the true CCPs. In practice, the CCPs must be estimated somehow in a preliminary step. However, because there are many possible methods for doing so, and because they tend to be application and data specific, we simply present the results for the second-stage parameters as if the true CCPs were known. We have estimated the CCPs nonparametrically using locally weighted averages with little change in the results.

Appendix C

Proofs for Chapter 2

C.1 Notation and Preliminary Results

C.1.1 Notation

First we introduce some notation. We shall make use of a modified signum function $\operatorname{sgn}(x)$ where

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0. \end{cases}$$

This definition, which is standard in the maximum score literature, differs from the common definition only at zero, where we define $\operatorname{sgn}(0) = 1$ instead of $\operatorname{sgn}(0) = 0$. We write $a \vee b$ to denote $\max\{a, b\}$ and $a \wedge b$ to denote $\min\{a, b\}$.

In a slight abuse of notation, define the distance between a point x and a set B to be

$$d(x, B) = \inf_{x' \in B} d(x, x'),$$

where d denotes the Euclidean distance. For any set B, we let B^ε denote an ε-expansion of B, defined as

$$B^\varepsilon = \{x \in B : d(x, B) \leq \varepsilon\}.$$

We write $a_n \downarrow a$ to a sequence which eventually equals a, or in other words, a sequence for which there exists an $N < \infty$ such that $a_n = a$ for all $n \geq N$. We also say that such a sequence converges *arbitrarily fast* to a since for any sequence r_n, $r_n |a_n - a| \to 0$. This includes all polynomials of n such as $r_n = n^{1/2}$. In particular, when a_n is a stochastic process, we say a_n converges arbitrarily fast to a in probability, or a_n is eventually a in probability, when $P\{\omega \in \Omega : a_n(\omega) = a\} \to 1$. In such cases we write $a_n \downarrow a$ in probability.

C.1.2 Preliminary Results

Lemma C.1. *Let f and g be bounded real functions on $A \subset \mathbb{R}^n$. Then*

$$\left| \sup_{x \in A} f(x) - \sup_{x \in A} g(x) \right| \leq \sup_{x \in A} \left| f(x) - g(x) \right|.$$

Proof of Lemma C.1. First, note that for all $x \in A$,

$$f(x) - \sup_{y \in A} g(y) \leq f(x) - g(x) \leq \left| f(x) - g(x) \right| \tag{C.1}$$

and

$$\sup_{y \in A} f(y) - g(x) \geq f(x) - g(x) \geq -\left| f(x) - g(x) \right|. \tag{C.2}$$

We prove the result by showing that

$$-\sup_{x \in A} \left| f(x) - g(x) \right| \leq \sup_{x \in A} f(x) - \sup_{x \in A} g(x) \leq \sup_{x \in A} \left| f(x) - g(x) \right|.$$

For the right hand side:

$$\sup_{x \in A} f(x) - \sup_{x \in A} g(x) = \sup_{x \in A} \left[f(x) - \sup_{y \in A} g(y) \right] \leq \sup_{x \in A} \left| f(x) - g(x) \right|.$$

The equality holds since $\sup g$ is constant with respect to x and the inequality follows

from (C.1), since it holds for all x. Similarly, the left hand side follows from (C.2):

$$\begin{aligned}\sup_{x\in A} f(x) - \sup_{x\in A} g(x) &= \sup_{x\in A} f(x) + \inf_{x\in A}\left(-g(x)\right) \\ &= \inf_{x\in A}\left[\sup_{y\in A} f(y) - g(x)\right] \\ &\geq \inf_{x\in A} -\left|f(x) - g(x)\right| \\ &= -\sup_{x\in A}\left|f(x) - g(x)\right|\end{aligned}$$

Together, these two inequalities imply the result. ■

C.2 Consistent Estimation

Proof of Theorem 2.3. The proof proceeds in two steps. In the first step, we show that $\sup_{\theta\in\hat{\Theta}_n} d(\theta,\Theta_I) \xrightarrow{p} 0$. The second step shows that $\lim_{n\to\infty} P(\Theta_I \subset \hat{\Theta}_n) = 1$. Combining these steps and using the definition of the Hausdorff distance yields the final conclusion of the theorem. Let B^ε denote an ε-expansion of a set B, as defined in Section C.1.1.

Step 1 For any $\varepsilon > 0$,

$$\sup_{\Theta\setminus\Theta_I^\varepsilon} Q_n \leq \sup_{\Theta\setminus\Theta_I^\varepsilon} Q + o_p(1) \leq \sup_{\Theta} Q - \delta_\varepsilon + o_p(1),$$

where $\delta_\varepsilon > 0$. The first inequality above follows from Assumption 2.4.d, giving uniform convergence in probability of Q_n to Q, and the second inequality follows from Assumption 2.4.c, since Θ_I maximizes Q. Similarly,

$$\inf_{\hat{\Theta}_n} Q_n \geq \sup_{\Theta} Q_n - \tau_n \geq \sup_{\Theta} Q - \tau_n + o_p(1)$$

The first inequality follows from the definition of $\hat{\Theta}_n$ and the second follows again from uniform convergence. By assumption, $\tau_n = o_p(1)$, and since $\delta_\varepsilon > 0$, with probability approaching one, $\tau_n < \delta_\varepsilon$, or equivalently, $\sup_\Theta Q - \tau_n + o_p(1) \geq \sup_\Theta Q - \delta_\varepsilon + o_p(1)$.

Given the inequalities above, this implies $\inf_{\hat{\Theta}_n} Q_n \geq \sup_{\Theta \setminus \Theta_I^\varepsilon} Q_n$, which in turn implies that $\Theta_n \subseteq \Theta_I^\varepsilon$, and so $\sup_{\theta \in \hat{\Theta}_n} d(\theta, \Theta_I) \leq \varepsilon$.

Step 2 By definition of $\hat{\Theta}_n$ and τ_n, we know that if $b_n \tau_n \geq \sup_\Theta b_n Q_n - \inf_{\Theta_I} b_n Q_n$, then $\Theta_I \subseteq \hat{\Theta}_n$. We have

$$\begin{aligned}
\sup_\Theta Q_n - \inf_{\Theta_I} Q_n &= \left[\sup_\Theta Q_n - \sup_\Theta Q\right] + \left[\sup_\Theta Q - \inf_{\Theta_I} Q_n\right] \\
&\leq \left|\sup_\Theta Q_n - \sup_\Theta Q\right| + \left|\sup_\Theta Q - \inf_{\Theta_I} Q_n\right| \\
&= \left|\sup_\Theta Q_n - \sup_\Theta Q\right| + \left|\sup_{\Theta_I} Q - \inf_{\Theta_I} Q_n\right| \\
&\leq \sup_\Theta |Q_n - Q| + \sup_{\Theta_I} |Q_n - Q| \\
&\leq \sup_\Theta |Q_n - Q| + \sup_\Theta |Q_n - Q|
\end{aligned}$$

These steps follow by, respectively, adding and subtracting $\sup_\Theta Q$, taking the absolute value, noting that Θ_I maximizes Q, using the fact that $\inf f = -\sup -f$, and applying Lemma C.1 (see Section C.1) twice, noting that $\Theta_I \subseteq \Theta$. By Assumption 2.4.d, $\sup_\Theta |Q_n - Q| = O_p(1/b_n)$ and so the requirement that $\tau_n b_n \xrightarrow{\text{p}} \infty$ (i.e., that τ_n approaches zero in probability slower than $1/b_n$) implies that $\tau_n \geq 2\sup_\Theta |Q_n - Q| \geq \sup_\Theta Q_n - \inf_{\Theta_I} Q_n$ with probability approaching one. ■

Proof of Theorem 2.4. From Theorem 2.3, $\lim_{n\to\infty} P(\Theta_I \subseteq \hat{\Theta}_n) = 1$. We will prove the result by showing that $\lim_{n\to\infty} P(\hat{\Theta}_n \subseteq \Theta_I) = 1$ and therefore the Hausdorff distance $d_H(\hat{\Theta}_n, \Theta_I)$ eventually equals zero with probability approaching one.

Uniform convergence at the $1/b_n$ rate (Assumption 2.4.d) implies $Q_n(\theta) \leq Q(\theta) + O_p(1/b_n)$ and $Q(\theta) \leq Q_n(\theta) + O_p(1/b_n)$. It follows that

$$\sup_{\Theta \setminus \Theta_I} Q_n \leq \sup_{\Theta \setminus \Theta_I} Q + O_p(1/b_n) \leq \sup_\Theta Q\delta + O_p(1/b_n) \leq \sup_\Theta Q_n - \delta + O_p(1/b_n),$$

where the second inequality follows from Assumption 2.5.

Since τ_n converges to zero in probability and $\delta > 0$ is constant, with probability approaching one, $\tau_n < \delta$. Thus, with probability approaching one, $-\delta < -\tau_n$, $\sup_{\Theta \setminus \Theta_I} Q_n \leq \sup_\Theta Q_n - \tau_n + O_p(1/b_n) \leq \inf_{\hat{\Theta}_n} Q_n + O_p(1/b_n)$, and therefore, $\hat{\Theta}_n \subseteq \Theta_I$. ■

Proof of Theorem 2.5. For any $\varepsilon > 0$, let δ, κ, γ_1, γ_2, κ_ε, and n_ε satisfy Assumption 2.6 and define

$$\nu_n \equiv \left(\frac{\kappa \cdot \kappa_\varepsilon \vee b_n \cdot \tau_n}{b_n \cdot \kappa} \right)^{\frac{1}{\gamma_2}}$$

where b_n is given by Assumption 2.4.d. Then, since $\nu_n = o_p(1)$, $\nu_n = O_p(\tau_n^{1/\gamma_2})$, and $\tau_n b_n \xrightarrow{p} \infty$, there is an $n'_\varepsilon > n_\varepsilon$ such that for all $n > n'_\varepsilon$, with probability at least $1 - \varepsilon$, we have both $\nu_n \leq \delta$ and $\nu_n \geq (\kappa_\varepsilon / b_n)^{1/\gamma_2}$. On a set $\Theta \setminus \Theta_I^{\nu_n}$, the distance satisfies $d(\theta, \Theta_I) \geq \nu_n$, so $\min\{d(\theta, \Theta_I), \delta\} \geq \min\{\nu_n, \delta\}$. Therefore, by Assumption 2.6,

$$\begin{aligned}
\sup_{\Theta \setminus \Theta_I^{\nu_n}} Q_n &\leq \sup_\Theta Q_n - \kappa \cdot (\nu_n \wedge \delta)^{\gamma_1} \\
&\leq \sup_\Theta Q_n - \kappa \cdot \nu_n^{\gamma_1} \\
&\leq \sup_\Theta Q_n - \left(\frac{\kappa \cdot \kappa_\varepsilon}{b_n} \vee \tau_n \right)^{\gamma_1/\gamma_2} \\
&\leq \sup_\Theta Q_n - \tau_n^{\gamma_1/\gamma_2} \\
&\leq \inf_{\hat{\Theta}_n} Q_n.
\end{aligned}$$

The above implies that $\hat{\Theta}_n \cap (\Theta \setminus \Theta_I^{\nu_n})$ is empty, or equivalently, that $\hat{\Theta}_n \subseteq \Theta_I^{\nu_n}$. Therefore, in light of Step 1 of the proof of Theorem 2.3, which shows that $\lim_{n \to \infty} P(\Theta_I \subseteq \hat{\Theta}_n) = 1$, we have $d_H(\hat{\Theta}_n, \Theta_I) = O_p(\tau_n^{1/\gamma_2})$ (since τ_n is slower than $1/b_n$ by assumption). ■

C.3 Confidence Regions

Proof of Lemma 2.6. Observe that

$$P\{\Theta_I \subseteq C_n(\hat{c}_n)\} = P\{\mathcal{Q}_n \leq \hat{c}_n\} = P\{\mathcal{Q} \leq c(1-\alpha)\} + o_p(1) \geq (1-\alpha) + o_p(1).$$

The first equality holds by definition of C_n and $\mathcal{Q}_n$, the second by Assumption 2.7 and $\hat{c}_n \xrightarrow{\text{p}} c(1-\alpha)$, and the third by definition of $c(1-\alpha)$. ■

Proof of Theorem 2.11. The proof proceeds in three steps. First, we derive upper and lower bounds for $\hat{\mathcal{Q}}_{n,m,j}$ such that $\underline{\mathcal{Q}}_{n,m,j} \leq \hat{\mathcal{Q}}_{n,m,j} \leq \overline{\mathcal{Q}}_{n,m,j}$ with probability approaching one. Next, we prove that the empirical distribution function of $\hat{\mathcal{Q}}_{n,m,j}$ converges in probability to the distribution function of $\mathcal{Q}$, the limiting distribution of $\mathcal{Q}_n$. Finally, we show that $\hat{c}_n$ converges in probability to $c(1-\alpha)$, the desired quantile of the distribution of $\mathcal{Q}$.

Step 1 By Theorem 2.4, we have $d_H(C_n(\kappa_n), \Theta_I) = 0$ with probability approaching one. Thus, $d_H(C_n(\kappa_n), \Theta_I) \leq \varepsilon_n$ for some sequence $\varepsilon_n \downarrow 0$ with probability approaching one. For a fixed subsample j, let $\underline{\mathcal{Q}}_{n,m,j} \equiv \sup_{\theta\in\Theta} b_m Q_{n,m,j}(\theta) - \inf_{\theta\in\Theta_I^{\varepsilon_n}} b_m Q_{n,m,j}(\theta)$. Let $\mathcal{K}_n$ be the collection of all subsets $K \subseteq \Theta$ such that $d_H(K, \Theta_I) \leq \varepsilon_n$ and define $\overline{\mathcal{Q}}_{n,m,j} \equiv \sup_{K\in\mathcal{K}_n} \left[\sup_{\theta\in\Theta} b_m Q_{n,m,j}(\theta) - \inf_{\theta\in K} b_m Q_{n,m,j}(\theta)\right]$. There exists a set $\Theta_{n,m,j} \in \mathcal{K}_n$ such that $\mathcal{Q}_{n,m,j}$ is equal to $\inf_{\theta\in\Theta_{n,m,j}} b_m Q_{n,m,j}(\theta)$. With probability approaching one, since $C_n(\kappa_n) \subseteq \Theta_I^{\varepsilon_n}$ and $C_n(\kappa_n) \in \mathcal{K}_n$, we have $\underline{\mathcal{Q}}_{n,m,j} \leq \hat{\mathcal{Q}}_{n,m,j} \leq \overline{\mathcal{Q}}_{n,m,j}$ for all $j = 1, \ldots, M_n$.

Step 2 From Step 1, with probability approaching one,

$$\underline{G}_{n,m}(x) \equiv M_n^{-1} \sum_{j=1}^{M_n} 1\{\overline{\mathcal{Q}}_{n,m,j} \leq x\} \leq \hat{G}_{n,m}(x) \equiv M_n^{-1} \sum_{j=1}^{M_n} 1\{\hat{\mathcal{Q}}_{n,m,j} \leq x\}$$

$$\leq \overline{G}_{n,m}(x) \equiv M_n^{-1} \sum_{j=1}^{M_n} 1\{\overline{\mathcal{Q}}_{n,m,j} \leq x\}.$$

We will show that $\underline{G}_{n,m}(x) \xrightarrow{p} P\{\mathcal{Q} \leq x\}$ and $\overline{G}_{n,m}(x) \xrightarrow{p} P\{\mathcal{Q} \leq x\}$ as $n \to \infty$ (and thus, $m \to \infty$). Therefore, $\hat{G}_{n,m}(x) \xrightarrow{p} P\{\mathcal{Q} \leq x\}$ for each $x \in \mathbb{R}$.

Let $\overline{J}_m(x)$ denote the cdf of $\overline{\mathcal{Q}}_{n,m,j}$. Note that $\underline{G}_{n,m}(x)$ is a U-statistic of degree m with $0 \leq \underline{G}_{n,m}(x) \leq 1$ (i.e., it is bounded). Furthermore, $\mathrm{E}[\underline{G}_{n,m}(x)] = \mathrm{E}[1\{\overline{\mathcal{Q}}_{n,m,j} \leq x\}] = \overline{J}_m(x)$, where the last equality holds by nonreplacement sampling, since each subsample of size m is itself an iid sample. By the Hoeffding inequality for bounded U-statistics for iid data (Serfling, 1980, Theorem A, p. 201), for any $t > 0$,

$$P\left\{\underline{G}_{n,m}(x) - \overline{J}_m(x) \geq t\right\} \leq \exp\left[-2t^2 \frac{n}{m}\right].$$

A similar inequality follows for $t < 0$ by considering the U-process $-\underline{G}_{n,m}(x)$. Therefore, $\underline{G}_{n,m}(x) = \overline{J}_m(x) + o_p(1)$ for fixed m. Finally, since $\overline{\mathcal{Q}}_{n,m,j}$ is obtained from sets satisfying Assumption 2.8, $\overline{J}_m(x) = P\{\overline{\mathcal{Q}}_{n,m,j} \leq x\} = P\{\overline{\mathcal{Q}} \leq x\} + o_p(1)$.

A similar argument shows that $\overline{G}_{n,m}(x) \xrightarrow{p} P\{\mathcal{Q} \leq x\}$ as well, and therefore, $\hat{G}_{n,m}(x) \xrightarrow{p} P\{\mathcal{Q} \leq x\}$.

Step 3 Convergence of the distribution function at continuity points implies convergence of the quantile function at continuity points (cf. Shorack, 2000, Proposition 3.1). Therefore, $\hat{c}_n = \inf\{x : \hat{G}(x) \geq 1 - \alpha\} \xrightarrow{p} c(1-\alpha)$. ■

C.4 Fixed Effects Model

Below we provide proofs for results pertaining to Model 2.1. First we present results which are independent of assumptions on state space $\mathcal{X}$, followed by results for discrete regressors and continuous regressors.

C.4.1 General Results

Proof of Theorem 2.1. For the proof, let Θ_I denote the identified set as defined in (2.4) and let $\tilde{\Theta}_I$ denote the set on the right side of (2.5). We first show $\Theta_I \subseteq \tilde{\Theta}_I$, and then

$\tilde{\Theta}_I \subseteq \Theta_I$.

Step 1 Let $\theta \in \Theta_I$. By definition of Θ_I, there exist distributions $F_{u_0|xc}$ and $F_{c|x}$ such that $\pi(y_t = 1 \mid x; \beta, F_{u_0|xc}, F_{c|x}) = P(y_t = 1 \mid x)$ F_x-almost surely for $t = 0, 1$. Conditioning on c, we have $P(y_0 = 1 \mid x, c) = 1 - F_{u_0|xc}(-x_0'\beta - c)$ and $P(y_1 = 1 \mid x, c) = 1 - F_{u_0|xc}(-x_1'\beta - c)$. By the monotonicity of $F_{u_0|xc}$,

$$\begin{aligned} P(y_1 = 1 \mid x, c) \geq P(y_0 = 1 \mid x, c) &\iff 1 - F_{u_0|xc}(-x_1'\beta - c) \geq 1 - F_{u_0|xc}(-x_0'\beta - c) \\ &\iff F_{u_0|xc}(-x_1'\beta - c) \leq F_{u_0|xc}(-x_0'\beta - c) \\ &\iff -x_1'\beta - c \leq -x_0'\beta - c \\ &\iff (x_1 - x_0)'\beta \geq 0 \end{aligned}$$

Since this event is independent of c, we have

$$P(y_1 = 1 \mid x) - P(y_0 = 1 \mid x) \geq 0 \iff (x_1 - x_0)'\beta \geq 0,$$

or, equivalently,

$$\operatorname{sgn}\big(P(y_1 = 1 \mid x) - P(y_0 = 1 \mid x)\big) = \operatorname{sgn}\big((x_1 - x_0)'\beta\big).$$

Therefore, $\theta \in \Theta_I \Rightarrow \theta \in \tilde{\Theta}_I$.

Step 2 Now, suppose $\theta \in \tilde{\Theta}_I$. We will show that for each such θ, given population distributions $P(y_t \mid x)$ for $t = 0, 1$, there are values of the remaining free model primitives—the cdfs $F_{u_0|xc}$ and $F_{c|x}$—such that the implications of the model coincide with the true population values $P(y_0 = 0 \mid x)$ and $P(y_1 = 0 \mid x)$.

First, note that we do not need to consider the events $y_0 = 1$ or $y_1 = 1$ since in each time period, the (binary) choice probabilities must sum to one. Thus, we need to show that there exist distributions $F_{u_0|xc}$ and $F_{c|x}$ such that for F_x-almost every x the model implications align with the population choice probabilities:

$$P(y_0 = 0 \mid x) = \pi(y_0 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x})$$

$$P(y_1 = 0 \mid x) = \pi(y_1 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x})$$

For a given x and for primitives $(\theta, F_{u_0|xc}, F_{c|x})$, the model implications are:

$$\pi(y_0 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) = \int F_{u_0|xc}(-x_0'\beta - c)\, dF_{c|x}$$

$$\pi(y_1 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) = \int F_{u_0|xc}(-x_1'\beta - c)\, dF_{c|x}$$

Fix x. It will suffice to construct a distribution $F_{c|x}$ with only a single mass point $c^*(x)$ (conditional on each fixed value of x):

$$F_{c|x}(c) = \begin{cases} 0 & \text{if } c < c^*(x), \\ 1 & \text{if } c \geq c^*(x). \end{cases}$$

Suppose that $P(y_1 = 1 \mid x) < P(y_0 = 1 \mid x)$ (the opposite case follows similarly). Then our choice of $\theta \in \tilde{\Theta}_I$ guarantees that β is such that $x_1'\beta < x_0'\beta$. We can rewrite these two inequalities equivalently as $P(y_0 = 0 \mid x) < P(y_1 = 0 \mid x)$ and $-x_0'\beta < -x_1'\beta$. Thus, the following choice for $F_{u_0|xc}$ is a valid cdf:

$$F_{u_0|xc}(u) = \begin{cases} 0 & \text{if } u < -x_1'\beta - c^*(x), \\ P(y_0 = 0 \mid x) & \text{if } -x_0'\beta - c^*(x) \leq u \leq -x_1'\beta - c^*(x), \\ P(y_1 = 0 \mid x) & \text{if } -x_1'\beta - c^*(x) \leq u < \bar{u}, \\ 1 & \text{if } u \geq \bar{u}, \end{cases}$$

for any $\bar{u} > -x_1'\beta - c^*(x)$. Essentially, we only need to choose a cdf that passes through the two points $\left(-x_0'\beta - c^*(x), P(y_0 = 0 \mid x)\right)$ and $\left(-x_1'\beta - c^*(x), P(y_1 = 0 \mid x)\right)$ and there are an infinite number of such cdfs, as illustrated by Figure C.1.

Given the above cdfs, we have:

$$\pi(y_0 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) = F_{u_0|xc}(-x_0'\beta - c^*(x)) = P(y_0 = 0 \mid x),$$

$$\pi(y_1 = 0 \mid x; \theta, F_{u_0|xc}, F_{c|x}) = F_{u_0|xc}(-x_1'\beta - c^*(x)) = P(y_1 = 0 \mid x).$$

Therefore $\theta \in \Theta_I$, and since $\theta \in \tilde{\Theta}_I$ was chosen arbitrarily, $\tilde{\Theta}_I \subseteq \Theta_I$. ■

Proof of Lemma 2.1. Define $w = x_1 - x_0$, $z = y_1 - y_0$, and $\Theta^* = \arg\max_{\theta \in \Theta} Q(\theta)$.

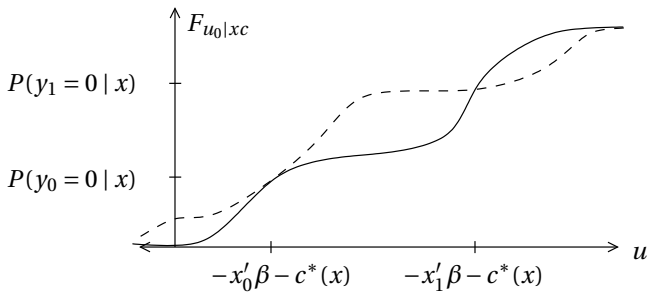

FIGURE C.1: Two distributions $F_{u_0|xc}$ with equivalent observable implications under $F_{c|x}$.

Step 1 Let $\theta_1 \in \Theta_I$ and $\theta_2 \in \Theta$. We will show that $\Theta_I \subseteq \Theta^*$ by proving that, for arbitrary choices of θ_1 and θ_2, $Q(\theta_1) \geq Q(\theta_2)$.

Consider the difference

$$\begin{aligned} Q(\theta_1) - Q(\theta_2) &= \mathrm{E}\left[z \operatorname{sgn}(w'\beta_1)\right] - \mathrm{E}\left[z \operatorname{sgn}(w'\beta_2)\right] \\ &= \mathrm{E}\left[z\left(\operatorname{sgn}(w'\beta_1) - \operatorname{sgn}(w'\beta_2)\right)\right] \\ &= 2\int_{D(\theta_1,\theta_2)} \operatorname{sgn}(w'\beta_1)\,\mathrm{E}\left[z \mid x, c\right]\, dF_{xc} \end{aligned}$$

where $D(\theta_1,\theta_2) = \{(x,c) : \operatorname{sgn}(w'\beta_1) \neq \operatorname{sgn}(w'\beta_2)\}$ is the set of values of x and c where $\operatorname{sgn}(w'\beta_1)$ and $\operatorname{sgn}(w'\beta_2)$ differ. The last equality above follows from the fact that the integrand vanishes on complement of $D(\theta_1,\theta_2)$, and that on $D(\theta_1,\theta_2)$ we have $\operatorname{sgn}(w'\beta_1) = -\operatorname{sgn}(w'\beta_2)$, implying that $\operatorname{sgn}(w'\beta_1) - \operatorname{sgn}(w'\beta_2) = 2\operatorname{sgn}(w'\beta_1)$. Since $\theta_1 \in \Theta_I$, Theorem 2.1 guarantees that

$$\operatorname{sgn}(w'\beta_1) = \operatorname{sgn}\left(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\right) = \operatorname{sgn}\mathrm{E}(z \mid x, c)$$

F_{xc}-almost surely. Rewriting the above difference,

$$Q(\theta_1) - Q(\theta_2) = 2\int_{D(\theta_1,\theta_2)} |\mathrm{E}[z \mid x, c]|\, dF_{xc} \geq 0$$

for all θ_2. Therefore, $\Theta_I \subseteq \Theta^*$.

Step 2 Now, let $\theta_1 \in \Theta_I$ and suppose there exists a $\theta_2 \in \Theta_I^c \cap \Theta^*$, where Θ_I^c denotes the complement of Θ_I. We will use the definition of Θ_I to show that $Q(\theta_2) < Q(\theta_1)$, contradicting the assumption that $\theta_2 \in \Theta^*$, and guaranteeing that $\Theta_I^c \cap \Theta^* = \emptyset$, or equivalently, $\Theta^* \subseteq \Theta_I$.

First, note that we can rewrite $Q(\theta)$ as follows:

$$
\begin{aligned}
Q(\theta) &= \mathrm{E}[z \operatorname{sgn}(w'\beta)] \\
&= \mathrm{E}_{xc}\, \mathrm{E}_{z|wc}[z \operatorname{sgn}(w'\beta)] \\
&= \mathrm{E}_{xc}\left[\big(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\big)\big(1\{w'\beta \geq 0\} - 1\{w'\beta < 0\}\big)\right] \\
&= \int_{\{w'\beta \geq 0\}} \big(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\big)\, dF_{xc} \\
&+ \int_{\{w'\beta < 0\}} \big(P(y_0 = 1 \mid x, c) - P(y_1 = 1 \mid x, c)\big)\, dF_{xc}
\end{aligned}
$$

The first equality is definitional, the second is an application of the law of iterated expectations, and the third follows from the definition of z and the signum function. In the fourth line, the expectations of the indicator functions are expressed as integrals over the corresponding regions of the support of x.

Now, consider the difference $Q(\theta_2) - Q(\theta_1)$:

$$
\begin{aligned}
Q(\theta_2) - Q(\theta_1) = &\int_{\{w'\beta_2 \geq 0\}} \big(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\big)\, dF_{xc} \\
&+ \int_{\{w'\beta_2 < 0\}} \big(P(y_0 = 1 \mid x, c) - P(y_1 = 1 \mid x, c)\big)\, dF_{xc} \\
&- \int_{\{w'\beta_1 \geq 0\}} \big(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\big)\, dF_{xc} \\
&- \int_{\{w'\beta_1 < 0\}} \big(P(y_0 = 1 \mid x, c) - P(y_1 = 1 \mid x, c)\big)\, dF_{xc}
\end{aligned}
$$

Over regions where $w'\beta_2$ and $w'\beta_1$ have the same sign, the difference is zero, therefore

$$Q(\theta_2) - Q(\theta_1) = \int_{\{w'\beta_2 \geq 0, w'\beta_1 < 0\}} \big(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\big)\, dF_{xc}$$
$$- \int_{\{w'\beta_2 < 0, w'\beta_1 \geq 0\}} \big(P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c)\big)\, dF_{xc}$$

From the proof of Theorem 2.1, we know that for $\theta_1 \in \Theta_I$,

$$P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c) \geq 0 \iff w'\beta_1 \geq 0$$

and for $\theta_2 \in \Theta_I^c$,

$$P(y_1 = 1 \mid x, c) - P(y_0 = 1 \mid x, c) < 0 \iff w'\beta_2 \geq 0.$$

This implies that the first term in the difference above is strictly negative and the second term, which is being subtracted, is weakly non-negative. Thus, $Q(\theta_2) < Q(\theta_1)$. This contradicts the choice of θ_2, meaning that $\Theta_I^c \cap \Theta^* = \emptyset$ and therefore it must be the case that $\Theta^* \subseteq \Theta_I$. ■

Proof of Lemma 2.2. Let $\mathscr{D} \subset \mathbb{R}^d$ denote the support of w and let $\mathscr{X} = \{-1, 0, 1\} \times \mathscr{D}$ denote the support of (z, w). For each $(z, w) \in \mathscr{X}$ and for each real number t, α, and γ, and real vector $\delta \in \mathbb{R}^d$, define

$$g(z, w, t, \alpha, \gamma, \delta) = \alpha t + \gamma z + \delta' w$$

and define

$$\mathscr{G} = \Big\{ g(\cdot, \cdot, \cdot, \alpha, \gamma, \delta) : \alpha, \gamma \in \mathbb{R} \text{ and } \delta \in \mathbb{R}^d \Big\}.$$

Since $\mathscr{G}$ is a vector space of real-valued functions on $\mathscr{X} \times \mathbb{R}$, by Lemma 2.4 of Pakes and Pollard (1989), classes of sets of the form $\{g \geq r\}$ or $\{g > r\}$ with $g \in \mathscr{G}$ and $r \in \mathbb{R}$ are VC classes. We will show that $\mathscr{F}$ is Euclidean by showing that it is a VC subgraph class, that is, that the collection of subgraphs of functions in $\mathscr{F}$ is a VC class. To accomplish this, we

will use Lemma 2.5 of Pakes and Pollard (1989) which states that, in particular, if $\mathscr{C}_1$ and $\mathscr{C}_2$ are VC classes, then so are $\{C_1 \cap C_2 : C_1 \in \mathscr{C}_1, C_2 \in \mathscr{C}_2\}$, $\{C_1 \cup C_2 : C_1 \in \mathscr{C}_1, C_2 \in \mathscr{C}_2\}$, and $\{C_1^c : C_1 \in \mathscr{C}_1\}$.

First, note that we can rewrite f as

$$\begin{aligned} f(z,w,\theta) &= (1\{z>0\} - 1\{z<0\}) \cdot \left(1\{w'\beta \geq 0\} - 1\{w'\beta < 0\}\right) \\ &= 1\{z>0,\, w'\beta \geq 0\} - 1\{z>0,\, w'\beta < 0\} \\ &\quad - 1\{z<0,\, w'\beta \geq 0\} + 1\{z<0,\, w'\beta < 0\}. \end{aligned}$$

Now, for any $\theta \in \Theta$,

$$\begin{aligned} \text{subgraph}(f(\cdot,\cdot,\theta)) &= \{(z,w,t) \in \mathscr{X} \times \mathbb{R} : 0 < t < f(z,w,\theta) \text{ or } 0 > t > f(z,w,\theta)\} \\ &= \left(\{z>0\} \cap \{w'\beta \geq 0\} \cap \{t \geq 1\}^c \cap \{t>0\}\right) \\ &\quad \cup \left(\{z>0\} \cap \{w'\beta \geq 0\}^c \cap \{t \geq -1\} \cap \{t \geq 0\}^c\right) \\ &\quad \cup \left(\{z \geq 0\}^c \cap \{w'\beta \geq 0\} \cap \{t \geq -1\} \cap \{t \geq 0\}^c\right) \\ &\quad \cup \left(\{z \geq 0\}^c \cap \{w'\beta \geq 0\}^c \cap \{t \geq 1\}^c \cap \{t>0\}\right) \\ &= \left(\{g_1>0\} \cap \{g_2 \geq 0\} \cap \{g_3 \geq 1\}^c \cap \{g_3>0\}\right) \\ &\quad \cup \left(\{g_1>0\} \cap \{g_2 \geq 0\}^c \cap \{g_3 \geq -1\} \cap \{g_3 \geq 0\}^c\right) \\ &\quad \cup \left(\{g_1 \geq 0\}^c \cap \{g_2 \geq 0\} \cap \{g_3 \geq -1\} \cap \{g_3 \geq 0\}^c\right) \\ &\quad \cup \left(\{g_1 \geq 0\}^c \cap \{g_2 \geq 0\}^c \cap \{g_3 \geq 1\}^c \cap \{g_3>0\}\right) \end{aligned}$$

where $g_k(z,w,t) = \alpha_k t + \gamma_k z + \delta_k' w \in \mathscr{G}$ for each k with, $\alpha_1 = 0$, $\gamma_1 = 1$, $\delta_1 = 0$, $\alpha_2 = 0$, $\gamma_2 = 0$, $\delta_2 = \beta$, $\alpha_3 = 1$, $\gamma_3 = 0$, and $\delta_3 = 0$. The collection of sets of the form $\{g \geq 0\}$ or $\{g>0\}$ is a VC class by Lemma 2.4 of Pakes and Pollard (1989). Furthermore, this property is preserved over complements, unions, and intersections of VC classes by their Lemma 2.5. Therefore, $\{\text{subgraph}(f) : f \in \mathscr{F}\}$ is a VC class, and by Lemma 2.12 of Pakes and Pollard (1989), $\mathscr{F}$ is Euclidean for any envelope. In particular, $\mathscr{F}$ is Euclidean for the constant envelope $F = 1$. ■

Proof of Theorem 2.6. We shall verify each of the conditions of Assumption 2.4. Condition a is satisfied by definition of Model 2.1, condition b holds as a result of Lemma 2.1, and condition d is satisfied with $b_n = \sqrt{n}$ as a result of Lemma 2.3. ■

C.4.2 Discrete Regressors

Proof of Lemma 2.7. Note that the first part of this proof is independent of assumption Assumption 2.1.

Verification of Assumption 2.7 First, note that we can rewrite $n^{1/2}Q_n$ as

$$n^{1/2}Q_n(\theta) = n^{1/2}(P_n f_\theta - P f_\theta) + n^{1/2} P f_\theta = \mathbb{G}_n(f_\theta) + n^{1/2} P f_\theta,$$

and therefore,

$$\mathscr{Q}_n \equiv \inf_{\theta \in \Theta_I} n^{1/2} Q_n(\theta) = \inf_{\theta \in \Theta_I} \left(\mathbb{G}_n(f_\theta) + n^{1/2} P f_\theta \right).$$

Supposing Q is normalized so that it is identically zero on Θ_I, since the map $\inf_{\Theta_I}$, which takes real functions on Θ into $\mathbb{R}$, is continuous in $\ell^\infty(\mathscr{F})$, the continuous mapping theorem gives $\mathscr{Q}_n \xrightarrow{\text{d}} \inf_{\theta \in \Theta_I} \mathbb{G}(f_\theta) \equiv \mathscr{Q}$.

Verification of Assumption 2.8 For any sequence of subsets Θ_n of Θ such that $d_H(\Theta_n, \Theta_I) \downarrow 0$ in probability, define $\mathscr{Q}'_n \equiv \inf_{\theta \in \Theta_n} n^{1/2} Q_n(\theta)$. For all $\varepsilon > 0$, there exists an n_ε such that for all $n \geq n_\varepsilon$, $P(\Theta_n = \Theta_I) \geq 1 - \varepsilon$. Then, $P\left(\inf_{\Theta_n} n^{1/2} Q_n = \inf_{\Theta_I} n^{1/2} Q_n\right) \geq 1 - \varepsilon$. Recall from above that $\inf_{\theta \in \Theta_I} n^{1/2} Q_n \xrightarrow{\text{d}} \mathscr{Q}$. Therefore, $\mathscr{Q}'_n \xrightarrow{\text{d}} \mathscr{Q}$. ■

C.4.3 Continuous Regressors

Proof of Theorem 2.8. Lemma 2.2 established that $\mathscr{F}$ is Euclidean and the conditions of Assumption 2.4 have been shown to hold previously with $b_n = n^{1/2}$. We will show that Assumption 2.6 holds with $\gamma_1 = 2$ and $\gamma_2 = 3/2$ and then use Theorem 2.5 to obtain the resulting rate.

Abrevaya and Huang (2005) show that $\nabla_{\theta\theta'} Q(\theta_0) = -V(\theta_0)$. Generalizing their argument to the set identified case yields $\nabla_{\theta\theta'} Q(\theta) = -V(\theta)$ for all $\theta \in \text{bd}(\Theta_I)$. Therefore, in a neighborhood $\mathcal{N}$ of Θ_I, Q is approximately quadratic and for some $C > 0$, $Q(\theta) \leq \sup Q - C \cdot d^2(\theta, \Theta_I)$.

Let $\eta > 0$ and define $\mathcal{F}_\eta \equiv \{f_\theta \in \mathcal{F} : d(\theta, \Theta_I) < \eta\}$. Again, following the arguments of Abrevaya and Huang (2005), $\mathcal{F}_\eta$ is a VC subgraph class with envelope F_η such that $PF_\eta^2 = O_p(\eta)$. Then, by Lemma 4.1 of Kim and Pollard (1990), for all $\epsilon > 0$, there exists a sequence $M_n = O_p(1)$ such that

$$\text{(C.3)} \quad P_n f_\theta - P f_\theta \leq \epsilon d^2(\theta, \Theta_I) + n^{-2/3} M_n^2$$

for $d(\theta, \Theta_I) \leq \eta$.

Let $\mathbb{G}_n(\theta) \equiv n^{1/2}(P_n f_\theta - P f_\theta)$ denote the standardized empirical process. For $\theta \in \mathcal{N}^c$,

$$\begin{aligned} Q_n(\theta) &\leq n^{-1/2}\mathbb{G}_n(\theta) + \sup_\Theta Q - \delta \\ &\leq n^{-1/2} O_p(1) + \sup_\Theta Q - \delta \\ &\leq \sup_\Theta Q - \tilde{\delta} \end{aligned}$$

for sufficiently large n. The final inequality is a result of the Donsker property which implies $\sup_{\theta\in\Theta} \left|\mathbb{G}_n(f_\theta)\right| = O_p(1)$.

For $\theta \in \mathcal{N}$, we can choose $\epsilon = \frac{1}{2}C$ in (C.3) and combine this with the quadratic approximation above to obtain

$$\begin{aligned} Q_n(\theta) &= (P_n f_\theta - P f_\theta) + P f_\theta \\ &\leq (P_n f_\theta - P f_\theta) + \sup_\Theta Q - C \cdot d^2(\theta, \Theta_I) \\ &\leq n^{-2/3} M_n^2 + \sup_\Theta Q - \frac{1}{2} C \cdot d^2(\theta, \Theta_I). \end{aligned}$$

Note that the term $n^{-2/3} M_n^2$ is smaller than $\frac{1}{4} C d^2(\theta, \Theta_I)$ whenever $d(\theta, \Theta_I) \geq \frac{4M_n^2}{Cn^{-1/3}}$.

For any $\varepsilon > 0$ we can choose δ, κ_ε and n_ε such that for all $n \geq n_\varepsilon$, with probability at least $1 - \varepsilon$,

$$Q_n(\theta) \leq \sup_{\Theta} Q - C \cdot (d(\theta, \Theta_I) \wedge \delta)^2$$

uniformly on $\{\theta \in \Theta : d(\theta, \Theta_I) \geq (\kappa_\varepsilon / n^{1/2})^{2/3}\}$. This follows since we can choose n_ε large enough to guarantee that set $D_n \equiv \{\theta : d(\theta, \Theta_I) \geq (\kappa_\varepsilon / b_n)^{1/\gamma_2}\}$ intersects the neighborhood $\mathcal{N}$.

Thus, we have verified Assumption 2.6 with $\gamma_1 = 2$, $\gamma_2 = 3/2$, and $b_n = n^{1/2}$. The conclusion then follows by applying Theorem 2.5. ■

C.5 Lagged Dependent Variable Model

Proof of Lemma 2.4. This proof parallels the proof of Lemma 2.1, the corresponding result for Model 2.1. Let $\Theta^* \equiv \arg\max_\Theta Q$ and define $w_t \equiv x_t - x_{t-1}$, $z \equiv y_2 - y_1$, and $v \equiv y_3 - y_0$.

Step 1 Let $\theta_1 \in \Theta_I$ and $\theta_2 \in \Theta$. We will show that $Q(\theta_1) \geq Q(\theta_2)$ and therefore, $\theta_1 \in \Theta^*$. We have

$$\begin{aligned} Q(\theta_1) - Q(\theta_2) &= \mathrm{E}\left[1\{w_3 = 0\} \cdot z \cdot \left(\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v) - \operatorname{sgn}(w_2'\beta_2 + \gamma_2 v)\right)\right] \\ &= \int \mathrm{E}[z \mid x, c, y_0, y_3, w_3 = 0] \left(\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v) - \operatorname{sgn}(w_2'\beta_2 + \gamma_2 v)\right) dF_{x,c,y_0,y_3|w_3=0} \\ &= 2 \int_{D(\theta_1,\theta_2)} \operatorname{sgn}(w_2'\beta_1 + \gamma_1 v) \, \mathrm{E}[z \mid x, c, y_0, y_3, w_3 = 0] \, dF_{x,c,y_0,y_3|w_3=0} \end{aligned}$$

where $D(\theta_1, \theta_2)$ is defined as the set of all (x, c, v) where $\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v)$ and $\operatorname{sgn}(w_2'\beta_2 + \gamma_2 v)$ differ. The first equality follows by definition of Q, the second is an application of the law of iterated expectations, and the third is due to the fact that on $D(\theta_1, \theta_2)$, $\operatorname{sgn}(w_2'\beta_2 + \gamma_2 v) = -\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v)$. Note that on the integrand above vanishes on the complement of $D(\theta_1, \theta_2)$.

Now, since $\theta_1 \in \Theta_I$, from Theorem 2.2 we have that for all d_0, d_3,

$$\begin{aligned}\operatorname{sgn}(w_2'\beta_1 + \gamma_1 v) &= \operatorname{sgn}\left(P(A \mid x, x_2 = x_3) - P(B \mid x, x_2 = x_3)\right) \\ &= \operatorname{sgn}\big(P(y_1 = 0, y_2 = 1 \mid x, x_2 = x_3, y_0 = d_0, y_3 = d_3) \\ &\qquad - P(y_1 = 1, y_2 = 0 \mid x, x_2 = x_3, y_0 = d_0, y_3 = d_3)\big)\end{aligned}$$

for all $d_0, d_3 \in \{0,1\}$. The second line follows because the common factor which was removed, $P(y_0 = d_0, y_3 = d_3 \mid x, x_2 = x_3)$, is always positive. Furthermore, we can write

$$\begin{aligned}\mathrm{E}[z \mid x, c, y_0, y_3, w_3 = 0] = P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0) \\ - P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0).\end{aligned}$$

So, the sign above times the conditional expectation of z simplifies to the absolute value of the conditional expectation. Returning to the objective function,

$$Q(\theta_1) - Q(\theta_2) = 2\int_{D(\theta_1,\theta_2)} \left|E[z \mid x, c, y_0, y_3, w_3 = 0]\right| \, dF_{x,c,y_0,y_3 \mid w_3 = 0} \geq 0.$$

Step 2 Let $\theta_1 \in \Theta_I$ and suppose there exists a $\theta_2 \in \Theta_I^c \cap \Theta^*$. We will show that this implies $Q(\theta_2) < Q(\theta_1)$, which is a contradiction of the choice of $\theta_2 \in \Theta^*$, and therefore $\Theta_I^c \cap \Theta^*$ must be empty.

Note that we can express Q as

$$\begin{aligned}Q(\theta) = &\int_{\{w_3'\beta + \gamma v \geq 0\}} \big[P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0) \\ &\qquad - P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0)\big] \, dF_{x,c,y_0,y_3 \mid w_3 = 0} \\ + &\int_{\{w_3'\beta + \gamma v < 0\}} \big[P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0) \\ &\qquad - P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0)\big] \, dF_{x,c,y_0,y_3 \mid w_3 = 0}.\end{aligned}$$

Again we consider a difference $Q(\theta_2) - Q(\theta_1)$. Using the linearity of integrals, we can partition the range of each integral into disjoint sets and subtract the corresponding integrands on each set. When $w_3'\beta_1 + \gamma_1 v$ and $w_3'\beta_2 + \gamma_2 v$ have the same sign, the difference

is zero, so we only need to consider regions where the sign differs:

$$D_1 \equiv \{(x, c, y_0, y_3) : w_3'\beta_2 + \gamma_2 v \geq 0, w_3'\beta_1 + \gamma_1 v < 0\},$$

$$D_2 \equiv \{(x, c, y_0, y_3) : w_3'\beta_2 + \gamma_2 v < 0, w_3'\beta_1 + \gamma_1 v \geq 0\}.$$

Hence,

$$\begin{aligned} Q(\theta_2) - Q(\theta_1) = &\int_{D_1} \big[P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0) \\ &\qquad - P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0)\big]\, dF_{x,c,y_0,y_3|w_3=0} \\ &+ \int_{D_2} \big[P(y_1 = 1, y_2 = 0 \mid x, c, y_0, y_3, w_3 = 0) \\ &\qquad - P(y_1 = 0, y_2 = 1 \mid x, c, y_0, y_3, w_3 = 0)\big]\, dF_{x,c,y_0,y_3|w_3=0}. \end{aligned}$$

Since $\theta_1 \in \Theta_I$ and $\theta_2 \notin \Theta_I$, first term is strictly negative and the second term is weakly non-positive. ■

Proof of Lemma 2.5. We follow the same strategy as in the proof of Lemma 2.2. Define $w_t \equiv x_t - x_{t-1}$, $z \equiv y_2 - y_1$, and $v \equiv y_3 - y_0$, and let $f(w_2, w_3, z, v, \theta) = 1\{w_3 = 0\} \cdot z_2 \cdot \big[2 \cdot 1\{w_2'\beta + \gamma v \geq 0\} - 1\big]$. First, note that f can be rewritten as

$$\begin{aligned} f(w_2, w_3, z, v, \theta) = &1\{w_3 \geq 0\} \cdot 1\{w_3 \leq 0\} \cdot (1\{z_2 > 0\} - 1\{z < 0\}) \\ &\cdot \big(1\{w_2'\beta + \gamma v \geq 0\} - 1\{w_2'\beta + \gamma v < 0\}\big). \end{aligned}$$

Upon expanding this expression, it is clear that, as before, for any θ we can express subgraph $f(\cdot, \cdot, \cdot, \cdot, \theta)$ as series of intersections and unions (and complements thereof) of the form $\{g \geq 0\}$ and $\{g > 0\}$ for specific coefficient values α of some polynomial

$$g(w_2, w_3, z, v, t, \alpha) = \alpha_1 t + \alpha_2 w_2 + \alpha_3 w_3 + \alpha_4 z + \alpha_5 v.$$

It then follows that $\{\text{subgraph}(f) : f \in \mathscr{F}\}$ is a VC class, and, therefore, $\mathscr{F}$ is Euclidean for any envelope. In particular, it is Euclidean for the envelope $F = 1$. ■

Proof of Theorem 2.9. We verify each of the conditions of Assumption 2.4. Condition a is satisfied by definition of Model 2.2, condition b holds as a result of Lemma 2.4, and condition d is satisfied with $b_n = \sqrt{n}$ as a result of Lemma 2.3, since the objective function is of the same form as that of Model 2.1—only the indexing class of functions $\mathscr{F}$ is different but both are Euclidean with envelope $F = 1$. ■

Proof of Theorem 2.10. When the support of x is a finite set, henceforth $\mathscr{X}$, the objective function $Q(\theta)$ can be rewritten as follows:

$$\begin{aligned} Q(\theta) = \sum_{y_0 \in \{0,1\}} \sum_{y_3 \in \{0,1\}} \sum_{x \in \mathscr{X}} & P(x)P(y_0 \mid x)P(y_3 \mid x, y_0) \\ & \times \left[P(y_2 = 1 \mid x, y_0, y_3) - P(y_1 = 1 \mid x, y_0, y_3)\right] \\ & \times \operatorname{sgn}\left((x_2 - x_1)'\beta + \gamma(y_3 - y_0)\right). \end{aligned}$$

Therefore, $Q(\theta)$ is a step function and there exists a real number $\delta > 0$ with

$$\delta \geq \inf_{(x, y_0, y_3)} P(x)P(y_0 \mid x)P(y_3 \mid x, y_0)\left[P(y_2 = 1 \mid x, y_0, y_3) - P(y_1 = 1 \mid x, y_0, y_3)\right]$$

such that for all $\theta \in \Theta \setminus \Theta_I$, $Q(\theta) \leq \sup_\Theta Q - \delta$. This verifies Assumption 2.5. Since conditions of Assumption 2.4 were already established under the current assumptions in the proof of Theorem 2.9, the result follows by applying Theorem 2.4. ■

Appendix D

Auxiliary Results for Chapter 3

Lemma D.1. *Under the assumptions of Theorem 3.2, the functional mapping*

$$\Gamma\left(v(0,\cdot)\right)(s) \equiv \mathrm{E}_{\eta}\left\{\sup_{c}\left[u(0,c,s,\eta)+\beta\,\mathrm{E}_{s'|d=0,c,s}\left[H\left(\Delta(1,s),\ldots,\Delta(K,s)\right)+v(0,s)\right]\right]\right\}$$

is a contraction with modulus β.

Proof of Lemma D.1. First we simplify the notation, defining $w(s) \equiv v(0,s)$ and

$$\Gamma(w)(s) = \sup_{c}\left[\psi(0,c,s)+\beta\,\mathrm{E}_{s'|d=0,c,s}\,w(s')\right],$$

where

$$\psi(d,c,s) \equiv \beta\,\mathrm{E}_{s'|d,c,s}\left[H\left(\Delta(1,s'),\ldots,\Delta(K,s')\right)\right]$$

Note that in light of the utility normalization, this expression no longer depends on η and so we can drop the outer expectation. Furthermore, under the assumptions of Theorem 3.2, the function $\psi(d,c,s)$ is identified.

We must show that for any two functions w and $\tilde{w}$, $\|\Gamma w - \Gamma\tilde{w}\| \leq k\,\|w-\tilde{w}\|$ for some

$0 < k < 1$ where $\|\cdot\|$ is the sup norm, $\|f\| \equiv \sup_{s \in \mathscr{S}} |f(s)|$. We have

$$\begin{aligned}
\|\Gamma w - \Gamma \tilde{w}\| &= \sup_s |\Gamma w(s) - \Gamma \tilde{w}(s)| \\
&= \sup_s \left| \sup_c \left[\psi(0,c,s) + \beta \mathrm{E}_{s'|d=0,c,s}\, w(s') \right] - \sup_c \left[\psi(0,c,s) + \beta \mathrm{E}_{s'|d=0,c,s}\, \tilde{w}(s') \right] \right| \\
&\leq \sup_s \sup_c \left| \left[\psi(0,c,s) + \beta \mathrm{E}_{s'|d=0,c,s}\, w(s') \right] - \left[\psi(0,c,s) + \beta \mathrm{E}_{s'|d=0,c,s}\, \tilde{w}(s') \right] \right| \\
&= \beta \sup_s \sup_c \left| \mathrm{E}_{s'|d=0,c,s} \left[w(s') - \tilde{w}(s') \right] \right| \\
&\leq \beta \sup_s \sup_c \mathrm{E}_{s'|d=0,c,s} \left| w(s') - \tilde{w}(s') \right| \\
&\leq \beta \sup_{s'} \left| w(s') - \tilde{w}(s') \right| \\
&= \beta \| w - \tilde{w} \| .
\end{aligned}$$

The first two equalities follow by definition while the third line follows from the properties of the supremum: $\left|\sup f(x) - \sup g(x)\right| \leq \sup \left| f(x) - g(x) \right|$. The fourth line is a simplification using the linearity of the expectation operator. The next two lines follow from properties of the integral: we know that $\left| \int f \right| \leq \int |f|$ and that for any measure μ, $\int_E f \, d\mu \leq \mu(E) \sup_{x \in E} f(x)$. The last equality holds by definition of the norm. ■

Lemma D.2. *Under the assumptions of Theorem 3.3, the functional mapping*

$$\begin{aligned}
&\Gamma\left(v_i(0,\cdot)\right)(s) \\
&\quad \equiv \mathrm{E}_{\eta_i|d_i=0,s}\, \mathrm{E}_{d_{-i}|d_i=0,s} \sup_{c_i} \beta \mathrm{E}_{c_{-i}|d,c_i,s}\, \mathrm{E}_{s'|d,c,s} \left[H(\Delta_i(1,s'), \ldots, \Delta_i(K,s')) + v_i(0,s') \right].
\end{aligned}$$

is a contraction with modulus β.

Proof of Lemma D.1. For simplicity, define $w(s) \equiv v(0,s)$ so that the functional equation becomes

$$\Gamma(w)(s) \equiv \mathrm{E}_{\eta_i|d_i=0,s}\, \mathrm{E}_{d_{-i}|d_i=0,s} \sup_{c_i} \mathrm{E}_{c_{-i}|d,c_i,s} \left\{ \psi(0,c,s) + \beta \mathrm{E}_{s'|d,c,s}\, w(s') \right\},$$

where

$$\psi(d,c,s) \equiv \beta \mathrm{E}_{c_{-i}|d,c_i,s} \mathrm{E}_{s'|d,c,s} \left[H\left(\Delta(1,s'),\ldots,\Delta(K,s')\right)\right]$$

Note that in light of the utility normalization, this expression no longer depends on η_i and so we can drop the outermost expectation. Furthermore, under the assumptions of Theorem 3.3, the function $\psi(d,c,s)$ is identified.

We must show that for any two functions w and $\tilde{w}$, $\|\Gamma w - \Gamma \tilde{w}\| \leq k \|w - \tilde{w}\|$ for some $0 < k < 1$ where $\|\cdot\|$ is the sup norm, $\left\|f\right\| \equiv \sup_{s\in\mathscr{S}} \left|f(s)\right|$. We have

$$\begin{aligned}
\|\Gamma w - \Gamma \tilde{w}\| &= \sup_s |\Gamma w(s) - \Gamma \tilde{w}(s)| \\
&= \beta \sup_s \left| \mathrm{E}_{d_{-i}|d_i=0,s} \left[\sup_{c_i} \mathrm{E}_{c_{-i}|d,c_i,s} \mathrm{E}_{s'|d,c,s}\, w(s') - \sup_{c_i} \mathrm{E}_{c_{-i}|d,c_i,s} \mathrm{E}_{s'|d,c,s}\, \tilde{w}(s') \right] \right| \\
&\leq \beta \sup_s \mathrm{E}_{d_{-i}|d_i=0,s} \left| \sup_{c_i} \mathrm{E}_{c_{-i}|d,c_i,s} \mathrm{E}_{s'|d,c,s}\, w(s') - \sup_{c_i} \mathrm{E}_{c_{-i}|d,c_i,s} \mathrm{E}_{s'|d,c,s}\, \tilde{w}(s') \right| \\
&\leq \beta \sup_s \mathrm{E}_{d_{-i}|d_i=0,s} \sup_{c_i} \mathrm{E}_{c_{-i}|d,c_i,s} \mathrm{E}_{s'|d,c,s} \left| w(s') - \tilde{w}(s') \right| \\
&\leq \beta \sup_s \mathrm{E}_{d_{-i}|d_i=0,s} \sup_{c_i} \mathrm{E}_{c_{-i}|d,c_i,s} \sup_{s'} \left| w(s') - \tilde{w}(s') \right| \\
&\leq \beta \sup_{s'} \left| w(s') - \tilde{w}(s') \right| \\
&= \beta \|w - \tilde{w}\|.
\end{aligned}$$

The first equality follows by definition of the norm. The second follows since ψ is known and by the linearity of $\mathrm{E}_{d_{-i}|d_i=0,s}$. The remaining inequalities from properties of the integral, the uniform continuity of sup, and the definition of the norm. ■

Bibliography

Abrevaya, J. (2000). Rank estimation of a generalized fixed-effects regression model. *Journal of Econometrics 95*, 1–23.

Abrevaya, J. and J. Huang (2005). On the bootstrap of the maximum score estimator. *Econometrica 73*, 1175–1204.

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

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

Aguirregabiria, V. and P. Mira (2009). Dynamic discrete choice structural models: a survey. Unpublished manuscript, University of Toronto.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Andrews, D. W. K., S. Berry, and P. Jia (2004). Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store location. Unpublished manuscript, Yale University.

Andrews, D. W. K. and X. Shi (2009). Inference based on conditional moment inequalities. Unpublished manuscript, Yale University.

Aradillas-Lopez, A. and E. Tamer (2008). The identification power of equilibrium in simple games. *Journal of Business and Economic Statistics 26*, 261–283.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica 73*, 1477–1524.

Arcidiacono, P. and R. A. Miller (2008). CCP estimation of dynamic discrete choice models with unobserved heterogeneity. Unpublished manuscript, Duke University.

Bajari, P., C. L. Benkard, and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica 75*, 1331–1370.

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2007). Semiparametric estimation of dynamic games of incomplete information. Unpublished manuscript, University of Minnesota.

Bajari, P., H. Hong, and S. Ryan (2007). Identification and estimation of discrete games of complete information. Unpublished manuscript, University of Minnesota.

Beresteanu, A., I. Molchanov, and F. Molinari (2009). Sharp identification regions in models with convex predictions: games, individual choice, and incomplete data. Working Paper CWP27/09, CEMMAP.

Beresteanu, A. and F. Molinari (2008). Asymptotic properties for a class of partially identified models. *Econometrica 76*, 763–814.

Bierens, H. J. and J. Hartog (1988). Non-linear regression with discrete explanatory variables, with an application to the earnings function. *Journal of Econometrics 38*, 269–299.

Blundell, R., X. Chen, and D. Kristensen (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica 75*, 1613–1669.

Bugni, F. A. (2008). Bootstrap inference in partially identified models. Unpublished manuscript, Northwestern University.

Byrd, R. H., P. Lu, and J. Nocedal (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing 16*, 1190–1208.

Caplin, A. and B. Nalebuff (1991). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica 59*, 25–59.

Chamberlain, G. (1985). Heterogeneity, omitted variable bias, and duration dependence. In J. J. Heckman and B. S. Singer (Eds.), *Longitudinal Analysis of Labor Market Data*. New York: Cambridge University Press.

Chen, S. (2002). Rank estimation of transformation models. *Econometrica 70*, 1683–1697.

Chen, X. and Y. Hu (2006). Identification and inference of nonlinear models using two samples with aribrary measurement errors. Discussion Paper 1590, Cowles Foundation.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2009). Identification and estimation of marginal effects in nonlinear panel models. Unpublished manuscript, Massachusetts Institute of Technology.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*, 1243–1284.

Ciliberto, F. and E. Tamer (2009). Market structure and multiple equilibria in the airline industry. *Econometrica 77*, 1791–1828.

Darolles, S., J.-P. Florens, and E. Renault (2007). Nonparametric instrumental regression. Unpublished manuscript, University of North Carolina, Chapel Hill.

Doraszelski, U. and K. L. Judd (2008). Avoiding the curse of dimensionality in dynamic stochastic games. Unpublished manuscript, Harvard University.

Doraszelski, U. and S. Markovich (2007). Advertising dynamics and competitive advantage. *The RAND Journal of Economics 38*, 557–592.

Doraszelski, U. and A. Pakes (2007). A framework for applied dynamic analysis in IO. In M. Armstrong and R. H. Porter (Eds.), *Handbook of Industrial Organization*, Volume 3, Chapter 30, pp. 1887–1966. North Holland.

Doraszelski, U. and M. Satterthwaite (2010). Computable Markov-perfect industry dynamics. Unpublished manuscript, Harvard University.

Eckstein, Z. and K. I. Wolpin (1989). The specification and estimation of dynamic stochastic discrete choice models: A survey. *The Journal of Human Resources 24*, 562–598.

Eckstein, Z. and K. I. Wolpin (1999). Why youth drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica 67*, 1295–1339.

Elfving, G. (1937). Zur theorie der Markoffschen ketten. *Acta Societatis Scientiarum Fennicæ Nova Series A 2*(8), 1–17.

Ericson, R. and A. Pakes (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economics and Statistics 62*, 53–82.

Gotz, G. A. and J. J. McCall (1980). Estimation in sequential decision-making models: A methodological note. *Economics Letters 6*, 131–136.

Gowrisankaran, G. (1999a). A dynamic model of endogenous horizontal mergers. *The RAND Journal of Economics 30*, 56–83.

Gowrisankaran, G. (1999b). Efficient representation of state spaces for some dynamic models. *Journal of Economic Dynamics and Control 23*, 1077–1098.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics 35*, 303–316.

Harsanyi, J. (1973). Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points. *International Journal of Game Theory 2*, 1–23.

Heckman, J. J. and S. Navarro (2005). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Honoré, B. E. and E. Kyriazidou (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica 68*, 839–874.

Honoré, B. E. and A. Lleras-Muney (2006). Bounds in competing risks models and the war on cancer. *Econometrica 74*, 1675–1698.

Honoré, B. E. and E. Tamer (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica 74*, 611–629.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica 60*, 505–531.

Horowitz, J. L. (1998). *Semiparametric Methods in Econometrics*. New York: Springer.

Horowitz, J. L. and C. F. Manski (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association 95*, 77–88.

Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies 60*, 497–529.

Hotz, V. J., R. A. Miller, S. Sanders, and J. Smith (1994). A simulation estimator for dynamic models of discrete choice. *Review of Economic Studies 61*, 265–289.

Hu, Y. and S. M. Schennach (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica 76*, 195–216.

Hu, Y. and M. Shum (2008a). Identifying dynamic games with serially-correlated unobservables. t-wp, Johns Hopkins University.

Hu, Y. and M. Shum (2008b). Nonparametric identification of dynamic models with unobserved state variables. Unpublished manuscript, Johns Hopkins University.

Imbens, G. W. and C. F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica 72*, 1845–1857.

Jofre-Bonet, M. and M. Pesendorfer (2003). Estimation of a dynamic auction game. *Econometrica 71*, 1443–1489.

Johansen, S. (1974). Some results on the imbedding problem for finite Markov chains. *Journal of the London Mathematical Society s2-8*, 345–351.

Karlin, S. and H. M. Taylor (1975). *A First Course in Stochastic Processes* (Second ed.). Academic Press.

Keane, M. and K. I. Wolpin (1997). The career decisions of young men. *Journal of Political Economy 105*, 473–522.

Khan, S. and E. Tamer (2009). Inference on endogenously censored regression models using conditional moment inequalities. *Journal of Econometrics 152*, 104–119.

Kim, J. and D. Pollard (1990). Cube root asymptotics. *Annals of Statistics 18*, 191–219.

Kim, K. (2009). Set estimation and inference with models characterized by conditional moment inequalities. Unpublished manuscript, University of Minnesota.

Kingman, J. F. C. (1962). The imbedding problem for finite Markov chains. *Probability Theory and Related Fields 1*, 14–24.

Komarova, T. (2008). Binary choice models with discrete regressors: Identification and misspecification. Unpublished manuscript, London School of Economics.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

Magnac, T. and E. Maurin (2008). Partial identification in monotone binary models: Discrete regressors and interval data. *Review of Economic Studies 75*, 835–864.

Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete decision processes. *Econometrica 70*, 801–816.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*, 205–228.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*, 313–333.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica 55*, 357–362.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag.

Manski, C. F. and E. Tamer (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica 70*, 519–546.

McFadden, D. (1981). Econometric models of probabilistic choice. In C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data*. Cambridge, MA: MIT Press.

McFadden, D. L. (1974). Conditional logit analysis of qualitative choice analysis. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. Academic Press.

McFadden, D. L. (1984). Econometric analysis of qualitative response models. In Z. Griliches and M. Intrilligator (Eds.), *Handbook of Econometrics*, Volume 2, Chapter 24, pp. 1396–1456. Amsterdam: Elsevier.

Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political Economy 92*(6), 1086–1120.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*, 1565–1578.

Pakes, A. (1986). Patents as options: Some estimates of the value of holding european patent stocks. *Econometrica 54*, 755–784.

Pakes, A., G. Gowrisankaran, and P. McGuire (1993). Implementing the Pakes-McGuire algorithm for computing Markov perfect equilibria in Gauss. Unpublished manuscript, Harvard University.

Pakes, A. and P. McGuire (1994). Computing Markov-perfect Nash equilibria: Numerical implications of a dynamic differentiated product model. *The RAND Journal of Economics 25*, 555–589.

Pakes, A. and P. McGuire (2001). Stochastic algorithms, symmetric Markov perfect equilibrium, and the 'curse' of dimensionality. *Econometrica 69*, 1261–1281.

Pakes, A., M. Ostrovsky, and S. Berry (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *The RAND Journal of Economics 38*, 373–399.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*, 1027–1057.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2006). Moment inequalities and their application. Unpublished manuscript, Harvard University.

Pesendorfer, M. and P. Schmidt-Dengler (2007). Asymptotic least squares estimators for dynamic games. *Review of Economic Studies 75*, 901–928.

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*. Springer.

Pollard, D. (1989). Asymptotics via empirical processes. *Statistical Science 4*, 341–366.

Puterman, M. L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (Second ed.). Wiley-Interscience.

Ridder, G. (1990). The non-parametric identification of generalized accelerated failure-time models. *Review of Economic Studies 57*, 167–182.

Romano, J. P. and A. M. Shaikh (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference 138*, 2786–2807.

Romano, J. P. and A. M. Shaikh (2009). Inference for the identified set in partially identified econometric models. Unpublished manuscript, University of Chicago.

Rosen, A. (2009). Set identification via quantile restrictions in short panels. Working Paper CWP26/09, CEMMAP.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica 55*, 999–1013.

Rust, J. (1994a). Estimation of dynamic structural models, problems and prospects: Discrete decision processes. In C. Sims (Ed.), *Advances in Econometrics: Sixth World Congress*, Volume 2, Chapter 4, pp. 119–170. Cambridge University Press.

Rust, J. (1994b). Structural estimation of Markov decision processes. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Amsterdam. North Holland.

Serfling, R. J. (1980). *Approximation theorems of Mathematical Statistics*. Wiley.

Shorack, G. R. (2000). *Probability for Statisticians*. Springer.

Sidje, R. B. (1998). Expokit: A software package for computing matrix exponentials. *ACM Transactions on Mathematical Software 24*, 130–156.

Singer, B. and S. Spilerman (1976). The representation of social processes by markov models. *The American Journal of Sociology 82*(1), 1–54.

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica 77*, 1299–1315.

Su, C.-L. and K. L. Judd (2008). Constrained optimization approaches to estimation of structural models. Unpublished manuscript, University of Chicago.

Tamer, E. (2003, January). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies 70*, 147–165.

Tamer, E. (2009). Partial identification in econometrics. Unpublished manuscript, Northwestern University.

Timmins, C. (2002). Measuring the dynamic efficiency costs of regulators' preferences: Municipal water utilities in the arid west. *Econometrica 70*, 603–629.

Weintraub, G. Y., C. L. Benkard, and B. Van Roy (2008). Markov perfect industry dynamics with many firms. *Econometrica 76*, 1375–1411.

Wolpin, K. I. (1984). An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political Economy 92*, 852–874.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal (1997). Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software 23*, 550–560.

Biography

Jason Ryan Blevins was born in Boone, North Carolina on December 15, 1981. He attended North Carolina State University from 2000 to 2004 where he earned a B.S. in applied mathematics, *summa cum laude*, with minors in computer science and economics. He participated in a National Science Foundation (NSF) funded Research Experiences for Undergraduates (REU) program hosted by the Department of Mathematical and Computer Sciences at the Colorado School of Mines in the summer of 2002 as well as the Budapest Semesters in Mathematics program in the fall of 2002. He was elected to Phi Beta Kappa, Zeta chapter of North Carolina, in 2003.

Jason later attended Duke University where he earned an M.A. in economics in 2006 and a Ph.D. in economics in 2010. He is a 2006 fellow of the Institute on Computational Economics, hosted by the University of Chicago and Argonne National Laboratory. Jason married Kelly Lauren McNeil on June 21, 2008 at Duke Chapel. He will begin an appointment as Assistant Professor of Economics at The Ohio State University in the autumn quarter of 2010.