P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

TWO

Recent Developments in Empirical IO: Dynamic Demand and Dynamic Games

Victor Aguirregabiria and Aviv Nevo

1.0 Introduction

Important aspects of competition in oligopoly markets are dynamic. Demand can be dynamic if products are storable or durable, or if utility from consumption is linked intertemporally. On the supply side, dynamics can be present as well. For example, investment and production decisions have dynamic implications if there is "learning-by-doing" or if there are sunk costs. Identifying the factors governing the dynamics is key to understanding competition and the evolution of market structure and for the evaluation of public policy. Advances in econometric methods and modeling techniques and the increased availability of data have led to a large body of empirical papers that study the dynamics of demand and competition in oligopoly markets.

A key lesson learned early by most researchers is the complexity and challenges of modeling and estimating dynamic structural models. The complexity and "curse of dimensionality" are present even in relatively simple models but are especially problematic in oligopoly markets in which firms produce differentiated products or have heterogeneous costs. These sources of heterogeneity typically imply that the dimension of these models, and the computational cost of solving and estimating them, increases exponentially with the number of products and the number of firms. As a result, much of the recent work in structural econometrics in IO focuses on finding ways to make dynamic problems more tractable in terms of computation and careful modeling to reduce the state space while properly accounting for rich heterogeneity, dynamics, and strategic interactions.

We thank Manuel Arellano, Gautam Gowrisankaran, Igal Hendel, Pedro Mira, Martin Pesendorfer, Marc Rysman, Matt Shum, and Junichi Suzuki for thoughtful comments.

We cannot provide here a complete survey of the large body of recent work. Instead, we focus on three main challenges discussed in the literature that we consider particularly important for applied work. Two challenges are common to models of dynamic demand and dynamic games: (1) the dimensionality problem and ways to reduce the state space and the computational burden; and (2) the treatment of heterogeneity in firm, consumer, and market characteristics. The empirical application of dynamic games also must deal with (3) the challenge of multiplicity of equilibria in estimation and prediction.

A key focus in dynamic structural models is on ways to reduce the state space. A problem that is tractable in an example used to illustrate a method might quickly become intractable when applied to answering questions in real markets. For instance, even static models of demand for differentiated products face a significant dimensionality problem due to the large number of products. The dimensionality problem becomes a difficult issue when we try to extend the methods of Rust (1987) originally applied to a durable-good decision – the replacement of a bus engine – to demand for durable differentiated products. The dimension of the state space increases (exponentially) with the number of products. A concept that has proved useful in reducing the state space in the modeling of both dynamic demand and dynamic games is the inclusive value (McFadden 1974). We show different examples of how the inclusive value is used to reduce the state space and the assumptions needed to justify these approaches. We also show how we can, by the correct conditioning, estimate many of the model parameters without the need to solve a dynamic-programming problem.

We cannot overemphasize the importance of allowing for heterogeneity across consumers, firms, products and markets to explain microdata. Not accounting for this heterogeneity can generate significant biases in parameter estimates and in our understanding of competition among firms. For instance, in the estimation of dynamic games of oligopoly competition, ignoring unobserved market heterogeneity when present can lead to serious biases in our estimates of the degree of strategic interaction among firms. Unfortunately, some of the methods used to reduce the state space and ease the computational burden limit the ability to estimate observed and unobserved heterogeneity. At times, this creates a trade-off between estimation methods that are faster – and potentially allows for the estimation of models that are richer in observed variables and have more flexible parametric forms – and methods that can handle only simpler models but can allow for richer unobserved heterogeneity. It is interesting that the two literatures we survey take somewhat different approaches in handling this trade-off. Using our examples, we highlight the trade-offs and the ways that they are addressed.

Multiple equilibria is a prevalent feature in dynamic games. We focus on several of the practical problems that this multiplicity introduces in estimation and prediction. A key way that the literature has addressed multiple equilibria is to assume that a unique game is being played in the data. One potential issue is whether this equilibrium is stable, in the sense that we define herein. As we show, it turns out this has important implications for the performance of many common methods. In addition, we review recent methods for the estimation of dynamic games that can deal with multiple equilibria and unobserved heterogeneity. We also examine the implementation of counterfactual experiments in models with multiple equilibria.

The chapter is organized in two parts. Section 2.0 discusses dynamic models of demand of differentiated products, and Section 3.0 addresses dynamic games of oligopoly competition.

2.0 Dynamic Demand for Differentiated Products

2.1 Overview

In the last 30 or so years, demand estimation has been a key part of studies in empirical industrial organization (IO). The key idea is to estimate demand and use the estimates to recover unobserved costs by inverting a pricing decision. Once cost has been recovered, the estimated demand and cost can be used to study the form of competition, understand firm behavior, generate counterfactuals (e.g., the likely effect of a merger), and quantify welfare gains (e.g., from the introduction of new products).[1]

Much of the literature relies on static demand models for this type of exercise. However, in many markets, demand is dynamic in the sense that (1) consumers' current decisions affect their future utility (equivalently, current utility depends on past decisions); or (2) consumers' current decisions depend on expectations about the evolution of future states. The exact effect of dynamics differs depending on the circumstances and can be generated for different reasons. The literature focuses on several cases including storable products, durable products, habit formation, switching costs, and learning. Because our goal is to demonstrate key challenges faced by

[1] See, for example, Bresnahan (1981, 1987); Porter (1983); Hausman, Leonard, and Zona (1994); Berry, Levinsohn, and Pakes (1995); Goldberg (1995); and Nevo (2001).

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

empirical researchers and not to provide a complete survey, we focus on the first two cases: storable and durable products.[2]

In the case of storable products, if storage costs are not too large and the current price is low relative to future prices (i.e., the product is on sale), there is an incentive for consumers to store the product and consume it in the future. Dynamics arise because consumers' past purchases and consumption decisions impact their current inventory and, therefore, may impact both the costs and benefits of purchasing today. Furthermore, consumers' expectations about future prices and the availability of products also impact the perceived trade-offs of buying today versus in the future.

In the case of durable products, dynamics arise due to similar trade-offs. The existence of transaction costs in the resale market of durable goods (e.g., because of adverse selection; Akerlof 1970) implies that a consumer's decision today of whether to buy a durable good and which product to buy is costly to change in the future and, for that reason, it will impact her future utility. Therefore, when a consumer makes a purchase, she is influenced by her current holdings of the good and by her expectations about future prices and attributes of available products. For instance, a consumer who currently owns a one-year-old automobile is likely to make a different purchasing decision than an identical consumer who owns a 10-year-old automobile. The dynamics are most important in industries in which prices and available products are changing rapidly over time, such as many consumer electronic goods, or in which there are policies that have dynamic effects, such as scrapping subsidies in the automobile industry.

Ignoring ~~the~~ dynamics and using the data to estimate a static demand model generates biased and inconsistent estimates. In addition to the econometric bias, it is important to realize that in many cases, static estimation does not recover desired features and thus fails to address many interesting questions. For example, in many applications, it is important to separate between a short-run price elasticity in response to a temporary price change and a long-run elasticity in response to a permanent price change. In general, due to econometric bias, static estimation does not recover short-run

[2] As noted in the introduction, our goal is not to provide a complete survey, so we will not offer a comprehnsive discussion of this wide litterature. For examples, in addition to the papers discussed herein, see Hartmann (2006); Carranza (2006); Esteban and Shum (2007); Nair (2007); Rossi (2007); Shcherbakov (2008); Sweeting (2008); Lou, Prentice, and Yin (2008); Osborne (2009); Perrone (2009); ~~R.~~ Lee (2011); and Schiraldi (2011), among many others.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

responses but, even if it does, in some special cases, it cannot recover separately the long-run response.

Computing price responses obviously is important to fields such as IO and marketing, but the possible uses of the models discussed herein are much wider and include many fields in economics; following are a few examples. Recently, macroeconomists looked at microlevel price data to study price rigidities.[3] A central issue in this literature is how to treat temporary price reductions, or "sales." A key to understanding sales and why they exist is to understand consumer response. Similarly, a key issue in trade is the pass-through of exchange rates. Here again, separating between short- and long-run price responses is critical. In another example, obesity and unhealthy eating habits are plaguing many countries and have led to suggestions of taxing unhealthy high-fat and high-calorie foods. To evaluate the effectiveness of these policies, it is crucial to estimate the heterogeneity in price response: If a tax were to be imposed, who responds and by how much? Furthermore, it is probably important to estimate the degree of habit persistence in the consumption of these unhealthy food products. Adoption of energy-efficient cars and appliances is an important aspect of environmental economics. To the extent that demand is dynamic, as discussed previously, modeling the dynamics is crucial. Modeling the dynamics of durable-good purchases has important implications for evaluating scrapping policies and computing price indices.

The dynamic factors impacting demand have long been recognized[4] and, indeed, many different models to capture these dynamics are offered in the literature ranging from models in which the dynamics decisions are modeled explicitly to modeling approaches in which the dynamics are handled by including lags and leads of variables (e.g., prices). The IO literature mostly takes the approach of explicit modeling, often referred to as a "structural" approach.

To implement these approaches in markets with differentiated products and address important applied questions, researchers must deal with several issues, including large state spaces, unobserved (endogenous) state variables, and heterogeneity. In this section, we survey the approaches taken to address these issues.

[3] See, for example, Kehoe and Midrigan (2008); Eichenbaum, Jaimovich, and Rebelo (2008); and Nakamura and Steinsson (2008).

[4] For an early contribution, see General Motors Corproration (1939), a volume developed from papers sponsored by General Motors and presented in a joint session of the American Statistical Association and the Econometric Society in 1938.

2.1.1 Background: Static Demand for Differentiated Products

Key lessons to learn from static demand estimation is the importance of allowing for heterogeneity and the difficulty of dealing with the dimensionality of the problem while still allowing for flexible functional forms. We consider a classical (static) demand system for J products, $\mathbf{q} = D(\mathbf{p}; \mathbf{r})$, where $\mathbf{q}$ is a J-dimensional vector of quantities demanded, $\mathbf{p}$ is a J-dimensional vector of prices, and $\mathbf{r}$ is a vector of exogenous variables. A key problem in estimating this system is the dimensionality – due to the large number of products, the number of parameters is too large to estimate. Several solutions are offered in the IO literature,[5] but the most common solution is to rely on a discrete-choice model (McFadden 1974; Berry, Levinsohn, and Pakes 1995).

The "workhorse" discrete-choice model used in IO has a consumer i choosing option j from one of $J + 1$ options (J brands and a no-purchase option). The (conditional indirect) utility that the consumer obtains from option j at time t is given by:

$$u_{ijt} = a_{jt}\,\beta_i - \alpha_i\; p_{jt} + \xi_{jt} + \varepsilon_{ijt} \tag{1}$$

where p_{jt} is the price of option j at time t, a_{jt} is a $1 \times K$ vector of observable attributes of product j, ξ_{jt} is an unobserved (by the econometrician) product characteristic, ε_{ijt} is a stochastic term. α_i represents the consumer's marginal utility of income, and β_i is a $K \times 1$ vector of individual-specific marginal utilities associated to the attributes in the vector a_{jt}. In this model, a product is viewed as a bundle of characteristics and, therefore, the relevant dimension is the number of characteristics, K, and not the number of products. Flexible substitution patterns are achieved by allowing for consumer heterogeneity in the willingness to pay and in the valuation of characteristics.

The model can be estimated using consumer-level data. However, the wider availability of market-level data and the development of appropriate econometric techniques made estimation using market-level data the more popular choice. The estimates from aggregate-level data generally are considered more credible if the data come from many different markets with variation in the observed attributes of consumers, or if the aggregate data

[5] For example, a common approach in the trade and macro literature is to use the constant elasticity of substitution (CES) demand system, which is economical on parameters. This model, however, is not flexible enough to explain microlevel data. An alternative approach is to use the multilevel demand system developed by Hausman, Leonard, and Zona (1994). See Nevo (2011) for a discussion.

are supplemented with so-called micromoments, which are basically the purchasing patterns of different demographic groups.

2.1.2 Dynamic Demand: Key Ingredients

In building dynamic demand models, the IO literature continues to rely heavily on the discrete-choice model.

Reducing the Dimensionality. If demand is dynamic, the dimensionality problem is even worse. The basic idea of a discrete-choice model – to project the products onto a characteristics space – that essentially solved the problem in the static context is not sufficient in the dynamic context. For example, we consider the problem of a forward-looking consumer trying to form expectations about future price and characteristics of products. In principle, this consumer must form expectations about the future $K + 1$ attributes of all products, the number of which could be changing, using the information of the current and past values of these attributes for all the products. Even if we assume that variables follow a first-order Markov process and that the number of products is fixed, the size of the state space is $(K + 1) * J$.

A useful concept, used in our examples, is the *inclusive value.* McFadden (1974) defined the inclusive value (or *social surplus*) as the expected utility of a consumer, from several discrete options, prior to observing $(\varepsilon_{i0t}, \dots \varepsilon_{ijt})$, and knowing that the choice will be made to maximize utility. When the idiosyncratic shocks ε_{ijt} are distributed i.i.d. extreme value, the inclusive value from a subset $A \subseteq \{1, 2, \dots, J\}$ of the choice alternatives is defined as:

$$\omega_{it}^{A} = \ln\left(\sum_{j\in A} \exp\left\{a_{jt}\,\beta_i - \alpha_i\ p_{jt} + \xi_{jt}\right\}\right) \tag{2}$$

When $\beta_i = \beta$ and $\alpha_i = \alpha$, the inclusive value captures the average utility in the population, up to a constant, averaging over the individual draws of ε; hence, the term *social surplus.*

The inclusive value has a key role in reducing the state space. In forming expectations, the consumer must form expectations about the future inclusive value or, in some cases, a low number of inclusive values for subsets of products, rather than expectations about the realizations of all attributes of all products. To reduce significantly the state space, this property is coupled

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

with a behavioral assumption on the information that consumers use to form these expectations.

Heterogeneity. As in static models, allowing for heterogeneity is key for explaining the data and retaining flexible demand systems. In some cases, however, a degree of unobserved heterogeneity must be sacrificed to deal with the dimensionality problem. As we show herein, the trade-off in some cases is between a richer model that includes more observed heterogeneity and a model that relies on unobserved heterogeneity.

Data. Similar to the static model, the dynamic model can be estimated using consumer- or market-level data. The advantages of consumer-level data seem more obvious in the dynamic setting: Consumer-level data allow us to see how individual consumers behave over time. However, this is exactly the reason why consumer-level datasets are difficult to collect, especially for products (e.g., some durable) that are purchased infrequently. For this reason, a number of applications rely on aggregate data. We informally discuss identification and estimation with market-level data.

2.2 Storable Products

Many of the products purchased by consumers are storable so that consumers can buy them for future consumption. A typical pricing pattern in these markets involves short-lived price reductions with a return to the regular price. This pattern of prices generates an incentive for consumers to store the product when the price is low. Boizot, Robin, and Visser (2001) and Pesendorfer (2002) were among the first to study the effects of temporary price reductions and storability in economics.[6]

2.2.1 Evidence

There is ample evidence that once faced with temporary price reductions, consumers store for future consumption. For example, using data for ketchup, Pesendorfer (2002) found that holding the current price constant, aggregate quantity sold depends on duration from previous sales. Hendel and Nevo (2006a, 2010) found similar evidence for other products.

[6] An earlier marketing literature examined the same issues, but the treatment there was generally not consistent with optimal dynamic behavior. See, for example, Shoemaker (1979); Blattberg, Eppen, and Lieberman (1981); Gupta (1988); and Chiang (1991).

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

Additional evidence for the existence of demand accumulation was provided by Hendel and Nevo (2006a), who used household-level data to document patterns that are consistent with consumer stockpiling behavior. For example, they showed that the household's propensity to purchase on sales is correlated with proxies of storage costs and that households in areas where houses are larger (with cheaper storage) buy more on sale. They also showed that when purchasing on sale, duration to the next purchase is longer. This is true both within households – for a given household when buying on sale, the duration is longer – and across households – households that purchase more on sale also purchase less frequently. Finally, proxies for inventory are negatively correlated with quantity purchased and the probability of purchasing.

2.2.2 Implications

Given the evidence on demand accumulation, it is natural to ask what are the implications. The primary implication is for demand estimation, which is an input for addressing important economic questions discussed herein.

Once we recognize that consumers can store the product, we must separate between the short-run response to a temporary price change and the long-run response to either a temporary or a permanent price change. For most economic applications, we are concerned about long-run changes. If price changes in the data are permanent, then static estimation yields consistent estimates of the long-run demand responses. Indeed, one way to estimate long-run responses is to use only permanent price changes and ignore – to the extent possible – the temporary prices changes. In many datasets, the temporary price changes constitute most or even all of the variation in prices. Therefore, dropping these price changes means a significant loss of efficiency, possibly even completely wiping out any price variation.

Conversely, if price changes in the data are temporary, then static demand estimates overestimate own-price effects. The (large) demand response to a sale is attributed to an increase in consumption (which in a static model equals purchase) and not to an increase in storage. The decline in purchases after a sale coincides with an increase in price and is misattributed as a decline in consumption. At the same time, static estimation underestimates cross-price effects. During a sale, the quantity of competing products sold decreases, but static estimation misses an additional effect: the decrease in the quantity sold in the future. Intuitively, when a competing product was on sale in the past, consumers purchased to consume today and, therefore,

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

the relevant or "effective" cross price is not the current cross price. The current price is (weakly) higher. Furthermore, when a (cross) product is on sale, the current (cross) price is more likely to be the effective price. Both of these effects bias the estimated cross-price effect toward zero.

2.2.3 A Model of Consumer Stockpiling

Hendel and Nevo (2006b) proposed the following model of consumer stockpiling, which we use to demonstrate some key issues faced by applied researchers.

The starting point is similar to the discrete-choice model discussed in Section 2.1.1. The consumer can purchase one of $J+1$ brands, which come in different sizes and which we index by $x \in \{1, 2, \ldots, X\}$. We let d_{jxt} be equal to 1 if the consumer purchases brand j of size x at time t, and 0 otherwise. Because the choice is discrete, stockpiling is achieved by buying larger sizes and adding to existing inventory rather than by buying multiple units on any given shopping trip. This assumption seems reasonable for the data used by Hendel and Nevo (2006b), in which there were few purchases of multiple units. In other contexts, this might not be reasonable and we would need to model the choice of multiple units.

The consumer also must decide how much to consume each period.[7] The per-period utility that consumer i obtains from consuming in t is:

$$u_i(\mathbf{c}_t, \boldsymbol{\nu}_t) + \alpha_i m_t \tag{3}$$

where $\mathbf{c}_t$ is a J-dimensional vector of the quantities consumed of each brand, $\boldsymbol{\nu}_t$ is a J-dimensional vector of shocks to utility that change the marginal utility from consumption, and m_t is the utility from the outside good. In addition to utility from consumption, the one-period utility has two other components. We assume that the consumer pays a cost $C_i(\mathbf{i}_t)$ for holding inventories $\mathbf{i}_t$, where $\mathbf{i}_t$ is a vector of inventories by brand. There also is an instantaneous utility associated with preference for the purchased brand.

[7] An alternative of assuming that consumption is constant over time but varying across households seems attractive, especially for the type of products usually modeled. A slightly more general model than constant consumption allows for random shocks, $\mathbf{v}_t$, that determine consumption. Both of these models are nested within our model and, in principle, can be tested. The results in Hendel and Nevo (2006b) suggested that consumption is mostly constant but, when inventory runs low, consumers reduce consumption. This behavior is required to explain long periods of no purchase followed by periods of frequent purchases observed in the data. Indeed, it is this variation in interpurchase time that identifies the utility from consumption.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

At period $t = 1$, the purchase and consumption decisions, $\{\mathbf{c}, j, x\}$, are made to maximize:

$$\sum_{t=1}^{\infty} \delta^{t-1} \mathbb{E}[u_i(\mathbf{c}_t, \boldsymbol{\nu}_t) - C_i(\mathbf{i}_t) + a_{jxt}\beta_i - \alpha_i p_{jxt} + \xi_{jxt} + \varepsilon_{ijxt} \mid \mathbf{s}_1]$$

$$s.t. \quad 0 \leq \mathbf{i}_t, \quad 0 \leq \mathbf{c}_t \quad \sum\nolimits_{j,x} d_{jxt} = 1, \tag{4}$$

$$i_{j,t+1} = i_{j,t} + \sum\nolimits_{x} d_{jxt} x_t - c_{j,t} \quad j = 1, \ldots, J$$

where $\mathbf{s}_t$ is the information set at time t; δ is the discount factor; p_{jxt} is the price of purchasing quantity x of brand j; ξ_{jxt} is an unobserved (to the researcher) brand-specific quality; a_{jxt} are observed product attributes; and ε_{ijxt} is a random shock. We allow ξ_{jxt} to vary by brand to capture differentiation across products and across sizes, for reasons we discuss herein. In principle, the brand preference also can vary across consumers.

The expectation $\mathbb{E}(.)$ is taken with respect to the uncertainty regarding future shocks in the vectors $\boldsymbol{\nu}_t$ and $\boldsymbol{\varepsilon}_t$ as well as future prices (and other time-varying attributes). We assume that ε_{jxt} is i.i.d. extreme value and that $\boldsymbol{\nu}_t$ is i.i.d. over time and across consumers with a known parametric distribution. Prices and observed characteristics evolve according to a first-order Markov process.

Some aspects of the specification of this consumer-decision problem warrant further explanation. First, we assume no physical depreciation of the product, although this assumption is easy to relax if needed. Second, we assume that a decision is made each period with perfect knowledge of current prices. Implicitly, we are assuming that the consumer visits the store every period. This assumption also helps us in the specification of consumer expectations regarding future prices. If consumers do not visit the store every period, we must model the process by which they arrive at the store to determine the next set of prices that they should expect.

At the moment, even with the simplifying assumptions already made, the vector of state variables is quite large and includes a J-dimensional vector of inventory holdings by brand, $\mathbf{i}_t$; a $(K+1) * J * X$-dimensional vector of prices and characteristics, $\mathbf{p}_t$[8]; a J-dimensional vector of consumption shocks, $\boldsymbol{\nu}_t$; and a $J * X$-dimensional vector of i.i.d. extreme-value shocks, $\boldsymbol{\varepsilon}_t$. The vector of state variables at period t is $\mathbf{s}_t = (\mathbf{i}_t, \mathbf{p}_t, \boldsymbol{\nu}_t, \boldsymbol{\varepsilon}_t)$. Without

[8] To keep notation simple, we use $\mathbf{p}_t$ to denote the observed variables at time t. These variables include prices and other observed variables.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

a first-order Markov assumption, the state space would be even larger and would include additional lags of prices and characteristics.

We let $V_i(s_t)$ be the value function of consumer i. As usual in a dynamic-programming problem, this value function can be obtained as the unique solution of a Bellman equation:

$$V_i(\mathbf{s}_t) = \max_{\{\mathbf{c}, j, x\}} \Big\{ u_i(\mathbf{c}, \boldsymbol{\nu}_t) - C_i(\mathbf{i}_t) + a_{jxt}\beta_i - \alpha_i p_{jxt} + \xi_{jxt} + \varepsilon_{ijxt} + \delta \int V_i(\mathbf{s}_{t+1}) dF_s(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{c}, j, x) \Big\} \tag{5}$$

where F_s represents the transition probability of the vector of state variables. Given that the state variables $(\boldsymbol{\nu}_t, \boldsymbol{\varepsilon}_t)$ are distributed independently over time, it is convenient to reduce the dimensionality of this dynamic-programming problem by using a value function that is integrated over these i.i.d. random variables. The *integrated value function*, also called the *ex-ante value function*, is defined as $EV_i(\mathbf{i}_t, \mathbf{p}_t) \equiv \int V_i(\mathbf{s}_t) dF_\varepsilon(\boldsymbol{\varepsilon}_t) dF_{\mathbf{v}}(\mathbf{v})$, where F_ε and $F_{\mathbf{v}}$ represent the Cumulative Distribution Functions (CDFs) of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\nu}_t$, respectively. The value function EV_i is the unique solution of the integrated Bellman equation. Given the distributional assumptions on the shocks $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\nu}_t$, the integrated Bellman equation is:

$$EV_i(\mathbf{i}_t, \mathbf{p}_t) = \max_{\mathbf{c}, x} \int \ln \left(\sum_j \exp \left\{ \begin{array}{l} u_i(\mathbf{c}, \boldsymbol{\nu}_t) - C_i(\mathbf{i}_t) + a_{jxt}\beta_i - \alpha_i p_{jxt} + \xi_{jxt} \\ + \delta\, \mathbb{E}\left[EV_i(\mathbf{i}_{t+1}, \mathbf{p}_{t+1}) \mid \mathbf{i}_t, \mathbf{p}_t, \mathbf{c}, j, x \right] \end{array} \right\} \right) dF(\boldsymbol{\nu}_t) \tag{6}$$

The main computational cost is to compute the functions EV_i. We now explore ways to reduce this cost.

2.2.4 Reducing the Dimension of the State Space

As it stands, the state space is quite large and not workable for anything except a small number of products J. To reduce the state space, several additional assumptions are needed.

Inventories and Consumption. We first explore ways to reduce the dimension of inventories needed to be tracked. One possible assumption is to assume that products are perfect substitutes in consumption and storage.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

Assumption A1: $U_i(\mathbf{c}_t, \mathbf{v}_t) = U_i(c_t, \nu_t)$ and $C_i(\mathbf{i}_t) = C_i(i_t)$, where $c_t = \mathbf{1}'\mathbf{c}_t$, $\nu_t = \mathbf{1}'\boldsymbol{\nu}_t$, $i_t = \mathbf{1}'\mathbf{i}_t$, and $\mathbf{1}$ is a vector of 1's.

Under this assumption, the inventory and the consumption shocks reduce to a scalar: We only need to keep track of a single inventory and a single consumption shock. Formally, now:

$$E\,V_i(\mathbf{i}_t, \mathbf{p}_t) = E\,V_i(i_t, \mathbf{p}_t) \tag{7}$$

This assumption not only reduces the state space but, as shown herein, it also allows us to modify the dynamic-programming problem, which can aid significantly in estimation of the model.

Taken literally, this assumption implies that there is no differentiation in consumption: The product is homogeneous in use. We note that through ξ_{jxt} and ε_{ijxt}, we allow differentiation in purchase, as is standard in the IO literature. Indeed, it is well known that this differentiation is needed to explain purchasing behavior. This seemingly creates a tension in the model: Products are differentiated at purchase but not in consumption. Before explaining how this tension is resolved, we note that the tension is not only in the model but potentially in reality as well. Many products seem to be highly differentiated at the time of purchase, but it is difficult to imagine that they are differentiated in consumption. For example, households tend to be extremely loyal to the laundry-detergent brand they purchase – a typical household buys only two or three brands of detergent over a significant horizon – yet, it is difficult to imagine that the usage and consumption are very different for different brands. One way to think of the model is to assume that there is a brand-specific utility in consumption. As long as the utility in this component is linear and we can ignore discounting, to a first order, then the brand-specific utility in consumption is captured by ξ_{jxt}.[9] This is the reason we want to let ξ_{jxt} vary by size; indeed, this suggests that $\xi_{jxt} = \xi_{jt} * x$.

Assumption A1 implies that the optimal consumption does not depend on which brand is purchased. Formally, we let $c^*_k(\mathbf{s}_t; x, k)$ be the optimal consumption of brand k conditional on state $\mathbf{s}_t$ and on purchase of size x of that brand. Lemma 1 in the appendix of Hendel and Nevo (2006b) showed that $c^*_k(\mathbf{s}_t; x, k) = c^*_j(\mathbf{s}_t; x, j) = c^*(\mathbf{s}_t; x)$. In words, the optimal consumption does not depend on the brand purchased, only on the size.

[9] See Hendel and Nevo (2006a) for details of the argument. Erdem, Imai, and Keane (2003) offered an alternative model that allowed for two inventories. We can show that under these assumptions, their model is a private case of the one discussed herein.

This result implies that the (integrated) Bellman equation in (6) can be written as:

$$EV_i(i_t, \mathbf{p}_t) = \max_{c,x} \int \ln\left(\sum_x \exp\{u_i(c_t, v_t) - C_i(i_t) + \omega_{ixt} + \delta\, \mathbb{E}[EV_i(i_{t+1}, \mathbf{p}_{t+1}) \mid i_t, \mathbf{p}_t, c, x]\}\right) dF_v(v_t) \qquad (8)$$

where ω_{xit} is the inclusive value from all brands of size x, as defined by Equation (2); that is, $\omega_{ixt} = \ln(\sum_j \exp(a_{jxt}\beta_i - \alpha_i p_{jxt} + \xi_{jxt}))$. In words, the problem now can be seen as a choice between sizes, each with a utility given by the size-specific inclusive value and extreme-value shock. The dimension of the state space is still large and includes all prices because we need all of the prices to compute the evolution of the inclusive value. However, in combination with additional assumptions, the modified problem is easier to estimate.

Finally, we note that if needed, we could reduce the inventory to several types of products rather than to a scalar. For example, suppose we are studying the breakfast-cereal market; we could divide the brands into children's and adults' cereals such that within a group, products are perfect substitutes. In this case, we would need to keep two inventories – for adults' and children's cereal – still significantly less than the number of brands.

Prices. As noted, even with Assumption A1, the state space is still large and includes all prices. Therefore, for a realistic number of products, the state space is still too large to be manageable. To further reduce it, we make an additional assumption (see Assumption A4 in Hendel and Nevo 2006b). We let $\boldsymbol{\omega}_{it}$ be a vector of inclusive values for the different sizes.

Assumption A2: $F(\boldsymbol{\omega}_{i,t+1} \mid \mathbf{s}_t) = F(\boldsymbol{\omega}_{i,t+1} \mid \boldsymbol{\omega}_{it}(\mathbf{p}_t))$

In words, the vector $\boldsymbol{\omega}_{it}$ contains all of the relevant information in $\mathbf{s}_t$ to obtain the probability distribution of $\boldsymbol{\omega}_{i,t+1}$ conditional on $\mathbf{s}_t$. Instead of all of the prices and attributes, we need only a single index for each size. Two vectors of prices that yield the same vector of current inclusive values imply the same distribution of future inclusive values. This assumption is violated if individual prices have predictive power beyond the predictive power of $\boldsymbol{\omega}_{it}$. Therefore, if the inclusive values can be estimated outside of the dynamic-demand model, the assumption can be tested and somewhat relaxed by including additional statistics of prices in the state space. We note that $\boldsymbol{\omega}_{it}$ is consumer-specific: Different consumers value a given set of products differently; therefore, this assumption does not further restrict the distribution of heterogeneity.

Given Assumptions A1 and A2, we can show (Hendel and Nevo 2006b) that:

$$E\,V_i(i_t, \mathbf{p}_t) = E\,V_i(i_t, \boldsymbol{\omega}_{it}(\mathbf{p}_t)) \tag{9}$$

In words, the expected future value depends on only a lower dimensional statistic of the full state vector.

2.3 Estimation

In this section, we discuss the identification and estimation of the model. We assume that the researcher has access to consumer-level data. Such data are widely available from several data-collection companies; researchers in several countries recently have been able to gain access to such data for academic use.[10] The data include the history of shopping behavior of a consumer in a period of one to three years. The researcher knows whether a store was visited; if a store was visited, then which one; and which product (i.e., brand and size) was purchased and at what price. In many cases, the most difficult information to gather is the prices of products not purchased. From the viewpoint of the model, the unobserved key information is consumer inventory and consumption decisions.

The most straightforward way to estimate the model follows an algorithm similar to the one suggested by Rust (1987).[11] For a given set of parameters, we solve the dynamic-programming problem and obtain deterministic decision rules for purchases and consumption as a function of the state variables, including the unobserved random shocks. Assuming a distribution for these shocks, we derive a likelihood of observing each consumer's

[10] See, for example, the ERIM data available at http://research.chicagobooth.edu/marketing/databases/erim/index.aspx; the so-called Stanford Basket described in Bell and Lattin (1998); or the IRI Marketing Data Set discussed by Bronnenberg, Kruger, and Mela (2008). For more recent datasets, see, for example, Griffith, Leicester, Leibtag, and Nevo (2009) for a use of UK data; Einav, Leibtag, and Nevo (2010) for U.S. data; Bonnet and Dubois (2010) for French data; and Browning and Carro (2010) for Danish data.

[11] For computational reasons, methods based on conditional-choice probabilities (Hotz and Miller 1993; Hotz, Miller, Sanders, and Smith 1994; Aguirregabiria and Mira 2002) have become popular. Because the model includes unobserved endogenous time-varying state variables, these methods cannot be directly applied herein. However, the method of Arcidiacono and Miller (2011) potentially could be applied to the estimation of this model. See Section 3.0 for further discussion of these methods.

decision conditional on prices and inventory. We nest this computation of the likelihood into the search for the values of parameters that maximize the likelihood of the observed sample.

We face two hurdles in implementing the algorithm. First, consumption (a decision variable) and inventory (a state variable) are not observed. As shown herein, this can be solved by using the model to derive the optimal consumption and the implied inventory. The second problem is dimensionality of the state space. We discussed several assumptions that can be used to reduce the state space; nevertheless, the computational problem still is quite difficult. We show how the computation can be simplified significantly by dividing the estimation into estimation of the brand choice conditional on size, which does not require solving the dynamic problem, and estimating the choice of size, which requires solving a much simpler dynamic problem.

For the purpose of inference, because in some specifications we want to allow for household fixed effects, we usually must assume that the number of observations per household is very large.

As noted previously, it is quite common in the IO literature to estimate static demand models using market-level data. We are unaware of any paper that estimated the model we propose here using aggregate data. Hendel and Nevo (2010) estimated a simpler model using aggregate data.

2.3.1 Identification

Before discussing the details of estimation, we informally discuss identification. If inventory and consumption were observed, then identification using consumer-level data follows standard arguments (Rust 1994; Magnac and Thesmar 2002; Aguirregabiria 2010). However, we do not observe inventory or consumption, so the question is: Which features of the data allow us to identify functions of these variables?

The correlations and patterns described in Section 2.2.1 to suggest that dynamics are relevant are those that identify the dynamic model. In particular, the individual-level data provide the probability of purchase conditional on current prices and past purchases of the consumer (i.e., amounts purchased and duration from previous purchases). We suppose that we see that this probability is not a function of past behavior; we then would conclude that dynamics are not relevant and that consumers are purchasing for immediate consumption and not for inventory. Conversely, if we observe that the purchase probability is a function of past behavior and we assume that preferences are stationary, then we conclude that there is dynamic behavior.[12] Regarding the identification of storage costs, we consider the following example. We suppose that we observe two consumers who face the same price process and purchase the same amount in a given period. However, one consumer purchases more frequently than the other. This variation leads us to conclude that this consumer has higher storage costs. Therefore, the storage costs are identified from the average duration between purchases. The utility from consumption is identified from the variation in these duration times, holding constant the amount purchased. For example, a model of constant consumption cannot explain large variation in the duration times.

In some cases, the researcher may not have consumer-level data, only store- or market-level data. We are unsure if the model presented here is identified from aggregate data. Given this discussion, it might seem unlikely. However, a slightly simpler dynamic-demand model for storable goods can be identified from aggregate-store-level data, as long as the aggregation corresponds to the timing of price changes (i.e., if we have weekly data, we need the prices to be constant within the week). The variation in the data that identify the model is dependent on total quantity sold on the duration from the last sale (Hendel and Nevo 2010).

A key emphasis in static demand estimation is the potential endogeneity of prices. The concern is that prices, and sometimes other variables, are correlated with ξ_{jxt}. Conversely, some researchers who estimate dynamic demand have dismissed this concern stating that papers that focus on endogeneity have "missed the mark" (Erdem, Imai, and Keane 2003, p. 11) because it is unlikely that prices respond to aggregate shocks. Others claimed that endogeneity is not an issue when using household-level data to estimate demand, static or dynamic, "since the demand of the consumer does not usually affect market price" (Train 2003, p. 8).

In our view, whether we should be concerned with endogeneity depends on the data structure, what is included in the model, and the institutional knowledge of the industry. Broad statements such as "endogeneity is not an issue in dynamic models or when using consumer-level data" generally are not correct. For example, if prices are higher for higher quality products, which in the model is captured by higher ξ_{jxt}, then prices will be correlated

12 Serial correlation in v_t also might generate a dependence of the purchase probability on past behavior. However, positive serial correlation in v_t generates positive dependence between past and current purchases, whereas the stockpiling model generates negative dependence between past and current purchases.

with ξ_{jxt}. Of course, with enough data, we could control for the higher quality of some products (e.g., with a product fixed effect). ~~Of course,~~ if quality is time varying, then the fixed effect will not fully capture its variation. Repeated observations and consumer-level data allow us to control for factors for which we cannot control otherwise, but they do not imply that prices are automatically exogenous. Furthermore, in many dynamic models, due to computational constraints, we are limited in the number of controls. In the next section, we show how the estimation can be simplified to allow for richer controls.

In the following estimation, as in the literature on dynamic demand that uses household-level data, we address endogeneity by (1) assuming that $\xi_{jxt} = \xi_{jx}$ (i.e., does not vary over time) and control for it with fixed effects; and (2) using the simplified computational problem to control for time-varying variables such as advertising and promotions. In the discussion of durable goods, we review a Generalized Method of Moments (GMM) method using market-level data that closely follows the static demand estimation.

2.3.2 Estimation

The parameters of the model can be estimated via maximum likelihood following an algorithm similar to Rust (1987).[13] Because inventory, one of the state variables, is not observed, we must impute it as part of the estimation. This can be accomplished in the following way:

(i) Guess an initial inventory distribution and draw from it for each consumer.
(ii) For a given value of the parameters, solve the consumer problem and obtain the value and policy functions.
(iii) Using the draws of inventory from (i), the computed consumption policy from (ii), and observed purchases, obtain the sequence of inventory and compute the likelihood of the observed purchases.
(iv) Repeat Steps (ii) and (iii) to choose the parameters that maximize the likelihood of the observed data, possibly omitting some of the initial observations to let the inventory process settle.
(v) Update the initial guess of the distribution of inventory and repeat Steps (i)–(iv).

[13] This is subject to the caveat regarding the endogeneity of prices; see the discussion in the previous section.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

The likelihood in Step (iii) of observing a sequence of purchasing decisions, $(\mathbf{d}_1, \ldots, \mathbf{d}_T)$, as a function of the observed state variables, $(\mathbf{p}_1, \ldots, \mathbf{p}_T)$, and observed demographic variables, D_i, is:

$$\begin{aligned}
&\mathbb{P}(\mathbf{d}_1, \ldots, \mathbf{d}_T | \mathbf{p}_1, \ldots, \mathbf{p}_T, D_i) \\
&\quad = \int \prod_{t=1}^{T} \mathbb{P}\left(\mathbf{d}_t | \mathbf{p}_t, i_t(\mathbf{d}_{t-1}, \ldots, \mathbf{d}_1, v_{t-1}, \ldots, v_1, i_1), v_t, D_i\right) \\
&\quad \times dF(v_1, \ldots, v_T)dF(i_1) \qquad (10)
\end{aligned}$$

Inventory is a function of previously observed purchase (or no purchase) decisions, the previous consumption shocks, and the initial inventory. The exact functional form of the dependence of inventory on past consumption shocks depends on the consumption policy. The probability inside the integral represents the integration over the set of epsilons that induce $\mathbf{d}_t$ as the optimal choice. Using Assumptions A1 and A2 and the results from Section 2.2.4, this probability is given by:

$$\begin{aligned}
&\mathbb{P}(j, x | \mathbf{p}_t, i_t, v_t, D_i) \\
&= \frac{\exp\left\{a_{jxt}\beta_i - \alpha_i p_{jxt} + \xi_{jx} + \max_c \left(u_i(c, v_t) - C_i(i_t) + \delta\, \mathbb{E}\left[EV_i(i_{t+1}, \boldsymbol{\omega}_{t+1}) \mid i_t, \boldsymbol{\omega}_t, c, j, x\right]\right)\right\}}{\sum_{k,y} \exp\left\{a_{kyt}\beta_i - \alpha_i p_{kyt} + \xi_{ky} + \max_c \left(u_i(c, v_t) - C_i(i_t) + \delta\, \mathbb{E}\left[EV_i(i_{t+1}, \boldsymbol{\omega}_{t+1}) \mid i_t, \boldsymbol{\omega}_t, c, k, y\right]\right)\right\}} \\
&\qquad (11)
\end{aligned}$$

Hence, to compute the likelihood, we need to solve the dynamic problem only in the reduced state space.

Splitting the Likelihood. It is important to note that to this point, we use the stochastic structure of the problem but we do not restrict the distribution of consumer heterogeneity. In particular, we can allow for the taste coefficients, α_i and β_i, to vary with both observed and unobserved factors, and we can estimate their distribution using the joint likelihood of brand and size choice.

We now show that if we are willing to place restrictions on the unobserved heterogeneity, we can significantly simplify the computational problem.

As discussed previously, the optimal consumption is not brand-specific, so $M_i(\boldsymbol{\omega}_t, i_t, v_t, y) = \max_c(u_i(c, v_t) - C_i(i_t) + \delta\mathbb{E}[EV_i(i_{t+1}, \boldsymbol{\omega}_{t+1}) | i_t, \boldsymbol{\omega}_t, c, j, y])$ does not vary by brand j, conditional on a size y. Thus, we note that this probability can be written as:

$$
\begin{aligned}
&\mathbb{P}(j, x|\mathbf{p}_t, i_t, \nu_t, D_i)\\
&= \frac{\exp\{a_{jxt}\beta_i - \alpha_i p_{jxt} + \xi_{jx}\}}{\sum_k \exp\{a_{kxt}\beta_i - \alpha_i p_{kxt} + \xi_{kx}\}} \frac{\exp\{\omega_t^x + M(\boldsymbol{\omega}_t, i_t, \nu_t, x)\}}{\sum_y \exp\{\omega_t^y + M(\boldsymbol{\omega}_t, i_t, \nu_t, y)\}} \quad (12)\\
&= \mathbb{P}(j|x, \mathbf{p}_t, D_i)\, \mathbb{P}(x|\boldsymbol{\omega}_t, i_t, D_i)
\end{aligned}
$$

For this factorization to be useful in reducing the computational cost, we need a conditional independence assumption.

Assumption A3 (Conditional Independence of Heterogeneity): $F(\alpha_i, \beta_i| x_t, \mathbf{p}_t, D_i) = F(\alpha_i, \beta_i|\mathbf{p}_t, D_i)$, where x_t represents the chosen size.

This assumption is satisfied if heterogeneity is a function of only observed demographics, including possible "fixed effects." If this assumption holds, then:

$$
\begin{aligned}
\mathbb{P}(j|x_t, \mathbf{p}_t, D_i) &= \int \mathbb{P}(j|x_t, \mathbf{p}_t, \alpha_i, \beta_i)\, dF(\alpha_i, \beta_i|x_t, \mathbf{p}_t, D_i)\\
&= \int \mathbb{P}(j|x_t, \mathbf{p}_t, \alpha_i, \beta_i)\, dF(\alpha_i, \beta_i|\mathbf{p}_t, D_i)
\end{aligned}
$$

Conversely, if the assumption does not hold, we must compute $F(\alpha_i, \beta_i|x_t, p_t, D_i)$, which, in general, requires us to solve the dynamic-programming problem.

To illustrate what Assumption A3 rules out, we consider the following example. We suppose that there are two brands, A and B, offered in two sizes, L and S. There are two types of consumers, each with equal mass. Type a prefers brand A; type b prefers brand B. We suppose that brand A goes on sale in size L but not size S. Now we consider the conditional-choice probabilities:

$$
\mathbb{P}(A|L, \mathbf{p}_t) = \mathbb{P}(A|L, \mathbf{p}_t, a)\, \mathbb{P}(a|L, \mathbf{p}_t) + \mathbb{P}(A|L, \mathbf{p}_t, b)\, \mathbb{P}(b|L, \mathbf{p}_t)
$$

Unconditionally, $\mathbb{P}(a) = \mathbb{P}(b) = 0.5$. However, because brand A size L was on sale, it is likely that conditional on purchasing size L, the mass of type a is higher than the mass of type b. To determine how much higher, we must compute for each type the probability that they purchase size L. In general, this requires solving the dynamic-programming problem.

If Assumption A3 holds, we can compute the likelihood in the following three steps:

1. Estimate the parameters governing brand choice, α_i and β_i, by maximizing $\mathbb{P}(j|x_t, \mathbf{p}_t)$. This results in estimating a static conditional logit using only the options with size x_t.[14] This estimation is static, can be accomplished at low computational cost, and can include many controls – which, among other benefits, help with concerns about the endogeneity of prices.
2. Use the estimated parameters to compute ω_{xit} and estimate the transition probability function $F(\boldsymbol{\omega}_{i,t+1} \mid \boldsymbol{\omega}_{it})$. Because this step is accomplished once and outside the dynamic-programing problem, the transition probability can be estimated flexibly (and Assumption A2 can be tested by testing whether elements of $\mathbf{p}_t$ have power in predicting $\boldsymbol{\omega}_{i,t+1}$ beyond $\boldsymbol{\omega}_t$).
3. Estimate the dynamic parameters – governing the utility from consumption, storage cost, and the distribution of ν_t – using $\mathbb{P}(x|\boldsymbol{\omega}_t, i_t)$, which requires solving the modified dynamic-programming problem.

The split of the likelihood significantly reduces the computational cost and, as a result, a much richer model can be estimated, allowing for additional variables and rich patterns of observed heterogeneity. Among other benefits, the control for additional variables somewhat reduces the concerns of price endogeneity. The results in Hendel and Nevo (2006b) suggested that this additional richness is important.

A final point worth emphasizing is that the split of the likelihood is separate from the simplification of the state space. The simplification of the state space relied on Assumptions A1 and A2. The split in the utility also required Assumption A3.

2.4 Durable Products

Another area that has seen a lot of recent work on dynamics is the estimation of demand for durable products. There is a long tradition in IO of estimating static demand for durable products. Indeed, some of the "classic" IO papers involved estimation of demand for durable goods (e.g., Bresnahan 1981; Berry, Levinsohn, and Pakes 1995; among many others). In durable-goods markets, dynamics arise naturally because products are used in multiple

[14] The idea is similar to the computation of fixed effects in a logit model estimated with panel data: The fixed effects can be partialied out with the correct conditioning. Here, the conditioning eliminates the dynamics.

periods. The durability of the product alone does not imply that a static model cannot properly capture demand. For example, if consumers hold only a single variety (e.g., single automobile) and there are no transaction costs in resale (i.e., products can be sold and purchased costlessly) and no uncertainty about future resale prices, then a purchase of a durable good can be seen as a static period-by-period "rental." However, if these conditions do not hold, then current products owned impact purchases. Furthermore, consumers' expectations about future prices, as well as quality of the available products, impact current decisions.

Several pricing patterns can drive dynamics for durable goods. As with storable products, there can be temporary price changes that arise – for example, in the case of automobiles – if gasoline prices temporarily increase or there are temporary discounts.[15] However, a more common pattern, observed across a wide range of industries, is of declining prices and increased quality. This means that the trade-off consumers face is between delaying purchase and obtaining a lower price or higher quality in the future. This is the pattern on which we focus.

2.4.1 Implications for Static Estimation

The implications for demand estimation of ignoring dynamics, if they are present in the data, depend on exact details of the data-generating process. For example, a temporary price cut, as in the case of storable goods, causes static estimation to overestimate the own-price elasticity and underestimate the cross-price elasticity. Conversely, if gasoline prices temporarily spike, we usually think that static estimates underestimate the impacts of a permanent price increase.

If the key dynamics are declining prices, then – in general – it is more difficult to sign the direction of the bias in static estimation. It is useful to separate between two cases: with and without repeat purchase.

Without repeat purchase – once consumers purchase, they leave the market forever – there are two problems with static demand estimation: Demand is changing over time because some consumers leave the market, and consumers might be forward-looking. In the standard static random-coefficients discrete-choice model, this manifests in the following way. First, the distribution of the random coefficients is likely to change over time as some consumers purchase and exit the market. For example, if prices fall

15 Busse, Simester, and Zettelmeyer (2010) studied the 2005 Employee Discount Pricing and showed that its main effect was to induce consumers to purchase earlier.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

over time, it is likely that fewer price-sensitive consumers purchase initially. Second, if consumers are forward-looking, then they realize that there is an option value to not purchasing today. This option value is reflected in the value of the outside option, which in the static model is assumed to be constant.

To demonstrate the bias this will generate in static demand estimation, we consider the following simple example. We suppose that consumers have a willingness to pay that is distributed uniformly on the unit interval and a total mass of 100. Consumers are myopic (therefore, we are shutting off the second effect) and buy the product if the price is below their willingness to pay. Once consumers buy the product, they are out of the market forever. This yields a well-defined linear demand curve $Q = 100 - 100P$. Suppose that we observe a sequence of prices equal to (0.9, 0.8, 0.7, . . . , 0.1). Given the ~~previous~~ demand structure, the quantity sold over that same time horizon equals 10 units per period. A static demand model leads a researcher to conclude that consumers are not sensitive to price because the same quantity is sold as the price declines and to estimate an own-price elasticity of 0. So, in this example, the static model underestimates the price sensitivity. More generally, however, even in this example, as we change the distribution of willingness to pay and the sequence of observed prices, the conclusions may change. Of course, signing the effect is more difficult once we consider more general models with forward-looking consumers.

With repeat purchases, the issues are somewhat different. First, the distribution of consumers does not change because consumers do not exit. However, consumers who previously purchased a product have a different value of no purchase because their alternative is to stay with their current product. Therefore, the problem with static estimation is that it does not account for the different value, across consumers and over time, of the outside option. Second, when purchasing now, consumers do not forgo the option to purchase in the future. Indeed, consumers might find it optimal to buy an inferior option only to replace it shortly thereafter.

2.4.2 A Model of Demand for Durable Goods

We now present a basic model of demand for durable differentiated products. Our presentation follows closely Gowrisankaran and Rysman (2009).[16] The framework extends the static discrete-choice model presented in Section 2.1.1 and is similar to the inventory model presented in the previous

[16] See also Melnikov (2001) and Conlon (2009).

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

section. Indeed, to some extent, the role of inventory is equivalent to the role of the quality of the product already owned. So, in the durable-goods model, "stockpiling" means buying a higher-quality product. The difference is in the trade-off faced by consumers. The typical price pattern for durable goods is a decreasing quality-adjusted price. Faced with this price pattern for storable goods, consumers would not stockpile; rather, they would buy a small amount for current consumption and buy in the future, when the price is lower, for future consumption. In durable-goods markets, consumers can buy a "small" amount only if they can rent, lease, or resell the used product with low transaction costs. If these options are not available, the consumer's trade-off is between waiting for a lower price or higher quality product and either forgoing consumption until then or purchasing a product now and retiring it earlier than needed.

The conditional indirect utility consumer i obtains from product j at time t is given by:

$$u_{ijt} = \gamma^f_{ijt} - \alpha_i p_{jt} + \varepsilon_{ijt} \tag{13}$$

where $\gamma^f_{ijt} = a_{jt}\beta_i + \xi_{jt}$ defines flow utility. The notation follows definitions of the static model in Section 2.1.1. We note that, implicitly, the price also includes finance costs. If a consumer does not purchase, she gets the utility $u_{i0t} = \gamma^f_{i0t} + \varepsilon_{i0t}$, where:

$$\gamma^f_{i0t} = \begin{cases} 0 & \text{if no previous purchase} \\ \gamma^f_{i\widehat{j}\widehat{t}} & \text{if last purchase was product } \widehat{j} \text{ at time } \widehat{t} \end{cases}. \tag{14}$$

This definition of the utility from the outside option is the main difference between the static and dynamic models. Once consumers purchase, it changes their outside option. Thus, previous purchases impact current decisions – a fact that forward-looking consumers realize when they make current choices. We note that implicitly in the definition of the no-purchase option, there is an assumption of repeated purchase: Consumers are still on the market even after purchase, just with a different outside option.

Assuming that (1) the consumer holds at most a single product at any time, and (2) there is no resale market, then the Bellman equation of the consumer problem is given by:

$$V_i(\boldsymbol{\varepsilon}_{it}, \gamma^f_{i0t}, \mathbf{p}_t) = \max_{j=0,\ldots J} \left\{ u_{ijt} + \delta\, \mathbb{E}[E V_i(\gamma^f_{ijt}, \mathbf{p}_{t+1} | \mathbf{p}_t] \right\} \tag{15}$$

where $E V_i(\gamma^f_{ijt}, \mathbf{p}_t) = \int V_i(\boldsymbol{\varepsilon}_{it}, \gamma^f_{ijt}, \mathbf{p}_t) d F_\varepsilon(\boldsymbol{\varepsilon}_{it})$ and $\mathbf{p}_t$ represents the set of prices and other product characteristics at period t. The expectation

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

is taken with respect to the uncertainty regarding future vector $\boldsymbol{\varepsilon}_t$, future products, prices, and attributes.

If there are no repeat purchases and no resale, then the consumer's problem is slightly different.[17] Because there is no resale, without loss of generality, the utility, γ^f_{ijt}, can be seen as capturing the lifetime value from the product and there is no continuation value. Also, there is no need to keep track of the consumer's stock and $\gamma^f_{i0t} = 0$. The dynamics arise because of the option value of not purchasing. The value function in the no-repeat-purchase case is given by:

$$V_i(\boldsymbol{\varepsilon}_{it}, \mathbf{p}_t) = \max \left\{ \varepsilon_{i0t} + \delta \, \mathbb{E}[\, E\, V_i(\mathbf{p}_{t+1}|\mathbf{p}_t] \;, \; \max_{j=1,\ldots,J} u_{ijt} \right\} \tag{16}$$

where now $E\, V_i(\mathbf{p}_t) = \int V_i(\boldsymbol{\varepsilon}_{it}, \mathbf{p}_t) dF_\varepsilon(\boldsymbol{\varepsilon}_{it})$. The first term within the brackets represents the value of waiting to purchase in the future; the second term is the value of purchasing today. Because we do not have to keep track of the current holding, the state space is reduced.

2.4.3 Reducing the Dimension of the State Space

The main computational cost is computing the expected value function, which in the repeat-purchase model equals $E\, V_i(\gamma^f_{ijt}, \mathbf{p}_t)$. The state space is similar to what we observed in the storable-goods problem and consists of the quality of the currently held product – which is equivalent to the inventory in the storable-goods problem – and the matrix of prices and current attributes required to form expectations regarding the future.

Holdings. It is useful to briefly consider a somewhat more general model of the consumer's holding. In this model, the consumer can hold several varieties of the products, and the utility from the different varieties interact with one another.[18] There are several ways to model the flow utility in this case[19] but, in all of them, the state variable includes a vector describing the consumer's current holdings and not a scalar. By assuming that the consumer holds only a single option at any point in time, we reduce the state

[17] See Melnikov (2001) and Conlon (2009) for applied examples and further discussion of the no-repeat-purchase model.

[18] We note that even in static models, the issue of multiple purchases and the interaction in utility, or through a budget constraint, is mostly an open question and usually ignored. The few exceptions are Hendel (1999); Dube (2004); Nevo, McCabe, and Rubinfeld (2005); and Genztkow (2007).

[19] For example, the utility can be a function of the products held or it can be a function of the characteristics of the products held.

space to a scalar value of the holding in the repeat-purchase model or avoided it all together in the no-repeat-purchase model. Thus, the assumption of holding only a single product serves the same purpose as Assumption A1 in the storable-goods model and reduces the dimension of the holdings or inventory variable.

To further understand the differences between the storable- and durable-goods models, we consider that a product can be characterized by two dimensions that can be consumer-specific: its quality (i.e., utility per use) and its quantity (i.e., how many times it can be used). In the storable-goods model, we simplified the model by making assumptions on the quality (see the discussion following Assumption A1) and focused on the quantity. Here, we leave the quality unrestricted but make assumptions on the quantity by assuming a single good and no depreciation. Allowing for depreciation that is a function of endogenously chosen usage makes the durable-goods model closer to the storable-goods model.

We note that there is another similarity with the storable-goods model. Here, the utility carried forward is γ^f_{ijt} and not $\gamma^f_{ijt} + \varepsilon_{ijt}$. Thus, as in the inventory model, there is a separation between the utility at the time of purchase and at the time of usage.

Finally, we note that an alternative way to reduce the state space is to allow for multiple purchases but to assume no interaction in the utility and to continue to assume no resale.

Prices. Even after reducing the dimension of the holding vector, for a realistic number of products, the state space is still too large to be manageable. As before, we rely on the inclusive value to reduce the state space. We define the *dynamic inclusive value* as from the *J*-inside alternatives as:

$$\Omega_{it}(\mathbf{p}_t) = \ln\left(\sum_{j=1}^{J} \exp(\gamma^f_{ijt} - \alpha_i p_{jt} + \delta\, \mathbb{E}[E V_i(\gamma^f_{ijt}, \mathbf{p}_{t+1}) \mid \mathbf{p}_t])\right) \quad (17)$$

We note that this definition is different in an important way from the definition given in Section 2.1.1. It provides the expected value, including the future value, from the J options. The definition in Section 2.1.1 provides the expected flow utility, not accounting for the future value. The difference is not just semantic. The static definition basically provides a utility-consistent welfare statistic that is a summary of prices and attributes of available products. The dynamic definition also includes endogenous future behavior of the agent. Once we impose a particular stochastic structure on the evolution of Ω_{it}, a natural question is whether the imposed structure is consistent with the consumer-optimization problem. Gowrisankaran and Rysman (2009)

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

discussed whether this is restrictive, but generally little is known on which behavioral assumptions are consistent with the imposed structure.

To reduce the state space, we change the modified version of Assumption A2 (Assumption A1, Inclusive Value Sufficiency, in Gowrisankaran and Rysman 2009) to the following:

Assumption A2′: $F(\Omega_{i,t+1} \mid \mathbf{p}_t) = F(\Omega_{i,t+1} \mid \Omega_{it}(\mathbf{p}_t))$

As before, the assumption assumes that the inclusive value is sufficient to compute the transition probabilities, but now it is the dynamic inclusive value, Ω_{it}. Furthermore, now there is a single inclusive value rather than a vector of size-specific inclusive values. Using this assumption, we now can write in the repeat-purchase model:

$$\begin{aligned} EV_i(\gamma^f_{i0t}, \mathbf{p}_t) &= EV_i(\gamma^f_{i0t}, \Omega_{it}) \\ &= \ln\Big(\exp(\Omega_{it}) + \exp\Big(\gamma^f_{i0t} + \delta\,\mathbb{E}[EV_i(\gamma^f_{i0t}, \Omega_{it+1}|\Omega_{it}]\Big)\Big) \end{aligned} \tag{18}$$

As in the storable-goods problem, Assumption A2′ allows us to reduce the state space; however, unlike the storable-goods problem, we do not modify the dynamic-programming problem. In the storable-goods problem, Assumption A1 allowed us to modify the dynamic-programming problem into a choice among sizes rather than a choice among brand-size combinations. The reason that we could do this is because under Assumption A1, choices of the same size impacted the dynamics in the same way. Here, we cannot modify the problem because we cannot generate such equivalence classes for the dynamics.

The situation is somewhat different in the no-repeat-purchase model. First, the state space can be reduced but the relevant definition of the inclusive value is the static one, given in Section 2.1.1, and not the one given in Equation (17). Assuming a version of A2 ′ for the inclusive values, we can show that $EV_i(\mathbf{p}_t) = EV_i(\Omega_{it})$. Second, in the no-repeat-purchase, the dynamics involves a decision on when to buy; however, conditional on purchase, the decision of which product to buy is static. As in the storable-products model, the dynamic-programming problem can be simplified.

2.5 Estimation and Identification

This section discusses identification and estimation of the model. Several studies estimated demand for durable products using household-level data.[20] However, many studies of demand for durable goods recently relied on aggregate data. For this reason, we focus our discussion on estimation with aggregate data.

2.5.1 Identification

If consumer-level data are observed, then – in principle – identification follows the standard arguments (Rust 1994; Magnac and Thesmar 2002; Aguirregabiria 2010).[21] With aggregate data, we do not observe the purchase history of each consumer, which makes identification significantly more difficult. To see this, we consider the example in Section 2.4.1. The sequence of quantities can be explained perfectly using a static model with zero price sensitivity or with the no-repeat-purchase dynamic model, which generated the example. We suppose that there are multiple products available in each period; then, the model must explain not only the time-series variation in shares but also the cross-sectional variation. The key to identifying the model and to separating the different alternative models is the ability of the models to explain both the cross-sectional variation, across markets and products, and the time-series variation.

We are unaware of a formal identification proof and obtaining one may be difficult. Standard identification proofs for static models require some form of substitution (e.g., what Berry, Haile, and Gandhi [2011] called *connected substitutes*) among products. In static models, the substitution is among products in a given period; here, however, the requirement is for substitution over time and across products. This need not be satisfied; for example, if the price of a high-quality product falls at time t, it actually could increase the demand for a low-quality product at $t - 1$ because some consumers might buy it for one period.

As previously discussed, a key issue in static demand estimation is the potential endogeneity of price. In dynamic demand models estimated using aggregate data, the solution follows closely the static literature using GMM and moment conditions similar to the static models.

2.5.2 Estimation

The estimation follows closely the method proposed by Berry, Levinsohn, and Pakes (1995) but nests a solution of the dynamic-programming problem

[20] Many of these studies estimated static demand. For examples of dynamic demand, see Erdem, Keane, Oncu, and Strebel (2005) and Prince (2008).

[21] The standard arguments must be adjusted for the existence of ξ_{jt}; however, with enough observations, these could be controlled for, and then we are back in the standard case.

inside the inner loop.[22] The basic steps are as follows (see Gowrisankaran and Rysman 2009 for details):

1. For a given value of the parameters and a vector of mean-flow-utility levels ($\gamma^f_{jt} = a_{jt}\beta + \xi_{jt}$), compute the predicted market shares by the following:
   (a) For a number of simulated consumers, each with a α_i and β_i, calculate the dynamic inclusive value given by Equation (17) (begin the process with an initial guess on EV_i).
   (b) Use these inclusive values to compute $F(\Omega_{it+1} \mid \Omega_{it}(\mathbf{p}_t))$.[23]
   (c) Use the estimated process to update EV_i.
   (d) Iterate the previous three steps until convergence.
   (e) Use the estimated policy to simulate for each consumer the purchase path, assuming that all consumers initially hold the outside good.
   (f) Aggregate the consumer's purchase decision to obtain market shares.
2. For a given value of the parameters, use the iteration proposed by Berry, Levinsohn, and Pakes (1995) to compute the vector of γ^f_{jt} values. The iteration uses the markets shares computed in Step 1.
3. As in Berry, Levinsohn, and Pakes (1995), use the vector of γ^f_{jt} to compute moment conditions and search for the parameters than minimize a GMM objective function.

In the no-repeat-purchase model, the computation can be simplified if we add an assumption like Assumption A3, which limits the heterogeneity (Melnikov 2001).

3.0 Dynamic Games of Oligopoly Competition

3.1 Overview

The study of firm behavior, especially in oligopoly, is at the heart of IO. In many industries, a firm's current actions affect its future profits, as well as the current and future profits of other firms in the industry. Supply-side

[22] For an alternative computational method, see Su and Judd (2012); Dube, Fox, and Su (2012); and Conlon (2010).

[23] In principle, the process here can be quite general. In reality, however, because the computation is nested with the computation, the process must be fairly simple. Gowrisankaran and Rysman (2009) assumed that $\omega_{i,t+1} = \gamma_{1i} + \gamma_{2i}\omega_{i,t+1} + v_{it}$.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

dynamics can arise from different sources, including sunk costs of entry, partially irreversible investments, product-repositioning costs, and learning-by-doing. Ignoring supply-side dynamics potentially can lead to biases in our estimates of structural parameters. More substantially, accounting for dynamics can change our view of the impact of competition in some industries, as well as our evaluation of public policies. The following examples illustrate these points.[24]

Example 1: Product Repositioning in Differentiated Product Markets. Sweeting (2007) and Aguirregabiria and Ho (2012) are two examples of empirical applications that endogenize product attributes using a dynamic game of competition in a differentiated-products industry. Sweeting (2007) estimated a dynamic game of oligopoly competition in the U.S. commercial-radio industry. The model endogenizes the choice of radio-station format (i.e., genre) and estimated product-repositioning costs. Aguirregabiria and Ho (2012) studied the contribution of different factors to explain airlines' adoption of hub-and-spoke networks. They proposed and estimated a dynamic game of airline-network competition in which the number of direct connections that an airline has in an airport is an endogenous product characteristic. These studies highlighted the two potential limitations of static models. First, a common assumption in many static and dynamic demand models is that product characteristics, other than prices, are exogenous. This assumption, if violated, can generate biases in the estimated parameters. The dynamic game acknowledges the endogeneity of some product characteristics and exploits the dynamic structure of the model to generate valid moment conditions for the consistent estimation of the structural parameters. A second important limitation of a static model of firm behavior is that it cannot recover the costs of repositioning-product characteristics. As a result, the static model cannot address important empirical questions, such as the effect of a merger on product repositioning.

Example 2: Evaluating the Effects of Regulation. Ryan (2012) provided another example of how ignoring the endogeneity of market structure and its dynamics can lead to misleading results. He studied the effects of the 1990 Amendments to the Clean Air Act on the U.S. cement industry. This environmental regulation added new categories of regulated emissions and introduced the requirement of an environmental certification that cement

[24] As before, we cannot provide a complete survey of the literature. Other examples of papers that studied similar questions are Einav (2009); Collard-Wexler (2006); Macieira (2007); Kryukov (2008); Hashmi and Van Biesebroeck (2010); Snider (2009); Suzuki (2010); Gowrisankaran et al. (2010); Walrath (2010); and Finger (2008), among others.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

plants must pass before starting their operation. Ryan estimated a dynamic game of competition in which the sources of dynamics are sunk-entry costs and adjustment costs associated with changes in installed capacity. The estimated model showed that the new regulation had negligible effects on variable production costs, but it increased significantly the sunk cost of opening a new cement plant. A static analysis that ignores the effects of the policy on firms' entry–exit decisions would conclude that the regulation had negligible effects on firms' profits and consumer welfare. In contrast, the dynamic analysis shows that the increase in sunk-entry costs caused a reduction in the number of plants that, in turn, implied higher markups and a decline in consumer welfare.

Initial attempts to answer many of these questions used entry models in the spirit of Bresnahan and Reiss (1990, 1991a, 1991b) and Berry (1992). The simplest forms of these models use a reduced-form profit function in the sense that variable profits are not derived from explicit models of price or quantity competition, and static in the sense that firms are not forward-looking. These models were used to explain cross-market variation in market structure that is assumed to be an equilibrium of an entry game. It is possible to include predetermined variables in the payoff function (e.g., firm size, capacity, and incumbent status) and to interpret the payoff function as an intertemporal value function (Bresnahan and Reiss 1993). Indeed, we could use panel data to estimate some of the parameters or use price and quantity data to estimate the variable profits. These models typically are much easier to estimate than the dynamic games discussed herein; therefore, at times, they might serve as a useful first cut of the data. The main limitation of this approach is that the parameters often do not have a clear economic interpretation in terms of costs or demand, and the model cannot be used for counterfactual-policy experiments. Furthermore, empirical questions in IO that are related to the effects of uncertainty on firm behavior and competition or that try to distinguish between short- and long-run effects of exogenous shocks typically require the specification and estimation of dynamic structural models that explicitly take into account firms' forward-looking behavior. For these reasons, most of the recent work in IO addressing industry dynamics relies on more explicit modeling of dynamics, as in the model of Ericson and Pakes (1995). Sections 3.2 and 3.3 briefly describe a simple version of this model that allows us to demonstrate our key points.

Sections 3.4–3.7 discuss some of the main econometric, computational, and modeling issues faced by applied researchers who want to estimate a dynamic game. The standard nested fixed-point algorithm, which was used successfully in the estimation of single-agent models, is computationally infeasible in actual applications of dynamic games. As a result, researchers turned to alternative methods based on the ideas of Hotz and Miller (1993) and Aguirregabiria and Mira (2002) (i.e., estimation methods based on conditional choice probabilities [CCP]). We survey some of the methods that were proposed to implement these ideas, and we focus on several issues. First, we discuss the impact of multiple equilibria on identification and present sufficient conditions for point identification of the structural parameters. We then discuss the properties of an iterative procedure that was proposed by Aguirregabiria and Mira (2009) to address a potential shortcoming of two-step CCP methods: finite-sample bias. A paper by Pesendorfer and Schmidt-Dengler (2008) showed that, indeed, in some cases, finite-sample bias is reduced but in other cases, the iterative procedure actually increases the bias. We provide stability conditions on the equilibrium that guarantee the performance of the method, and we explain the results of Pesendorfer and Schmidt-Dengler (2008).

Another main shortcoming of the CCP approach is the lack of unobserved firm- or market-level heterogeneity beyond a firm level i.i.d. shock. Section 3.5 briefly discusses new CCP methods that allow us to relax this assumption. Section 3.6 returns to a theme that was a major part of our discussion of dynamic demand: methods to reduce the dimension of the state space. We show how the inclusive-value approach discussed herein can be extended to dynamic games to reduce the computational burden in the solution and estimation of this class of models. Section 3.7 concludes with a description of a homotopy method that can be used to implement counterfactual experiments given the estimated model.

3.2 The Structure of Dynamic Games of Oligopoly Competition

We use a simple dynamic game of market entry–exit to illustrate different issues and methods. Time is discrete and indexed by t. The game is played by N firms that we index by i. We let a_{it} be the decision variable of firm i at period t. In the entry–exit model we consider, the decision variable is a binary indicator of the event "firm i is active in the market at period t." The action is taken to maximize the expected and discounted flow of profits in the market, $\mathbb{E}_t\left(\sum_{r=0}^{\infty} \delta^r\ \Pi_{it+r}\right)$, where $\delta \in (0, 1)$ is the discount factor and Π_{it} is firm i's profit at period t.

The profits of firm i at time t are given by $\Pi_{it} = VP_{it} - FC_{it} - EC_{it}$, where VP_{it} represents variable profits, FC_{it} is the fixed cost of operating, and EC_{it} is a one-time entry cost. Following the standard structure in the

Ericson and Pakes (1995) framework, incumbent firms in the market at period t compete in prices or quantities in a static Cournot or Bertand model. For example, the variable profit function can take on the form[25]:

$$VP_{it}(a_{it}, a_{-it}) = a_{it}\ H_t\ \sum_{n=0}^{N-1} 1\left\{\sum_{j\neq i} a_{jt} = n\right\}\ \theta_{i,n}^{VP} \tag{19}$$

where H_t is a measure of market size; $1\{.\}$ is the indicator function; and $\sum_{j\neq i} a_{jt}$ is the number of active competitors of firm i at period t. The vector of parameters $\{\theta_{i,n}^{VP} : n = 0, 1, \ldots, N-1\}$ represents firm i's variable profit per-capita when there are other n competitors active in the market. We expect $\theta_{i,0}^{VP} \geq \theta_{i,1}^{VP} \geq \ldots \geq \theta_{i,N-1}^{VP}$. The fixed cost is paid every period that the firm is active in the market, and it has the following structure: $FC_{it} = a_{it}\left(\theta_i^{FC} + \varepsilon_{it}\right)$, where θ_i^{FC} is a parameter that represents the mean value of the fixed operating cost of firm i; and ε_{it} is a zero-mean shock that is private information of firm i. The entry cost is paid only if the firm was not active in the market at the previous period: $EC_{it} = a_{it}(1 - s_{it})\,\theta_i^{EC}$, where s_{it} is a binary indicator that is equal to 1 if firm i was active in the market in period $t-1$ (i.e., $s_{it} \equiv a_{i,t-1}$), and θ_i^{FC} is a parameter that represents the entry cost of firm i. The specification of the primitives of the model is completed with the transition rules of the state variables. Market size follows an exogenous Markov process with transition-probability function $F_H(H_{t+1}|H_t)$. The transition of the incumbent status is trivial, $s_{it+1} = a_{it}$. Finally, the private-information shock ε_{it} is i.i.d. over time and independent across firms with CDF G_i.[26]

Somewhat in contrast to static entry models, in which both games of complete and incomplete information are studied, the recent literature on empirical dynamic games focuses solely on games of incomplete information. The introduction of private-information shocks ensures the existence of an equilibrium in pure strategies (Doraszelski and Satterthwaite 2010). In addition, these random shocks are a convenient way to allow for

[25] This indirect variable-profit function may come from the equilibrium of a static Bertrand game with differentiated product, as in the framework presented in Section 2.1. We suppose that all firms have the same marginal cost and that product differentiation is symmetric. For instance, consumer utility of buying product i is $u_{it} = \nu - \alpha\, p_{it} + \varepsilon_{it}$, where ν and α are parameters and ε_{it} is a consumer-specific i.i.d. random variable. Then, the equilibrium variable profit of an active firm depends on only the number of firms active in the market.

[26] In this example, we consider that firms' entry–exit decisions are made at the beginning of period t and that they are effective during the same period. An alternative timing considered in some applications is that there is a one-period time-to-build (i.e., the decision is made at period t and entry costs are paid at period t but the firm is not active in the market until period $t+1$). The latter is, in fact, the timing of decisions in Ericson and Pakes (1995).

econometric unobservables that can explain how agents with the same observable characteristics make different decisions.

Following Ericson and Pakes (1995), most of the recent IO literature studying industry dynamics focuses on studying a Markov Perfect Equilibrium (MPE), as defined by Maskin and Tirole (1987, 1988a, 1988b). The key assumption in this solution concept is that players' strategies are functions of only payoff-relevant state variables. We use the vector $\mathbf{x}_t$ to represent all of the common-knowledge state variables at period t (i.e., $\mathbf{x}_t \equiv (H_t, s_{1t}, s_{2t}, \ldots, s_{Nt})$). In this model, the payoff-relevant state variables for firm i are $(\mathbf{x}_t, \varepsilon_{it})$.[27] We let $\boldsymbol{\alpha} = \{\alpha_i(\mathbf{x}_t, \varepsilon_{it}) : i \in \{1, 2, \ldots, N\}\}$ be a set of strategy functions, one for each firm. A MPE is a set of strategy functions $\boldsymbol{\alpha}^*$ such that every firm is maximizing its value given the strategies of other players. For given strategies of the other firms, the decision problem of a firm is a single-agent dynamic-programming problem. We let $V_i^{\boldsymbol{\alpha}}(\mathbf{x}_t, \varepsilon_{it})$ be the value function of this dynamic-programming problem. This value function is the unique solution to the Bellman equation:

$$V_i^{\boldsymbol{\alpha}}(\mathbf{x}_t, \varepsilon_{it}) = \max_{a_{it} \in \{0,1\}} \left\{ a_{it}\,(\Pi_i^{\boldsymbol{\alpha}}(\mathbf{x}_t) - \varepsilon_{it}) + \delta \int V_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}, \varepsilon_{it+1})\, dG_i(\varepsilon_{it+1})\, F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|a_{it}, \mathbf{x}_t) \right\} \quad (20)$$

where $a_{it}\,(\Pi_i^{\boldsymbol{\alpha}}(\mathbf{x}_t) - \varepsilon_{it})$ and $F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|a_{it}, \mathbf{x}_t)$ are the expected one-period profit and the expected transition of the state variables, respectively, for firm i given the strategies of other firms. By definition, the expected one-period profit $\Pi_i^{\boldsymbol{\alpha}}(\mathbf{x}_t)$ is:

$$\Pi_i^{\boldsymbol{\alpha}}(\mathbf{x}_t) = H_t \sum_{n=0}^{N-1} \Pr\left(\textstyle\sum_{j \neq i} \alpha_j(\mathbf{x}_t, \varepsilon_{jt}) = n \mid \mathbf{x}_t\right) \theta_{i,n}^{VP} - \theta_i^{FC} - (1 - s_{it})\theta_i^{EC} \quad (21)$$

The expected transition of the state variables is:

$$F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|a_{it}, \mathbf{x}_t) = 1\{s_{it+1} = a_{it}\} \left[\prod_{j \neq i} \Pr\left(s_{j,t+1} = \alpha_j(\mathbf{x}_t, \varepsilon_{jt}) \mid \mathbf{x}_t\right)\right] F_H(H_{t+1} \mid H_t) \quad (22)$$

[27] If private-information shocks are correlated serially, the history of previous decisions contains useful information to predict the value of a player's private information, and it should be part of the set of payoff-relevant state variables. Therefore, the assumption that private information is distributed independently over time has implications for the set of payoff-relevant state variables.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

A player's best-response function gives his optimal strategy if the other players behave – now and in the future – according to their respective strategies. In this model, the best-response function of player i is $1\{v_i^{\boldsymbol{\alpha}}(\mathbf{x}_t) - \varepsilon_{it} \geq 0\}$, where $v_i^{\boldsymbol{\alpha}}(\mathbf{x}_t)$ is the difference between the value of firm i if it chooses Alternative 1 (without including ε_{it}) and its value if it chooses Alternative 0, given that the other players play their strategies in $\boldsymbol{\alpha}$. According to our model:

$$v_i^{\boldsymbol{\alpha}}(\mathbf{x}_t) \equiv \Pi_i^{\boldsymbol{\alpha}}(\mathbf{x}_t) + \delta \int V_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}, \varepsilon_{it+1})\, dG_i(\varepsilon_{it+1}) \times [F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|1, \mathbf{x}_t) - F_i^{\boldsymbol{\alpha}}(\mathbf{x}_{t+1}|0, \mathbf{x}_t)] \tag{23}$$

A MPE in this game is a set of strategy functions $\boldsymbol{\alpha}^*$ such that for any player i and for any $(\mathbf{x}_t, \varepsilon_{it})$, we have that $\alpha_i^*(\mathbf{x}_t, \varepsilon_{it}) = 1\{\varepsilon_{it} \leq v_i^{\boldsymbol{\alpha}^*}(\mathbf{x}_t)\}$.

3.3 Conditional-Choice Probabilities

This section introduces the concept of the *conditional-choice probability (CCP)* function and defines players' strategies, value functions, and best responses in terms of this probability function. This representation is useful in the empirical analysis of dynamic games because CCPs can be seen as conditional expectations involving players' actions and state variables observed in the data.

Given a strategy function $\alpha_i(\mathbf{x}_t, \varepsilon_{it})$, we define the corresponding CCP function as:

$$P_i(\mathbf{x}_t) \equiv \Pr(\alpha_i(\mathbf{x}_t, \varepsilon_{it}) = 1 \mid \mathbf{x}_t) = \int \alpha_i(\mathbf{x}_t, \varepsilon_{it})\, dG_i(\varepsilon_{it}) \tag{24}$$

Because choice probabilities are integrated over the continuous variables in ε_{it}, they are lower-dimensional objects than the strategies $\boldsymbol{\alpha}$. For instance, when both a_{it} and $\mathbf{x}_t$ are discrete, CCPs can be described as vectors in a finite-dimensional Euclidean space. In our entry–exit model, $P_i(\mathbf{x}_t)$ is the probability that firm i is active in the market given the state $\mathbf{x}_t$. By definition, given $\alpha_i(\mathbf{x}_t, \varepsilon_{it})$, the CCP $P_i(\mathbf{x}_t)$ is uniquely determined. If the private-information shock ε_{it} (1) is i.i.d. over time; (2) does not enter in the transition probability of $\mathbf{x}_t$ (i.e., conditional-independence assumption); and (3) enters additively in the expected one-period profit (i.e., additive separability), then given a CCP function $P_i(\mathbf{x}_t)$, there is a unique strategy function $\alpha_i(\mathbf{x}_t, \varepsilon_{it})$ compatible with it.[28] Therefore, there is a one-to-one

[28] Under conditions (1), (2), and (3), the best-response function has the single-threshold form $1\{\varepsilon_{it} \leq v_i^{\boldsymbol{\alpha}}(\mathbf{x}_t)\}$. Therefore, we can limit our analysis to the set of strategy functions

relationship between strategy functions and CCPs. From now on, we use CCPs to represent players' strategies, and we use the terms *strategy* and *CCP* interchangeably. We also use $\Pi_i^{\mathbf{P}}$ and $F_i^{\mathbf{P}}$ instead of $\Pi_i^{\boldsymbol{\alpha}}$ and $F_i^{\boldsymbol{\alpha}}$ to represent the expected-profit function and the transition-probability function, respectively.

Based on the CCP concept, we describe a representation of the equilibrium mapping and of a MPE that is particularly useful for the econometric analysis.[29] This representation has two main features: (1) a MPE is a vector of CCPs, and (2) a player's best response is an optimal response not only to other players' strategies but also to his own strategy in the future. A MPE is a vector of CCPs, $\mathbf{P} \equiv \{P_i(\mathbf{x}_t) : i = 1, 2, \ldots, N; \mathbf{x}_t \in \mathcal{X}\}$, such that for every firm and any state $\mathbf{x}_t$, the following equilibrium condition is satisfied:

$$P_i(\mathbf{x}_t) = G_i\left(\Pi_i^{\mathbf{P}}(\mathbf{x}_t) + \delta \textstyle\sum_{\mathbf{x}_{t+1}} \left[F_i^{\mathbf{P}}(\mathbf{x}_{t+1}|1, \mathbf{x}_t) - F_i^{\mathbf{P}}(\mathbf{x}_{t+1}|0, \mathbf{x}_t)\right] V_i^{\mathbf{P}}(\mathbf{x}_{t+1})\right) \tag{25}$$

The right-hand side of Equation (25) is a best-response probability function. $V_i^{\mathbf{P}}$ is the valuation operator of player i if every player behaves now and in the future according to their respective strategies in $\mathbf{P}$. We can obtain $V_i^{\mathbf{P}}$ as the unique solution of the recursive expression:

$$V_i^{\mathbf{P}}(\mathbf{x}_t) = P_i(\mathbf{x}_t)\ \left[\Pi_i^{\mathbf{P}}(\mathbf{x}_t) + e_i\left(P_i(\mathbf{x}_t)\right)\right] + \delta \sum_{\mathbf{x}_{t+1}} V_i^{\mathbf{P}}(\mathbf{x}_{t+1})\ F^{\mathbf{P}}(\mathbf{x}_{t+1}|\mathbf{x}_t) \tag{26}$$

where $e_i\left(P_i(\mathbf{x}_t)\right)$ is the expectation $\mathbb{E}_{\varepsilon_{it}}(-\varepsilon_{it}|\varepsilon_{it} \leq G_i^{-1}(P_i(\mathbf{x}_t)))$ and the form of the function $e_i(.)$ depends on the probability distribution of ε_{it}.[30]

with this threshold structure. This implies that CCP functions should have the form $P_i(\mathbf{x}_t) = G_i(v_i^{\boldsymbol{\alpha}}(\mathbf{x}_t))$. Because the CDF function G_i is invertible everywhere, for given $P_i(\mathbf{x}_t)$, there is a unique threshold $v_i^{\boldsymbol{\alpha}}(\mathbf{x}_t)$ that is compatible with this choice probability; that is, $v_i^{\boldsymbol{\alpha}}(\mathbf{x}_t) = G_i^{-1}(P_i(\mathbf{x}_t))$, where $G_i^{-1}(.)$ is the inverse function of G_i. Thus, given a CCP function $P_i(\mathbf{x}_t)$, the unique strategy function compatible with it is $\alpha_i(\mathbf{x}_t, \varepsilon_{it}) = 1\{\varepsilon_{it} \leq G_i^{-1}(P_i(\mathbf{x}_t))\}$. This result can be extended to a multinomial discrete-choice model with a general number of J choice alternatives and to dynamic games with continuous decision variables.

29 For the general results, see the Representation Lemma in Aguirregabiria and Mira (2009).

30 We consider the event $\varepsilon \leq c$, where c is a constant and ε is a random variable with CDF G. We let P be the probability of that event, $P \equiv \Pr(\varepsilon_{it} \leq c)$, such that $P = G(c)$. We define the function $e(P) \equiv \mathbb{E}(-\varepsilon|\varepsilon \leq c) = \mathbb{E}(-\varepsilon|\varepsilon \leq G^{-1}(P))$. If ε is normally distributed with zero mean and variance σ^2, then $e(P) = \sigma\ \phi(\Phi^{-1}(P))$, where ϕ is the PDF and Φ^{-1} is the inverse CDF of the standard normal. If ε is extreme-value Type 1 with dispersion parameter σ, we have that $e(P) = \sigma(\gamma + \ln(P))$, where γ is Euler's constant.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

When the space $\mathcal{X}$ is discrete and finite, we can obtain $V_i^{\mathbf{P}}$ as the solution of a system of linear equations of dimension $|\mathcal{X}|$. In vector form, $\mathbf{V}_i^{\mathbf{P}} = \left(\mathbf{I} - \delta\, \mathbf{F}^{\mathbf{P}}\right)^{-1} P_i * [\mathbf{\Pi}_i^{\mathbf{P}} + e_i^{\mathbf{P}}]$, where $\mathbf{I}$ is the identity matrix; P_i is the $|\mathcal{X}| \times 1$ vector of CCPs for player i; $\mathbf{V}_i^{\mathbf{P}}$, $\mathbf{\Pi}_i^{\mathbf{P}}$, and $e_i^{\mathbf{P}}$ also are $|\mathcal{X}| \times 1$ vectors; $\mathbf{F}^{\mathbf{P}}$ is the $|\mathcal{X}| \times |\mathcal{X}|$ transition matrix with elements $F^{\mathbf{P}}(\mathbf{x}_{t+1}|\mathbf{x}_t)$; and $*$ is the element-by-element product. We represent the equilibrium mapping in matrix form as $\Psi(\mathbf{P}, \boldsymbol{\theta})$, such that a MPE associated with a vector of structural parameters $\boldsymbol{\theta}$ is a fixed point $\mathbf{P} = \Psi(\mathbf{P}, \boldsymbol{\theta})$.

The valuation and the best-response operators can be simplified further for the class of models in which the expected-profit function is multiplicatively separable in the structural parameters. In our entry-exit model, the profit function can be written as $\Pi_{it} = a_{it}\ (\mathbf{z}_i(a_{-it}, \mathbf{x}_t)\ \boldsymbol{\theta}_i - \varepsilon_{it})$, where $\boldsymbol{\theta}_i$ is the column vector of structural parameters $\left(\theta_{i,0}^{VP}, \theta_{i,1}^{VP}, \ldots, \theta_{i,N-1}^{VP},\ \theta_i^{FC},\ \theta_i^{EC}\right)'$ and $\mathbf{z}_i(a_{-it}, \mathbf{x}_t)$ is the row vector of known functions $\{H_t\ 1\{\sum_{j\neq i} a_{jt} = 0\},\ H_t\ 1\{\sum_{j\neq i} a_{jt} = 1\}, \ldots,\ H_t\ 1\{\sum_{j\neq i} a_{jt} = N-1\},\ -1,\ -(1 - s_{it})\}$. Therefore, the expected-profit function also is multiplicatively separable in the structural parameters: $\Pi_i^{\mathbf{P}}(\mathbf{x}_t) = \mathbb{E}(\mathbf{z}_i(a_{-it}, \mathbf{x}_t) \mid \mathbf{x}_t, \mathbf{P}(\mathbf{x}_t))\ \boldsymbol{\theta}_i$, where the expectation is over the distribution of a_{-it} conditional on $\mathbf{x}_t$ under the condition that firms behave according to their respective CCPs in $\mathbf{P}(\mathbf{x}_t)$. In this example, we have that $\mathbb{E}(\mathbf{z}_i(a_{-it}, \mathbf{x}_t) \mid \mathbf{x}_t, \mathbf{P}(\mathbf{x}_t)) = \{\ H_t \Pr(\sum_{j\neq i} a_{jt} = 0 \mid \mathbf{x}_t, \mathbf{P}), \ldots, H_t \Pr(\sum_{j\neq i} a_{jt} = N-1 \mid \mathbf{x}_t, \mathbf{P}), -1, -(1 - s_{it})\}$. The $|\mathcal{X}| \times 1$ vector with expected one-period payoffs, $\mathbf{\Pi}_i^{\mathbf{P}}$, can be represented as $\mathbf{\Pi}_i^{\mathbf{P}} = \mathbf{Z}_i^{\mathbf{P}}\, \boldsymbol{\theta}_i$, where $\mathbf{Z}_i^{\mathbf{P}}$ is a matrix with rows the vectors $\mathbb{E}(\mathbf{z}_i(a_{-it}, \mathbf{x}_t) \mid \mathbf{x}_t, \mathbf{P}(\mathbf{x}_t))$ for each value of $\mathbf{x}_t$ in the state space $\mathcal{X}$. If ε_{it} is normally distributed with zero mean and variance σ_i^2, we have that $e_i\,(P_i(\mathbf{x}_t)) = \sigma_i\ \phi(\Phi^{-1}\,(P_i(\mathbf{x}_t)))$. Thus, the valuation operator is multiplicatively separable in the structural parameters and it has the following structure: $\mathbf{V}_i^{\mathbf{P}} = \mathbf{W}_{z,i}^{\mathbf{P}}\, \boldsymbol{\theta}_i - \mathbf{W}_{e,i}^{\mathbf{P}}\, \sigma_i$, where $\mathbf{W}_i^{\mathbf{P}}$ is the matrix $\left(\mathbf{I} - \delta\, \mathbf{F}^{\mathbf{P}}\right)^{-1} P_i * \mathbf{Z}_i^{\mathbf{P}}$ and $\mathbf{W}_{e,i}^{\mathbf{P}}$ is the vector $\left(\mathbf{I} - \delta\, \mathbf{F}^{\mathbf{P}}\right)^{-1} P_i * e_i^{\mathbf{P}}$, with $e_i^{\mathbf{P}}$ being the vector with elements $\phi(\Phi^{-1}\,(P_i(\mathbf{x}_t)))$. Then, the best-response probability function is:

$$P_i(\mathbf{x}_t) = \Phi\left(\tilde{\mathbf{z}}_i^{\mathbf{P}}(\mathbf{x}_t)\ \frac{\boldsymbol{\theta}_i}{\sigma_i} + \tilde{e}_i^{\mathbf{P}}(\mathbf{x}_t)\right) \tag{27}$$

where $\tilde{\mathbf{z}}_i^{\mathbf{P}}(\mathbf{x}_t)$ is equal to $\mathbf{Z}_i^{\mathbf{P}}(\mathbf{x}_t) + \delta \sum_{\mathbf{x}_{t+1}} [F_i^{\mathbf{P}}(\mathbf{x}_{t+1}|1, \mathbf{x}_t) - F_i^{\mathbf{P}}(\mathbf{x}_{t+1}|0, \mathbf{x}_t)]\ W_{z,i}^{\mathbf{P}}(\mathbf{x}_{t+1})$, and $\tilde{e}_i^{\mathbf{P}}(\mathbf{x}_t)$ is $\delta \sum_{\mathbf{x}_{t+1}}\ [F_i^{\mathbf{P}}(\mathbf{x}_{t+1}|1, \mathbf{x}_t) - F_i^{\mathbf{P}}(\mathbf{x}_{t+1}|0, \mathbf{x}_t)]\ W_{e,i}^{\mathbf{P}}(\mathbf{x}_{t+1})$. The vector $\tilde{\mathbf{z}}_i^{\mathbf{P}}(\mathbf{x}_t)$ has a more intuitive interpretation as the difference between two expected-discounted values of current and future firm i's profits:

$$\begin{aligned}\tilde{\mathbf{z}}_i^{\mathbf{P}}(\mathbf{x}_t) = \mathbb{E}\left(\sum_{r=0}^{\infty} \delta^r\ a_{it+r}\ \mathbf{z}_i(a_{-it+r}, \mathbf{x}_{t+r}) \mid a_{it} = 1, \mathbf{x}_t, \mathbf{P}\right) \\ - \mathbb{E}\left(\sum_{r=0}^{\infty} \delta^r\ a_{it+r}\ \mathbf{z}_i(a_{-it+r}, \mathbf{x}_{t+r}) \mid a_{it} = 0, \mathbf{x}_t, \mathbf{P}\right)\end{aligned} \tag{28}$$

That is, $\tilde{\mathbf{z}}_i^{\mathbf{P}}(\mathbf{x}_t)$ is equal to the difference in the expected value of $\sum_{r=0}^{\infty} \delta^r$ a_{it+r} $\mathbf{z}_i(a_{-it+r}, \mathbf{x}_{t+r})$ when firm i chooses to be in the market at period t $(a_{it} = 1)$ and when it decides not to be in the market $(a_{it} = 0)$. $\tilde{e}_i^{\mathbf{P}}(\mathbf{x}_t)$ has a similar interpretation, but it applies to the component ε of the profit function and it includes only future values of ε and not the current value; that is:

$$\tilde{e}_i^{\mathbf{P}}(\mathbf{x}_t) = \mathbb{E}\left(\sum_{r=1}^{\infty} \delta^r\ \varepsilon_{it+r} \mid a_{it} = 1, \mathbf{x}_t, \mathbf{P}\right) - \mathbb{E}\left(\sum_{r=1}^{\infty} \delta^r\ \varepsilon_{it+r} \mid a_{it} = 0, \mathbf{x}_t, \mathbf{P}\right) \tag{29}$$

3.4 Data, Identification, and Estimation

3.4.1 Data

In most applications of dynamic games in empirical IO, the researcher observes a random sample of M markets, indexed by m, over T periods of time, where the observed variables consist of players' actions and state variables. In the standard application in IO, the values of N and T are small but M is large. Two aspects of the data warrant comment. For the moment, we consider that the industry and the data are such that (1) each firm is observed making decisions in every of the M markets; and (2) the researcher knows all of the payoff-relevant market characteristics that are common knowledge to the firms. We describe condition (1) as a dataset with *global players*. For instance, this is the case in a retail industry characterized by competition between large retail chains that are potential entrants in any of the local markets that constitute the industry. With this type of data, we can allow for rich-firm heterogeneity that is fixed across markets and time by estimating firm-specific structural parameters, $\boldsymbol{\theta}_i$. This "fixed-effect" approach to address firm heterogeneity is not feasible in datasets in which most of the competitors can be characterized as *local players* (i.e., firms specializing in operating in a few markets). Condition (2) rules out the existence of unobserved market heterogeneity. Although it is a convenient assumption, it also is unrealistic for most applications in empirical IO. In Section 3.5, we present estimation methods that relax conditions (1) and (2) and address unobserved market and firm heterogeneity.

3.4.2 Identification with Multiple Equilibria

Multiple equilibria are the rule rather than the exception in most dynamic games. We now discuss the implications of multiple equilibria for identification. Equilibrium uniqueness is neither necessary nor sufficient condition for the identification of a model (Jovanovic 1989). To see this, we consider a model with vector of structural parameters $\theta \in \Theta$ and define the mapping $C(\theta)$ from the set of parameters Θ to the set of measurable predictions of the model. Multiple equilibria imply that the mapping $C(.)$ is a correspondence. A model is not point-identified if at the observed data the inverse mapping C^{-1} is a correspondence. In general, C being a function (i.e., equilibrium uniqueness) is neither a necessary nor sufficient condition for C^{-1} being a function (i.e., for point identification).

To illustrate the identification of a game with multiple equilibria, we start with a simple binary-choice game with identical players in which the equilibrium probability P is implicitly defined as the solution of the condition $P = \Phi(-1.8 + \theta\ P)$, where θ is a structural parameter and $\Phi(.)$ is the CDF of the standard normal. We suppose that the true value θ_0 is 3.5. It is possible to verify that the set of equilibria associated with θ_0 is $C(\theta_0) = \{\ P^{(A)}(\theta_0) = 0.054,\ P^{(B)}(\theta_0) = 0.551, \text{and } P^{(C)}(\theta_0) = 0.924\}$. The game has been played M times and we observe players' actions for each realization of the game $\{a_{im} : i, m\}$. We let P_0 be the population probability $\Pr(a_{im} = 1)$. Without further assumptions, the probability P_0 can be estimated consistently from the data. For instance, a simple frequency estimator $\hat{P}_0 = (NM)^{-1} \sum_{i,m} a_{im}$ is a consistent estimator of P_0. Without further assumption, we do not know the relationship between population probability P_0 and the equilibrium probabilities in $C(\theta_0)$. If all of the sample observations come from the same equilibrium, then P_0 should be one of the points in $C(\theta_0)$. However, if the observations come from different equilibria in $C(\theta_0)$, then P_0 is a mixture of the elements in $C(\theta_0)$. To obtain identification, we can assume that every observation in the sample comes from the same equilibrium. Under this condition, because P_0 is an equilibrium associated with θ_0, we know that $P_0 = \Phi(-1.8 + \theta_0\ P_0)$. Given that $\Phi(.)$ is an invertible function, we have that $\theta_0 = (\Phi^{-1}(P_0) + 1.8)/P_0$.

Provided that P_0 is not zero, it is clear that θ_0 is point-identified regardless of the existence of multiple equilibria in the model.[31]

The basic idea in this example can be extended to obtain identification in our class of dynamic games. We make the following assumptions.

Assumption: *Single equilibrium in the data.* Every observation in the sample comes from the same equilibrium; that is, for any observation (m, t), $\mathbf{P}^0_{mt} = \mathbf{P}^0$.

Assumption: *No unobserved common-knowledge variables.* The only unobservables for the econometrician are the private-information shocks ε_{imt} and the structural parameters $\boldsymbol{\theta}$.

The distribution of ε_{imt} is known up to a scale parameter. For the purpose of concreteness, we consider that ε_{imt} is a normal random variable with zero mean and variance σ_i. Under these assumptions, the vector of population CCPs $\mathbf{P}^0$ is an equilibrium of the model associated with $\boldsymbol{\theta}^0$ (i.e., it is not a mixture of equilibria) and it is identified from the data. Because $\mathbf{P}^0$ is an equilibrium, the condition $P_i^0(\mathbf{x}_{mt}) = \Phi(\tilde{\mathbf{z}}_i^{\mathbf{P}^0}(\mathbf{x}_{mt})\,\boldsymbol{\theta}_i^0/\sigma_i + \tilde{e}_i^{\mathbf{P}^0}(\mathbf{x}_{mt}))$ is satisfied for any firm i and any state $\mathbf{x}_{mt}$. We can rewrite this equilibrium condition as a linear-in-parameters model $Y_{imt} = Z_{imt}\,\boldsymbol{\theta}_i^0/\sigma_i$ where $Y_{imt} \equiv \Phi^{-1}\left(P_i^0(\mathbf{x}_{mt})\right) - \tilde{e}_i^{\mathbf{P}^0}(\mathbf{x}_{mt})$ and $Z_{imt} \equiv \tilde{\mathbf{z}}_i^{\mathbf{P}^0}(\mathbf{x}_{mt})$. A necessary and sufficient condition for the identification of $\boldsymbol{\theta}_i^0/\sigma_i$ is that the variance–covariance matrix $\mathbb{E}(Z'_{imt}Z_{imt})$ is nonsingular or, equivalently, that the variables in Z_{imt} are not perfectly collinear. The variables in Z_{imt} are expected-present values of the variables in the one-period expected profit, $\mathbf{z}_i^{\mathbf{P}^0}(\mathbf{x}_{mt})$. In general, these variables and their expected-present values are not collinear; therefore, $\boldsymbol{\theta}_i^0/\sigma_i$ is identified. Under the single-equilibrium-in-the-data assumption, the multiplicity of equilibria in the model does not have any role in the identification of the structural parameters.

The assumption of a single equilibrium in the data is less restrictive than it may appear. The vector of observable state variables $\mathbf{x}_{mt}$ can include discrete and time-invariant market characteristics that we can use to define

[31] The single-equilibrium-in-the-data assumption has a key role in this identification result. We suppose that a fraction λ of the observations comes from the stable equilibrium $P^{(A)}(\theta_0) = 0.054$ and a fraction $1 - \lambda$ from the other stable equilibrium $P^{(C)}(\theta_0) = 0.924$, but the researcher does not know which observation comes from which equilibrium. Therefore, $P_0 = \lambda P^{(A)}(\theta_0) + (1 - \lambda) P^{(C)}(\theta_0)$. We note that P_0 is not an equilibrium of the model but rather a mixture of two equilibria. If the researcher ignores this mixture and imposes the assumption of a single-equilibrium-in-the-data assumption, the estimator of θ_0 will be inconsistent.

a finite number of *market types* (e.g., urban versus rural markets or markets in different geographic regions). As long as the observed number of market types does not increase with the number of markets M, we can allow for different equilibria at each market type. Therefore, the assumption of a single equilibrium in the data is basically a restriction on unobserved heterogeneity. Under this assumption, there is no unobserved heterogeneity that implies the selection of a different type of equilibrium in markets with the same observable characteristics (i.e., within the same observable type). We note that the single-equilibrium-in-the-data assumption is sufficient for identification but it is not necessary. Sweeting (2009), Aguirregabiria and Mira (2009), and De Paula and Tang (2012) presented conditions for the identification of static games of incomplete information when there are multiple equilibria in the data. These papers assumed that there is no payoff-relevant unobserved heterogeneity for the researcher. As far as we know, there are not yet identification results of games that allow for both payoff-relevant unobserved heterogeneity and multiple equilibria in the data (i.e., non–payoff-relevant unobserved heterogeneity).

3.4.3 Estimation: Maximum Likelihood and Two-Step Methods

In principle, estimation of dynamic games could follow the same methods as the estimation of single-agent dynamic structural models. For example, we could imagine using a nested fixed-point algorithm that maximizes a sample criterion function over the space of structural parameters and solves for the equilibrium of the model – assuming it is unique – for each trial value of the parameters. Although this approach has been successful in single-agent problems, it is problematic in games. The existence of multiple equilibria significantly increases the computational burden, especially if we use standard estimation methods such as maximum likelihood or GMM. In this section, we discuss how the literature addresses this issue.

The use of an "extended" or "pseudo" likelihood (or, alternatively, GMM criterion) function has an important role in the different estimation methods. For arbitrary values of the vector of structural parameters $\boldsymbol{\theta}$ and firms' strategies $\mathbf{P}$, we define the following likelihood function of observed players' actions $\{a_{imt}\}$ conditional on observed state variables $\{\mathbf{x}_{mt}\}$:

$$Q(\boldsymbol{\theta}, \mathbf{P}) = \sum_{i,m,t} a_{imt} \ln \Phi\left(\tilde{\mathbf{z}}^{\mathbf{P}}_{imt}\, \boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}}_{imt}\right) + (1 - a_{imt}) \ln \Phi\left(-\tilde{\mathbf{z}}^{\mathbf{P}}_{imt}\, \boldsymbol{\theta}_i - \tilde{e}^{\mathbf{P}}_{imt}\right) \tag{30}$$

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

where, for the porpose of concreteness, we consider that private-information shocks are normally distributed. For notational simplicity, we use $\boldsymbol{\theta}_i$ to represent $\boldsymbol{\theta}_i/\sigma_i$, $\tilde{\mathbf{z}}^{\mathbf{P}}_{imt} \equiv \tilde{\mathbf{z}}^{\mathbf{P}}_i(\mathbf{x}_{mt})$ and $\tilde{e}^{\mathbf{P}}_{imt} \equiv \tilde{e}^{\mathbf{P}}_i(\mathbf{x}_{mt})$. We call $Q(\boldsymbol{\theta}, \mathbf{P})$ a pseudolikelihood function because players' CCPs in $\mathbf{P}$ are arbitrary and do not represent the equilibrium probabilities associated with $\boldsymbol{\theta}$ implied by the model. An implication of using arbitrary instead of equilibrium CCPs is that likelihood Q is a function and not a correspondence.[32] To compute this pseudolikelihood, a useful construct is the representation of equilibrium in terms of CCPs, which we presented previously.

Full Maximum Likelihood. The dynamic game imposes the restriction that the strategies in $\mathbf{P}$ should be in equilibrium. The maximum likelihood (ML) estimator is defined as the pair $(\hat{\boldsymbol{\theta}}_{MLE}, \hat{\mathbf{P}}_{MLE})$ that maximizes the pseudolikelihood subject to the constraint that the strategies in $\hat{\mathbf{P}}_{MLE}$ are equilibrium strategies associated with $\hat{\boldsymbol{\theta}}_{MLE}$. That is:

$$\begin{gathered}(\hat{\boldsymbol{\theta}}_{MLE}, \hat{\mathbf{P}}_{MLE}) = \arg\max_{(\boldsymbol{\theta},\mathbf{P})} \; Q(\boldsymbol{\theta}, \mathbf{P}) \\ \text{s.t. } P_i(\mathbf{x}_{mt}) = \Phi\left(\tilde{\mathbf{z}}^{\mathbf{P}}_i(\mathbf{x}_{mt})\boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}}_i(\mathbf{x}_{mt})\right) \text{ for any } (i, \mathbf{x}_{mt}) \in I \times \mathcal{X}\end{gathered} \tag{31}$$

This is a constrained ML estimator that satisfies the standard regularity conditions for consistency, asymptotic normality, and efficiency of ML estimation.[33] The numerical solution of the constrained-optimization problem that defines these estimators requires us to search over an extremely large dimensional space. In the empirical applications of dynamic-oligopoly games, the vector of probabilities $\mathbf{P}$ includes thousands or millions of elements. Searching for an optimum in that type of space is computationally demanding. Su and Judd (2012) proposed to use a Mathematical Programming with Equilibrium Constraints (MPEC) algorithm, which is a general-purpose algorithm for the numerical solution of constrained-optimization problems. However, even using the most sophisticated algorithm such as MPEC, the optimization with respect to $(\mathbf{P}, \boldsymbol{\theta})$ can be extremely demanding when $\mathbf{P}$ has a high dimension.

Two-Step Methods. To avoid this large computational cost, alternative two-step methods were explored. In this class of models, for given $\mathbf{P}$, the best-response probability function $G_i(\tilde{\mathbf{z}}^{\mathbf{P}}_i(\mathbf{x}_t)\,\boldsymbol{\theta}_i + \tilde{e}^{\mathbf{P}}_{imt})$ has the structure in

[32] In fact, Q is the likelihood function of a standard probit model for a_{imt} with explanatory variables $\tilde{\mathbf{z}}^{\mathbf{P}}_{imt}$ and $\tilde{e}^{\mathbf{P}}_{imt}$.

[33] Similarly, we could define a pseudo-GMM criterion function. The GMM estimator is defined as the pair $(\hat{\boldsymbol{\theta}}_{GMM}, \hat{\mathbf{P}}_{GMM})$ that minimizes the pseudo-GMM criterion subject to the constraint that the strategies in $\hat{\mathbf{P}}_{GMM}$ are equilibrium strategies associated with $\hat{\boldsymbol{\theta}}_{GMM}$.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

a standard binary-choice model with an index that is linear in parameters. A pseudolikelihood function based on these best-response probabilities is globally concave in the structural parameters; therefore, optimization of $Q(\boldsymbol{\theta}, \mathbf{P})$ with respect to $\boldsymbol{\theta}$ for given $\mathbf{P}$ is a simple task. Furthermore, the multiplicative separability of the valuation operator $\mathbf{V}_i^{\mathbf{P}}$ in the structural parameters[34] implies that although it is costly to compute $\mathbf{V}_i^{\mathbf{P}}$ for multiple values of $\mathbf{P}$, it is much cheaper to compute it for multiple values of $\boldsymbol{\theta}_{\mathbf{i}}$. Two-step estimation methods exploit this particular structure of the model.

We let $\mathbf{P}^0$ be the vector with the population values of the probabilities $P_i^0(\mathbf{x}) \equiv \Pr(a_{imt} = 1|\mathbf{x}_{mt} = \mathbf{x})$ for every firm i and any value of $\mathbf{x}$. Under the assumptions of "no unobserved common knowledge variables" and "single equilibrium in the data," the CCPs in $\mathbf{P}^0$ also represent firms' strategies in the only equilibrium that is played in the data. These probabilities can be estimated consistently using standard nonparametric methods. We let $\hat{\mathbf{P}}^0$ be a consistent nonparametric estimator of $\mathbf{P}^0$. The two-step estimator of $\boldsymbol{\theta}^0$ is defined as $\hat{\boldsymbol{\theta}}_{2S} = \arg\max_{\boldsymbol{\theta}}\ Q(\boldsymbol{\theta}, \hat{\mathbf{P}}^0)$. Under standard regularity conditions, this two-step estimator is root-M consistent and asymptotically normal. This idea originally was exploited for estimation of single-agent problems by Hotz and Miller (1993) and Hotz, Miller, Sanders, and Smith (1994). It was expanded to the estimation of dynamic games by Aguirregabiria and Mira (2009); Bajari, Benkard, and Levin (2007); Pakes, Ostrovsky, and Berry (2007); and Pesendorfer and Schmidt-Dengler (2008). Pesendorfer and Schmidt-Dengler (2008) showed that different estimators can be described as a general class of two-step estimators of dynamic games. An estimator within this class can be described using the following minimum distance (or asymptotic least squares) approach:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\ \left[\hat{\mathbf{P}}^0 - \Psi\left(\hat{\mathbf{P}}^0, \boldsymbol{\theta}\right)\right]' \mathbf{A}_M \left[\hat{\mathbf{P}}^0 - \Psi\left(\hat{\mathbf{P}}^0, \boldsymbol{\theta}\right)\right] \tag{32}$$

where $\mathbf{A}_M$ is a weighting matrix. Each estimator within this general class is associated with a particular choice of the weighting matrix. The asymptotically optimal estimator within this class has a weighting matrix equal to the inverse of $[\mathbf{I} - \partial\Psi(\mathbf{P}^0, \boldsymbol{\theta}^0)/\partial\mathbf{P}']'\ \Sigma_{\hat{\mathbf{P}}}\ [\mathbf{I} - \partial\Psi(\mathbf{P}^0, \boldsymbol{\theta}^0)/\partial\mathbf{P}']$, where $\Sigma_{\hat{\mathbf{P}}}$ is the variance matrix of the initial nonparametric estimator $\hat{\mathbf{P}}^0$. Pesendorfer and Schmidt-Dengler (2008) showed that this estimator is asymptotically equivalent to the maximum ML estimator. Therefore, there is no loss of

[34] The vector of values $\mathbf{V}_i^{\mathbf{P}}$ is equal to $\mathbf{W}_{z,i}^{\mathbf{P}}\ \boldsymbol{\theta}_i + \mathbf{W}_{e,i}^{\mathbf{P}}\ \sigma_i$, where $\mathbf{W}_i^{\mathbf{P}}$ is the matrix $\left(\mathbf{I} - \delta\,\mathbf{F}^{\mathbf{P}}\right)^{-1} P_i * \mathbf{Z}_i^{\mathbf{P}}$ and $\mathbf{W}_{e,i}^{\mathbf{P}}$ is the vector $\left(\mathbf{I} - \delta\,\mathbf{F}^{\mathbf{P}}\right)^{-1} P_i * e_i^{\mathbf{P}}$. Calculating $\mathbf{W}_{z,i}^{\mathbf{P}}$ and $\mathbf{W}_{e,i}^{\mathbf{P}}$ is significantly simpler than solving for the dynamic-progamming problem of one player and much simpler than solving for an equilibrium of the dynamic game.

asymptotic efficiency by using a two-step estimator of the structural parameters instead of the MLE.

The main advantage of these two-step estimators is their computational simplicity. The first step is a simple nonparametric regression; the second step is the estimation of a standard discrete-choice model with a criterion function that in most applications is globally concave (e.g., the likelihood of a standard probit model in the entry–exit example). The main computational burden comes from the calculation of the present values $W^{\hat{\mathbf{P}}}_{z,i}(\mathbf{x})$ and $W^{\hat{\mathbf{P}}}_{e,i}(\mathbf{x})$. Although the computation of these present values may be subject to a curse of dimensionality (see Section 3.6), the cost of obtaining a two-step estimator is several orders of magnitude smaller than solving (just once) for an equilibrium of the dynamic game. In most applications, this makes the difference between being able to estimate the model or not.

These two-step estimators have important limitations. First is the restriction imposed by the assumption of no unobserved common-knowledge variables. Ignoring persistent unobservables, if present, can generate important biases in the estimation of structural parameters. We discuss this issue in Section 3.5.

A second problem is finite-sample bias. Whereas two-step CCP methods address the curse of dimensionality in the solution of dynamic games, they also create a different curse of dimensionality that is not present in full-solution maximum likelihood or GMM methods. This is the so-called curse of dimensionality in nonparametric estimation. As the dimension of the state space increases, so does the asymptotic variance and finite-sample variance and bias of the initial nonparametric estimator of CCPs.[35] The initial nonparametric estimator can be imprecise in the samples available in actual applications, which can generate serious finite-sample biases in the two-step estimator of structural parameters. The source of this bias is well understood in two-step methods: $\hat{\mathbf{P}}$ enters nonlinearly in the sample-moment conditions that define the estimator, and the expected value of a nonlinear function of $\hat{\mathbf{P}}$ is not equal to that function evaluated at the expected value of $\hat{\mathbf{P}}$. The larger the variance or the bias of $\hat{\mathbf{P}}$, the larger is the bias of the two-step estimator of $\boldsymbol{\theta}_0$. This problem is particularly serious in dynamic

[35] If the model includes state variables with continuous support, the rate of convergence of the nonparametric estimator of CCPs declines with the number of continuous state variables. However, in this class of models and under standard regularity conditions, the lower rate of convergence of the initial nonparametric estimator of CCP functions does not affect the root-M consistency of the estimator of structural parameters in the second step (Kasahara and Shimotsu 2008a; Linton and Srisuma 2012). Of course, the slower rate of convergence of the initial nonparametric estimator has important implications on the finite-sample properties of the two-step estimator.

games with heterogeneous players because in these models, the number of observable state variables is proportional to the number of players in the game.[36]

3.4.4 Recursive K-Step Estimators

To address finite sample bias, Aguirregabiria and Mira (2002, 2009) considered a recursive K-step extension. Given the two-step estimator $\hat{\boldsymbol{\theta}}_{2S}$ and the initial nonparametric estimator of CCPs, $\hat{\mathbf{P}}^0$, we can construct a new estimator of CCPs, $\hat{\mathbf{P}}^1$, such that $\hat{P}_i^1(\mathbf{x}) = \Phi\left(\tilde{\mathbf{z}}_i^{\hat{\mathbf{P}}^0}(\mathbf{x})\, \hat{\boldsymbol{\theta}}_{i,2S} + \tilde{e}_i^{\hat{\mathbf{P}}^0}(\mathbf{x})\right)$. This estimator exploits the parametric structure of the model and the structure of best-response functions. It seems intuitive that this new estimator of CCPs has better statistical properties than the initial nonparametric estimator (i.e., smaller asymptotic variance and smaller finite-sample bias and variance). As explained herein, this intuition is correct as long as the equilibrium that generated the data is Lyapunov stable. Under this condition, it seems natural to obtain a new two-step estimator by replacing $\hat{\mathbf{P}}^0$ with $\hat{\mathbf{P}}^1$ as the estimator of CCPs. The same argument can be applied recursively to generate a sequence of K-step estimators. Given an initial consistent nonparametric estimator $\hat{\mathbf{P}}^0$, the sequence of estimators $\{\hat{\boldsymbol{\theta}}^K, \hat{\mathbf{P}}^K : K \geq 1\}$ is defined as $\hat{\boldsymbol{\theta}}^K = \arg\max_{\boldsymbol{\theta}}\ Q(\boldsymbol{\theta}, \hat{\mathbf{P}}^{K-1})$, where $\hat{\mathbf{P}}^K = \Psi(\hat{\mathbf{P}}^{K-1}, \hat{\boldsymbol{\theta}}^K)$. To study the properties of these K-step estimators, it is convenient to represent the sequence $\{\hat{\mathbf{P}}^K : K \geq 1\}$ as the result of iterating in a fixed-point mapping. For arbitrary $\mathbf{P}$, we define the mapping $\varphi(\mathbf{P}) \equiv \Psi(\mathbf{P}, \hat{\boldsymbol{\theta}}(\mathbf{P}))$, where $\hat{\boldsymbol{\theta}}(\mathbf{P}) \equiv \arg\max_{\boldsymbol{\theta}}\ Q(\boldsymbol{\theta}, \mathbf{P})$. The mapping $\varphi(\mathbf{P})$ is called the nested pseudo-likelihood (NPL) mapping. The sequence of estimators $\{\hat{\mathbf{P}}^K : K \geq 1\}$ can be obtained by successive iterations in the mapping φ starting with the nonparametric estimator $\hat{\mathbf{P}}^0$ (i.e., for $K \geq 1$, $\hat{\mathbf{P}}^K = \varphi(\hat{\mathbf{P}}^{K-1})$).

Monte Carlo experiments in Aguirregabiria and Mira (2002, 2009) and Kasahara and Shimotsu (2008a, 2009) showed that iterating in the NPL

[36] The asymptotically efficient two-step estimator proposed by Pesendorfer and Schmidt-Dengler (2008) did not address the finite-sample-bias problem. In fact, as it is well known in the literature of covariance-structure models, the finite-sample bias of this asymptotically optimal estimator can be significantly more severe than the standard two-step estimator because the estimation of the optimal weighting matrix also is contaminated by the imprecise nonparametric estimator and it contributes to increase the finite-sample bias (see Altonji and Segal 1996 and Horowitz 1998). In the context of two-step estimation of dynamic games, Pakes, Ostrovsky, and Berry (2007) presented Monte Carlo experiments supporting this concern.

mapping can reduce significantly the finite-sample bias of the two-step estimator. The Monte Carlo experiments in Pesendorfer and Schmidt-Dengler (2008) presented a different, more mixed picture. Whereas for some of their experiments, NPL iteration reduced the bias, in other experiments, the bias remained constant or even increased. The Monte Carlo experiments in Pesendorfer and Schmidt-Dengler (2008) showed that the NPL iterations provide poor results in those cases in which the equilibrium that generates the data is not Lyapunov stable. As explained herein, this is not a coincidence. It turns out that the computational and statistical properties of the sequence of K-step estimators depend critically on the stability of the NPL mapping around the equilibrium in the data. Lyapunov stability of the NPL mapping also is important for the properties of the methods proposed so far to address unobserved heterogeneity in the estimation of dynamic games (see Section 3.5). Therefore, it is important to analyze this issue in more detail here.

Lyapunov Stability. We let $\mathbf{P}^*$ be a fixed point of the NPL mapping such that $\mathbf{P}^* = \varphi(\mathbf{P}^*)$. We say that the mapping φ is Lyapunov stable around the fixed point $\mathbf{P}^*$ if there is a neighborhood of $\mathbf{P}^*$, $\mathcal{N}$, such that successive iterations in the mapping φ starting at $\mathbf{P} \in \mathcal{N}$ converge to $\mathbf{P}^*$. A necessary and sufficient condition for Lyapunov stability is that the *spectral radius* of the Jacobian matrix $\partial\varphi(\mathbf{P}^*)/\partial\mathbf{P}'$ is smaller than 1. The neighboring set $\mathcal{N}$ is denoted the dominion of attraction of the fixed point $\mathbf{P}^*$.[37] Similarly, if $\mathbf{P}^*$ is an equilibrium of the mapping $\Psi(., \boldsymbol{\theta})$, we say that this mapping is Lyapunov stable around $\mathbf{P}^*$ if and only if (iff) the spectral radius of the Jacobian matrix $\partial\Psi(\mathbf{P}^*, \boldsymbol{\theta})/\partial\mathbf{P}'$ is smaller than 1.

There is a relationship between the stability of the NPL mapping and the equilibrium mapping $\Psi\left(., \boldsymbol{\theta}^0\right)$ around $\mathbf{P}^0$ (i.e., the equilibrium that generates the data). The Jacobian matrices of the NPL and equilibrium mapping are related by the following expression (Kasahara and Shimotsu 2009): $\partial\varphi\left(\mathbf{P}^0\right)/\partial\mathbf{P}' = M(\mathbf{P}^0)\,\partial\Psi(\mathbf{P}^0, \boldsymbol{\theta}^0)/\partial\mathbf{P}'$, where $M(\mathbf{P}^0)$ is an idempotent projection matrix.[38] In single-agent dynamic-programming models, the Jacobian matrix $\partial\Psi\left(\mathbf{P}^0, \boldsymbol{\theta}^0\right)/\partial\mathbf{P}'$ is zero (i.e., the zero Jacobian matrix property; Aguirregabiria and Mira 2002). Therefore, for that class of models, $\partial\varphi\left(\mathbf{P}^0\right)/\partial\mathbf{P}' = 0$ and the NPL mapping is Lyapunov stable around

[37] The spectral radius of a matrix is the maximum absolute eigenvalue. If the mapping φ is twice continuously differentiable, then the spectral radius is a continuous function of $\mathbf{P}$. Therefore, if φ is Lyapunov stable at $\mathbf{P}^*$, for any $\mathbf{P}$ in the dominion of attraction of $\mathbf{P}^*$, we have that the spectral radius of $\partial\varphi(\mathbf{P})/\partial\mathbf{P}'$ also is smaller than 1.

[38] The idempotent matrix $M(\mathbf{P}^0)$ is $I - \Psi_\theta(\Psi_\theta' \; diag\{\mathbf{P}^0\}^{-1} \; \Psi_\theta)^{-1}\Psi_\theta' \; diag\{\mathbf{P}^0\}^{-1}$, where $\Psi_\theta \equiv \partial\Psi(\mathbf{P}^0, \boldsymbol{\theta}^0)/\partial\theta'$.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

$\mathbf{P}^0$. In dynamic games, $\partial\Psi\left(\mathbf{P}^0, \boldsymbol{\theta}^0\right)/\partial\mathbf{P}'$ is not zero. However, Aguirregabiria and Mira (2011) showed that a strong stability condition of $\mathbf{P}^0$ in the equilibrium mapping Ψ implies Lyapunov stability of $\mathbf{P}^0$ in the NPL mapping φ.

Convergence of NPL Iterations. We suppose that the true equilibrium in the population, $\mathbf{P}^0$, is Lyapunov stable with respect to the NPL mapping. This implies that with probability approaching 1, as M goes to infinity, the sample NPL mapping is stable around a consistent nonparametric estimator of $\mathbf{P}^0$. Therefore, the sequence of K-step estimators converges to a limit $\hat{\mathbf{P}}^0_{\text{lim}}$ that is a fixed point of the NPL mapping; that is, $\hat{\mathbf{P}}^0_{\text{lim}} = \varphi(\hat{\mathbf{P}}^0_{\text{lim}})$. It is possible to show that this limit $\hat{\mathbf{P}}^0_{\text{lim}}$ is a consistent estimator of $\mathbf{P}^0$ (Kasahara and Shimotsu 2009). Therefore, under Lyapunov stability of the NPL mapping, if we begin with a consistent estimator of $\mathbf{P}^0$ and iterate in the NPL mapping, we converge to a consistent estimator that is an equilibrium of the model. It is possible to show that this estimator is asymptotically more efficient than the two-step estimator (Aguirregabiria and Mira 2007).

Pesendorfer and Schmidt-Dengler (2010) presented an example in which the sequence of K-step estimators converges to a limit estimator that is not consistent. As implied by the results presented herein, the equilibrium that generated the data in their example is not Lyapunov stable.

The concept of Lyapunov stability of the best-response mapping at an equilibrium means that if we marginally perturb players' strategies and then allow players to best-respond to the new strategies, then we converge to the original equilibrium. To us, this seems like a plausible equilibrium-selection criterion. However, it should be clear that Lyapunov stability of an equilibrium is not a regularity condition but rather a testable restriction that we can impose (or not) in our model. Ultimately, whether an unstable equilibrium is interesting depends on the application and the researchers' taste. Nevertheless, at the end of this section, we present simple modified versions of the NPL method that can address data generated from an equilibrium that is not stable.

Reduction of Finite-Sample Bias. Kasahara and Shimotsu (2008a, 2009) derived a second-order approximation to the bias of the K-step estimators. They showed that the key component in this bias is the distance between the first-step and the second-step estimators of $\mathbf{P}^0$ (i.e., $\|\varphi\left(\hat{\mathbf{P}}^0\right) - \hat{\mathbf{P}}^0\|$). An estimator that reduces this distance is an estimator with lower finite-sample bias. Therefore, based on our previous discussion, the sequence of K-step estimators is decreasing in their finite-sample bias iff the NPL mapping is Lyapunov stable around $\mathbf{P}^0$ (see Lemma 2 in Kasahara and Shimotsu 2009).

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

The Monte Carlo experiments in Pesendorfer and Schmidt-Dengler (2008) illustrated this point. They implemented experiments using different Data Generating Process (DGP) in some, the data are generated from a stable equilibrium; in others, the data come from a nonstable equilibrium. It is simple to verify (Aguirregabiria and Mira 2011) that the experiments in which NPL iterations do not reduce the finite-sample bias are those in which the equilibrium that generates the data is not Lyapunov stable.

Modified NPL Algorithms. We note that Lyapunov stability can be tested after obtaining the first NPL iteration. Once we have obtained the two-step estimator, we can calculate the Jacobian matrix $\partial\varphi(\hat{\mathbf{P}}^0)/\partial\mathbf{P}'$ and its eigenvalues and then check whether Lyapunov stability holds at $\hat{\mathbf{P}}^0$. If the applied researcher considers that his data may have been generated by an equilibrium that is not stable, then it is worthwhile to compute this Jacobian matrix and its eigenvalues. If Lyapunov stability holds at $\hat{\mathbf{P}}^{\mathbf{0}}$, then we know that NPL iterations reduce the bias of the estimator and converge to a consistent estimator.[39]

When the condition does not hold, the solution to this problem is not simple. Although the researcher may choose to use the two-step estimator, the nonstability of the equilibrium also has important negative implications on the properties of this simple estimator.[40] In this context, Kasahara and Shimotsu (2009) proposed alternative recursive estimators based on fixed-point mappings other than the NPL that, by construction, are stable. Iterating in these alternative mappings is significantly more costly than iterating in the NPL mapping, but these iterations guarantee reduction of the finite-sample bias and convergence to a consistent estimator.

Aguirregabiria and Mira (2011) proposed two modified versions of the NPL algorithm that are simple to implement and that always converge to a consistent estimator with better properties than two-step estimators. The

[39] We note that stability testing followed by NPL estimation amounts to a pretest estimator, possibly with different statistical properties than simply applying NPL estimation without a pretest. As far as we know, the properties of this particular type of pretest estimator have not been studied yet in this literature.

[40] Nonstability of the NPL mapping at $\mathbf{P}^0$ implies that the asymptotic variance of the two-step estimator of $\mathbf{P}^0$ is larger than asymptotic variance of the nonparametric reduced-form estimator. To see this, we note that the two-step estimator of CCPs is $\hat{\mathbf{P}}^1 = \varphi(\hat{\mathbf{P}}^0)$; applying the delta method, we have that $Var(\hat{\mathbf{P}}^1) = [\partial\varphi\left(\mathbf{P}^0\right)/\partial\mathbf{P}']\ Var(\hat{\mathbf{P}}^{\mathbf{0}})\ [\partial\varphi\left(\mathbf{P}^0\right)/\partial\mathbf{P}']'$. If the spectral radius of $\partial\varphi\left(\mathbf{P}^0\right)/\partial\mathbf{P}'$ is greater than 1, then $Var(\hat{\mathbf{P}}^1) > Var(\hat{\mathbf{P}}^0)$. This is a puzzling result because the estimator $\hat{\mathbf{P}}^0$ is nonparametric, whereas the estimator $\hat{\mathbf{P}}^1$ exploits most of the structure of the model. Therefore, the nonstability of the equilibrium that generates the data is an issue for this general class of two-step or sequential estimators.

first modified-NPL algorithm applies to dynamic games. The first NPL iteration is standard but, in every successive iteration, best-response mappings are used to update guesses of each player's own future behavior without updating beliefs about the strategies of the other players. This algorithm always converges to a consistent estimator, even if the equilibrium generating the data is not stable and it reduces monotonically the asymptotic variance and the finite-sample bias of the two-step estimator. The *second modified-NPL algorithm* applies to static games and it consists in the application of the standard NPL algorithm, to both the best-response mapping and the inverse of this mapping. If the equilibrium that generates the data is unstable in the best-response mapping, it should be stable in the inverse mapping. Therefore, the NPL applied to the inverse mapping should converge to the consistent estimator and should have the largest value of the pseudolikelihood that the estimator to which we converge when applying the NPL algorithm to the best-response mapping. Aguirregabiria and Mira (2010) illustrated the performance of these estimators using the examples in Pesendorfer and Schmidt-Dengler (2008, 2010).

3.5 Addressing Unobserved Heterogeneity

So far, we maintain the assumption that the only unobservables for the researcher are the private-information shocks that are i.i.d. over firms, markets, and time. In most applications in IO, this assumption is not realistic and it can be rejected easily by the data. Markets and firms are heterogeneous in terms of characteristics that are payoff-relevant for firms but unobserved to the researcher. Not accounting for this heterogeneity may generate significant biases in parameter estimates and in our understanding of competition in the industry. For instance, Aguirregabiria and Mira (2007) and Collard-Wexler (2006) presented empirical applications of dynamic games and compared the estimation results with and without controlling for unobserved market heterogeneity. They found that the model without unobserved market heterogeneity implies estimates of strategic interaction between firms (i.e., competition effects) that are close to zero – or even have the opposite sign than the one expected under competition. In both applications, including unobserved heterogeneity in the models results in estimates that show significant and strong competition effects.

Aguirregabiria and Mira (2007); Aguirregabiria, Mira, and Roman (2007); and Arcidiacono and Miller (2011) proposed methods for the estimation of dynamic games that allow for persistent unobserved heterogeneity in players or markets. Here, we concentrate on the case of permanent unobserved market heterogeneity in the profit function. Arcidiacono and Miller (2011) proposed a method that combines the NPL method, presented herein, with an expectation-maximization (EM) algorithm, and they considered a more general framework that included unobserved heterogeneity that can vary over time according to the Markov chain process and that can enter in both the payoff function and the transition of state variables.[41]

We consider our entry – exit model, but now the profit of firm i, if active in market m, includes a term ξ_m that is unobserved to the researcher:

$$\Pi_{imt} = H_{mt} \sum_{n=0}^{N-1} 1\left\{\sum_{j\neq i} a_{jmt} = n\right\} \theta_{i,n}^{VP} - \theta_i^{FC} - (1 - s_{imt})\theta_i^{EC} - \sigma_{\xi i}\,\xi_m - \varepsilon_{imt} \tag{33}$$

where $\sigma_{\xi i}$ is a parameter and ξ_m is a time-invariant "random effect" that is common knowledge to the players but unobserved to the researcher.[42] The distribution of this random effect has the following properties:

(A.1) It has a discrete and finite support $\left\{\xi^1, \xi^2, \ldots, \xi^L\right\}$, each value in the support of ξ represents a "market type," and we index market types by $\ell \in \{1, 2, \ldots, L\}$.
(A.2) It is i.i.d. over markets with probability-mass function $\lambda_\ell \equiv \Pr(\xi_m = \xi^\ell)$.
(A.3) It does not enter into the transition probability of the observed state variables (i.e., $\Pr(\mathbf{x}_{mt+1} \mid \mathbf{x}_{mt}, \mathbf{a}_{mt}, \xi_m) = F_x(\mathbf{x}_{mt+1} \mid \mathbf{x}_{mt}, \mathbf{a}_{mt})$).

Without loss of generality, ξ_m has mean zero and unit variance because the mean and the variance of ξ_m are incorporated in the parameters θ_i^{FC} and $\sigma_{\xi i}$, respectively. Also, without loss of generality, the researcher knows the points of support $\left\{\xi^\ell : \ell = 1, 2, \ldots, L\right\}$ although the probability mass function $\{\lambda_\ell\}$ is unknown.

Assumptions (A.1) and (A.2) define a *finite-mixture model*.[43] Assumption (A.1) is common when dealing with permanent unobserved heterogeneity in dynamic structural models. The discrete support of the unobservable

[41] In fact, the framework that we present herein can be generalized to include unobserved market heterogeneity that varies over time according to a Markov chain with finite support.
[42] In this example, we include unobserved heterogeneity only in the fixed cost. However, the estimation methods presented here can address richer forms of unobserved heterogeneity (e.g., in fixed costs, variable profits, and entry costs).
[43] A finite mixture is a general class of semiparametric model for distribution of a random variable that is convenient for its flexibility and simplicity (McLachlan and Peel 2000). In econometrics, the influential work of Heckman and Singer (1984) made finite mixtures a useful tool to incorporate time-invariant unobserved heterogeneity in panel data and duration models. More closely related to the literature in this paper, Wolpin et al. estimated

implies that the contribution of a market to the likelihood (or pseudolikelihood) function is a finite mixture of likelihoods under the different possible best responses that we would have for each possible market type. With continuous support, we would have an infinite mixture of best responses, which could complicate significantly the computation of the likelihood. Nevertheless, as we illustrate herein, using a pseudolikelihood approach and a convenient parametric specification of the distribution of ξ_m simplifies this computation such that we can consider many values in the support of this unobserved variable at a low computational cost. Assumption (A.2) also is standard when addressing unobserved heterogeneity. Unobserved spatial correlation across markets does not generate inconsistency of the estimators that we present here because the likelihood equations that define the estimators are still valid moment conditions under spatial correlation. Incorporating spatial correlation in the model, if present in the data, would improve the efficiency of the estimator but at a significant computational cost. Assumption (A.3) can be relaxed; in fact, the method by Arcidiacono and Miller (2011) addressed unobserved heterogeneity in both payoffs and transition probabilities.

Each market type ℓ has its own equilibrium mapping (with a different level of profits given ξ^{ℓ}) and its own equilibrium. We let $\mathbf{P}_{\ell}$ be a vector of strategies (i.e, CCPs) in market-type ℓ: $\mathbf{P}_{\ell} \equiv \{P_{i\ell}(\mathbf{x}_t) : i = 1, 2, \ldots, N; \mathbf{x}_t \in \mathcal{X}\}$.[44] It is straightforward to extend the description of an equilibrium mapping in CCPs to this model. A vector of CCPs $\mathbf{P}_{\ell}$ is a MPE for market type ℓ iff for every firm i and every state $\mathbf{x}_t$ we have that $P_{i\ell}(\mathbf{x}_t) = \Phi\left(\tilde{\mathbf{z}}_i^{\mathbf{P}_\ell}(\mathbf{x}_t, \xi^{\ell})\,\boldsymbol{\theta}_i + \tilde{e}_i^{\mathbf{P}_\ell}(\mathbf{x}_t, \xi^{\ell})\right)$, where now the vector of structural parameters $\boldsymbol{\theta}_i$ is $\left\{\theta_{i,0}^{VP}, \ldots, \theta_{i,N-1}^{VP}, \theta_i^{FC}, \theta_i^{EC}, \sigma_{\xi i}\right\}$ that includes $\sigma_{\xi i}$ and the vector $\tilde{\mathbf{z}}_i^{\mathbf{P}_\ell}(\mathbf{x}_t, \xi^{\ell})$ has a similar definition as before, with the only difference that it has one more component associated with $-\xi^{\ell}$. Because the points of support $\left\{\xi^{\ell} : \ell = 1, 2, \ldots, L\right\}$ are known to the researcher, he can construct the equilibrium mapping for each market type.

We let λ be the vector of parameters in the probability-mass function of ξ (i.e., $\lambda \equiv \{\lambda_{\ell} : \ell = 1, 2, \ldots, L\}$), and we let $\mathbf{P}$ be the set of CCPs for every market type, $\{\mathbf{P}_{\ell} : \ell = 1, 2, \ldots, L\}$. The (conditional) pseudolog-likelihood function of this model is $Q(\boldsymbol{\theta}, \lambda, \mathbf{P}) = \sum_{m=1}^{M} \log \Pr(\mathbf{a}_{m1}, \mathbf{a}_{m2},$

dynamic discrete-choice structural models with a rich structure of unobserved heterogeneity with a finite-mixture structure (Eckstein and Wolpin 1990; Keane and Wolpin 1997).

[44] The introduction of unobserved market heterogeneity also implies that we can relax the assumption of only "a single equilibrium in the data" to allow for different market types to have different equilibria.

$\ldots, \mathbf{a}_{mT} \mid \mathbf{x}_{m1}, \mathbf{x}_{m2}, \ldots, \mathbf{x}_{mT}; \boldsymbol{\theta}, \lambda, \mathbf{P})$. We can write this function as $\sum_{m=1}^{M} \log q_m(\boldsymbol{\theta}, \lambda, \mathbf{P})$, where $q_m(\boldsymbol{\theta}, \lambda, \mathbf{P})$ is the contribution of market m to the pseudolikelihood:

$$q_m(\boldsymbol{\theta}, \boldsymbol{\lambda}, \mathbf{P}) = \sum_{\ell=1}^{L} \lambda_{\ell|\mathbf{x}_{m1}} \left[\prod_{i,t} \Phi\left(\tilde{\mathbf{z}}_{im\ell t}^{\mathbf{P}_\ell}\, \boldsymbol{\theta}_i + \tilde{e}_{im\ell t}^{\mathbf{P}_\ell}\right)^{a_{imt}} \Phi\left(-\tilde{\mathbf{z}}_{im\ell t}^{\mathbf{P}_\ell}\, \boldsymbol{\theta}_i - \tilde{e}_{im\ell t}^{\mathbf{P}_\ell}\right)^{1-a_{imt}} \right] \tag{34}$$

where $\tilde{\mathbf{z}}_{im\ell t}^{\mathbf{P}_\ell} \equiv \tilde{\mathbf{z}}_{i}^{\mathbf{P}_\ell}(\mathbf{x}_{mt}, \xi^{\ell})$, $\tilde{e}_{im\ell t}^{\mathbf{P}_\ell} \equiv \tilde{e}_{i}^{\mathbf{P}_\ell}(\mathbf{x}_{mt}, \xi^{\ell})$, and $\lambda_{\ell|\mathbf{x}}$ is the conditional probability $\Pr(\xi_m = \xi^{\ell} | \mathbf{x}_{m1} = \mathbf{x})$. The conditional-probability distribution $\lambda_{\ell|x}$ is different than the unconditional distribution λ_ℓ. In particular, ξ_m is not independent of the predetermined endogenous state variables that represent market structure. For instance, we expect a negative correlation between the indicators of incumbent status, s_{imt}, and the unobserved component of the fixed cost ξ_m (i.e., markets in which it is more costly to operate tend to have a smaller number of incumbent firms). This is the so-called initial-conditions problem (Heckman 1981). In short panels (for T relatively small), not considering this dependence between ξ_m and $\mathbf{x}_{m1}$ can generate significant biases, similar to those associated with ignoring the existence of unobserved market heterogeneity. There are different ways to address the initial-conditions problem in dynamic models (Heckman 1981). One possible approach is to derive the joint distribution of $\mathbf{x}_{m1}$ and ξ_m implied by the equilibrium of the model. That is the approach proposed and applied in Aguirregabiria and Mira (2007) and Collard Wexler (2006). We let $\mathbf{p}^{\mathbf{P}_\ell} \equiv \{p^{\mathbf{P}_\ell}(\mathbf{x}_t) : \mathbf{x}_t \in \mathcal{X}\}$ be the ergodic or steady-state distribution of $\mathbf{x}_t$ induced by the equilibrium $\mathbf{P}_\ell$ and the transition F_x. This stationary distribution can be obtained simply as the solution to the following system of linear equations: For every value $\mathbf{x}_t \in \mathcal{X}$, $p^{\mathbf{P}_\ell}(\mathbf{x}_t) = \sum_{\mathbf{x}_{t-1} \in \mathcal{X}} p^{\mathbf{P}_\ell}(\mathbf{x}_{t-1}) F_x^{\mathbf{P}_\ell}(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ or, in vector form, $\mathbf{p}^{\mathbf{P}_\ell} = \mathbf{F}_x^{\mathbf{P}_\ell}\, \mathbf{p}^{\mathbf{P}_\ell}$ subject to $\mathbf{p}^{\mathbf{P}_\ell\prime}\mathbf{1} = 1$. Given the ergodic distributions for the L market types, we can apply Bayes', rule to obtain:

$$\lambda_{\ell|\mathbf{x}_{m1}} = \frac{\lambda_\ell\; p^{\mathbf{P}_\ell}(\mathbf{x}_{m1})}{\sum_{\ell'=1}^{L} \lambda_{\ell'}\; p^{\mathbf{P}_{\ell'}}(\mathbf{x}_{m1})} \tag{35}$$

We note that given the CCPs $\{\mathbf{P}_\ell\}$, this conditional distribution does not depend on parameters in the vector $\boldsymbol{\theta}$, only on the distribution λ. Given this expression for the probabilities $\{\lambda_{\ell|\mathbf{x}_{m1}}\}$, we have that the pseudolikelihood

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

in Equation (34) depends on only the structural parameters $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ and the incidental parameters $\mathbf{P}$.

For the estimators discussed here, we maximize $Q(\boldsymbol{\theta},\lambda,\mathbf{P})$ with respect to $(\boldsymbol{\theta},\lambda)$ for given $\mathbf{P}$. Therefore, the ergodic distributions $p^{\mathbf{P}_\ell}$ are fixed during this optimization. This implies a significant reduction in the computational cost associated with the initial-conditions problem. Nevertheless, in the literature of finite-mixture models, it is well known that optimization of the likelihood function with respect to the mixture probabilities λ is a complicated task because the problem is plagued with many local maxima and minima. To address this problem, Aguirregabiria and Mira (2007) introduced an additional parametric assumption on the distribution of ξ_m that simplifies significantly the maximization of $Q(\boldsymbol{\theta},\lambda,\mathbf{P})$ for fixed $\mathbf{P}$. They assumed that the probability distribution of unobserved market heterogeneity is such that the only unknown parameters for the researcher are the mean and the variance, which are included in θ_i^{FC} and $\sigma_{\xi i}$, respectively. Therefore, they assumed that the distribution of ξ_m (i.e., the points of support and the probabilities λ_ℓ) are known to the researcher. For instance, we may assume that ξ_m has a discretized standard normal distribution with an arbitrary number of points of support L. Under this assumption, the pseudolikelihood function is maximized only with respect to $\boldsymbol{\theta}$ for given $\mathbf{P}$. Avoiding optimization with respect to λ simplifies importantly the computation of the different estimators described herein.

NPL Estimator. As defined previously, the NPL mapping $\varphi(.)$ is the composition of the equilibrium mapping and the mapping that provides the maximand in $\boldsymbol{\theta}$ to $Q(\boldsymbol{\theta},\mathbf{P})$ for given $\mathbf{P}$. That is, $\varphi(\mathbf{P}) \equiv \Psi(\hat{\theta}(\mathbf{P}),\mathbf{P})$, where $\hat{\theta}(\mathbf{P}) \equiv \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta},\mathbf{P})$. By definition, an NPL fixed point is a pair $(\hat{\boldsymbol{\theta}},\hat{\mathbf{P}})$ that satisfies two conditions: (1) $\hat{\boldsymbol{\theta}}$ maximizes $Q(\boldsymbol{\theta},\hat{\mathbf{P}})$; and (2) $\hat{\mathbf{P}}$ is an equilibrium associated to $\hat{\boldsymbol{\theta}}$. The NPL estimator is defined as the NPL fixed point with the maximum value of the likelihood function. The NPL estimator is consistent under standard regularity conditions (Aguirregabiria and Mira 2007; Proposition 2).

When the equilibrium that generates the data is Lyapunov stable, we can compute the NPL estimator using a procedure that iterates in the NPL mapping (see Section 3.4) to obtain the sequence of K-step estimators (i.e., NPL algorithm). The main difference is that now we must calculate the steady-state distributions $\mathbf{p}(\mathbf{P}_\ell)$ to address the initial-conditions problem. However, the pseudolikelihood approach also reduces significantly the cost of addressing the initial-conditions problem. This NPL algorithm proceeds as follows. We start with L arbitrary vectors of players'-choice probabilities,

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

one for each market type: $\{\hat{\mathbf{P}}_\ell^0 : \ell = 1, 2, \ldots, L\}$. Then, we perform the following steps:

Step 1: For every market type, we obtain the steady-state distributions and the probabilities $\{\lambda_{\ell|x_{m1}}\}$.
Step 2: We obtain a PML estimator of $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}}^1 = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \hat{\mathbf{P}}^0\})$.
Step 3: We update the vector of players'-choice probabilities using the best-response probability mapping. That is, for market type ℓ, firm i, and state $\mathbf{x}$, $\hat{P}^1_{i\ell}(\mathbf{x}) = \Phi(\tilde{\mathbf{z}}_i^{\hat{\mathbf{P}}^0_\ell}(\mathbf{x}, \xi^\ell)\hat{\boldsymbol{\theta}}^1_i + \tilde{e}_i^{\hat{\mathbf{P}}^0_\ell}(\mathbf{x}, \xi^\ell))$.
Step 4: If, for every type ℓ, $||\hat{\mathbf{P}}^1_\ell - \hat{\mathbf{P}}^0_\ell||$ is smaller than a predetermined small constant, then we stop the iterative procedure and keep $\hat{\boldsymbol{\theta}}^1$ as a candidate estimator; otherwise, we repeat Steps 1 through 4 using $\hat{\mathbf{P}}^1$ instead of $\hat{\mathbf{P}}^0$.

The NPL algorithm, on convergence, finds an NPL fixed point. To guarantee consistency, the researcher needs to start the NPL algorithm from different CCPs in case there are multiple NPL fixed points. This situation is similar to using a gradient algorithm, designed to find a local root, to obtain an estimator that is defined as a global root. Of course, this global-search aspect of the method renders it significantly more costly than the application of the NPL algorithm in models without unobserved heterogeneity. This is the additional computational cost that we must pay for dealing with unobserved heterogeneity. We note, however, that this global search can be parallelized in a computer with multiple processors.

Arcidiacono and Miller (2011). They extended this approach in several interesting and useful ways. Their paper made two main contributions to the literature of dynamic discrete-choice structural models. The first contribution is the *finite-state representation* of optimal-decision rules in dynamic discrete-choice models. We suppose that the model is such that there are two sequences of agents' choices, sequence $A^1 = \{a_t^1, a_{t+1}^1, \ldots, a_{t+k}^1\}$ and sequence $A^2 = \{a_t^2, a_{t+1}^2, \ldots, a_{t+k}^2\}$ with k relatively small (e.g., $k = 1$) that satisfy the following conditions: (1) the sequences have different initial choices, $a_t^1 \neq a_t^2$; and (2) the sequences lead to the same distribution of the state variables at period $t + k + 1$. Under these conditions and a Generalized Extreme Value (GEV) distribution of the unobservable $\varepsilon' s$, there is a simple transformation of players' best-response functions between periods t and $t + k$ such that this transformation provides a closed-form expression that includes *only* structural parameters and conditional-choice probabilities at the states visited between periods t and $t + q$. Using this expression,

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

we can define a "new" best response or equilibrium mapping in the space of conditional-choice probabilities. This mapping is much simpler to evaluate than our original mapping because it involves probabilities at only a few states. For instance, in the well-known bus-replacement model in Rust (1987), we can obtain the following "equilibrium" condition for the probability of bus replacement: $P(x) = \Psi^*(\theta, P(0), P(x+1))$, where x represents the cumulative mileage since the last replacement and Ψ^* is the *finite-state representation* of the equilibrium mapping:

$$\Psi^*(\theta_1, P(0), P(x+1)) = \frac{\exp\{-\theta_1 x - \theta_2 + \delta \ \ln P(x+1) - \delta \ \ln P(0)\}}{1 + \exp\{-\theta_1 x - \theta_2 + \delta \ \ln P(x+1) - \delta \ \ln P(0)\}} \tag{36}$$

In this example, the evaluation of the equilibrium mapping Ψ^* at state x involves only CCPs $P(0)$ and $P(x+1)$ and it does not require any matrix inversion to compute the inclusive values $\tilde{\mathbf{z}}^{\mathbf{P}}$ and $\tilde{e}^{\mathbf{P}}$. The new equilibrium conditions can be used to define a pseudolikelihood in a similar way as described previously. Arcidiacono and Miller (2011) showed that conditions (1) and (2) are satisfied in a class of dynamic decision models that includes but it is not limited to optimal stopping problems. We note that the finite-state representation applies to models with or without permanent unobserved heterogeneity.

A second contribution is that Arcidiacono and Miller (2011) proposed a new algorithm that reduces substantially the complexity in the optimization of the likelihood function with respect to the distribution of the finite mixture. Their algorithm combined the NPL method with an EM algorithm. They considered a general class of finite-mixture models in which unobserved heterogeneity may enter in both the payoff function and the transition of state variables; it also can be time-invariant or follow a Markov chain. We note that Lyapunov stability of each equilibrium type that generates the data is a necessary condition for the NPL and for the Arcidiacono – Miller algorithms to converge to a consistent estimator.

Kasahara and Shimotsu (2008b). The estimators of finite-mixture models in Aguirregabiria and Mira (2007) and Arcidiacono and Miller (2011) considered that the researcher cannot obtain consistent nonparametric estimates of market-type CCPs $\{\mathbf{P}_\ell^0\}$. Kasahara and Shimotsu (2008b), based on previous work by Hall and Zhou (2003), derived sufficient conditions for the nonparametric identification of market-type CCPs, $\{\mathbf{P}_\ell^0\}$, and the probability distribution of market types, $\{\lambda_\ell^0\}$. Given the nonparametric identification of market-type CCPs, it is possible to estimate structural

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

parameters using a two-step approach similar to that described herein. However, this two-step estimator has three limitations that do not appear in two-step estimators without unobserved market heterogeneity. First, the conditions for nonparametric identification of $\mathbf{P}^0$ may not hold. Second, the nonparametric estimator in the first step is a complex estimator from a computational point of view. In particular, it requires that the minimization of a sample criterion function with respect to the large-dimensional object $\mathbf{P}$.[45] This is, in fact, the type of computational problem that we want to avoid by using two-step methods instead of standard ML or GMM. Finally, the finite-sample bias of the two-step estimator can be significantly more severe when $\mathbf{P}^0$ incorporates unobserved heterogeneity and we estimate it nonparametrically.

3.6 Reducing the State Space

Although two-step and sequential methods are computationally much less expensive than full-solution-estimation methods, they still are impractical for applications in which the dimension of the state space is large. The cost of computing exactly the matrix of present values $\mathbf{W}^{\mathbf{P}}_{z,i}$ increases cubically with the dimension of the state space. In the context of dynamic games, the dimension of the state space increases exponentially with the number of heterogeneous players. Therefore, the cost of computing the matrix of present values may become intractable even for a relatively small number of players.

A simple approach to deal with this curse of dimensionality is to assume that players are homogeneous and the equilibrium is symmetric. For instance, in our dynamic game of market entry–exit, when firms are heterogeneous, the dimension of the state space is $|H| * 2^N$, where $|H|$ is the number of values in support of market size H_t. To reduce the dimensionality of the state space, we must assume that (1) only the number of competitors (and not their identities) affects the profit of a firm; (2) firms are homogeneous in their profit function; and (3) the selected equilibrium is symmetric. Under these conditions, the payoff-relevant state variables for a firm i are $\{H_t, s_{it}, n_{t-1}\}$, where s_{it} is its own incumbent status and n_{t-1} is the total number of active firms at period $t-1$. The dimension of the state space is $|H| * 2 * (N+1)$ that increases only linearly with the number of players.[46]

[45] Furthermore, the criterion function is not globally concave/convex and its optimization requires a global search.

[46] This is a particular example of the "exchangeability assumption" proposed by Pakes and McGuire (2001).

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

It is clear that the assumption of homogeneous firms and symmetric equilibrium can reduce substantially the dimension of the state space, and it can be useful in empirical applications. Nevertheless, there are many applications in which this assumption is too strong (e.g., in applications in which firms produce differentiated products).

To address this issue, Hotz, Miller, Sanders, and Smith (1994) proposed an estimator that uses Monte Carlo simulation techniques to approximate the values $\mathbf{W}^{\mathbf{P}}_{z,i}$. Bajari, Benkard, and Levin (2007) extended this method to dynamic games and models with continuous-decision variables. This approach proved useful in some applications; nevertheless, it is important to be aware that in those applications with large state spaces, simulation error can be sizable and can induce biases in the estimation of structural parameters. In those cases, it is worthwhile to reduce the dimension of the state space by making additional structural assumptions. This is the general idea in the inclusive-value approach discussed in Section 2.0, which can be extended to the estimation of dynamic games. Different versions of this idea were proposed and applied by Nevo and Rossi (2008), Macieira (2007), Rossi (2009), and Aguirregabiria and Ho (2012).

To present the main ideas, we consider here a dynamic game of quality competition in the spirit of Pakes and McGuire (1994): the differentiated-product version of the Ericson and Pakes (1995) model. There are N firms in the market that we index by i and B brands or differentiated products that we index by b. The set of brands sold by firm i is $\mathcal{B}_i \subset \{1, 2, \ldots, B\}$. Demand is given by a model similar to that in Section 2.1.1: Consumers choose one of the B products offered in the market or the outside good. The utility that consumer h obtains from purchasing product b at time t is $U_{hbt} = x_{bt} - \alpha\, p_{bt} + u_{hbt}$, where x_{bt} is the quality of the product, p_{bt} is the price, α is a parameter, and u_{hbt} represents consumer-specific taste for product b. These idiosyncratic errors are i.i.d. over (h, b, t) with Type I extreme-value distribution. If the consumer decides to not purchase any of the goods, she chooses the outside option that has a mean utility normalized to zero. Therefore, the aggregate demand for product b is $q_{bt} = H_t \exp\{x_{bt} - \alpha\, p_{bt}\}\ [1 + \sum_{b'=1}^{B} \exp\{x_{b't} - \alpha\, p_{b't}\}]^{-1}$, where H_t represents market size at period t. The market structure of the industry at time t is characterized by the vector $\mathbf{x}_t = (H_t, x_{1t}, x_{2t}, \ldots, x_{Bt})$. Every period, firms take as given current market structure and decide simultaneously their current prices and their investment in quality improvement. The one-period profit of firm i can be written as:

$$\Pi_{it} = \sum_{b\in\mathcal{B}_i} (p_{bt} - mc_b)\, q_{bt} - FC_b - (c_b + \varepsilon_{bt})\, a_{bt} \tag{37}$$

where $a_{bt} \in \{0, 1\}$ is the binary variable that represents the decision to invest in quality improvement of product b; mc_b, FC_b, and c_b are structural parameters that represent marginal cost, fixed operating cost, and quality investment cost for product b, respectively; and ε_{bt} is an i.i.d. private-information shock in the investment cost. Product quality evolves according to a transition probability $f_x(x_{bt+1}|a_{bt}, x_{bt})$. For instance, in the Pakes–McGuire (2001) model, $x_{bt+1} = x_{bt} - \zeta_t + a_{bt}\ v_{bt}$, where ζ_t and v_{bt} are two independent and non-negative random variables that are i.i.d. over (b, t).

In this model, price competition is static. The Nash–Bertrand equilibrium determines prices and quantities as functions of market structure $\mathbf{x}_t$ (i.e., $p_b^*(\mathbf{x}_t)$ and $q_b^*(\mathbf{x}_t)$). Firms' quality choices are the result of a dynamic game. The one-period profit function of firm i in this dynamic game is $\Pi_i(\mathbf{a}_{it}, \mathbf{x}_t) = \sum_{b \in \mathcal{B}_i} (p_i^*(\mathbf{x}_t) - mc_b)\ q_b^*(\mathbf{x}_t) - FC_b - (c_b + \varepsilon_{bt})\ a_{bt}$, where $\mathbf{a}_{it} \equiv \{a_{bt} : b \in \mathcal{B}_i\}$. This dynamic game of quality competition has the same structure as the game described in Section 3.2, and it can be solved and estimated using the same methods. However, the dimension of the state space increases exponentially with the number of products, and the solution and estimation of the model becomes impractical even when B is not too large.

We define the *cost-adjusted inclusive value* of firm i at period t as $\omega_{it} \equiv \log[\sum_{b \in \mathcal{B}_i} exp\{x_{bt} - \alpha\ mc_b\}]$. This value is closely related to the inclusive value discussed in Section 2.2.4. It can be interpreted as the net-quality level, or a value-added of sorts, that the firm is able to produce in the market. Under the assumptions of the model, the variable profit of firm i in the Nash–Bertrand equilibrium can be written as a function of the vector of inclusive values $\boldsymbol{\omega}_t \equiv (\omega_{1t}, \omega_{2t}, \ldots, \omega_{Nt}) \in \Omega$; that is $\sum_{b \in \mathcal{B}_i} (p_i^*(\mathbf{x}_t) - mc_b)\ q_b^*(\mathbf{x}_t) = vp_i(\boldsymbol{\omega}_t)$. Therefore, the one-period profit Π_{it} is a function $\tilde{\Pi}_i(\mathbf{a}_{it}, \boldsymbol{\omega}_t)$. The following assumption is similar to Assumption A2 in Section 2.0 and it establishes that given vector $\boldsymbol{\omega}_t$, the remainder of the information contained in the $\mathbf{x}_t$ is redundant for the prediction of future values of $\boldsymbol{\omega}$.

Assumption: The transition probability of the vector of inclusive values $\boldsymbol{\omega}_t$ from the point of view of a firm (i.e., conditional on a firm's choice) is such that $\Pr(\boldsymbol{\omega}_{t+1} \mid \mathbf{a}_{it}, \mathbf{x}_t) = \Pr(\boldsymbol{\omega}_{t+1} \mid \mathbf{a}_{it}, \boldsymbol{\omega}_t)$.

Under these assumptions, $\boldsymbol{\omega}_t$ is the vector of payoff-relevant state variables in the dynamic game. The dimension of the space Ω increases exponentially with the number of firms but not with the number of brands. Therefore, the dimension of Ω can be much smaller than the dimension of the original state space of $\mathbf{x}_t$ in applications in which the number of brands is large relative to the number of firms.

Of course, the assumption of sufficiency of $\boldsymbol{\omega}_t$ in the prediction of next-period $\boldsymbol{\omega}_{t+1}$ is not trivial. To justify it, we can place strong restrictions on the stochastic process of quality levels. Alternatively, it can be interpreted in terms of limited information, and/or bounded rationality. For instance, a possible way to justify this assumption is that firms face the same type of computational burdens that we do. Limiting the information that they use in their strategies reduces firms' computational costs of calculating a best response.

We note that the dimension of the space of $\boldsymbol{\omega}_t$ still increases exponentially with the number of firms. To deal with this curse of dimensionality, Aguirregabiria and Ho (2012) considered a stronger inclusive value/sufficiency assumption. We let vp_{it} be the variable profit of firm i at period t.

Assumption: $\Pr(\omega_{it+1}, vp_{it+1} \mid \mathbf{a}_{it}, \mathbf{x}_t) = \Pr(\omega_{it+1}, vp_{it+1} \mid \mathbf{a}_{it}, \omega_{it}, vp_{it})$.

Under this assumption, the vector of payoff-relevant state variables in the decision problem of firm i is (ω_{it}, vp_{it}) and the dimension of the space of (ω_{it}, vp_{it}) does not increase with the number of firms.

3.7 Counterfactual Experiments with Multiple Equilibria

One of the attractive features of structural models is that they can be used to predict the effects of new counterfactual policies. This is a challenging exercise in a model with multiple equilibria. Under the assumption that our data are generated by a single equilibrium, we can use the data to identify which of the multiple equilibria is the one that we observe. However, even under this assumption, we still do not know which equilibrium will be selected when the values of the structural parameters are different than those we estimated from the data. For some models, a possible approach to address this issue is to calculate all of the equilibria in the counterfactual scenario and then draw conclusions that are robust to whatever equilibrium is selected. However, this approach is of limited applicability in dynamic games of oligopoly competition because the different equilibria typically provide contradictory predictions for the effects we want to measure.

For instance, Aguirregabiria and Ho (2012) used their estimated dynamic game of airline-network competition to disentangle the contribution of demand, cost, and strategic factors to explain airlines' propensity to operate using hub-and-spoke networks. To measure the relative contribution of each factor, the authors implemented four counterfactual experiments in which they shut down (i.e., set to zero) different structural parameters of the model such as hub-size effects in variable profits, in fixed costs, and in entry costs (i.e., Experiments 1 through 3) and the entry deterrence motive of hub-and-spoke networks (i.e., Experiment 4). For each scenario, they compared the actual airlines' networks observed in the data with the airlines' networks that result from "the" equilibrium in counterfactual scenario. When implementing each counterfactual experiment, the authors should have dealt with multiplicity of equilibria. The dynamic game has many equilibria in the counterfactual scenario and some of these equilibria implied very different airline networks. Which of these equilibria should we select to compare it with the actual equilibrium in the data?

Here, we describe a simple homotopy method that was proposed in Aguirregabiria (2012) and applied in the empirical application in Aguirregabiria and Ho (2012). Under the assumption that the equilibrium-selection mechanism, which is unknown to the researcher, is a smooth function of the structural parameters, we showed how to obtain a Taylor approximation to the counterfactual equilibrium. Despite the fact that the equilibrium-selection function is unknown, a Taylor approximation of that function, evaluated at the estimated equilibrium, depends on objects that the researcher knows.

We let $\Psi(\boldsymbol{\theta}, \mathbf{P})$ be the equilibrium mapping such that an equilibrium associated with $\boldsymbol{\theta}$ can be represented as a fixed point $\mathbf{P} = \Psi(\boldsymbol{\theta}, \mathbf{P})$. We suppose that there is an equilibrium-selection mechanism in the population under study, but we do not know that mechanism. We let $\boldsymbol{\pi}(\boldsymbol{\theta})$ be the selected equilibrium given $\boldsymbol{\theta}$. The approach here is agnostic with respect to this equilibrium-selection mechanism: It assumes only that there is such a mechanism and that it is a smooth function of $\boldsymbol{\theta}$. Because we do not know the mechanism, we do not know the form of the mapping $\boldsymbol{\pi}(\boldsymbol{\theta})$ for every possible $\boldsymbol{\theta}$. However, we know that the equilibrium in the population, $\mathbf{P}^0$, and the vector of the structural parameters in the population, $\boldsymbol{\theta}^0$, belong to the graph of that mapping (i.e., $\mathbf{P}^0 = \boldsymbol{\pi}(\boldsymbol{\theta}^0)$).

We let $\boldsymbol{\theta}^*$ be the vector of parameters under the counterfactual experiment that we want to analyze. We want to know the counterfactual equilibrium $\boldsymbol{\pi}(\boldsymbol{\theta}^*)$ and compare it to the factual equilibrium $\boldsymbol{\pi}(\boldsymbol{\theta}^0)$. We suppose that Ψ is twice continuously differentiable in θ and P. The following is the key assumption to implement the homotopy method described here.

Assumption: The equilibrium-selection mechanism is such that $\pi(.)$ is a continuous-differentiable function within a convex subset of Θ that includes $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}^*$.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

That is, the equilibrium-selection mechanism does not "jump" among the possible equilibria when we move over the parameter space from $\boldsymbol{\theta}^0$ to $\boldsymbol{\theta}^*$. This seems a reasonable condition when the researcher is interested in evaluating the effects of a change in the structural parameters but "keeping constant" the same equilibrium type as the one that generates the data.

Under these conditions, we can make a Taylor approximation to $\boldsymbol{\pi}(\boldsymbol{\theta}^*)$ around $\boldsymbol{\theta}^0$ to obtain:

$$\boldsymbol{\pi}(\boldsymbol{\theta}^*) = \boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right) + \frac{\partial \boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right)}{\partial \boldsymbol{\theta}'}\left(\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\right) + O\left(\left\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\right\|^2\right) \tag{38}$$

We know that $\boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right) = \mathbf{P}^0$. Furthermore, by the implicit-function theorem, $\partial\boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right)/\partial\boldsymbol{\theta}' = \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial\boldsymbol{\theta}' + \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial\mathbf{P}'\, \partial\boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right)/\partial\boldsymbol{\theta}'$. If $\mathbf{P}^0$ is not a singular equilibrium, then $I - \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial\mathbf{P}'$ is not a singular matrix and $\partial\boldsymbol{\pi}\left(\boldsymbol{\theta}^0\right)/\partial\boldsymbol{\theta}' = (I - \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial\mathbf{P}')^{-1}\, \partial\Psi(\boldsymbol{\theta}^0, \mathbf{P}^0)/\partial\boldsymbol{\theta}'$. Solving this expression in the Taylor approximation, we have the following approximation to the counterfactual equilibrium:

$$\hat{\mathbf{P}}^* = \hat{\mathbf{P}}^0 + \left(I - \frac{\partial\Psi(\hat{\boldsymbol{\theta}}^0, \hat{\mathbf{P}}^0)}{\partial\mathbf{P}'}\right)^{-1} \frac{\partial\Psi(\hat{\boldsymbol{\theta}}^0, \hat{\mathbf{P}}^0)}{\partial\boldsymbol{\theta}'}\left(\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}^0\right) \tag{39}$$

where $(\hat{\boldsymbol{\theta}}^0, \hat{\mathbf{P}}^0)$ represents our consistent estimator of $(\boldsymbol{\theta}^0, \mathbf{P}^0)$. It is clear that $\hat{\mathbf{P}}^*$ can be computed given the data and $\boldsymbol{\theta}^*$. Under our assumptions, $\hat{\mathbf{P}}^*$ is a consistent estimator of the linear approximation to $\boldsymbol{\pi}(\boldsymbol{\theta}^*)$.

As in any Taylor approximation, the order of magnitude of the error depends on the distance between the value of the structural parameters in the factual and counterfactual scenarios. Therefore, this approach can be inaccurate when the counterfactual experiment implies a substantial change in some of the parameters. For these cases, we can combine the Taylor approximation with iterations in the equilibrium mapping. We suppose that $\mathbf{P}^*$ is a Lyapunov stable equilibrium. We also suppose that the Taylor approximation $\hat{\mathbf{P}}^*$ belongs to the dominion of attraction of $\mathbf{P}^*$. Then, by iterating in the equilibrium mapping $\Psi(\boldsymbol{\theta}^*, .)$ starting at $\hat{\mathbf{P}}^*$, we obtain the counterfactual equilibrium $\mathbf{P}^*$. We note that this approach is substantially different to iterating in the equilibrium mapping $\Psi(\boldsymbol{\theta}^*, .)$ starting with the equilibrium in the data $\hat{\mathbf{P}}^0$. This approach returns the counterfactual equilibrium $\mathbf{P}^*$ iff $\hat{\mathbf{P}}^0$ belongs to the dominion of attraction of $\mathbf{P}^*$. This condition is stronger than the one that establishes that the Taylor approximation $\hat{\mathbf{P}}^*$ belongs to the dominion of attraction of $\mathbf{P}^*$.[47]

[47] Aguirregabiria (2012) provided examples in which iterating in $\Psi(\boldsymbol{\theta}^*, .)$ starting from $\hat{\mathbf{P}}^0$ returns an equilibrium that is not $\boldsymbol{\pi}(\boldsymbol{\theta}^*)$ (i.e., it is not of the same "type" as the

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

4.0 Concluding Comments

In this chapter, we survey several challenges that we consider particularly important for applied work in estimation of dynamic demand and dynamic games. Our discussions of the two areas are mostly separate, reflecting to a large extent that these two literatures developed almost separately. In our view, an interesting area for future work is better integration and cross fertilization. We see several directions in which future work might proceed.

Further Model Simplification. Although we discuss several ways to simplify the computation and estimation of the models, the methods and computation are still complex and have limited applications. For demand for storable goods, Hendel and Nevo (2010) offered an alternative simple model that can be estimated easily using aggregate data. They made several nontrivial assumptions; the most important for simplifying the computation is that consumers can store at most for a known and predetermined number of periods. With these assumptions, they showed that the storable-goods model is identified from aggregate data and does not require solving the dynamic-programming problem. Thus, the computational cost is of the same order as that of a static demand model. We think this type of careful economic modeling is potentially useful in both the modeling of dynamic demand and dynamic games.

Integration of Dynamic Demand and Supply Models. Another promising avenue for future research is the combination of dynamic demand and dynamic supply. Most of the literature on estimation of dynamic games concentrate on dynamics in supply but ignores dynamics in demand, and most of the literature on dynamic demand does not allow for dynamic supply.[48] This obviously is an important limitation in the current state of the literature. As we move toward combining the two areas, we believe that the modeling simplifications discussed in the previous paragraph would be a particularly useful way to proceed. For example, with a simpler demand model, Hendel and Nevo (2011) were able to add a supply side to the dynamic demand.

Identification. Our discussion of identification of dynamic demand was informal, which reflects the state of the literature. A productive future avenue for research is to formally derive identification conditions, especially for estimation using aggregate data.

equilibrium $\mathbf{P}^0$), whereas the iterations starting at $\hat{\mathbf{P}}^*$ converge to the desired counterfactual equilibrium.

48 Goettler and Gordon (2012) is one of the few exceptions.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

Estimation Methods. We see several directions for future work in estimation methods. First, estimation based on CCP has been applied successfully elsewhere but has been used rarely for estimating dynamic demand, in large part because the first generation of these estimators could not allow for persistent unobserved heterogeneity. With the emergence of new estimators (see Section 3.5), we suspect that we will see more use of these methods in estimation of dynamic demand.[49] Second, Bayesian estimation methods are particularly efficient from a computational point of view when multiple integration is cheaper than optimization. As shown for some of the estimators presented herein, optimization is particularly costly because it requires a global search over a large dimensional space. This seems to be a good scenario to which to apply Bayesian estimation methods. Although Bayesian methods were proposed for the estimation of single-agent dynamic structural models (Norets 2009; Imai, Nair, and Ching 2009; Norets and Tang 2010), this type of method has not been extended yet to address games with multiple equilibria.

Multiple Equilibria. We see a couple of directions for future work here. First, in the macroeconometric literature of Dynamic Stochastic General Equilibrium (DSGE) models, the standard approach to address multiple equilibria is to linearize the equilibrium mapping.[50] This seems a reasonable approach when we consider that multiple equilibria do not comprise an important feature of the model that is needed to explain the data but is more of a nuisance associated to the nonlinearity of the model. Although the idea of linearizing the equilibrium mapping is related to two-step methods presented herein, it is a different approach and it will be interesting to explore it. Alternatively, instead of treating multiple equilibrium as a nuisance, we might consider whether it actually may aid in identification. Sweeting (2009) exploited multiple equilibria in a static entry game to gain identification.[51] This idea has not been explored for dynamic games.

[49] The recent paper by J. Lee (2011) is an important step in this direction. Lee proposed a nested pseudo GMM algorithm, in the spirit of the NPL method, for the estimation of the static Berry-Levinsohn-Pakes (BLP) model of demand. He showed that the *zero Jacobian property* in Aguirregabiria and Mira (2002) holds in this class of models. Based on this result, Lee showed that his nested pseudo GMM algorithm has good computational and statistical properties in the estimation of the static BLP model. The extension of this result to dynamic versions of the BLP model is an interesting and promising area for future research.

[50] See the survey on the econometrics of DSGE models by Fernandez-Villaverde (2010) and the references there. Section 4.1 of that survey describes the linearization (or log-linearization) of equilibrium conditions as the most common approach for approximating the solution and likelihood function of DSGE models.

[51] See also de Paula and Tang (2010).

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

Strategic Uncertainty and Beliefs out of Equilibrium. If the researcher believes that multiplicity of equilibria is a real issue in competition in actual markets, then firms may face significant strategic uncertainty in the sense that they may not know the strategies that other firms are playing. This strategic uncertainty can be particularly important in the context of oligopoly competition. Firms tend to be secretive about their own strategies, and it can be in their own interest to hide or even misrepresent them. The identification and estimation of dynamic oligopoly games when firms' beliefs are out of equilibrium is an interesting area of further research.

References

Aguirregabiria, V. (2010), "Another Look at the Identification of Dynamic Discrete Decision Processes: An Application to Retirement Behavior," *Journal of Business and Economic Statistics*, 28(2), 201–18.

Aguirregabiria, V. (2012), "A Method for Implementing Counterfactual Experiments in Models with Multiple Equilibria," *Economics Letters*, 114 (February), 190–4.

Aguirregabiria, V., and C.-Y. Ho (2012), "A Dynamic Oligopoly Game of the U.S. Airline Industry: Estimation and Policy Experiments," *Journal of Econometrics*, 168, 156–73.

Aguirregabiria, V., and P. Mira (2002), "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Markov Decision Models," *Econometrica*, 70(4), 1519–43.

Aguirregabiria, V., and P. Mira (2007), "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

Aguirregabiria, V., and P. Mira (2009), "Structural Estimation of Games with Multiple Equilibria in the Data," University of Toronto, manuscript.

Aguirregabiria, V., and P. Mira (2011), "Stability of Equilibria and Properties of Sequential Estimators: Problems and Solutions," University of Toronto, manuscript.

Aguirregabiria, V., P. Mira, and H. Roman (2007), "An Estimable Dynamic Model of Entry, Exit and Growth in Oligopoly Retail Markets," *American Economic Review*, 97, 449–54.

Akerlof, G. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84(3), 488–500.

Altonji, J., and L. Segal (1996), "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14(3), 353–66.

Arcidiacono, P., and R. Miller (2011), "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 79(6), 1823–67.

Bajari, P., L. Benkard, and J. Levin (2007), "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331–70.

Bell, D., and J. Lattin (1998), "Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers Prefer EDLP," *Marketing Science*, 17(1), 66–88.

Berry, S. (1992), "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60(4), 889–917.

Berry, S., P. Haile, and A. Gandhi (2011), "Connected Substitutes and Invertibility of Demand," NBER, Working Paper, No. 17193.

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–90.

Blattberg, R., G. Eppen, and J. Lieberman (1981), "A Theoretical and Empirical Evaluation of Price Deals in Consumer Nondurables," *Journal of Marketing*, 45, 116–29.

Boizot, C., J-M. Robin, and M. Visser (2001), "The Demand for Food Products: An Analysis of Interpurchase Times and Purchased Quantities," *Economic Journal*, 111, 391–419.

Bonnet, C., and P. Dubois (2010), "Inference on Vertical Contracts between Manufacturers and Retailers Allowing for Nonlinear Pricing and Resale Price Maintenance," *RAND Journal of Economics*, 41(1), 139–64.

Bresnahan, T. (1981), "Departures from Marginal-Cost Pricing in the American Automobile Industry: Estimates for 1977–1978," *Journal of Econometrics*, 17(2), 201–27.

Bresnahan, T. (1987), "Competition and Collusion in the American Automobile Industry: The 1955 Price War," *Journal of Industrial Economics*, 35(4), 457–82.

Bresnahan, T., and P. Reiss (1990), "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531–53.

Bresnahan, T., and P. Reiss (1991a), "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57–81.

Bresnahan, T., and P. Reiss (1991b), "Entry and Competition in Concentrated Markets," *Journal of Political Economy,* 99, 977–1009.

Bresnahan, T., and P. Reiss (1993), "Measuring the Importance of Sunk Costs," *Annales d'Economie et de Statistique*, 34, 181–217.

Bronnenberg, B., M. Kruger, and C. Mela (2008), "The IRI Marketing Data Set," *Marketing Science*, 27(4), 745–8.

Browning, M., and J. Carro (2010), "Heterogeneity in Dynamic Discrete Choice Models," *The Econometrics Journal*, 13, 1–39.

Busse, M., D. Simester, and E. Zettelmeyer (2010), "The Best Price You'll Ever Get: The 2005 Employee Discount Pricing Promotions in the US Automobile Industry," *Marketing Science*, 29(2), 268–90.

Carranza, J. (2006), "Demand for Durable Goods and the Dynamics of Quality," University of Wisconsin, unpublished manuscript.

Chiang, J. (1991), "A Simultaneous Approach to the Whether, What, and How Much to Buy Questions," *Marketing Science*, 10, 297–315.

Collard-Wexler, A. (2006), "Demand Fluctuations in the Ready-Mix Concrete Industry," New York University, manuscript.

Conlon, C. (2009), "A Dynamic Model of Costs and Margins in the LCD TV Industry," Department of Economics, Yale University, manuscript.

Conlon, C. (2010), "An MPEC Approach to Empirical Likelihood Estimation of Demand," Department of Economics, Columbia University, manuscript.

De Paula, A., and X. Tang (2012), "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information," *Econometrica*, 80(1), 143–72.

Doraszelski, U., and M. Satterthwaite (2010), "Computable Markov-Perfect Industry Dynamics," *RAND Journal of Economics*, 41(2), 215–43.

Dubè, J.-P. (2004), "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science*, 23(1), 66–81.

Dubè, J.-P., J. Fox, and C.-L. Su (2012), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," *Econometrica*, forthcoming.

Dunne, T., S. Klimek, M. Roberts, and Y. Xu (2009), "Entry, Exit and the Determinants of Market Structure," Department of Economics, Pennsylvania State University, manuscript.

Eckstein, Z., and K. I. Wolpin (1990), "Estimating a Market Equilibrium Search Model from Panel Data on Individuals," *Econometrica*, 58(4), 783–808.

Eichenbaum, M., N. Jaimovich, and S. Rebelo (2008), "Reference Prices and Nominal Rigidities," NBER, Working Paper No. 13829.

Einav, L. (2009), "Not All Rivals Look Alike: An Empirical Model for Discrete Games with Asymmetric Rivals," *Economic Inquiry*, 48(2), 369–90.

Einav, L., E. Leibtag, and A. Nevo (2010), "Recording Discrepancies in Nielsen Homescan Data: Are They Present and Do They Matter?," *Quantitative Marketing and Economics*, 8(2), 207–39.

Erdem, T., S. Imai, and M. P. Keane (2003), "Brand and Quantity Choice Dynamics under Price Uncertainty," *Quantitative Marketing and Economics* 1, 5–64.

Erdem, T., M. Keane, S. Oncu, and J. Strebel (2005), "Learning about Computers: An Analysis of Information Search and Technology Choice," *Quantitative Marketing and Economics*, 3, 207–46.

Ericson, R., and A. Pakes (1995), "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies,* 62, 53–82.

Esteban, S., and M. Shum (2007), "Durable Goods Oligopoly with Secondary Markets: The Case of Automobiles," *RAND Journal of Economics*, 38, 332–54.

Fernandez-Villaverde, J. (2010), "The Econometrics of DSGE Models," *SERIEs: Journal of the Spanish Economic Association*, 1, 3–49.

Finger, S. (2008), "An Empirical Analysis of R&D Competition in the Chemicals Industry," University of South Carolina, manuscript.

General Motors Corprorațion (1939), "The Dynamics of Automobile Demand," based on papers presented at a joint meeting of the American Statistical Association and the Econometric Society in Detroit, Michigan, on December 27, 1938.

GentzKow, M. (2007), "Valuing New Goods in a Model with Complementarity: Online Newspapers," *American Economic Review*, 97(3), 713–44.

Goettler, R., and B. Gordon (2011), "Does AMD Spur Intel to Innovate More?," *Journal of Political Economy*, 119(6). 1141–1200.

Goldberg, P. (1995), "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 63, 891–951.

Gowrisankaran, G., C. Lucarelli, P. Schmidt-Dengler, and R. Town (2010), "Government Policy and the Dynamics of Market Structure: Evidence from Critical Access Hospitals," University of Arizona, manuscript.

Gowrisankaran, G., and M. Rysman (2009), "Dynamics of Consumer Demand for New Durable Goods," University of Arizona, manuscript.

Griffith, R., A. Leicester, E. Leibtag, and A. Nevo (2009), "Consumer Shopping Behavior: How Much Do Consumers Save?," *Journal of Economics Perspectives*, 23(2), 99–120.

Gupta, S. (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 24, 342–55.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

Hall, P., and X-H. Zhou (2003), "Nonparametric Estimation of Component Distributions in a Multivariate Mixture," *Annals of Statistics*, 31(1), 201–24.

Hartmann, W. (2006), "Intertemporal Effects of Consumption and Their Implications for Demand Elasticity Estimates," *Quantitative Marketing and Economics*, 4(4), 325–49.

Hashmi, A., and J. Van Biesebroeck (2010), "Market Structure and Innovation: A Dynamic Analysis of the Global Automobile Industry," NBER, Working Paper No. 15959.

Hausman, J., G. Leonard, and J. Zona (1994), "Competitive Analysis with Differentiated Products," *Annales D'Economie et de Statistique*, 34, 159–80.

Heckman, J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time – Discrete Data Stochastic Process," in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA, The MIT Press.

Heckman, J., and B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52(2), 271–320.

Hendel, I. (1999), "Estimating Multiple Disccrete–Choice Models: An Application to Computerization Returus," *Review of Economic Studies*, 66(2), 423–46.

Hendel, I., and A. Nevo (2006a), "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74, 1637–74.

Hendel, I., and A. Nevo (2006b), "Sales and Consumer Inventory," *RAND Journal of Economics*, 37(3), 543–61.

Hendel, I., and A. Nevo (2010), "A Simple Model of Demand Anticipation," Department of Economics, Northwestern University, manuscript.

Hendel, I., and A. Nevo (2011), "Intertemporal Price Discrimination in Storable Goods Markets," Department of Economics, Northwestern University, manuscript.

Horowitz, J. (1998), "Bootstrap Methods for Covariance Structures," *Journal of Human Resources*, 33, 39–61.

Hotz, J., and R. A. Miller (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies,* 60, 497–529.

Hotz, J., R. A. Miller, S. Sanders, and J. Smith (1994), "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61, 265–89.

Imai, S., N. Jain, and A. Ching (2009), "Bayesian Estimation of Dynamic Discrete Choice Models," *Econometrica*, 77(6), 1865–99.

Jovanovic, B. (1989), "Observable Implications of Models with Multiple Equilibria," *Econometrica*, 57, 1431–7.

Kasahara, H., and K. Shimotsu (2008a), "Pseudo-Likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models," *Journal of Econometrics*, 146(1), 92–106.

Kasahara, H., and K. Shimotsu (2008b), "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77(1), 135–75.

Kasahara, H., and K. Shimotsu (2009), "Sequential Estimation of Structural Models with a Fixed Point Constraint," University of Western Ontario, manuscript.

Keane, M., and K. I. Wolpin (1997), "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473–522.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

Kehoe, P., and V. Midrigan (2008), "Temporary Price Changes and the Real Effects of Monetary Policy," NBER, Working Papers No. 14392.

Kryukov, Y. (2008), "Dynamic R&D and the Effectiveness of Policy Intervention in the Pharmaceutical Industry," Tepper School of Business, Carnegie Mellon University, manuscript.

Lee, J. (2011), "A New Computational Algorithm for Estimating Random Coefficients Models with Aggregate-Level Data," Department of Economics, UCLA, manuscript.

Lee, R. (2011), "Dynamic Demand Estimation in Platform and Two-Sided Markets: The Welfare Cost of Software Incompatibility," Stern School of Business, New York University, manuscript.

Linton, O., and S. Srisuma (2012), "Semiparametric Estimation of Markov Decision Processes with Continuous State Space," *Journal of Econometrics*, 166, 320–41.

Lou, W., D. Prentice, and X. Yin (2008), "The Effects of Product Ageing on Demand: The Case of Digital Cameras," MPRA, Paper No. 13407.

Macieira, J. (2007), "Extending the Frontier: A Structural Model of Investment and Technological Competition in the Supercomputer Industry," Virginia Tech, manuscript.

Magnac, T., and D. Thesmar (2002), "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70(2), 801–16.

Maskin, E., and Tirole, J. (1987), "A Theory of Dynamic Oligopoly, III: Cournot Competition," *European Economic Review*, 31, 947–68.

Maskin, E., and Tirole, J. (1988a), "A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs," *Econometrica*, 56, 549–69.

Maskin, E., and Tirole, J. (1988b), "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica*, 56, 571–99.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics*, pp. 105–42, New York: Academic Press.

McLachlan, G., and D. Peel (2000), *Finite Mixture Models*, New York: John Wiley and Sons.

Melnikov, O. (2001), "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market," Department of Economics, Yale University, manuscript.

Nair, H. (2007). "Intertemporal Price Discrimination with Forward Looking Consumers: Application to the US Market of Console Video-Games," *Quantitative Marketing and Economics*, 5(3), 239–92.

Nakamura, E., and J. Steinsson (2008), "Five Facts about Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 123(4), 1415–64.

Nevo, A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69(2), 307–42.

Nevo, A. (2011), "Empirical Models of Consumer Behavior," *Annual Review of Economics*, 3, 51–75.

Nevo, A., M. McCabe and D. Rubinfeld (2005), "Academic Journal Pricing and the Demand of Libraries," *American Economic Review*, 95(2), 447–52.

Nevo, A., and Rossi, F. (2008), "An Approach for Extending Dynamic Models to Settings with Multi-Product Firms," *Economics Letters*, 100(1), 49–52.

Norets, A. (2009), "Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables," *Econometrica*, 77, 1665–82.

Norets, A., and X. Tang (2010), "Semiparametric Inference in Dynamic Binary Choice Models," Department of Economics, Princeton University, manuscript.

Osborne, M. (2011), "Consumer Learning, Switching Costs and Heterogeneity: A Structural Examination," *Quantitative Marketing and Economics*, 9(1), 25–70.

Pakes, A., and P. McGuire (1994), "Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model," *RAND Journal of Economics*, 25(4), 555–89.

Pakes, A., and P. McGuire (2001), "Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'Curse of Dimensionality,'" *Econometrica*, 69, 1261–81.

Pakes, A., M. Ostrovsky, and S. Berry (2007), "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)," *RAND Journal of Economics*, 38, 373–99.

Perrone, H. (2009), "Inventories, Unobservable Heterogeneity and Long-Run Price Elasticities," Department of Economics and Business at Universitat Pompeu Fabra, manuscript.

Pesendorfer, M. (2002), "Retail Sales: A Study of Pricing Behavior in Supermarkets," *Journal of Business*, 75(1), 33–66.

Pesendorfer, M., and P. Schmidt-Dengler (2008), "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75(3), 901–28.

Pesendorfer, M., and P. Schmidt-Dengler (2010), "Sequential Estimation of Dynamic Discrete Games: A Comment," *Econometrica*, 78(2), 833–42.

Porter, R. (1983), "A Study of Cartel Stability: The Joint Executive Committee, 1880–1886," *Bell Journal of Economics*, 14(2), 301–14.

Prince, J. (2008), "Repeat Purchase Amid Rapid Quality Improvement: Structure Estimation of Demand for Personal Computers," *Journal of Economics and Management Strategy*, 17, 1–33.

Rossi, F. (2008), "1$ Discount or 1$ Reward? The Effect of Consumers' Preferences on Reward Programs," Kellog School of Management, Northwestern University, manuscript.

Rossi, F. (2009), "A Dynamic Oligopoly Model of Product Entry, and Exit with Multi-Product Firms," University of North Carolina at Chapel Hill, manuscript.

Rust, J. (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55(5), 999–1033.

Rust, J. (1994), "Estimation of Dynamic Structural Models, Problems and Prospects: Discrete Decision Processes," in C. Sims (ed.), *Advances in Econometrics*, Sixth World Congress, Cambridge University Press.

Ryan, S. (2012), "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 80(3), 1019–61.

Schiraldi, P. (2011), "Automobile Replacement: A Dynamic Structural Approach," *RAND Journal of Economics*, 42(2), 266–91.

Shcherbakov, O. (2008), "Measuring Switching Costs in the Television Industry," University of Arizona, manuscript.

Shoemaker, R. (1979), "An Analysis of Consumer Reactions to Product Promotions," in N. Beakwith, M. Houston, R. Mittlestaedt, K. Monroe, and S. Ward (eds.), Educators' Conference Proceedings, 244–248. Chicago: American Marketing Association.

Snider, C. (2009), "Predatory Incentives and Predation Policy: The American Airlines Case," Department of Economics, UCLA, manuscript.

Su, C., and K. Judd (2012), "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, forthcoming.

P1: KpB Trim: 6in × 9in Top: 0.5in Gutter: 0.875in
CUUS1759-02 CUUS1757/Acemoglu Volume III 978 1 107 01606 4 November 5, 2012 15:17

Suzuki, J. (2010), "Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry," Department of Economics, University of Toronto, manuscript.

Sweeting, A. (2007), "Dynamic Product Repositioning in Differentiated Product Markets: The Case of Format Switching in the Commercial Radio Industry," NBER, Working Paper No. 13522.

Sweeting, A. (2009), "The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria," *RAND Journal of Economics*, 40(4), 710–42.

Sweeting, A. (2012), "Price Dynamics in Perishable Goods Markets: the Case of Secondary Markets for Major League Baseball Tickets," *Journals of Political Economy*, forthcoming.

Train, K. (2003), *Discrete Choice Models with Simulation*, New York: Cambridge University Press.

Walrath, M. (2010), "Religion as an Industry: Estimating a Strategic Entry Model for Churches," University of St. Thomas, manuscript.