Single Agent Dynamics

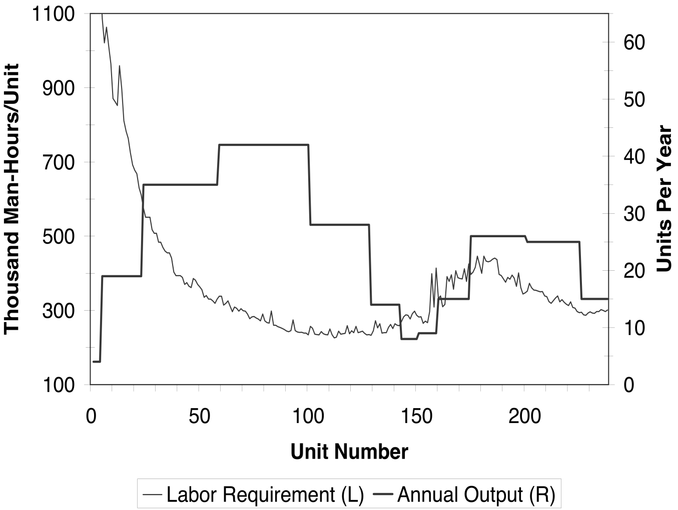
Charlie Murry

February 17, 2025

*Graduate IO. Parts of notes gratuitously borrowed from Tobias Salz, Nikhil Agarwal, John Asker, Nick Buchholz, Allan Collard-Wexler, Chris Conlon, Kei Kawai, Robin Lee, Ariel Pakes, Paul Scott, and Matt Shum.

Why Dynamics?

Production of Lockheed Tristar L-1011



Benkard (2000)

Learning and Forgetting: The Dynamics of Aircraft Production

Production of complex products may exhibit interesting dynamics

- As production ramps us, workers/managers may learn how to streamline processes, make fewer mistakes, spend less time correcting them.
- When production decreases, these abilities may diminish if workers are replaced.

Interesting implications for returns to scale/scope and product variety

- Pricing low at an early stage may improve cost efficiency $\longrightarrow$ low-prices are an investment.
- Product variety in unconstrained equilibrium may be detrimental to cost efficiency.

Canonical Dynamic Discrete Choice (DDC)

Dynamic Discrete Choice

Some High Level Comments

- This literature has a strong focus on computation.
- The main problem is that the estimation, as in BLP, solves a nested fixed point.
- Unlike the demand literature the DDC literature that deals with endogeneity and unobservable confounds is much less developed.
- Requires more restrictive assumption than static discrete choice.

Outline

Computing/estimating models of dynamic discrete choice.

- Many economic problems are inherently dynamic: *investment decisions*, *capacity choice*, *entry*, *R&D*, *certain types of demand*.
- Goal: Correct for **biases** using static methods in dynamic environments.
- Goal: Make more realistic counterfactual predictions.
- Surveys: Rust (1994), Aguirregabiria and Mira (2010).
- These methods can be used widely across fields!

Toolbox

Nested fixed point estimation (NFXP), conditional choice probabilities (CCP), mathematical programming with equilibrium constraints (MPEC), EM-Algorithm, simulation based estimation, inclusive value sufficiency, linear representation of DDC Model.

Outline

Road Map for Single Agent Dynamics

1. Introduction to canonical dynamic discrete choice (DDC) model, Rust 1994.
2. DDC Estimation 1 (NFXP): "Brute Force".
3. DDC Estimation 2 (CCP): Exploiting observed policies.
4. DDC Estimation 3 (MPEC): Exploiting modern computational methods.
5. Identification of DDC models.
6. DDC and persistent unobserved heterogeneity.
7. Linear representation of DDC models.
8. Simulation based estimation of DDC models.
9. Applications in dynamic demand: Durable upgrade decisions.

Basic model

Canonical Dynamic Discrete Choice

- Discrete time t, agent i, finite or infinite horizon
- s_{it}: vector of state variables
- $a_{it} \in A = \{0, 1, \dots, J\}$: discrete action
- Agent's beliefs $F(s_{it+1}|s_{it}, a_{it})$ (usually Markov and rational expectations (a key identification assumption)).
- Agent's optimization gives rise to dynamic program:

$$V(s_{it}) = \max_{a \in A} \{u(a, s_{it}) + \beta \cdot \int V(s_{it+1}) \cdot dF(s_{it+1}|s_{it}, a)\}$$

Basic model

Data and Primitives

Data typically used:

- N individuals forming a panel (i, t).
- Observables: actions a_{it}, observable state variables x_{it}
- Unobservables (econometric error): ϵ_{it}.
  - This error is "structural" in the sense that $s_{it} = (x_{it}, \epsilon_{it})$.

We may seek to estimate:

- Structural parameters governing preferences in $u(a, s_{it})$.
- Transition probabilities.
- Discount factor β (typically difficult to identify, more on this later).

Dynamic Programming

Assumptions on unobservables

- **Additive separability**:
  - $u(a, x_{it}, \epsilon_{it}) = u(a, x_{it}) + \epsilon_{it}(a)$.
  - One unobservable for each action, so dimensionality of ϵ_{it} is $J \times 1$.
- **Type-I Extreme Value (logit) errors**
  - The unobservable shocks $\{\epsilon_{it}(a), a = 1, 2, \dots, J\}$ are independently distributed as Type-I E.V.
- **IID**:
  - ϵ is i.i.d. distributed over agents and time according to G_ϵ.
- **Conditional independence of x_{it}**:
  - $F(s_{it+1}|s_{it}, a) = F(x_{it+1}|x_{it}, a) \cdot G_\epsilon$.

John Rust (1987)

John Rust (1987)

Harold Zurcher's Problem

- He is the superintendent of maintenance at the Madison, WI Metropolitan Bus Company.
- Each week, must decide to replace engine or keep running for another week.
- This is an **optimal stopping** problem:
  - Fixed costs of repair, but new engine has lower future maintenance costs.
  - Engine maintenance costs grows as engine gets older.
- Can show that there is a **cutoff** mileage threshold x^* above which a bus will have its engine replaced. Common type of optimal policy in many models, for example, job search.
- Note: as econometricians we are making a very strong assumption, namely that we observe the optimal *solution* to a complex dynamic problem. We want to learn the primitives governing the problem.

John Rust (1987)

Some Comments

Many of the empirical IO appeals, issues, and trade-offs showcased.

- Paper about one guy, who cares? The *yogurt/cereals* critique.
- Obviously, this is a paper about methodology, like many early empirical IO papers. *Recently the empirical IO literature has moved away from pure methodological work...*
- Model focuses on one particular decision and leaves out many related decisions that Harold Zurcher makes. **George E.P. Box:** *All models are wrong but some are useful...*
- However, your model should capture the crucial aspects that are material to the question. How do omitted aspects affect the conclusions drawn from the paper? It is good practice to think about all model assumptions in light of their effect on counterfactual results.

John Rust (1987)

Model

- θ_1: parameters of cost function and RC; θ_2: parameters of distribution ϵ (assumed to be known); θ_3: parameters that govern state transition;β imputed.
- $a_t = 1$ if engine is replaced at time t, otherwise $= 0$. Choose sequence of actions $a \equiv \{a_1, a_2, ..., a_t, ...\}$:
$$\max_a E \sum_{t=1}^{\infty} \beta^{t-1} u(x_t, \epsilon_t, a_t; \theta)$$
- Mileage x_{t+1} evolves according to $F(x_{t+1}|x_t; \theta_3)$ if $a_t = 0$, otherwise $x_{t+1} = 0$ (engine is brand new if replaced). This transition process can be estimated separately from the model.
- Utility is parameterized $u(x_t, \epsilon_t, a_t; \theta_1, RC) = -c(x_t \cdot (1 - a_t); \theta) - a_t \cdot RC + \epsilon_t(a_t)$, where RC is replacement cost and c is increasing in x_t.
- Want to estimate: parameters of cost function c, replacement cost RC, and parameters governing $F(x_{t+1}|x_t; \theta_3)$.

Rust (1987): Data & Likelihood

Conditional Independence Assumption

Allows to simplify **transition probabilities** and **value function**:

- Factor transition probabilities:
  $P(x_{t+1}, \epsilon_{t+1} | x_t, \epsilon_t, a_t; \theta) = P(x_{t+1} | x_t, a_t, \theta_3) \cdot G(\epsilon_{t+1} | x_{t+1}, \theta_2)$
- Do not need to treat unobserved ϵ's as state variables. Can write $EV_\theta(x_t, a_t)$ instead of $EV_\theta(x_t, \epsilon_t, a_t)$. Can express Bellman eq. in terms of EV.
- <u>Testable:</u> estimate alternative model where last periods decision enters utility, then perform likelihood ratio test.

Also, with **CI** the **likelihood function** can be simplified to:

$$\log \ell(x_1, \ldots, x_T, a_1, \ldots, a_T | x_0, a_0; \theta) = \sum_{t=1}^{T} \log F(x_t | x_{t-1}, a_{t-1}; \theta_3) + \sum_{t=1}^{T} \log P(a_t | x_t; \theta_1)$$

- Two components: the mileage transitions, and the probability of replacement.

Rust (1987)

Comments on the Specification

The likelihood function is not specified directly.

- Instead, it is *derived* from the solution to the underlying optimization problem

Rust distinguishes structural model from reduced form:

- The parametric specification occurs at the level of the primitive objects of the model (utility function and transition probabilities).

He discusses the importance of incorporating unobservables into the dynamic optimization:

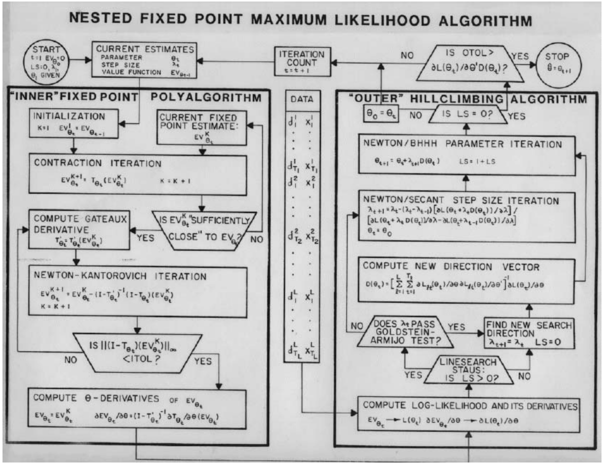
- Model without “error term”: due to cut-off rule in x_t model would be immediately rejected by the data.
- Econometric model where error term is simply added to the model solution lacks interpretation. Instead, view ϵ_t as a state variable which is unobserved to the econometrician but observed by agents (enters utility function).

Nested Fixed Point Algorithm

Rust (1987): Estimation

NFXP 1

Key idea: Estimation first solves an inner “loop” (the value function) and then an outer “loop”, the objective function. Hence, nested fixed point.



Rust (1987): Estimation

NFXP 2

- Note we can specify each agent's dynamic decision problem:

$$a_t^\star = argmax_a \{u(x_t, \epsilon_t, a; \theta) + \beta \cdot EV_\theta(x_t, a)\}$$

- Let $u(x, a; \theta) \equiv u(x, \epsilon, a; \theta) - \epsilon_a$ be part of the utility minus the error term.
- Then the **choice specific value functions** are defined as:

$$v_\theta(a, x) \equiv u(x, a; \theta_1) + \beta \cdot EV_\theta(x, a)$$

- Since we assume that ϵ are i.i.d. logit:

$$P(a_t|x_t; \theta_1) = \frac{exp(u(x_t, a_t; \theta_1) + \beta EV_\theta(x_t, a_t))}{\sum_{j=0,1} exp(u(x_t, a_j; \theta_1) + \beta EV_\theta(x_t, a_t))} = \frac{exp(v_\theta(a_t, x_t))}{\sum_{j=0,1} exp(v_\theta(a_j, x_t))}$$

This means that if we can solve for each agent's value function $V_\theta(\cdot)$, we can compute $P(a_t|x_t; \theta)$, and form the likelihood.

Rust (1987): Estimation

NFXP 3

For a given parameter value θ Rust iterates on the expected value function.

- Let x, a denote last period's mileage and choice.
- Let y, j denote next period's mileage and choice.

$$\begin{aligned} EV_\theta(x,a) &= E_{y,\epsilon}[V_\theta(y,\epsilon|x,a)] \\ &= E_{y,\epsilon}\Big[\max_{j=0,1}[u(y,j;\theta)+\epsilon+\beta\cdot EV_\theta(y,j)]\Big] \\ &= E_y \log\Big[\sum_{j=0,1} exp(u(y,j;\theta)+\beta\cdot EV_\theta(y,j))\Big] \end{aligned}$$

- Don't need to evaluate the value function for a grid of the ϵ draws.
- Works because of **conditional independence** assumption!
- Once convergence on $EV(\cdot)$ is obtained, the choice probabilities are obtained (as on the previous slide).

Rust (1987): Estimation

Recap and Comments

Estimation Steps:

- “Although the data clearly reject the myopic model, I was not able to precisely estimate the discount factor β.” Impute value for discount factor.
- Estimate θ_3 — transition function for mileage — can be done without behavioral model
- NFXP outer “loop”: search over (θ_1, RC) to maximize likelihood function.
- NFXP inner “loop”: for evaluating a likelihood for a given (θ_1, RC) need to find fixed point of bellman equation.

Comments:

- Heavy reliance on assumptions, particularly i.i.d. errors and conditional independence of state variables. Discrete states and actions, correct expectations. Will limit certain applications.
- Computationally costly: This is a full-solution method: for every evaluation of the parameter vector θ_1, the agent’s dynamic programming problem is explicitly solved (value-function iteration).
- Inherits all positive aspects of maximum likelihood. Asymptotically efficient full information estimator.

Hotz and Miller (1993)

+

Hotz, Miller, Sanders and Smith (1994)

Estimation: An alternative approach (CCP)

Idea

Key insight: we can express the value functions in terms of observed choice probabilities. This allows us to obtain the value function without having to solve the agent's optimization problem. This leads to an estimator that is:

- Computationally efficient but **not** econometrically efficient.
- How the mapping from choice probabilities to values is done depends on the application.

Important idea

We don't need to solve for policies whose state variables and outcomes are observed in the data.

Estimation: An alternative approach (CCP)

Hotz & Miller inversion with logit errors

Let $G(.|x,\theta)$ be the social surplus function:

$$G(v|x,\theta) \equiv \int_{\epsilon} \max_{a} \left\{ v(a) + \epsilon(a) \right\} \cdot q(d\epsilon|x,\theta)$$

Theorem 1 in **Rust (1987)**:

$$P(a|x,\theta) = \frac{\partial}{\partial u(x,a,\theta)} \cdot G\big(u(x,\theta) + \beta \cdot EV_{\theta}(x)|x,\theta\big)$$

where:

- $P(a|x,\theta)$ is the probability of action a conditional on state x.

Mapping from **choice-specific values** to **choice probabilities**.

Estimation: An alternative approach (CCP)

Hotz and Miller (1993) and Hotz, Miller, Sanders and Smith (1994)

- Notice that CCP's are unchanged when subtracting a constant from every value:

$$dv(x,a) \equiv v_\theta(x,a) - v_\theta(x,0),$$

  where $a = 0$ is some reference action.
- Let $Q : \mathbb{R}^{|\mathbf{J}|-1} \to \Delta^{|\mathbf{J}|}$ be the mapping from the deltas in conditional values to CCP's.
- We maintain the assumption that the ϵ are iid.

Hotz and Miller (1993) Proposition 1

Q is invertible.

Estimation: An alternative approach (CCP)

Hotz & Miller inversion with logit errors

Like in Rust, consider $\epsilon \sim$ T1EV, IID. Expression for CCP's:

$$P(a|x;\theta) = \frac{exp(v_\theta(a,x))}{\sum_{a\in\mathcal{A}} exp(v_\theta(a,x))}$$

The *HM inversion* follows by taking logs and differencing. Thus, with logit errors:

$$Q^{-1}(P) = v_\theta(x,a) - v_\theta(x,0)$$

Estimation: An alternative approach (CCP)

Using Inversion for Estimation

The **inversion result** can be used for estimation in a number ways:

1. Forward simulation. *Hotz, et Al. (1994)*
2. Finite dependence. *Scott (2016)*.
3. Solve a system of equations to express the vector of values in terms of probabilities and parameters. Pesendorfer and Schmidt-Dengler (2008) for dynamic games.

We will now discuss simulation and leave the former approach for when we discuss dynamic games.
For both approaches we **first** need choice probabilities and transition laws:

$$F(x'|x,a) = \begin{cases} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \frac{1(x_{i,t+1}<x', x_{it}=x, a_{it}=0)}{1(x_{it}=x, a_{it}=0)} & \text{if } a=0 \\ \sum_{i=1}^{N} \sum_{t=1}^{T-1} \frac{1(x_{i,t+1}<x', a_{it}=1)}{1(a_{it}=1)} & \text{if } a=1 \end{cases}$$

$$\hat{P}(a=1|x) \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1(a_{it}=1, x_{it}=x)}{1(x_{it}=x)}, \qquad \hat{P}(a=0|x) \equiv 1 - \hat{P}(a=1|x)$$

Estimation: An alternative approach (CCP)

Forward Simulation

If ϵ is i.i.d. logit, then $\mathbb{E}[\epsilon|a = a^*, x] = \gamma - \log(P(a|x))$, where γ is Euler's constant.

$$\mathbb{E}\left(u(a) + \varepsilon(a)\middle|u, a = a^*\right) = \ln\left(\sum_{j \in J} exp(u(a_j))\right) + \gamma \Leftrightarrow$$

$$\mathbb{E}\left(\varepsilon(a)\middle|u, a = a^*\right) = \ln\left(\sum_{j \in J} exp(u(a_j))\right) - u(a) + \gamma = \gamma - \log(P(a|x))$$

For more algebra tricks for discrete choice have a look at *Discrete Choice Theory of Product Differentiation* by **Anderson, de Palma, and Thisse**.

Estimation: An alternative approach (CCP)

Forward Simulation

With estimates of $\hat{F}(x'|x,a)$, $\hat{P}(a=1|x)$, $\hat{P}(a=0|x)$ in hand, we can express the choice specific value function (w/o ϵ) for any a taken this period as follows:

$$v_\theta(a_t, x_t) = u(x_t, a_t; \theta) + \beta \cdot \mathbb{E}\Big[u(x_{t+1}, a_{t+1}; \theta) + \epsilon_{t+1} + \\ \beta \cdot \mathbb{E}[u(x_{t+1}, a_{t+1}; \theta) + \epsilon_{t+2} + \beta^2 \cdot \mathbb{E}[\ldots]]\Big]$$

where expectations are conditional, i.e., the first expectation is

$$\mathbb{E}_{x_{t+1}|x_t,a_t}\mathbb{E}_{a_{t+1}|x_{t+1}}\mathbb{E}_{\epsilon_{t+1}|x_{t+1},a_{t+1}}.$$

Estimation: An alternative approach (CCP)

- So, for a given θ, simulate forward (for $a = 0, 1$):

$$v_\theta(a,x) \equiv \frac{1}{S}\sum_s \Big[u(a,x;\theta) + \beta \cdot \Big[u(a^s_{t+1}, x^s_{t+1};\theta) + \gamma - \log(\hat{P}(a^s_{t+1}|x^s_{t+1})) + \beta \cdot [u(a^s_{t+2}, x^s_{t+2};\theta) + \gamma - \log(\hat{P}(a^s_{t+2}|x^s_{t+2})) + \beta \cdot \ldots]\Big]\Big]$$

where

$$x^s_{t+1} \sim \hat{F}(\cdot|x_t, a_t), a^s_{t+1} \sim \hat{P}(\cdot|x^s_{t+1}), x^s_{t+2} \sim \hat{F}(\cdot|x^s_{t+1}, a^s_{t+1}), \ldots$$

- Given simulated choice-specific value functions, predict CCP's:

$$\tilde{P}(a=1|x;\theta) \equiv \frac{\exp(v_\theta(a=1,x))}{\exp(v_\theta(a=0,x)) + \exp(v_\theta(a=1,x))}$$

$$\hat{\theta} = argmin_\theta || \left[\log(\hat{P}(a=1|x)) - \log(\hat{P}(a=0|x))\right] - [v_\theta(a=1,x) - v_\theta(a=0,x)] ||$$

Aguirregabiria and Mira (2002)

From Two-Step to K-Step estimators

The CCP approach is sometimes referred to as **two-step** method. **First**, estimate choice probabilities. **Second**, estimate structural parameters.

- CCP estimators reduce computational burden at the expense of econometric efficiency.
- Aguirregabiria and Mira (2002) proposes a K-step estimator to leverage the CCP approach but obtain a more efficient estimator through iteration.
- **Idea:** After forming an initial $\widehat{V}(\widehat{P}_0; \theta)$ one can obtain a new set of choice probabilities $\hat{P}_1$ using current guess $\hat{V}$. One can iterate on this and form K-step estimator $\hat{P}_K$.
- The $K = 1$-estimator corresponds to Hotz and Miller and $K = \infty$ to Rust.
- They show that in small samples, each step leads to an increase in efficiency.
- They also show that asymptotically, all K-step estimators are equivalent to ML.

Estimation: An alternative approach (CCP)

Recap

- **Advantage:** Estimate dynamic parameters without solving dynamic programming problem for each evaluation of θ. The key here is having a consistent estimate of $\hat{P}(a|x)$.
- Leverages following idea:
  - Assume that agents choose the optimal policy in each state
  - We recover "reduced form" consistent estimates of these optimal policies $\hat{P}$ and transitions over state variables $\hat{F}$
  - At θ_0, simulated estimates of choice-specific value functions yield predicted optimal policies $\tilde{P}$ that coincide with observed policies $\hat{P}$.
- When S is discrete, solving for the value function given estimated policies is equivalent to solving a system of linear equations (matrix inversion); see Pesendorfer and Schmidt-Dengler (2008), or use finite dependence, both of which we will discuss later in class.
- CCP approach can be extended to continuous time models, Arcodiacono et Al. (2016).

Mathematical Programming with Equilibrium Constraints

Yet another approach for Estimation

Mathematical Programming with Equilibrium Constraints (MPEC)

Su and Judd (2012): *Constrained optimization approaches to estimation of structural models*.

- NFXP not computationally efficient, but CCP not asymptotically efficient.
- In a nested NFXP we require the equilibrium condition to hold for *every* guess of θ.
- **Key idea of MPEC**: bellman equation can be imposed as a constraint on the optimization.
- As optimization climbs the hill, this constraint can be slack and is only required to hold at the end. $\longrightarrow$ many fewer evaluations of bellman eq.
- Makes use of modern solver's (IPOPT (open source), KNITRO (commercial)).

MPEC

Main Idea

- For NFXP (Rust), likelihood is formulated as $\max_\theta \mathcal{L}(\theta, EV(\theta), \mathbf{X})$, where $V(\theta)$ is the value function, and $\mathbf{X}$ is the data.
- However, we know that $EV(\theta)$ is the unique solution to $EV = T(EV, \theta)$
- Use this insight to instead, solve:

$$\max_{EV,\theta} \mathcal{L}(\theta, EV, \mathbf{X}), \text{ s.t. } EV = T(EV, \theta)$$

- Same idea can be applied to BLP, using contraction mapping for δ.

Su and Judd Proposition 1

Let $\hat{\theta}$ be the maximum likelihood estimator and let $(\bar{\theta}, \bar{\sigma})$ be a solution of the constrained optimization problem. Define $\hat{\sigma}^\star(\theta) = argmax_{\hat{\sigma}(\theta)} L(\theta, \hat{\sigma}(\theta))$. Then $L(\bar{\theta}, \bar{\sigma}) = L(\hat{\theta}, \hat{\sigma}^\star(\hat{\theta}))$. If the model is identified, then we have $\hat{\theta} = \bar{\theta}$.

Concretely, this means that we can instead solve:

$$\max_{\theta^c, RC, \mathbf{EV}} \mathscr{L}(\theta, RC) = \Pi_{t=2}^{T} P(a_t|x_t, \theta_1) \cdot p(x_t|x_{t-1}, a_{t-1}, \theta_3)$$

subject to:

$$P(a|x, \theta_1, RC) = \frac{exp(u(x, a; \theta_1, RC) + \beta \cdot EV_\theta(x, a))}{\sum_{j=0,1} exp(u(x, j; \theta_1, RC) + \beta \cdot EV_\theta(x, j))}$$

and

$$EV_\theta(x, a) = T_\theta(EV_\theta)(x, a) = \int_0^\infty log\Big(\sum_{a'=0,1} exp\big(u(x, a; \theta_1, RC) + \beta \cdot EV_\theta(x', a')\big)\Big) \cdot p(dx'|x, a', \theta_3)$$

Advantages and Disadvantages:

- No natural way to obtain standard errors. Needs Bootstrap procedure.
- MPEC can be very burdensome on memory. Sparsity is key. In a demand setting sparsity is coming from the assumption that markets are independent.
- Becomes really useful when modern solvers are exploited.
  - Automatic differentiation.
- Easy to implement with Julia JuMP or AMPL.

Identification in DDC

Digression: Nonparametric Identification

Definition

- Let $m^{\star}$ be the true vector of functions and distributions.
- Let $P(m)$ denote the joint distribution of observable variables under the assumption that the data is generated under m

Definition of Identification

Definition 1: $m^{\star}$ is identified in M iff $\forall m \in M, m \neq m^{\star}, P(m) \neq P(m^{\star})$

Definition 2: $m^{\star}$ is not identified in M iff $\exists m \in M, m \neq m^{\star}, P(m) = P(m^{\star})$

Digression: Nonparametric Identification

Definition

- Let $m^\star$ be the true vector of functions and distributions.
- Let $P(m)$ denote the joint distribution of observable variables under the assumption that the data is generated under m

Definition of Identification

Definition 1: $m^\star$ is identified in M iff $\forall m \in M, m \neq m^\star, P(m) \neq P(m^\star)$

Definition 2: $m^\star$ is not identified in M iff $\exists m \in M, m \neq m^\star, P(m) = P(m^\star)$

- If M is a subset of a finite dimensional space, we say that the model M is **parametric**.
- If M is not a subset of a finite dimensional space, we say that M is:
  - **Semi-parametric**, if some of the functions, distributions lie inside a finite dimensional space.
  - **Nonparametric**, if none of the functions, distributions lie inside a finite dimensional space.
- Notice, that we are not worried about finite sample variation. Identification asks, if we had infinite data, can we recover the objects of interest.

Digression: Nonparametric Identification

Some Comments

- For many canonical models (such as BLP, Rust, and many types of auction setups) non-parametric identification has been shown.
- If you want to be creative and come up with your own model some people will ask how it is non-parametrically identified.

Identification and Modeling: When we make an economic model that we wish to take to the data, we have to think of two things:

(1) How well does the model capture important economic behavior?

(2) Can we identify and estimate the model using the data we have?

(1) tends to push your model to become more and more complicated. (2) pushes you in the opposite direction. Your model building exercise will be a constrained optimization problem: Maximize (1) subject to obtaining (2). Knowing the limits of identification allows you to max out on (1), which is a good thing.

Digression: Nonparametric Identification

Proof Techniques

Two popular ways to prove identification.

(1) Take m' and $m^{\star}$, and proceed as in definition, i.e., prove that $P(m') \neq P(m^{\star})$ if $m' \neq m^{\star}$.

(2) Express m as a functional of $P(m)$ for all $m \in M$, as $m = T(P(m))$ for some T.

The latter "constructive approach" is preferred since it often provides a natural estimator for free. Once we discuss auctions it will become clear how (2) works. In the meantime, let's see how this would work for discrete choice.

Digression: Nonparametric Identification

Discrete Choice, $\epsilon \sim T1EV$

Consider the following **Discrete Choice** model:

$$U_j = u_j(X) + \epsilon_j$$
$$Y = arg \max_{j \in J} \{U_j\}$$

Assume: $\epsilon \perp X$, and that we observe (Y, X_j), whereas ϵ_j is unobservable. For now, look at case where ϵ_j's are independent across j. Q: can we identify $(u(.), F_{\epsilon_1,...,\epsilon_J})$.? What we can conclude right away:

- We can not identify the scale of u_j separately from F_ϵ. Consider two models (u_j, F_ϵ) and $(C \cdot u_j, F_{C\epsilon})$, where $F_{C\epsilon}(\tau) = F_\epsilon(\tau/C)$.
- We can not identify u_j up to an additive function of X. Consider (u_j, F_ϵ) and $(u_j + g(X), F_\epsilon)$, where $g(.)$ does not depend on j. We can set $u_J = 0$ without loss of generality.

Let's therefore consider the case where $\mathbb{E}[\epsilon_j] = 0$, $\mathbb{V}[\epsilon_j] = \sigma = 1$, $\epsilon \sim T1EV$.

Claim: u_j is non-parametrically identified.

Digression: Nonparametric Identification

Discrete Choice, $\epsilon \sim T1EV$

Proof:

$$log(P(Y = J|X)) = log\left(\frac{exp(U_j)}{\sum_{k \in J} exp(U_k)}\right) = U_j - log\left(\sum_k exp(U_k)\right)$$

And therefore:

$$log(P(Y = j|X)) - log(P(Y = J|X)) = U_j - U_J = u_j(X) - u_J(X) = u_j(X)$$

Digression: Nonparametric Identification

Discrete Choice, $\epsilon \sim T1EV$

Proof:

$$log(P(Y = J|X)) = log\left(\frac{exp(U_j)}{\sum_{k \in J} exp(U_k)}\right) = U_j - log\left(\sum_k exp(U_k)\right)$$

And therefore:

$$log(P(Y = j|X)) - log(P(Y = J|X)) = U_j - U_J = u_j(X) - u_J(X) = u_j(X)$$

Estimation:

If X is discrete:

$$log(\widehat{P_j(x)}) = log\left(\frac{\sum 1\{Y = j \wedge X = x\}}{\sum 1\{X = x\}}\right) \xrightarrow{p} u_j(x),$$

else use a Kernel estimator or sieves.

Digression: Nonparametric Identification

Discrete Choice, $\epsilon \sim T1EV$

- Hotz & Miller (1993), as well as Magnac & Thesmar (2002) extend the above result for static models to dynamic models.
- We already knew this, since we have expressed value functions in terms of choice probabilities for estimation (**H+M** inversion).
- More specifically: we can specify a vector of utilities for the reference action π, a distribution for the idiosyncratic shocks G, and a discount factor β, and we will be able to find a model rationalizing the observed CCPs that features (π, β, G) (**Magnac and Thesmar (2002)**).
- This implies that the discount factor is not identified.

Identification, Dynamic Single Agent Models

Intuitive explanation

H+M: $ln(p_j(x)) - ln(p_J(x)) = v_i(x) - v_J(x) \;\; \forall \;\; j \in \mathcal{J}/J, x \in X$. This summarizes all the predictions the model makes about the data. Need β to decompose $v_j(x) - v_J(x)$ in a current utility and a continuation value component.

- $(J-1) \times K$ equations in **H+M**.
- $(J-1) \times K + 1$ unknown primitives, $\Delta\mathbf{v}$ and β.

Solution:

- Tests for forward-looking behavior exploit scenarios in which some variables which affect future utility are known in period t: consumers are deemed forward-looking if their period t decisions depends on these variables. (Example: Chevalier and Goolsbee (2005))
- Use exclusion restriction, shift continuation value component, holding current period utility fixed (Abbring and Oeystein (2018)).
- Typically, literature treats β as known.

Persistent Unobserved Heterogeneity

Arcidiacono and Miller (2011)

Persistent unobserved heterogeneity

So far: Unobserved heterogeneity (in form of action specific ϵ) has convenient form.

- Temporary, IID.

Distinguish two cases:

- Unobserved heterogeneity has some persistence and might change over time. We will see how to deal with this in estimation, Pakes 1986. If you are interested in a serious econometric account of such "Hidden Rust" models you can read Connault (2016).
- Unobserved heterogeneity fully persistent. Formal identification papers: Hall and Zhou (2003), Kasahara and Shimotsu (2009).

Arcodiacono and Miller (2011) provide model for fully persistent heterogeneity.

- Examples in Harold Zurcher setting:
  - Bus brand unobserved to econometrician.
  - Some decisions are made by Harold Zurcher's son.
- **CCP approach:** if there is an unobserved state s, $p(x, a)$ no longer an unbiased estimator.

Arcodiacono and Miller (2011)

EM-Algorithm

Setup: Start again from Rust (1987), add unobserved (to the econometrician) persistent state s.

- Persistent unobserved heterogeneity be at the bus level. Denote π_s the population frequency of type s.
- Let $a_n = (a_{n1}, ..., a_{n\mathcal{T}})$

The ML estimator is no longer additively separable:

$$\{\hat{\theta}, \hat{\pi}\} = arg \max_{\theta, \pi} \sum_{n=1}^{\mathcal{N}} ln\left(\sum_{s=1}^{S} \pi_s \prod_{t=1}^{\mathcal{T}} l(a_{nt}|x_{nt}, s, \hat{p}(a, x, s))\right)$$

EM-Algorithm: instead of directly estimating the above, iterate on the estimation of $\hat{\theta}$ and $\hat{\pi}$. Assume that we have some initial guess $\hat{p}_1(a, x, s)$ and $\hat{\pi}_s$. Note via Bayesian updating one can compute:

$$\hat{q}_{ns} = P(s_n = s|a_n, x_n; \hat{\theta}, \hat{\pi}, \hat{p}_1) = \frac{\hat{\pi}_s \cdot \prod_{t=1}^{\mathcal{T}} l(a_{nt}|x_{nt}, s, \hat{p}_1)}{\sum_{k=1}^{S} \hat{\pi}_k \cdot \prod_{t=1}^{\mathcal{T}} l(a_{nt}|x_{nt}, k, \hat{p}_1)}$$

and from that:

$$\hat{\pi}_s = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \hat{q}_{ns}$$

Arcodiacono and Miller (2011)

EM-Algorithm

Algorithm, Iterate on (2) to (5) until convergence.:

(1) Initialize with some weights $\hat{\pi}_0$ and probabilities p_1^0, θ_0

(2) The m'th step: $\theta^{(m+1)} = \arg\max_\theta \sum_{n=1}^{\mathcal{N}} \sum_{s=1}^{\mathcal{S}} \sum_{t=1}^{\mathcal{T}} q_{ns}^{(m+1)} \cdot ln\left(l(a_{nt}|x_{nt}, s, \hat{p}_1^{(m)}, \theta)\right)$

(3) $\hat{q}_{ns}^{m+1} = \frac{\hat{\pi} \cdot \prod_{t=1}^{\mathcal{T}} l(a_{nt}|x_{nt}, s, \hat{p}_1)}{\sum_{k=1}^{S} \hat{\pi}_k \cdot \prod_{t=1}^{\mathcal{T}} l(a_{nt}|x_{nt}, k, \hat{p}_1)}$

(4) $\hat{\pi}_s^{m+1} = \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \hat{q}_{ns}$

(5) $p_1^{(m+1)}(a, x, s) = \frac{\sum_{n=1}^{\mathcal{N}} \sum_{t=1}^{T} a_{1nt} \cdot q_{ns}^{(m+1)} \cdot I(x_{nt} = x)}{\sum_{n=1}^{\mathcal{N}} \sum_{t=1}^{T} q_{ns}^{(m+1)} \cdot I(x_{nt} = x)}$

Note on EM-algorithm:

- Has the property that each iteration increases the likelihood.
- But no guarantee of convergence, i.e. might get stuck in local max.
- Convergence can be very slow.
- Bonhomme, et Al. (2017) "Discretizing Unobserved Heterogeneity" as an alternative to EM-algorithm.

Finite Dependence

Arcidiacono and Miller's Lemma

Value Function $\longleftrightarrow$ Choice Specific Value Function

Choice Probabilities:

$$P(a|x) = \frac{exp(v(a,x))}{\sum_j exp(v(j,x))}$$

Inclusive value of state:

$$V(x) \equiv log\left(\sum_a exp(v(a,x))\right) + \gamma$$

Combining the two:

$$V(x) \equiv log\left(\sum_j exp(v(a,x))\right) + \gamma = log\left(\sum_a exp(v(a,x))\right)$$

$$+ log(exp(v(a,x))) - log(exp(v(a,x))) + \gamma =$$

$$v(a,x) - log(P(a|x)) + \gamma$$

Finite Dependence

An even simpler approach

We are interested in parameters θ governing current payoffs $u(x, a; \theta)$. The Hotz-Miller inversion does not directly get us to an estimate of θ. We went through the steps to get us there:

1. Express $V(\cdot; \theta)$ in terms of observed choice probabilities $P(\cdot)$ and known current payoffs $u(\cdot, \cdot; \theta)$
2. Construct objective function from implied choice probabilities

In some cases, this can be even simpler:

- Suppose there are sequences of actions which lead to the same state
- After getting to that state, the continuation values are the same
- Differences in the frequencies of those action sequences are related to differences in *current*-period payoffs
- General idea: **finite dependence** (Arcidiacono & Miller '11)

Finite Dependence

Renewal Action

Suppose $a = 0$ takes you to the same distribution of states, **regardless of the current state**. In Rust, replacing the engine is a renewal action:

- $x_t = mileage_t = 0$ if $a_t = replace$, regardless of previous state $x_{t-1} = mileage_{t-1}$

More generally, we require that $\forall x_t, x_t'$:

$$P(x_{t+1} \mid x_t, a_t = 0) = P(x_{t+1} \mid x_t', a_t = 0)$$

$\rightarrow$ $a_t = 0$ yields the same continuation value.

Finite Dependence

Apply to $v(x_t, 1)$ and $v(x_t, 0)$, taking renewal action at $t+1$:

$$v(x_t, 1) = u(x_t, 1) + \beta \cdot \sum_{x_{t+1}} P(x_{t+1} \mid x_t, 1) \cdot V(x_{t+1})$$

$$= u(x_t, 1) + \beta \cdot \sum_{x_{t+1}} P(x_{t+1} \mid x_t, 1) \cdot [\, v_{t+1}(x_{t+1}, 0) - \ln p_{t+1}(x_{t+1}, 0) + \gamma]$$

$$v_{t+1}(x_{t+1}, 0) = u(x_{t+1}, 0) + \beta \cdot \sum_{x_{t+2}} P(x_{t+2} \mid x_{t+1}, 0) \cdot [\, v_{t+2}(x_{t+2}, 0) - \ln p_{t+2}(x_{t+2}, 0) + \gamma]$$

Substituting the second equation into the first:

$$v(x_t, 1) = u(x_t, 1) + \beta \cdot \sum_{x_{t+1}} P(x_{t+1} \mid x_t, 1)[\, u_{t+1}(x_{t+1}, 0) - \ln p_{t+1}(x_{t+1}, 0) + \gamma]$$

$$+ \beta \cdot \sum_{x_{t+1}} P(x_{t+1} \mid x_t, 1) \cdot \left[\beta \sum_{x_{t+2}} P(x_{t+2} \mid x_{t+1}, 0)[\, v_{t+2}(x_{t+2}, 0) - \ln p_{t+2}(x_{t+2}, 0) + \gamma] \right]$$

Key: the term in the second line is **identical** for $v(x_t, 0)$, because of renewal action at $t+1$.

Finite Dependence

Renewal Action

We therefore have

$$v_t(x_t, 1) - v_t(x_t, 0) = u(x_t, 1) - u(x_t, 0)$$
$$+ \beta \cdot \sum_{x_{t+1}} (P(x_{t+1} \mid x_t, 1) - P(x_{t+1} \mid x_t, 0)) \cdot [u(x_{t+1}, 0) - \ln p_{t+1}(x_{t+1}, 0)]$$

Obtains a simple likelihood function:

$$P(a_t = 1 \mid x_t; \theta) = \frac{\exp\{v_t(x_t, 1; \theta) - v_t(x_t, 0; \theta)\}}{1 + \exp\{v_t(x_t, 1; \theta) - v_t(x_t, 0; \theta)\}},$$

given estimated $\hat{p}$'s, at each θ

DDC Estimation via OLS

Scott (2016)

Overview

Applied question

- What are the effects of biofuels policy?
- US biofuels mandate: $\approx$ 10% of gasoline must come from biofuels.
- Supply elasticities based on static models might be downwards-biased.

Estimate model using **linear regression**.

- Relies on Hotz-Miller (1993) inversion.
- Exploits finite dependence, Altug and Miller (1998).
- Can be seen as an **"Euler equation"** approach for DDC models.

Scott (2016)

Method

Key insight:

- Realizations of agents future utility are noisy measures of agent's expectation.

Advantages of approach:

- No need to specify how all state variables evolve over time.

Can be combined with **quasi-experimental** variation:

- Diamond, McQuade, Qian (2018): combine with diff-in-diff IV to study rent control.
- No need to explicitly model expectations.

Scott (2016)

Model + Assumptions

Binary crop choice model

A landowner's choice set: $\mathscr{A} = \{crops, other\}$. If field i is in state x at time t, then the *expected* profits to land use j are:

$$\pi(a, x, \omega_t, \nu_{it}) = \alpha_{0,a,x} + \alpha_R \cdot R_a(\omega_t) + \xi_{a,x}(\omega_t) + \nu_{iat}$$

Notation: i: field,a: land use, k: field state, ω: market state, R: expected returns, observable to econometrician, ξ: unobservable shock to returns, ν: idiosyncratic field-level shock, α: parameters to be estimated,

Assumptions:

1. Small fields, no externalities: the distribution of the market state ω_{t+1} conditional on ω_t is not affected by changing the land use in any single field.
2. Logit errors: The idiosyncratic error term ν_{iat} has a type 1 extreme value distribution, independently and identically distributed across i, a, and t.
3. Landowners have rational expectations.

Scott (2016)

Dynamics: Field-Level

Landowners maximize expected discounted profits. Field states evolve according to a simple deterministic process:

$$x_{i,t+1} = x^+(a_{it}, x_{it}) = \begin{cases} 0 & \text{if } a_{it} = crops \\ \min\{x_{it} + 1, \bar{x}\} & \text{if } a_{it} = other \end{cases}$$

Important details:

- Like in bus-engine replacement, there is a renewal action.
- No explicit assumptions on the evolution of R and ξ
- Estimating the process governing the evolution of the unobservable supply shock ξ is especially difficult.

Scott (2016)

Regression equation construction

Steps:

1. Start with condition for indifferent agent
2. Introduce expectational error (**“Euler equation”** error term)
3. Forward calculation of continuation values using conditional choice probabilities
4. Rearrange into regression equation

Scott (2016)

Dynamics: Definitions

Ex ante value function:

$$\bar{V}_t(x) \equiv \int V_t(x, \omega_t, \nu)\, dF(\nu),$$

the value function integrated over idiosyncratic shocks ν.

Choice specific value function:

$$v_t(a, x) \equiv \bar{\pi}_t(a, x) + \beta \cdot E_t\left[\bar{V}_{t+1}\left(x^+(a, x)\right)\right],$$

where $\bar{\pi}$ indicates ex ante profits (without idiosyncratic error). Conditional choice probabilities (with logit assumption):

$$P(a|x, t) = \frac{\exp(v_t(a, x))}{\sum_{j \in \mathscr{A}} \exp(v_t(j, x))}$$

Scott (2016)

Step 1: Indifferent agent condition (HM-inversion)

The **HM inversion** with logit errors:

$$\ln\left(\frac{P(a|x,t)}{P(a'|x,t)}\right) = v_t(a,x) - v_t(a',x)$$

Rewrite as relationship between current profits and continuation values:

$$\bar{\pi}_t(a,x) - \bar{\pi}_t(a',x) + \ln\left(\frac{P(a|x,t)}{P(a'|x,t)}\right) =$$

$$\beta\left(E_t\left[\bar{V}_{t+1}\left(x^+(a',x)\right)\right] - E_t\left[\bar{V}_{t+1}\left(x^+(a,x)\right)\right]\right)$$

Expectation over **unobserved** state variables!

Scott (2016)

Step 2: Expectational errors

Expectational error:

$$\varepsilon_t^{CV}(a,x) \equiv \beta \left(E_t \left[\bar{V}_{t+1}\left(x^+(a,x)\right)\right] - \bar{V}_{t+1}\left(x^+(a,x)\right)\right)$$

Adding and subtracting realizations of the value function, the condition can be rewritten:

$$\bar{\pi}_t(a,x) - \bar{\pi}_t(a',x) + \ln\left(\frac{P(a|x,t)}{P(a'|x,t)}\right) =$$

$$\beta\left(\bar{V}_{t+1}\left(x^+(a',x)\right) - \bar{V}_{t+1}\left(x^+(a,x)\right)\right)$$

$$+\varepsilon_t^{CV}(a',x) - \varepsilon_t^{CV}(a,x)$$

The expectational error terms are mean uncorrelated with any variables in the information set ω_t.

$\longrightarrow$ Next two slides show how continuation values can be replaced in this equation to transform the problem into a **linear** estimation equation.

Scott (2016)

Step 3: Forward calculation

Use Arcidiacono and Miller's lemma to replace continuation values:

$$\bar{V}_t(x) = -\ln\left(P(a^\star|x,t)\right) + v_t(a^*,x) + \gamma$$

which holds for **any** land use a^*. Recall $x^+(crops,x) = 0$ for all x (renewal action). By choosing $a^* = crops$, continuation values from $t+2$ onward will cancel:

$$\begin{aligned}
\bar{V}_{t+1}(x^+(a,x)) &= -\ln\left(P(a^*|x^+(a,x), t+1)\right) + \bar{\pi}_{t+1}(a^*, x^+(a,x)) + \beta\bar{V}_{t+2}(0) + \gamma \\
\bar{V}_{t+1}(x^+(a',x)) &= -\ln\left(P(a^*|x^+(a',x), t+1)\right) + \bar{\pi}_{t+1}(a^*, x^+(a',x)) + \beta\bar{V}_{t+2}(0) + \gamma
\end{aligned}$$

Scott (2016)

Step 4: rearrange into regression equation

$$Y_{x,t} = \tilde{\Delta}\alpha_0(x) + \alpha_R \Delta R_t + \tilde{\Delta}\xi_{xt} + \Delta\varepsilon_t^{CV}$$

where

$$\begin{aligned}
Y_{x,t} &= \ln\left(\frac{P(crops|x,t)}{P(other|x,t)}\right) + \beta \cdot \ln\left(\frac{P(crops|0,t+1)}{P(crops|x^+(other,x),t+1)}\right) \\
\tilde{\Delta}\alpha_0(x) &= \alpha_{0,crops,x} - \alpha_{0,other,x} \\
&\quad +\beta\left(\alpha_{0,crops,0} - \alpha_{0,crops,x^+(other,x)}\right) \\
\Delta R_t &= R_{crops,t} - R_{other,t} \\
\tilde{\Delta}\xi_t &= \xi_{crops,x,t} - \xi_{other,x,t} + \beta\left(\xi_{crops,0,t+1} + \xi_{crops,x^+(other,x),t+1}\right) \\
\Delta\varepsilon_t &= \varepsilon_t^{CV}(crops,x) - \varepsilon_t^{CV}(other,x)
\end{aligned}$$

Generalizes to multinomial setting and (almost) any distribution for ν.

Scott (2016)

Identification of Payoff Parameters

- The regression identifies one value for each x:

$$\begin{aligned}\tilde{\Delta}\alpha_0(x) &= \alpha_{0,crops,x} - \alpha_{0,other,x} \\ &\quad + \beta\left(\alpha_{0,crops,0} - \alpha_{0,crops,x^+(other,x)}\right)\end{aligned}$$

- However, we have $2|x|$ values of α_0 in the profit function for each x: $\alpha_{0,crops,x}$ and $\alpha_{0,other,x}$. For identification, assume that $\alpha_{0,other,x} = 0$ for all x.
- Such restrictions are often called “normalizations” in the literature, but they are substantive restrictions in principle (Magnac and Thesmar, 2002).
- However, more recent work (Norets and Tang, 2014; Kalouptsidi, Scott, and Souza-Rodrigues 2015) has considered whether these restrictions actually matter for counterfactuals.
- Turns out counterfactuals in this paper (i.e., long run elasticity calculations) are identified, meaning that the above restriction is harmless.

Scott (2016)

Conclusion

- Fully dynamic model estimated with regression equation
- Euler equation approach avoids the need for a full model of state variables; data limitations show up in an error term (like static applied micro models) rather than necessarily leading to model misspecification.
- Can be combined with quasi-experimental techniques.
- Depending on the size of the state space, not all cells in $p(.|x, t)$ might be populated. Might have to interpolate, for example, through kernel smoothing.

Autocorrelated Unobservables

Some other relevant work

Related work to take care of endogeneity issues caused by **autocorrelated** unobserved state variables:

Berry and Compiani (2017): *An Instrumental Variable Approach to Dynamic Models*

- Generalized instrumental variable approach (Chesher and Rosen (2015)).
- Set identification, estimator is discrete grid search.

Kalouptsidi et Al. (2017): *An Instrumental Variable Approach to Dynamic Models*

- Linear IV approach builds on Scott's Euler equations insight.

Simulation Estimation

Pakes (1986): Patents as Options

Overview

The patent literature:

- Intellectual property (what is the value of patents?)
- Measuring the causes, effects and distribution of benefits from innovation (often uses patent counts and author linkages).
- Important input to inform optimal length of patents.

Main idea

- In many countries, patent holders required to pay renewal fee to keep patents in force.
- More than 90% of patent holders let patents expire before the limit on patent lives.
- Use patent renewal decisions and costs of renewal to infer the distribution of patent values. Patents as options.

Methodological contributions

- Simulation estimator, serially correlated unobserved state variable (relaxes iid, conditional logit assumptions)

Pakes (1986): Patents as Options

TABLE I

CHARACTERISTICS OF THE DATA[a]

Country Characteristic	France	U.K.	Germany
1. f	2	5	3
2. L	20	16	18
3. Application dates of cohorts	1951–79	1950–74	1952–72
4. First/last year in which renewals are observed	1970/81	1955/78	1955/74
5. Patents studied from cohort: all patents	Applied for	Applied for	Granted
6. Estimated average ratio of patents granted to patents applied for[b]	.93	.83	.35
7. $\overline{NPAT} = N/J$	36,865	37,286	21,273

[a] Symbols are defined as follows: f is the first age for which a renewal fee is due; L is the last age at which an agent can keep the patent in force by payment of an annual renewal fee; and $\overline{NPAT}$ is the average number of patents per cohort.

[b] For France and the U.K. these estimates were obtained as follows. Let n_t be the number of patents applied for in year t, and $\tilde{n}_t$ be the number of patents granted. Then the ratio was calculated as $T^{-1}\sum_{t=1}^{T}[(\sum_{\tau=1}^{4} .25\tilde{n}_{t+\tau})/n_t]$. In Germany the ratio of the patents granted to those applied for from a given cohort was directly available, and these ratios were simply averaged over the cohorts studied.

Pakes (1986): Patents as Options

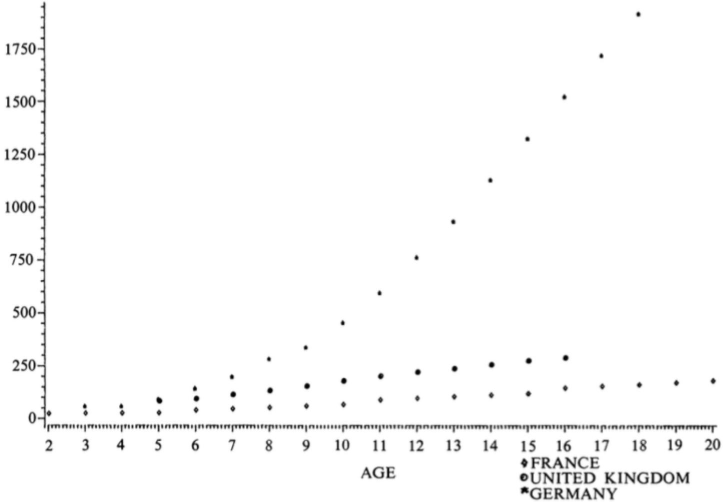


FIGURE 3.—Average of renewal fee schedules.

Pakes (1986): Patents as Options

Model Details

- This is, again, an optimal stopping problem!
- First year returns from patent protection are r_1. (May be only small part of total returns to the patented idea).
- Returns in future years $r_2, r_3, \ldots$ are random.
- Cost of renewing each year $c_1, c_2, \ldots$
- 2 period intuition:
  - 2nd period: renew if $r_2 > c_2$, so obtain $\max\{r_2 - c_2, 0\}$
  - First period (if renew):

$$r_1 - c_1 + \beta \int \max\{r_2 - c_2, 0\} P(dr_2 | r_1)$$

  - P stochastically increasing in r_1.
  - There is a cutoff $\bar{r}_1 < c_1$ s.t. if $r < \bar{r}_1$, patent is not renewed.
  - Second period cutoff is simply $\bar{r}_2 = c_2$.

Pakes (1986): Patents as Options

Model Details

Ω_a : history of returns up to age, $\{r_1, ..., r_a\}$. The expectation is over $r_{a+1}|\Omega_a$. The sequence of conditional distributions $F(a_{t+1}|\Omega_a)$, $a = 1, 2, ...$ is an important component of the model. Pakes' assumption:

$$r_{a+1} = \begin{cases} 0 & \text{with prob. } \exp(-\theta r_a) \\ \max(\delta r_a, z) & \text{with prob. } 1 - \exp(-\theta r_a) \end{cases}$$

where density of z is $q_a(z) = \frac{1}{\sigma_a} \exp[-(\gamma + z)/\sigma_a]$ and $\sigma_a = \phi^{a-1}\sigma$, $a = 1, \ldots, L-1$.

Sequence of renewal fees $\{c_a\}$, increasing in age. Gives rise to value function:

$$V(r, a) = \begin{cases} \max\{0, r - c_a + \beta E[V(r', a+1)|r, a]\} & \text{if } a < A \\ \max\{0, r - c_A\} & \text{if } a = A \end{cases}$$

Pakes (1986): Patents as Options

Model Details

A note on the nature of this problem: Since the maximal age is finite this is a finite horizon (non-stationary) dynamic optimization problem. Most dynamic problems fall into two camps (i) infinite horizon stationary problems and finite horizon, non-stationary problems. Stationarity just means that the value functions and optimal decision rules are *time-invariant* functions of the state variables.

Solution is a cut-off strategy:

- Note that agent renews if $r + \beta E[V(r', a+1)|r, a] > c_a$. Since this is strictly increasing in r at each a, there exists a unique cutoff $\bar{r}_a < c_a$ s.t. patent is renewed iff $r > \bar{r}_a$.
- $\bar{r}_a < \bar{r}_{a+1} < \ldots < \bar{r}_A = c_A$ (The fact that renewal fees are increasing in age, while the option value is decreasing, implies that cutoffs are increasing in age.). Solved for starting in the last period.

Pakes (1986): Patents as Options

Simulation Estimator

Outer loop: is concerned with evaluating likelihood that arises from a complicated integral:

- Maximize log-likelihood: $\log \mathscr{L}(\theta) = n^{-1} \sum_a s(a) \log \pi_a(\theta)$.
  - n is # of patents in cohort
  - s_a is the fraction of the original sample dropping out at age a (or surviving until terminal year if $a = A$.
  - $\pi_a(\theta)$ is probability of dropping out at age a.
- If $F(r, a; \theta)$ be the distribution of patent values at age a we have $\pi_a(\theta) = F(\bar{r}_a, a; \theta) - F(\bar{r}_{a-1}, a-1; \theta)$
- Issue: family $\{F(\cdot, \cdot; \theta)\}$ is complicated (not analytic).

Inner loop: solve the agent's problem.

Pakes (1986): Patents as Options

Inner loop procedure: backwards induction.

- At L there is no more continuation value, return is r_L.
- At $L - 1$ solve for a grid of interest rates...interpolate (expectations over returns depend on parameter guess).
- ...
- This procedure gives the cut-offs needed for the outer loop.

Outer loop procedure: simulation estimator.

- Start with a draw $r_1 \sim f(r, 1; \theta)$.
- Let $\bar{r}_t$ be cut-off from value functions, at iteration $t < A$:
  - If $r_t \geq \bar{r}_t(\theta)$, take a draw from $r_{t+1} \sim P(\cdot | t, r_t; \theta)$.
    - i.e., stayed in at t
  - if $r_t < \bar{r}_t(\theta)$, up counter for $\hat{\pi}_t(\theta)$ by one.
    - i.e., dropped out
- Use $\hat{\pi}_t(\theta)/(NSIM)$ as estimate of probability of dropping out at age a (conditional on making it to that point) to compute likelihood.

Pakes (1986)

TABLE V

PERCENTILES (pl) AND LORENZ CURVE COEFFICIENTS (lc) FROM THE DISTRIBUTION OF REALIZED PATENT VALUES[a]

	Country					
	France		U.K.		Germany	
Per cent p	pl ($)	lc per cent	pl ($)	lc per cent	pl ($)	lc per cent
.25	75.23	.544	355.55	.554	1,999.60	2.249
.50	533.96	1.833	1,516.84	3.247	6,252.93	7.341
.75	3,731.35	8.087	7,947.55	16.369	19,576.26	25.288
.85	10,292.06	19.575	15,357.09	31.721	32,428.14	41.001
.90	17,423.11	31.261	22,206.21	44.257	44,241.87	52.672
.95	31,609.59	52.461	34,740.07	62.960	65,753.61	69.223
.97	42,905.78	65.514	43,889.95	73.640	78,299.01	78.348
.98	51,215.84	73.729	51,277.22	80.072	94,842.63	83.800
.99	66,515.40	84.011	65,075.08	87.858	118,354.78	90.330
maximum	259,829.27	—	374,028.70	—	419,217.55	—
mean	5,631.03	—	7,357.05	—	16,169.48	—
NPAT	36,865		37,826		21,273	

[a] The realized value for patent i is $\sum_{\tau=1}^{\tau_i^*} \beta^{(\tau-1)}(r_{i,\tau}-c_\tau)$, where τ_i^* is the last age at which patent i was kept in force. See also the note to Table III.

[16] Of course some of these patents had negative (though small in absolute value) realized values, as they were patents on which early renewals were paid for options which did not materialize. If, for example, we had defined the realized values as the discounted sum of net returns from age two, rather than from age one (as in the table), the Lorenz curve coefficient corresponding to $p = .25$ would have been negative, though close to zero.

Pakes (1986)

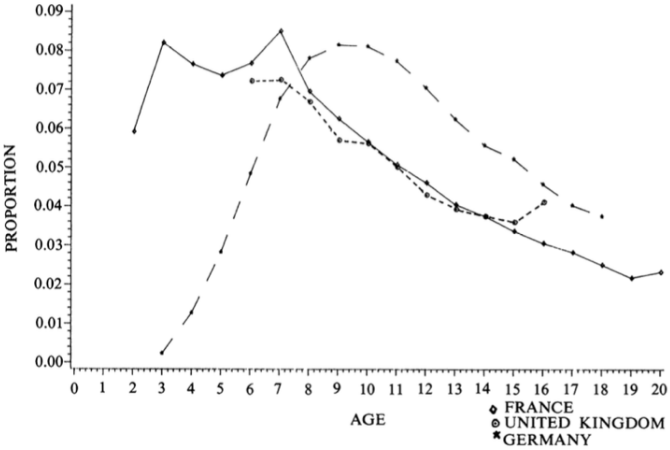


FIGURE 2.—Average drop out proportions.[a]

Pakes (1986): Patents as Options

Conclusion

- Important question, how to value patents absent transaction data that would pin down their value directly.
- Paper was one of the first using simulation for estimation in economics (Lerman and Manski, 1981).
- One limitation of the paper is that renewing a patent is cheap, but the aggregate value of patents is really driven by the upper tail of values (most patents turn out to be worth zero).
- This upper tail is identified only by the strong functional form assumption - but not obvious how else you might pin it down.
- Non-parametric identification for these types of models not fully worked out.

Dynamic Demand

Dynamic Demand

Two Examples

We have covered static models of demand, such as BLP (1995). We have also looked at single agent models of dynamic behavior such as Rust (1987). What if we could put those two together? Why?

Example: Durable Goods

- We are often interested in the case where a firms' products compete not only against products of other firms, but also with their own products over time. This is a feature of many durables.
- In high-tech products, we have 'S'-shaped penetration curve. At some point prices are cut and the sales go down rather than up. Why?

Example: Storable Goods

- Laundry detergent goes on discount: $\approx$ 1 week in every 4-5. A large fraction of overall quantity is sold during these discount weeks.
- This would imply highly elastic demand. (1) Laundry detergent business must be extremely competitive. (2) There would be a substantial response to a **permanent** price change. Seems unlikely that (1) and (2) are true.

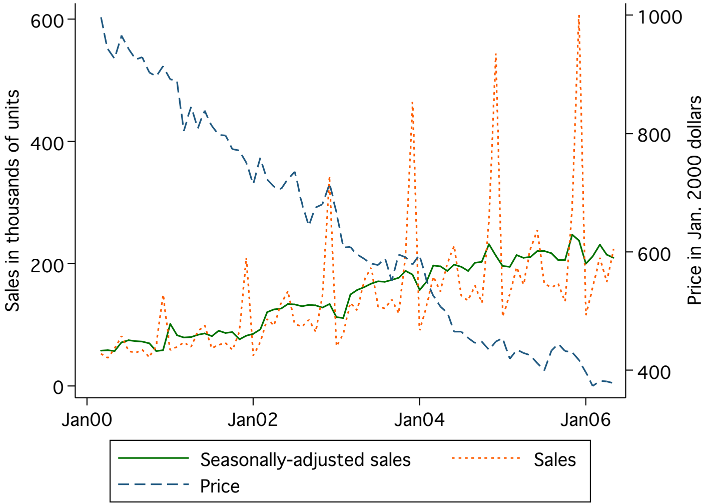
We already discussed **Hendel and Nevo (2006)** and will now look at **Gowrisankaran and Rysman (2012)**. Both are unusual for modern IO in that they are pure measurement exercises; no counterfactuals. They illustrate nicely how we would obtain biased demand estimates without considering dynamics.

Gowrisankaran and Rysman (2012)

Figure 1: Average non-indicator characteristics over time

Gowrisankaran and Rysman (2012)

Figure 3: Prices and sales for camcorders



Dynamic Demand

Some Observations about Consumer Electronics

- Today a 55” 4K LCD TV (8M pixels) is $330. In 2006, you could buy a 32” 720P (< 1M pixels) TV for $> \$10,000$.
- Over time consumers buy better cameras, smart phones, or larger TV’s
- The BLS tries to do *chaining* and *quality adjustments* but in high-tech products this can be very difficult.
- This has a potentially large impact on price indices (a small bias in the CPI can be billions of dollars in SSA/Medicare payments).

Gowrisankaran and Rysman (2012)

Setup

Each consumer type is subscripted by i, and chooses a product j in period t to maximize utility:

$$\begin{aligned} u_{ijt} &= \underbrace{\alpha_i^x x_{jt} + \xi_{jt}}_{f_{jt}} - \alpha_i^p p_{jt} + \varepsilon_{ijt} \\ u_{i0t} &= f_{i0t} + \varepsilon_{i0t} \end{aligned}$$

In the static model we have $f_{i0t} = 0$ for every i, t.

- Consumers are heterogeneous. Dynamic replacements will lead to a law of motion $w_{i,t+1} = h(w_{i,t}, s_{ijt})$.
- Where does the bias in the static model arise from? Correlation between f_{i0t} and prices.

We can think about dynamic models as having time varying utility for the outside option f_{i0t}.
Consumers are **forward looking** regarding durables they will own in the future.

Gowrisankaran and Rysman (2012)

Outside Good

Goal is to endogenize the utility of the outside option f_{i0t}:

- Previous choice(s): $f_{ij,t-1}$.
- Consumer's best option from tomorrow's market

Ad-Hoc approach:

- Just proxy with a time trend (Lou Prentice Ying 2012), (Eizenberg 2011) etc. That is $f_{i0t} = \gamma_{0i} + \gamma_{1i}t + \gamma_{2i}t^2 + \ldots$
- Can still yield the correct elasticity.
- Not structural. When is this a problem?

Here:

$$V_i(f_{i0t}, \varepsilon_{it}, \Omega_t) = \max\{f_{i0t} + \beta \cdot E_{\Omega}[E_{\varepsilon}V_i(f_{i0t}, \varepsilon_{it}, \Omega_{t+1})|\Omega_t], \\ \max_j f_{ijt} - \alpha_i p_{jt} + \beta \cdot E_{\Omega}[E_{\varepsilon}V_i(f_{ijt}, \varepsilon_{it}, \Omega_{t+1})|\Omega_t]\}$$

where Ω is state implied by product selection and prices in the market.

Gowrisankaran and Rysman (2012)

Replacement Problem

This Bellman defines a **Replacement Problem**.

- You own a single durable good with the option to *upgrade* in each period.
- When you upgrade you throw away the old durable and get nothing in exchange.
- After a purchase *j* you receive flow utility $f_{i0t+1} = f_{ijt}$ each period if you don't make a new purchase.
- We could add in depreciation if we wanted to.
- Other studies of durables have focused on *Secondary Markets* or *Perfect Rental Markets*.
- *Resale:* Important for cars, less relevant for high-tech.

Gowrisankaran and Rysman (2012)

Inclusive Value

It is helpful to use **Rust's Trick** and write: $EV_i(f_{i0t}, \Omega_t) = \int V_i(f_{i0t}, \varepsilon_{it}, \Omega_t) f(\varepsilon)$.

$$V_i(f_{i0t}, \varepsilon_{it}, \Omega_t) = \max\{f_{i0t} + \beta E_\Omega[EV_i(f_{i0t}, \Omega_{t+1})|\Omega_t] + \varepsilon_{i0t},$$

$$max_j f_{ijt} - \alpha_i p_{jt} + \beta E_\Omega[EV_i(f_{ijt}, \Omega_{t+1})|\Omega_t] + \varepsilon_{ijt}\}$$

We can write the **ex-ante** expected utility of purchasing in period t without having to condition on which good you purchase:

$$\delta_i(\Omega_t) = E_\varepsilon[\max_j f_{ijt} - \alpha_i p_{jt} + \beta E_\Omega[EV_i(f_{ijt}, \Omega_{t+1})|\Omega_t] + \varepsilon_{ijt}] =$$

$$\log\left(\sum_j \exp[f_{ijt} - \alpha_i p_{jt} + \beta E_\Omega[EV_i(f_{ijt}, \Omega_{t+1})|\Omega_t]]\right)$$

Gowrisankaran and Rysman (2012)

Inclusive Value Sufficiency

$$EV_i(f_{i0}, \Omega) = \log\left(\exp[f_{i0} + \beta E_{\Omega'}[EV_i(f_{i0}, \Omega')|\Omega]] + \exp(\delta_i(\Omega))\right) + \eta$$

where $\eta = 0.577215665$ *(Euler's Constant)*.

The fact that the expected value function depends recursively on itself and $\delta_i(\Omega_t)$ (Inclusive Value) leads to the following assumption:

Inclusive Value Sufficiency

If $\delta_i(\Omega) = \delta_i(\tilde{\Omega})$ then $g(\delta_i(\Omega')|\Omega) = g(\delta_i(\tilde{\Omega}')|\tilde{\Omega})$ for all $\Omega, \tilde{\Omega}$.

- The idea is that δ tells me everything about the future evolution of the states
- More restrictive than it looks. Is δ low because quality is low or because prices are high? Is this the result of a dynamic pricing equilibrium?
- Similar assumption made to keep games with many firms tractable. We will learn about this later.

Gowrisankaran and Rysman (2012)

Inclusive Value Sufficiency and Rational Expectations

Under **IVS** the problem reduces to:

$$\begin{aligned} EV_i(f_{i0}, \delta_i) &= \log\left[\exp(f_{i0} + \beta E_{\Omega'}[EV_i(f_{i0}, \delta_i')|\delta_i]) + \exp(\delta_i)\right] \\ \delta_i &= \log\left(\sum_j \exp[f_{ijt} - \alpha_i p_{jt} + \beta E_{\delta'}[EV_i(f_{ijt}, \delta_i')|\delta_i]\right) \end{aligned}$$

The idea is that the inclusive value δ_{it} IS the state space, along with his current holding of the durable f_{i0t}. How to deal with **expectations**?

$$E_{\delta'}[EV_i(f_{ijt}, \delta_i')|\delta_i]$$

We need to take a stand on $g_i(\delta_i'|\delta_i)$ the anticipated law of motion for δ_i. G&R assume it follows an $AR(1)$ process.

$$\delta_{it+1} = \gamma_0 + \gamma_1 \delta_{it} + \nu_{it} \text{ with } \nu_{it} \sim N(0, \sigma_\nu^2)$$

If we see δ_{it} we could just run the $AR(1)$ regression to get consumer belief's $\hat{\gamma}$

Gowrisankaran and Rysman (2012)

Rational Expectations-Interpolation

How to compute **complicated expectations**?

$$E_{\delta'}[EV_i(f_{ijt}, \delta'_i)|\delta_i, \gamma] = \int EV_i(f_{ijt}, \delta'_i) g(\delta'|\delta, \gamma)$$

- We need to integrate $EV(f_{ijt}, \delta_i)$ (a function) over a normal density.
- But we don't observe $EV(f_{ijt}, \delta_i)$ everywhere, only on the grid points of our state space.
- We need to *interpolate* $\widehat{EV}_i(\delta_i^s)$ (Linear, Cubic Spline, etc.)
- We might as well interpolate the function at the *Gauss-Hermite* quadrature nodes and weights, re-centered at $\gamma_0 + \gamma_1 \delta$ in order to reduce the number of places we interpolate $\widehat{EV}_i$.
- Tishara will discuss **quadrature methods** in recitation (?).

The Estimation Problem

Overview

We need to solve $\forall i, t$:

$$
\begin{aligned}
S_{jt} &= \sum_i w_i s_{ijt}(f_{i0t}, \delta_{it}) \\
f_{ijt} &= \overline{\alpha} x_{jt} + \xi_{jt} + \sum_l \sigma_l x_{jl} \nu_{il} \\
s_{ijt}(f_{i0t}, \delta_{it}) &= \frac{\exp[f_{ijt} - \alpha_i p_{jt} + \beta E_{\Omega'}[EV_i(f_{ijt}, \delta_i')|\delta_i]}{\exp[EV_i(f_{i0t}, \delta_{it})]} \\
EV_i(f_{i0}, \delta_i) &= \log\left[\exp(f_{i0} + \beta E_{\Omega'}[EV_i(f_{i0}, \delta_i')|\delta_i]) + \exp(\delta_i)\right] \\
\delta_i(EV_i) &= \log\left(\sum_j \exp[f_{ijt} - \alpha_i p_{jt} + \beta E_{\delta'}[EV_i(f_{ijt}, \delta_i')|\delta_i]\right) \\
E[\delta_{it+1}|\delta_{it}] &= \gamma_0 + \gamma_1 \delta_{it}
\end{aligned}
$$

Estimation

Algorithm

Like BLP we guess the nonlinear parameters of the model θ

1. **(Inner Loop)** Solve for EV_i by iteratively computing δ, and running the regression for each i and spline/interpolating to compute $E[EV_i]$.
2. **(Middle Loop)** G&R use a modified version of the BLP contraction mapping. We need to make sure to update the w_{it} via $h(.)$. (This is a TPM that tells maps the transition probabilities of type i holding $f_{i,0,t}$ to $f_{i,0,t+1}$).
3. **(Outer Loop)** Once we've solved this whole system of equations, we use ξ to form moments just like BLP and do GMM.

Gowrisankaran and Rysman (2012)

COLI Index

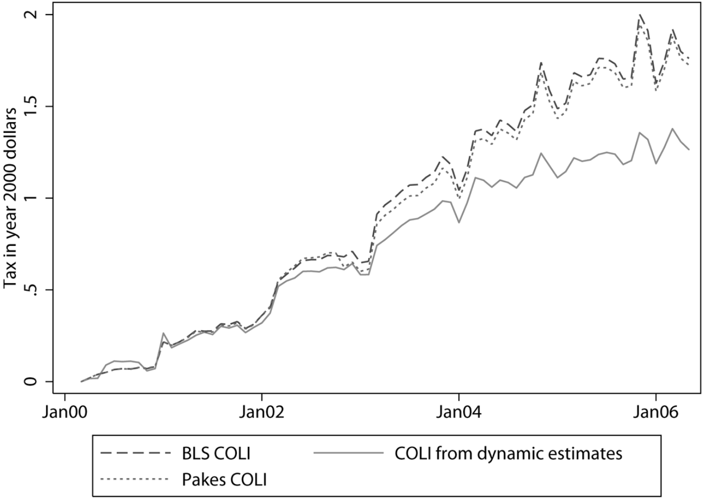


FIG. 14.—Average monthly value from camcorder market

Gowrisankaran and Rysman (2012)

Conclusion

Elasticity measurement:

- *Temporary price change, single camcorder:* long run and short run elasticity are close, camcorders are close substitute.
- *Temporary price change, all camcorders:* Long run elasticity is smaller than short run elasticity.
- *Permanent price change:* long run elasticity is larger.

The value of new goods:

- COLI: Income (tax) that would make household indifferent between buying today and some later period.
- G+R: People who buy later, value goods less. COLI is overstated if we interpret this as an increase in quality.

Other comments:

- No panel data, not clear what people are upgrading from.
- No supply side, but many interesting exercises without counterfactuals.

Demand for Storable Goods

Demand for Storable Goods

References:

- Pesendorfer (2002), Erdem, et al. (2003), Hendel and Nevo (2006).

Observation:

- When a supermarket cuts the price of laundry detergent for a week there is a large increase in sales. This leads us to conclude consumers are extremely elastic with respect to price
- When a supermarket makes a permanent price cut to laundry detergent, there is little sales impact in the long run. Now consumers look highly inelastic with respect to price

Purchasing vs. consumption elasticity

- In the short-run, store-level demand elasticity reflect stockpiling behavior.
- In the long-run, store-level demand elasticity reflects consumption decisions.

$\longrightarrow$ Ignoring stockpiling and forward-looking behavior can lead to **biased own and cross price elasticities**.

Demand for Storable Goods

Price discrimination theory of sales:

- If consumers differ in the storage costs (or credit constraints) firms have an incentive to price discriminate. Ketchup pricing example (Pesendorfer 2002):

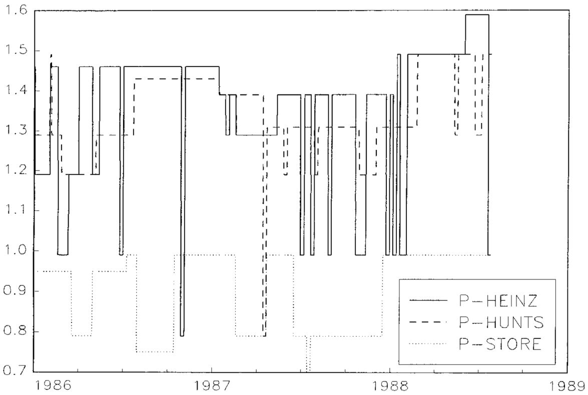


FIG. 1.—Prices for selected products in supermarket 1

Hendel and Nevo (2006)

A Note on the Data

- Dominick's, 9 Supermarkets in Chicago.
- Three years of data (1991-1993), store level and household panel.
- **Store level:** For each product in each store-week observe average price, average quantity, promotions.
- **Household level:** Sample of households, supermarket visits, total expenditure during each visit, detailed purchase information in 24 product categories.

- These days: *Nielsen Scanner Data*, available through the *Booth Kilts Center*.

Hendel and Nevo (2006)

Reduced Form Companion Paper

Results:

- Aggregate demand increases as a function of duration from previous sale, and this effect differs between sale and non-sale periods.
- Fraction of purchases on sale is higher in one market (the market that on average has larger houses), and if there is a dog in the house. Both of these measures could potentially be correlated with lower storage costs.
- When buying on sale, households tend to buy more quantity (by buying either more units or larger sizes), buy earlier, and postpone their next purchase.
- Inventory constructed under the assumption of fixed consumption over time is negatively correlated with quantity purchased and the probability of purchase.
- The patterns of sales across different product categories are consistent with the variation in storage costs across these products.

Hendel and Nevo (2006)

Data and Model Basics

- Store-level: for each brand 13 (j) size x: 32-256oz in each store, each week (t)
  - Price p_{jxt}
  - Quantity q_{jxt}
  - Promotions a_{jxt} (binary for feature/display)
- Consumers are of type h with utility: $u(c_{ht} + \nu_{ht}; \theta_h)$
- Current consumption is $c_{ht} = \sum_j c_{jht}$ not brand specific!
- There is a shock affected marginal utility of consumption ν_{ht}.
- Decision: $d_{hjxt} = 1$ is a purchase of h of brand j and size x at t. (includes outside option $= 0$).

Hendel and Nevo (2006)

Quantity Sold

	Quantity Discount (%)	Quantity Sold on Sale (%)	Weeks on Sale (%)	Average Sale Discount (%)	Quantity Share (%)
Liquid					
32 oz.	—	2.6	2.0	11.0	1.6
64 oz.	18.1	27.6	11.5	15.7	30.9
96 oz.	22.5	16.3	7.6	14.4	7.8
128 oz.	22.8	45.6	16.6	18.1	54.7
256 oz.	29.0	20.0	9.3	11.8	1.6
Powder					
32 oz.	—	16.0	7.7	14.5	10.1
64 oz.	10.0	30.5	16.6	12.9	20.3
96 oz.	14.9	17.1	11.5	11.7	14.4
128 oz.	30.0	36.1	20.8	15.1	23.2
256 oz.	48.7	12.9	10.8	10.3	17.3

Hendel and Nevo (2006)

The Maximization Problem

$$
\begin{aligned}
V(s_t) &= \max_{c_h(s_t), d_{jxt}(s_t)} \sum_t \beta^{t-1} E[u(c_{ht} + \nu_{ht}; \theta_h) - C_h(i_{h,t+1}; \theta_h) \\
&+ \sum_j d_{hjxt}(\alpha_h^p p_{jxt} + \xi_{hjx} + \alpha_h^a a_{jxt} + \epsilon_{hjxt})|s_t] \\
i_{h,t+1} &= i_{ht} + x_{ht} - c_{ht} \\
\sum_{j,x} d_{hjxt} &= 1
\end{aligned}
$$

- Abuse of notation: x_{ht} is size of the choice
- $C_h(i; \theta_h)$ is cost of storage
- ξ_{hjx} captures expected future differences in utility of x units of j **at time of purchase**. Strange interpretation, but valid as long as: discounting is low, brand-specific differences in utility (but not consumption) enter linearly.

Hendel and Nevo (2006)

Comments on the State

State variable:

- $s_t = \{i_t, \nu_t, \mathbf{p}_t, \mathbf{a}_t, \xi_t, \epsilon_t\}$.

Comments:

- State variable i_t is unobserved to the econometrician.
- Curse of dimensionality: $\{\mathbf{p}_t, \mathbf{a}_t, \xi_t, \epsilon_t\}$ is a $4 \times 5 \times J$ dimensional matrix.
- Rust (1987), also has unobserved state variables (i.e. logit errors). However, we assumed that it was conditionally independent of the observed state and separately additive, which allowed us to work with the expected maximum function. Here: it is serially correlated and nonseparable.

Assumptions

- 1: ν_t is independently distributed over time and across consumers. **No serial correlation!**
- 2: Prices p_{jxt} and advertising a_{jxt} follow an exogenous first-order Markov process. **Hard to justify this with a model of profit maximizing supply!**
- 3: ϵ_{jxt} is i.i.d. extreme value type 1.

Hendel and Nevo (2006)

Likelihood

Conditional on Household we can write the probability of a sequence of purchase decisions.

Beginning of period inventory depends on previous decisions, previous shocks, and initial inventory.

$$
\begin{aligned}
Pr(d_{jx}|p_t, i_t, \nu_t) &= \frac{\exp[\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + M(s_t, j, x)]}{\sum_{k,y} \exp[\alpha p_{kyt} + \xi_{jy} + \beta a_{kyt} + M(s_t, k, y)]} \\
M(s_t, j, x) &= \max_c [u(c + \nu_t) - C(i_{t+1}) + \beta E[V(s_{t+1}|d_{jx}, c, s_t]]
\end{aligned}
$$

State space has very **high dimension**. Keeping track of all brands/prices would be very costly

Hendel and Nevo (2006)

3-Step Proceure

To reduce complexity, Hendel and Nevo propose a 3-step estimator

- First, maximize likelihood of observed brand choice **conditional** on size in order to recover the (α, ξ) parameters.
- Second step: compute **inclusive values** for each size and transition probability matrix.
- Third, solve a quantity choice only nested fixed point problem. The key is that there is **only one "index price" per size**.

Hendel and Nevo (2006)

Step 1: Brand Choice

$$Pr(d_{jx}|p_t, i_t, \nu_t) = \frac{\exp[\alpha p_{jxt} + \xi_{jx} + \beta a_{jxt} + M(s_t, j, x)]}{\sum_{k,y} \exp[\alpha p_{kyt} + \xi_{jy} + \beta a_{kyt} + M(s_t, k, y)]}$$

$$M(s_t, j, x) = \max_c [u(c + \nu_t) - C(i_{t+1}) + \beta E[V(s_{t+1}|d_{jx}, c, s_t]]$$

- The trick is that $M(s_t, j, x)$ is the same for all products of the same size x.
- This means dynamics drop out of the brand-choice equation conditional on x.
- Recover (α, ξ) from static demand estimation:

$$\begin{aligned} Pr(d_{jx}|x_t, p_t, i_t, \nu_t) &= \frac{\exp[\alpha^p p_{jxt} + \xi_{jx} + \alpha^a a_{jxt}]}{\sum_{k,y} \exp[\alpha^p p_{kyt} + \xi_{jy} + \alpha^a a_{kyt}]} \\ &= Pr(d_{jx}|x_t, p_t) \end{aligned}$$

Hendel and Nevo (2006)

Step 2: Inclusive Values

Assumption 4, IVS: $F(\omega_t | s_{t-1}) = F(\omega_t | \omega_{t-1})$.

$$\omega_{xt} = \mathbb{E}[\max_j u_j] = \log\left(\sum_k \exp(\alpha^p p_{kxt} + \xi_{xt} + \alpha^a a_{kxt})\right)$$

- Compute ex-ante expected utility of purchasing size x in period t
- Does not depend on which j is purchased.
- Same as G&R two price vectors with same inclusive values must have same transition probabilities. Do individual prices still matter? (Test)

Hendel and Nevo (2006)

Step 3: Dynamic Choice of Size

$$V(i, \omega_t, \epsilon_t, \nu_t) = \max_{c,x} [u(c + \nu_t) - C(i_{t+1}) + \omega_{xt} + \epsilon_{xt} + \\ \beta E[V(i_{t+1}, \omega_{t+1}, \epsilon_{t+1}, \nu_{t+1}) | i_t, \omega_t, \epsilon_t, \nu_t, c, x]]$$

- Compute ex-ante expected utility of purchasing size x in period t
- Does not depend on which j is purchased.
- **IVS** means we can keep track of a lot less information.
- Same as G&R two price vectors with same inclusive values must have same transition probabilities.
- Do individual prices still matter? (Test)

Hendel and Nevo (2006)

Unobserved initial conditions

Remaining problem: i_t is unobserved. **Solution**: Integrate out the unobserved initial inventory levels.

- Let $i_{i0} = 0$.
- Simulate the model for the first T_0 weeks of the data.
- Record inventory level at T_0 : i_{T_0} (i.e. initial state variable).
- Evaluate the likelihood from weeks T_0 to T, as if inventory levels were observed.
- Repeat the process S times, and average the likelihood contribution of individual i.
- **Implicit assumption**: Stationary process at time T_0.

Hendel and Nevo (2006)

Elasticities Compared to Static Model

AVERAGE RATIOS OF ELASTICITIES COMPUTED FROM A STATIC MODEL TO LONG-RUN ELASTICITIES COMPUTED FROM THE DYNAMIC MODEL[a]

		64 oz.						128 oz.					
Brand	Size (oz.)	All[b]	Wisk	Surf	Cheer	Tide	Private Label	All[b]	Wisk	Surf	Cheer	Tide	Private Label
All[b]	64	1.03	0.13	0.14	0.12	0.13	0.15	0.14	0.17	0.17	0.18	0.21	0.34
	128	0.17	0.24	0.26	0.20	0.28	0.35	1.23	0.09	0.11	0.09	0.15	0.22
Wisk	64	0.14	1.20	0.13	0.17	0.12	0.13	0.16	0.22	0.14	0.22	0.25	0.20
	128	0.25	0.27	0.23	0.31	0.26	0.28	0.08	1.42	0.08	0.13	0.18	0.11
Surf	64	0.14	0.13	0.93	0.16	0.13	0.14	0.18	0.18	0.12	0.18	0.22	0.28
	128	0.25	0.22	0.18	0.27	0.25	0.18	0.12	0.11	1.20	0.08	0.15	0.14
Cheer	64	0.12	0.17	0.16	0.84	0.09	0.13	0.14	0.24	0.16	0.14	0.22	0.24
	128	0.25	0.26	0.26	0.12	0.23	0.22	0.09	0.12	0.06	0.89	0.15	0.07
Tide	64	0.16	0.17	0.13	0.13	1.26	0.15	0.22	0.28	0.16	0.26	0.22	0.37
	128	0.25	0.31	0.22	0.24	0.22	0.31	0.11	0.16	0.08	0.13	1.44	0.31
Solo	64	0.15	0.12	0.15	0.14	0.12	0.14	0.17	0.15	0.15	0.30	0.30	0.28
	128	0.23	0.20	0.24	0.21	0.21	0.25	0.07	0.07	0.06	0.16	0.17	0.21
Era	64	0.21	0.12	0.13	0.13	0.10	0.19	0.43	0.17	0.15	0.22	0.19	0.35
	128	0.31	0.22	0.24	0.25	0.17	0.38	0.19	0.08	0.09	0.11	0.10	0.22
Private	64	0.19	0.15	0.14	0.17	0.17	1.02	0.32	0.22	0.15	0.26	0.31	0.25
label	128	0.29	0.28	0.34	0.30	0.39	0.29	0.16	0.12	0.13	0.10	0.27	1.29
No purchase		2.12	1.13	1.15	1.40	1.27	2.39	1.80	7.60	2.26	14.11	2.38	10.86

[a]Cell entries i and j, where i indexes row and j indexes column, give the ratio of the (short-run) elasticities computed from a static model divided by the long-run elasticities computed from the dynamic model. The elasticities for both models are the percent change in market share of brand i with a 1 percent change in the price of j. The static model is identical to the model estimated in the first step, except that brands of all sizes are included as well as a no-purchase decision, not just products of the same size as the chosen option. The results from the dynamic model are based on the results presented in Tables IV–VI.

Hendel and Nevo (2006)

Elasticities Compared to Static Model

AVERAGE RATIOS OF ELASTICITIES COMPUTED FROM A STATIC MODEL TO LONG-RUN ELASTICITIES COMPUTED FROM THE DYNAMIC MODEL[a]

Brand	Size (oz.)	64 oz. All[b]	64 oz. Wisk	64 oz. Surf	64 oz. Cheer	64 oz. Tide	64 oz. Private Label	128 oz. All[b]	128 oz. Wisk	128 oz. Surf	128 oz. Cheer	128 oz. Tide	128 oz. Private Label
All[b]	64	1.03	0.13	0.14	0.12	0.13	0.15	0.14	0.17	0.17	0.18	0.21	0.34
	128	0.17	0.24	0.26	0.20	0.28	0.35	1.23	0.09	0.11	0.09	0.15	0.22
Wisk	64	0.14	1.20	0.13	0.17	0.12	0.13	0.16	0.22	0.14	0.22	0.25	0.20
	128	0.25	0.27	0.23	0.31	0.26	0.28	0.08	1.42	0.08	0.13	0.18	0.11
Surf	64	0.14	0.13	0.93	0.16	0.13	0.14	0.18	0.18	0.12	0.18	0.22	0.28
	128	0.25	0.22	0.18	0.27	0.25	0.18	0.12	0.11	1.20	0.08	0.15	0.14
Cheer	64	0.12	0.17	0.16	0.84	0.09	0.13	0.14	0.24	0.16	0.14	0.22	0.24
	128	0.25	0.26	0.26	0.12	0.23	0.22	0.09	0.12	0.06	0.89	0.15	0.07
Tide	64	0.16	0.17	0.13	0.13	1.26	0.15	0.22	0.28	0.16	0.26	0.22	0.37
	128	0.25	0.31	0.22	0.24	0.22	0.31	0.11	0.16	0.08	0.13	1.44	0.31
Solo	64	0.15	0.12	0.15	0.14	0.12	0.14	0.17	0.15	0.15	0.30	0.30	0.28
	128	0.23	0.20	0.24	0.21	0.21	0.25	0.07	0.07	0.06	0.16	0.17	0.21
Era	64	0.21	0.12	0.13	0.13	0.10	0.19	0.43	0.17	0.15	0.22	0.19	0.35
	128	0.31	0.22	0.24	0.25	0.17	0.38	0.19	0.08	0.09	0.11	0.10	0.22
Private	64	0.19	0.15	0.14	0.17	0.17	1.02	0.32	0.22	0.15	0.26	0.31	0.25
label	128	0.29	0.28	0.34	0.30	0.39	0.29	0.16	0.12	0.13	0.10	0.27	1.29
No purchase		2.12	1.13	1.15	1.40	1.27	2.39	1.80	7.60	2.26	14.11	2.38	10.86

[a]Cell entries i and j, where i indexes row and j indexes column, give the ratio of the (short-run) elasticities computed from a static model divided by the long-run elasticities computed from the dynamic model. The elasticities for both models are the percent change in market share of brand i with a 1 percent change in the price of j. The static model is identical to the model estimated in the first step, except that brands of all sizes are included as well as a no-purchase decision, not just products of the same size as the chosen option. The results from the dynamic model are based on the results presented in Tables IV–VI.

Hendel and Nevo (2006)

Important empirical lessons

- Two differences between *static* and *dynamic* model:
  - Static model implies larger price coefficient,
  - And ignores the inventory problem.
- Both lead to a <u>larger own price elasticity</u> with the static model, than the long-run own price elasticity with the dynamic model.
- Point two leads to a <u>lower cross price elasticity</u> (compared to static model):
  - In the data the response to sales is mostly coming from people going from not buying to buying the brand on sale.
  - This leads to predict small cross price elasticities with respect to other products, and large cross price elasticity with respect to the outside good.
  - The dynamic model rationalize this phenomenon high unobserved inventories, and conditional choice probabilities (conditional on purchasing). Everybody needs detergent...