MKT927: INTRO TO QUANTITATIVE MARKETING

Prof. Andrey Fradkin

Lecture 4: Demand Estimation

HOUSEKEEPING

- Discuss Assignment 1
- Schedule Coming Up
- Any Questions?

WHY DEMAND ESTIMATION?

DEMAND ESTIMATION IN MARKETING

- Figure out how to price a product.
- Figure out what product attributes are important and how to design products.
- Figure out how to segment customers.
- Ingredient in modeling an industry/marketplace/platform.
- Welfare / regulation / policy.

OUTLINE

- Some facts about consumer purchases
- Discrete choice models
- Diversion ratios
- Estimation in practice
- Moshary et al. (2025)

OUTLINE

- **Some facts about consumer purchases**
- Discrete choice models
- Diversion ratios
- Estimation in practice
- Moshary et al. (2025)

SOME FACTS ABOUT CONSUMER PURCHASES

- CPG industry is a valuable laboratory for studying consumer behavior
  - Widely available across store formats
  - High purchase frequency (1.7 grocery trips/week in 2017)
  - Significant portion of household budgets
  - Politically important → Salient for inflation.
- Market size
  - Global CPG sector: $8 trillion (2014) → $14 trillion (2025)
  - US households spent $407 billion on CPGs (2016)

NIELSEN HOMESCAN PANEL DATA (2012)

- Nationally representative panel
- Coverage (2012):
    - 52,093 households
    - 6.57 million shopping trips
    - 46 million products purchased

PRODUCT VARIETY IN CPG CATEGORIES

- Average category offers:
  - 402.8 unique products (UPCs)
  - 64.4 unique brands
  - 31.9 different pack sizes
- Brand packaging:
  - Average brand offers 5.4 different pack sizes
  - Products sold in pre-packaged, indivisible units
  - Suggests extensive use of non-linear pricing

CONSUMER PURCHASE PATTERNS

- Average category: 39,787 trips with purchases
- Per category-trip:
  - 94.3% purchase single brand
  - 67.3% purchase single pack size
  - Average of 1.07 brands purchased
- Implications:
  - Discrete brand choice assumption largely holds
  - Discrete quantity choice assumption moderately holds

Outline

- Some facts about consumer purchases
- **Discrete choice models**
- Diversion ratios
- Estimation in practice
- Moshary et al. (2025)

ESTIMATING DEMAND AND SUPPLY (HISTORY)

- Phillip. G. Wright (1928) - “The Tariff on Animal and Vegetable Oils”
- First use of instrumental variables.

SUPPLY AND DEMAND: LINEAR EXAMPLE (TAKEN FROM CONLON NOTES)

$$\begin{aligned}
\text{Demand:} \quad & Q_t^d = \alpha_0 + \alpha_1 P_t + U_t \\
\text{Supply:} \quad & Q_t^s = \beta_0 + \beta_1 P_t + V_t \\
\text{Market Clearing:} \quad & Q_t^d = Q_t^s
\end{aligned}$$

Solving for equilibrium price and quantity:

$$\begin{aligned}
P_t &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{V_t - U_t}{\alpha_1 - \beta_1} \\
Q_t &= \frac{\alpha_1 \beta_0 - \beta_1 \alpha_0}{\alpha_1 - \beta_1} + \frac{\alpha_1 V_t - \beta_1 U_t}{\alpha_1 - \beta_1}
\end{aligned}$$

UNDERSTANDING OLS BIAS IN SUPPLY AND DEMAND ESTIMATION

Covariance Relationships:

$$\mathrm{Cov}(P_t, Q_t) = \alpha_1 \mathrm{Var}(P_t) + \mathrm{Cov}(P_t, U_t)$$

$$\mathrm{Cov}(P_t, Q_t) = \beta_1 \mathrm{Var}(P_t) + \mathrm{Cov}(P_t, V_t)$$

Bias in OLS Estimates:

$$\mathrm{Bias}(\hat{\alpha}_1) = \frac{\mathrm{Cov}(P_t, U_t)}{\mathrm{Var}(P_t)} \quad \text{(Demand)}$$

$$\mathrm{Bias}(\hat{\beta}_1) = \frac{\mathrm{Cov}(P_t, V_t)}{\mathrm{Var}(P_t)} \quad \text{(Supply)}$$

When $\text{Cov}(U_t, V_t) = 0$:

$$\text{OLS Estimate} = \frac{\alpha_1 \text{Var}(V_t) + \beta_1 \text{Var}(U_t)}{\text{Var}(V_t) + \text{Var}(U_t)}$$

Key Insights:

- More variation in supply (V_t) → better demand estimates
- More variation in demand (U_t) → better supply estimates

INSTRUMENTAL VARIABLES AS A SOLUTION

- Find a variable that affects price but is not correlated with the demand error term.
- Use that variable to instrument for price in the demand equation.
- Get consistent estimates of the demand parameters using two-stage least squares (2SLS).
- Rossi (2014) - even the rich can make themselves poor. Argues that we can sometimes observe enough demand covariates so that we don't need to instrument, and at the same time that most instruments are weak.

WHY NOT USE LINEAR DEMAND?

- Linear demand model is not structural. Note, no “deep parameters” about the utility function.
- 2SLS identifies a local average treatment effect (LATE). Which instrument we use matters.
- Linear demand model is not consistent with discrete choice models (Jaffe and Weyl (2010)).
- Multiple products?

CONTINUOUS VS DISCRETE CHOICE MODELS

- Continuous choices models are commonly used to model total consumption in macroeconomics and consumer behavior. Examples:
  - Energy
  - Total food demand
  - Agricultural products
- Discrete choice models are commonly used to model product-level choices (most of this class).
- Level of aggregation is important, model each store trip or consumption over the course of a year?

DEMAND MODELING BASICS

- Utility function and a budget constraint.
- Constrained optimization problem allows one to derive the indirect utility function.
- The indirect utility function is a function of prices.
- Leads to the most common demand model in empirical marketing (abstracting from income effects):

$$u(x, p, z) = \beta' x - \alpha p + \varepsilon \tag{1}$$

- β is a vector of parameters, x is a vector of product attributes, p is a vector of prices, ε is an error term.

THE STANDARD LOGIT DISCRETE CHOICE MODEL

- We have a set of products and an outside option.
- Utility is a function of product characteristics and prices. The utility of the outside option is normalized to zero. (Note, a normalization is always necessary).

$$u(x, p, z) = \beta' x - \alpha p + \varepsilon \tag{2}$$

- ε is an i.i.d. error term distributed according to a type I extreme value distribution.
- The probability of choosing product j is given by:

$$P(y = j) = \frac{e^{\beta' x_j - \alpha p_j}}{1 + \sum_{k=1}^{J} e^{\beta' x_k - \alpha p_k}} \tag{3}$$

THE STANDARD LOGIT DISCRETE CHOICE MODEL

- Why is it so commonly used?
- Closed form solution for the probabilities.
- Easy to extend to random coefficients and random coefficients with fixed effects.
- Other models assume the error term is normal, leading to multinomial probit models.

OTHER POSSIBILITIES

- Notice that we are modeling the demand as a function of product characteristics and an error term.
- Why not model demand for products directly (e.g., demand for snickers vs demand for candy bar with X grams of sugar)?
  - Many more parameters to estimate. Creates identification and statistical power issues.
- Why do we need an error term? Otherwise, hard to rationalize the choice probabilities in the data. Some work on the “Pure Characteristics” model, but it is not used much.

THE IIA PROPERTY

- The standard logit model imposes the independence of irrelevant alternatives (IIA) property.
- The IIA property is that the ratio of the choice probabilities is constant across choice sets.

$$\frac{P(y = j \mid x, p, z)}{P(y = k \mid x, p, z)} = \frac{P(y = j \mid x, p_{-j}, z)}{P(y = k \mid x, p_{-j}, z)} \tag{4}$$

- Highly unrealistic. For example, those purchasing a product that is on sale are more likely to purchase other products on sale.

THE SEARCH FOR FLEXIBLE MODELS

- When we write down a demand model, we need to make assumptions about which parameters are constant across consumers and over time, and which are not.
- All else equal, prefer flexible models.
- But the more “flexible” the model, the harder it is to estimate precise parameters.
- Will come back to this.

Outline

- Some facts about consumer purchases
- Discrete choice models
- **Diversion ratios**
- Estimation in practice
- Moshary et al. (2025)

WHAT CAN WE ANSWER WITH AN EXPERIMENTAL APPROACH? (CONLON AND MORTIMER (2021))

- Consider the following randomized controlled trials:
  - Changing one price in a product category for a random set of users.
  - Randomly removing a product from the assortment for a set of users.
  - Randomly adding a product to the assortment for a set of users.
  - Randomly changing all prices in a product category for a set of users.
- What can we learn from these experiments? A lot, but not equilibrium responses.

DIVERSION RATIOS

- As the price of product j changes, we can observe how the quantity demanded of product j decreases and how the quantity demanded of other products increases.
- The diversion ratio D_{jk} is the ratio of switchers to stayers.
- If two products are close substitutes, we would expect D_{jk} to be high.
- This ratio is used frequently in horizontal mergers.

CONLON AND MORTIMER (2021)

- Study the properties of diversion ratios, and show several key insights.
- Constant diversion ratios (as implied by the logit model) are very implausible if people are heterogeneous.
- When people are heterogeneous, the diversion ratio is different depending on the size of the price change (or if the product is removed (price = ∞)). Similar for other product characteristics.
- Substitution to the outside option is also not constant but is important to consider.

UNDERSTANDING DIVERSION RATIOS

- Start with Wald estimator for price change $p_j \to p'_j$:

$$\text{Wald}(p_j, p'_j, x) = \frac{q_k(p'_j, x) - q_k(p_j, x)}{-(q_j(p'_j, x) - q_j(p_j, x))}$$

- Diversion ratio is the limit as price change becomes small:

$$D_{jk}(p_j, x) = \lim_{p'_j \to p_j} \text{Wald}(p_j, p'_j, x) = \frac{\partial q_k / \partial P_j}{-\partial q_j / \partial P_j}$$

Key Points:

- Measures how demand for product k changes relative to product j
- Related to Local Average Treatment Effects (LATE)
- Requires downward-sloping demand ($\partial q_j / \partial P_j < 0$)

CONLON AND MORTIMER (2021)

$$d_{ij}(p_j, x) = \begin{cases} 1 & \text{if } u_{ij}(p_j, x) > u_{ij'}(p_j, x) \text{ for all } j' \in J \text{ and } j' \neq j \\ 0 & \text{otherwise} \end{cases}$$

TABLE 1 Description of Compliance Types and Treatment-Effects Parameters

Compliance Type	$(d_{ij}(p_j, x), d_{ij}(p'_j, x))$	Description
Always Takers	(0,0)	Do not buy j at either price.
Never Takers	(1,1)	Buy j at either price
Compliers	(1,0)	Only buy j at lower price $p_j < p'_j$
Defiers	(0,1)	Only buy j at higher prices $p'_j > p_j$

The Wald estimator identifies the average diversion ratio among compliers.

TREATMENT EFFECT HETEROGENEITY IN DIVERSION RATIOS

Key Insight: Different consumers exhibit different diversion ratios

Aggregate diversion ratio is a weighted average across consumers:

$$D_{jk}(p_j, x) = \int_i w_i(p_j, x) D_{jk,i}(p_j, x)\, di$$

where consumer-specific weights are:

$$w_i(p_j, p_j', x) = \frac{q_{ij}(p_j', x) - q_{ij}(p_j, x)}{q_j(p_j', x) - q_j(p_j, x)}$$

Interpretation:

- $D_{jk,i}$ represents individual i's diversion ratio
- Weights w_i reflect relative contribution to aggregate demand change

OUTLINE

- Some facts about consumer purchases
- Discrete choice models
- Diversion ratios
- **Estimation in practice**
- Moshary et al. (2025)

MIXED LOGIT

$$U_{ij} = \beta_i x_j + \varepsilon_{ij}, \quad \beta_i \sim f(\beta_i|\theta), \quad \varepsilon_{ij} \sim \text{Type I EV}$$

We could have specified an error components structure on ε_i.

$$U_{ij} = \beta x_{ij} + \underbrace{\nu_i z_{ij} + \varepsilon_{ij}}_{\tilde{\varepsilon}_{ij}}$$

- x_{ij} are observed and ε_{ij} is IID Type I EV.
- The key is that ν_i is unobserved and mean zero (often normally distributed). Each draw represents a consumer with different preferences.

KINDS OF HETEROGENEITY

$$s_{ij}(\theta) = \int \frac{\exp[x_j \beta_i]}{1 + \sum_k \exp[x_k \beta_i]} f(\boldsymbol{\beta_i}|\theta) \approx \sum_{s=1}^{S} w_i^s \frac{\exp[x_j \beta_i^s]}{1 + \sum_k \exp[x_k \beta_i^s]}$$

- We can only **approximate** the integral with **weights** w_i^s and **nodes** β_i^s
  - We can allow for there to be two types of β_i in the population (high-type, low-type). **latent class model**.
  - We can allow β_i to follow an independent normal distribution for each component of x_{ij} such as $\beta_i = \overline{\beta} + \nu_i \sigma$.
  - We can allow for correlated normal draws $\beta_i \sim N(\mu_\beta, \Sigma_\beta)$.
  - Can allow for non-normal distributions too (lognormal, exponential).
  - If multiple choices per person, can allow for even more heterogeneity.

ESTIMATION DETAILS: MAXIMUM SIMULATED LIKELIHOOD

$$\theta_{MLE} = \arg\min_{\theta} - \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} \ln \widehat{s}_{ij}(\theta)$$

- We need to perform numerical integration to get $\widehat{s}_{ij}(\theta)$ and its derivative: $\frac{\widehat{\partial s_{ij}}}{\partial \theta}$.
- Accurate integration is important, can see how estimates are sensitive to the number of integration points.
- Automatic differentiation tools available in PyTorch or other modern machine learning libraries are very useful.

LET'S READ AN ESTIMATES TABLE FROM A DEMAND ESTIMATION (MUSOLF AND LEE (2024)).

	(1)	(2)	(3)	(4)
Price / MSRP	-1.37 (0.152)	-1.42 (0.161)	-0.61 (0.161)	-0.68 (0.136)
Dispatch Time (in Days)		-0.54 (0.096)		-0.05 (0.013)
100×Log(#Feedback+1)		2.53 (2.068)		-0.79 (0.187)
Fulfilled by Amazon?		0.24 (0.177)		0.05 (0.016)
Sold by Amazon?		4.75 (0.267)		0.16 (0.037)
Constant (Inside Options)	-3.49 (0.155)	-3.60 (0.321)	-2.60 (0.126)	-2.53 (0.140)
Nesting Coefficient (λ)			0.06 (0.020)	0.06 (0.012)
Sophisticates Fraction (ρ)				
Offer FE?	✗	✗	✗	✗
Elasticity (Rec. Fixed)	-1.44	-1.50	-8.21	-10.07

In marketing:

- We are often interested in estimating demand for a given set of individuals and then making a business decision: pricing, advertising, etc.
- We often have a panel of data of individuals.

But, suppose we want to study equilibrium, or we only have aggregate data.

- We need to use models and estimation that match the aggregate data.
- Berry 1994 (Nested Logit), Berry, Levinsohn, and Pakes (1995) - BLP model.

AGGREGATE DEMAND ESTIMATION - BLP

Add unobservable error for each $\mathfrak{s}_{jt}$ labeled ξ_{jt}.

$$u_{ijt} = \underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}} + \varepsilon_{ijt}, \quad \sigma_j(\boldsymbol{\delta_t}) = \frac{e^{\delta_{jt}}}{1 + \sum_k e^{\delta_{kt}}}$$

- ξ_{jt} is observed to everyone expect the econometricians. Works to match the aggregate data.
- Potentially correlated with price $\text{Corr}(\xi_{jt}, p_{jt}) \neq 0$
- But not characteristics $\mathbb{E}[\xi_{jt} \mid x_{jt}] = 0$.
  - This allows for products j to better than some other product in a way that is not fully explained by differences in x_j and x_k.
  - Consumers agree on its value (**vertical component**).

USING FLEXIBLE MODELS MATTERS (CONLON AND MORTIMER (2021))

BLP	Best Practices
Price/inc	−51.254
	(5.847)
σ_{cons}	2.052
	(1.111)
$\sigma_{\text{HP/weight}}$	1.785
	(2.061)
σ_{air}	1.899
	(0.439)
$\sigma_{\text{MP\$}}$	0.708
	(0.184)
σ_{size}	1.126
	(0.917)

BLP	Best Practices	Nested Logit
Median Own-Elasticity	−3.811	−1.600
Median Aggregate Elasticity	−0.096	−0.033
Median Outside-Good Diversion	0.201	0.197
Mean Top 5 Diversion	0.182	0.165
Mean Markup	0.334	0.936
Median Consumer Surplus	2.071	2.827

ESTIMATION OF BLP MODELS

- It's a pain in the butt numerically and there are lots of numerical tricks to get it to work.
- See Conlon and Gortmaker (2020) and the package 'PyBLP' in Python.
- Modern take: the best approach is to combine aggregate data with micro data, survey data, or another set of moments (for example include a supply model). See Grieco et al. (2024) and references therein.
- Two types of instruments are required.
  - Instruments for price: cost shocks or other supply side shocks.
  - Instruments for parameters related to unobserved heterogeneity: characteristics of other products in the choice set and so on (see Ghandi and Houde (2019), for example).

OTHER DEMAND MODELING INNOVATIONS / APPROACHES

- Fox et al. (2011) - semi-parametric models. Have seen reasonable amount of adoption in practice and they are quite simple to implement. Not great at dealing with endogeneity or fixed effects.
- Modern machine learning approaches: Donnelly et al. (2021), Ferrel, Liang, and Misra (2021).
- Bayesian approaches: The work of Allenby and Rossi.

HOW TO READ DEMAND ESTIMATION PAPERS

- Look at the utility function, the distributional assumptions, etc...
- What is the unit of observation. How is the data aggregated? Does it represent a subset of individuals, a market? Is an important aspect of demand missing?
- Where does the identification come from? Price and heterogeneity variation.

OUTLINE

- Some facts about consumer purchases
- Discrete choice models
- Diversion ratios
- Estimation in practice
- **Moshary et al. (2025)**

CONJOINT ANALYSIS

- Conjoint analysis uses survey responses to choice tasks in order to elicit preferences.
- Some advances: hierarchical Bayes, optimal question design, etc, software packages.

DISCRETE CHOICE CONJOINT ANALYSIS EXAMPLE

Which laptop would you choose?

Features	**Option 1**	**Option 2**
Brand	Dell Latitude	MacBook Pro
Price	$1,100	$2,100
Screen Size	14"	15"

Option 3: I would not purchase either laptop

ALLENBY, HARDT, AND ROSSI

- **Essential Tool:**
  "Conjoint analysis is an indispensable tool for predicting the effects of changes to marketplace offerings when observational data does not exist to inform an analysis" (P. 188)
- **Validity Requirements:**
  - A valid conjoint analysis requires a valid model and constructs for inference.
- **Method Comparisons:**
  - Choice-based conjoint is argued to be closer to revealed rather than stated preferences.
- **Demand Estimation:** Emphasis on techniques that incorporate individual heterogeneity, particularly using Bayesian econometrics.

EFFICIENT QUESTION DESIGN (TOUBIA ET AL 2003)

- "Adapting question design within a respondent, using that respondent's answers to previous questions, is a difficult dynamic optimization problem."
- Example
  - Three features of a camera: picture quality (# pixels), style (permanent or changeable), steps in picture taking (one or two)
  - Want to estimate u_1, u_2, u_3
  - Comparisons only have a relative meaning → choices generate inequalities
  - Want questions to shrink the feasible set of choices quickly.
- Relates to classic experimental design issues as in number of treatments required and how to run them at scale.

INCENTIVE COMPATIBILITY IN CONJOINT ANALYSIS (DING 2007)

- Key objective: Design mechanism where agents truthfully report preferences
  - Message space: What players can communicate
  - Outcome function: Maps messages to results
- Recasts conjoint as incomplete information game
- **Important Note:** Recent work by Allenby et al. challenges this view
  - Lab incentives may not improve marketplace predictions
  - Better approach: Screen for active category purchasers

THE BDM MECHANISM (BECKER, DEGROOT, MARSCHAK 1964)

- Participant states willingness to pay (WTP)
- Price drawn randomly from uniform distribution
- Outcome determined by comparison:
  - If drawn price > WTP: No purchase possible
  - If drawn price ≤ WTP: Purchase at drawn price

MOSHARY ET AL. (2025): MOTIVATION & RESEARCH QUESTION

- **Key Policy Challenge:** Crafting effective firearms regulation requires understanding consumer demand
- **Data Limitation:** No centralized database of gun purchases matched with prices
- **Research Questions:**
  - How price sensitive are different types of firearm buyers?
  - What are the substitution patterns across gun types?
  - How would different policies affect gun ownership?

METHODOLOGY

- **Novel Approach:** Stated-choice conjoint analysis
  - Survey instrument typically used in marketing research
  - Allows estimation of price sensitivity and substitution patterns
  - Experimentally manipulate prices and choice sets
- **Data:**
  - 22,522 US adults surveyed (March-April 2022)
  - 4,018 respondents interested in firearms
  - Each respondent completes 7 choice tasks

CONJOINT TASK

(a) Step 1

Which of the following firearms do you **most** prefer?

Glock G44

Pistol
Click here for product information
$477.00

Remington 870 Express

Shotgun
Click here for product information
$652.00

Colt 1911

Pistol
Click here for product information
$715.00

« Back | Continue »

DEMAND MODEL

- **Random Utility Model with Limited Consideration:**

$$u_{ijt} = X_j' \beta_i - \alpha_i \cdot p_{ijt} + \epsilon_{ijt}$$

$$u_{i0t} = 0$$

- **Key Features:**
  - Individual-level preferences
  - Explicit use of stated consideration sets
  - Observable and unobservable heterogeneity
  - Price coefficient constrained to be negative

DEMAND ESTIMATES

	Posterior Mean	SD	2.5%	97.5%
Price	-0.013	0.022	-0.077	-0.000
Revolver	0.824	1.211	-1.610	3.109
Pistol	1.888	1.205	-0.663	4.047
Rifle	0.887	1.038	-1.267	2.727
Shotgun	0.342	0.982	-1.718	2.157
Assault Weapon	1.151	1.222	-0.956	3.478

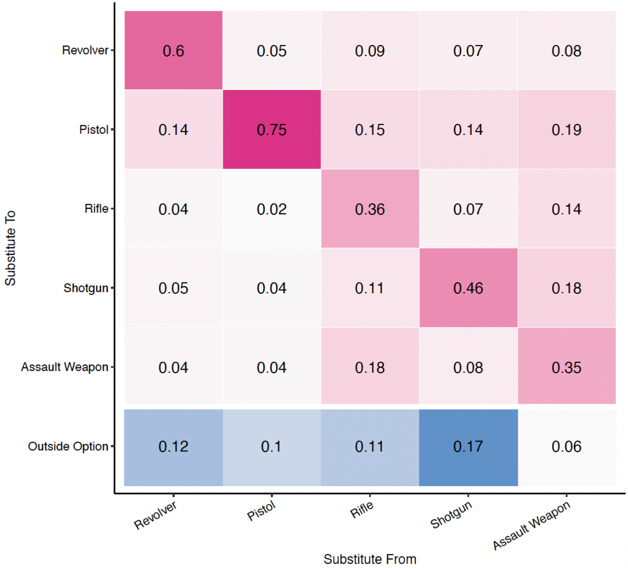
	Own-Price Elasticity	Market Share
Revolver	-1.134	15.134
	[-1.282, -0.992]	[14.586, 15.724]
Pistol	-1.046	37.336
	[-1.185, -0.940]	[36.533, 38.130]
Rifle	-0.871	7.482
	[-1.180, -0.693]	[6.827, 8.179]

KEY FINDINGS (1): PRICE SENSITIVITY

- **Overall Low Price Elasticity:**
  - Consumers generally price insensitive
  - Implies high valuations for firearms
- **Heterogeneous Effects:**
  - First-time buyers more price sensitive
  - Strong preference for handguns among new buyers

DIVERSION RATIOS

Figure 3. : Diversion Ratios for Firearms



KEY FINDINGS (2): SUBSTITUTION PATTERNS

- **Significant Cross-Category Substitution:**
  - Considerable substitution from assault weapons to handguns
  - Limited substitution in reverse direction
- **Policy Implications:**
  - Assault weapon restrictions may increase handgun sales
  - Handguns associated with more crimes per gun
  - Handgun regulations more likely to reduce overall ownership

EXTERNAL VALIDATION

- **Model Predictions Match:**
  - Aggregate background check volumes
  - Handgun vs. long gun market shares
  - Observed stock-out patterns
- **Limitations:**
  - Hypothetical nature of choices
  - Selection into survey sample
  - Cannot capture illegal market dynamics

CONTRIBUTIONS

- Novel application of marketing tools to policy question
- Assault weapon bans may have limited impact on overall ownership
- Price-based regulation most effective for new buyers
- Substitution effects critical for policy design

FOLLOW-ON WORK

- Rosenberg (2025) on the economics of gun control.
- Found administrative data on gun purchases in California.
- Role of retailers of guns on gun purchases and suicides.
- Effects of taxes.