Nonparametric Estimates of Demand in the California Health Insurance Exchange(ish)

Pietro Tebaldi, Alexander Torgovitsky, Hanbin Yang

Nadav Kunievsky
UChicago

What I'm Going to Talk About

- Mainly intuition (no proofs)
- Very few words on healthcare insurance and even less on California
- Get you to where I'm at in understanding this paper (no proofs!)
- It would be great if this presentation could be more of a discussion
- Demonstration (?)

Motivation

- Current (fake) MSM approaches for demand function identification and estimation are based heavily on distributional assumptions
  - Logit
  - BLP framework
- The basic parametric set for discrete choice models is

$$Y_{im} = arg \max_{j \in \mathcal{J}} X'_{i,j,m}\beta_{i,m} - \alpha_{i,m}P_{i,j,mj,m} + \epsilon_{i,j,m}$$

where $\epsilon \sim T1EV$ and there is some heterogeneity in the coefficient, which are drawn from some distribution

Motivation

- These strong distributional assumptions may not be innocuous.
- It's hard to separate how much of our inference comes from these assumptions.
- Currently, there aren't a lot of alternative methods of estimating discrete choice problems (can't do regression, thinking on causality is not clear in settings where there are multiple outcomes and non-scalar treatments, few papers, e.g., Compiani, 2019).

Motivation

- This paper constructs a framework for dealing with discrete choice problems while imposing a minimum amount of assumptions
- Unfortunately, we won't see any magic here

Basic Set-Up

- Each consumer chooses one option Y_i from a set of J+1 options $\mathcal{J} = \{0, 1, 2..., J\}$
- Each option has some price $P_{i,j}$, where price can be different between individuals
- Consumer i has a vector of unobserved $V_i = (V_{i,0}, V_{i,1}, ..., V_{i,J})$ valuation for each plan, with the standard normalization that $V_{i,0} = 0$
- The indirect utility is additively separable in prices and latent valuation $V_{i,j} - P_{i,j}$, and the consumer decision rule is given by

$$Y_i = arg \max_{j \in \mathcal{J}} V_{i,j} - P_{i,j}$$

Basic Set-Up - A Few Notes

- The additive separability assumption places limitations on substitution patterns, such as the absence of an income effect on the intensive margins However, we may enable $V_{i,j}$ to differ across income groups (provided that price changes do not cause consumers to shift between income categories or alter preferences, even if their income has changed).
- There are no restrictions on the relationship between the covariates and the utility, as we have in the BLP framework. Conditional on the x's, we allow for the distribution of $V_{i,j}$ to change freely
- This model permits the $V_{i,j}$ and the prices to be correlated, although to identify the effect of a price shift, we need to assume exogenous movement, conditional on observables

Counterfactuals and Parameters of interest

- The model primitive is the density $f(V_i|m,x)$
- We can construct counterfactual objectives by integrating over this distribution
- For instance, suppose we wish to determine the proportion of consumers who would purchase product j under a certain pricing scheme

$$\int \mathbb{I}\left[v_j - p_j^\star \geq v_k - p_k^\star \text{ for all } k\right] f(v|m,x)dv$$

- We can also investigate causal questions regarding changes in surplus resulting from change in prices

$$\underbrace{\int \left\{\max_{j\in\mathcal{J}} v_j - p_j^\star\right\} f(v|m,x)dv}_{\text{consumer surplus under } p^\star} - \underbrace{\int \left\{\max_{j\in\mathcal{J}} v_j - p_j\right\} f(v|m,x)dv}_{\text{consumer surplus under } p}$$

- We think of these as the target parameters, which are $\theta : \mathcal{F} \to \mathbb{R}^{d_\theta}$

Additional Assumptions

- Let $P_{i,j} = P(x_i, m_i, \epsilon_i)$
- **The Instrument Assumption**. W_i and Z_i are two subvectros of X and M, such that Z_i satisfy the exogeneity assumption

$$f_{V|WZ}(v|w,z) = f_{V|WZ}\left(v|w,z'\right) \quad \text{for all } z, z', w, \text{ and } v$$

- The distribution of valuations is invariant to shifts in Z_i , conditional on W_i. That is, Z_i is exogenous

Additional Assumptions

- **Support**. f is concentrated in a known set, such that

$$\int_{\mathcal{V}\bullet(w)} f_{V|WZ}(v|w,z)dv = 1 \quad \text{for all } w, z$$

  where $\mathcal{V} \bullet (w)$ is the support of f
- This can be satisfied by letting $\mathcal{V} \bullet (w) = R^J$

Some Notes

- This assumption is similar to the Heckman selection model, where we look for variation in the cost. Specifically, we can think of a regular Heckman setup:

$$Y_1 = \mu_1(x) + u_1$$
$$Y_0 = \mu_0(x) + u_0$$
$$D = \mathbb{I}\{Y_1 - \mu_{c,1}(z) > Y_0 - \mu_{c,0}(z)\}$$

  Here, we are focusing on the "first stage" to estimate the model.
- In contrast to the regular BLP setup, where we consider variation in prices across markets, here we are interested in variation within homogeneous groups. Therefore, we may want to condition on the market.
- The more prices vary with Z_i, the more information we will have to pin down different parts of the density of valuations, f, and therefore the target parameter, θ.

The identified Sets

- We are interested in the set of possible values that the target parameter $\theta(f)$ can take, given the observed data
- Let $s(m,x) = \mathrm{P}[Y_i = j | m_i, x_i]$
- As consumers choose the option that maximizes their surplus, we have that the shares are

$$s_j(m,x|f) = \int_{\mathcal{V}_j(p)} f(v|m,x)dv$$

where $\mathcal{V}_j(p) = \{(v_1, ...v_J) \in \mathbb{R}^j | v_j - p_j > v_k - p_k \text{ for all k}\}$

The identified Sets

- The identified set of valuation densities is the set of all f that matches the observed data

$$\mathcal{F}^* = \{f | \text{satisfies the assumptions and the share conditions}\}$$

- The identified set for θ is given by

$$\Theta^* = \{\theta(f) : f \in \mathcal{F}^*\}$$

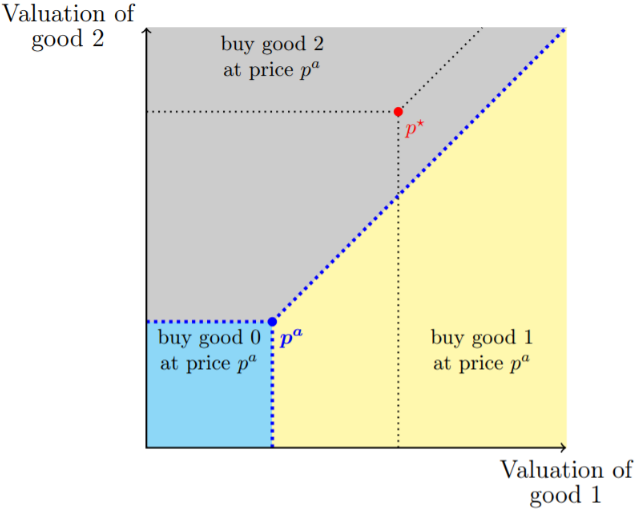
Minimal Relevant Partition

- Our goal is to compute Θ exactly using the observed data and the Minimal Relevant Partition (MRP)
- The MRP partitions the space of valuations such that any two consumers with valuations in the same set will exhibit the same choice behavior under different prices p^a and p

Definition MRP. *Let* $Y(v,p) \equiv \arg\max_{j \in \mathcal{J}} v_j - p_j$ *for any* $(v_1, \ldots, v_J), (p_1, \ldots, p_J) \in \mathbb{R}^J$, *where* $v \equiv (v_0, v_1, \ldots, v_J)$ *and* $p \equiv (p_0, p_1, \ldots, p_J)$ *with* $v_0 = p_0 = 0$. *The* minimal relevant partition of valuations *(MRP) is a collection* $\mathbb{V}$ *of sets* $\mathcal{V} \subseteq \mathbb{R}^J$ *for which the following property holds for almost every* $v, v' \in \mathbb{R}^J$ *(with respect to Lebesgue measure):*

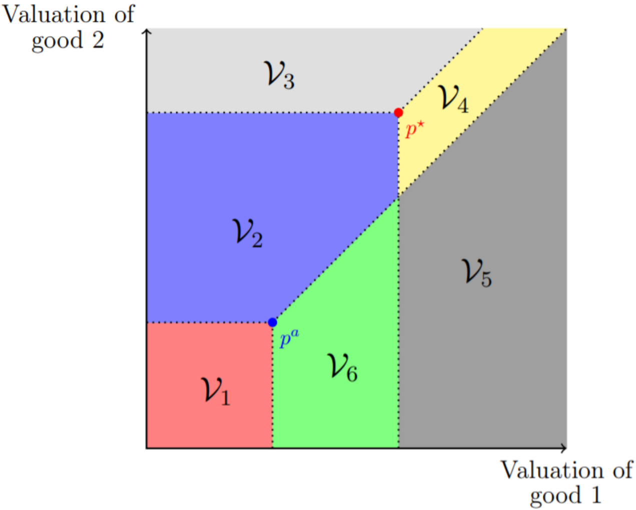
$$v, v' \in \mathcal{V} \text{ for some } \mathcal{V} \in \mathbb{V} \quad \Leftrightarrow \quad Y(v,p) = Y(v',p) \text{ for all } p \in \mathcal{P}. \tag{17}$$

MRP



***(a)** Choices if prices were p^a.*

MRP



(c) *The minimal relevant partition (MRP) constructed from p^a and $p^\star$.*

MRP - A Binary Example

- How can we use MRP to solve for the identified set?

MRP - a J=2 example

- Consider the figure from before, and assume that the choice probabilites, under price 1 are as follows

$$s_0\left(m, x^a\right)=.20, \quad s_1\left(m, x^a\right)=.14, \quad \text{and} \quad s_2\left(m, x^a\right)=.66$$

- Therefore, we know that

$$\begin{aligned}
&\int_{\mathcal{V}_1} f(v|m)dv = s_0\left(m, x^a\right) = .20 \\
&\int_{\mathcal{V}_5} f(v|m)dv + \int_{\mathcal{V}_6} f(v|m)dv = s_1\left(m, x^a\right) = .14 \\
&\int_{\mathcal{V}_2} f(v|m)dv + \int_{\mathcal{V}_3} f(v|m)dv + \int_{\mathcal{V}_4} f(v|m)dv = s_2\left(m, x^a\right) = .66
\end{aligned}$$

Solving the thing

- Assume we want to find the share of consumers who would buy product 2, under the new price.
- Let $\int_{\mathcal{V}_i} = \phi_i$ and notice that we can write this problem as a linear programming problem for the upper bound

$$
\begin{aligned}
&t^{\star}_{\mathrm{ub}} \equiv \max_{\phi \in \mathbb{R}^6} \phi_3 \\
&\text{subject to:} \quad \phi_1 = .20 \\
&\phi_5 + \phi_6 = .14 \\
&\phi_2 + \phi_3 + \phi_4 = .66 \\
&\quad \phi_l \geq 0 \quad \text{for } l = 1, \ldots, 6
\end{aligned}
$$

- $t_{ub} = 0.66$ and $t_{lb} = 0$.
- and similarly for the lower bound

Final Remarks

- In practice, when dealing with multiple markets and covariates, but interested in the aggregate, one can simply average the bounds over markets and covariates:

$$\theta_a = \sum \mathrm{P}(X = x, M = m|f)\Delta Share_j(m, x|f)$$

- To use this method in real-world settings, one needs:
  - Some exogenous variation in costs
  - A discrete choice setting (otherwise, just use instruments)
  - A strong desire to avoid distributional assumptions
- Note that getting strict bounds can be challenging, so it's important to have enough variation in costs
- Noise
- Imposing restrictions