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

Pietro Tebaldi, Alexander Torgovitsky, Hanbin Yang

Nadav Kunievsky
UChicago

April 30, 2020

What I'm Going to Talk About

- Mainly intuition (No Proofs)
- Very few words on health care insurance and even less on California
- Get you to where I'm at in understanding this paper (No Proofs!)
- It would 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 discreate 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 heavy 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 Compiani, 2019)

Motivation

- In this paper they construct a framework of dealing with discrete choices problem while imposing minimum amount of assumption
- (Unfortunately we won't see any magic here)

Basic Set-Up

- Each consumer chooses one option Y_i from 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 separablity assumption implies some restriction on substitution patterns. (No income effect on the intensive margins), but if we allow $V_{i,j}$ to change condition on income, then we can generate these income effects
- There is no restriction on the relation between the covariates and the utility, as we have in the BLP framework. Conditional on the $x's$ we allow for distribution of $V_{i,j}$ to change freely.
- This model allows for the $V_{i,j}$ and the prices to be correlated, although for identification of effect in price shift, we are going to need to assume exogenous movement, conditional on observables.

Counterfactuals and Parameters of interest

- The model primitive is $f(V_i|m,x)$
- We can construct counterfactual objectives by integrating over this distribution
- For example, consider we want to know the share of consumers who would buy product j, given some price 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 ask causal questions on the changes in surplus, given some price change

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

- 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

- The assumption is very "Heckmanish", where we look for variation in the cost. We can think of a regular Heckman setup
$$Y_1 = \mu_1(x) + u_1$$
$$Y_0 = \mu_0(x) + u_0$$
$$D = \mathbb{I}\{Y_1 - \mu_{c,1}(z) > Y_0 - \mu_{c,0}(z)\}$$
where now we are focusing on the "first stage"
- Where in the regular BLP setup we have considered variation in prices across markets, here we are thinking on variation within homogeneous groups. Therefore, we would probably like to condition on the market.
- The more that 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 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 matched the observed data

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

- The identified set for θ is given by

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

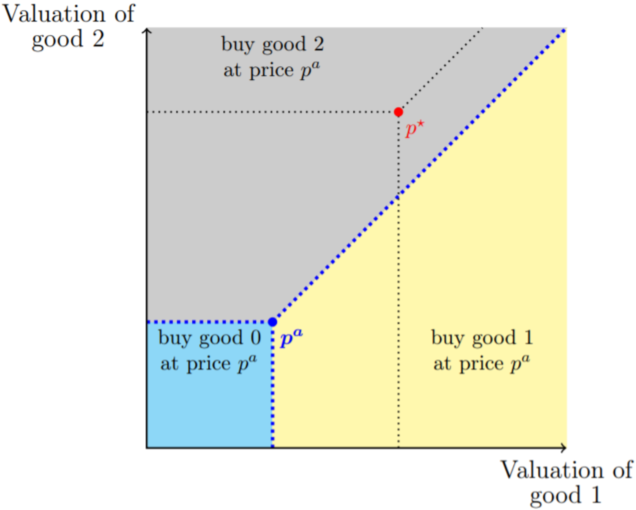
MRP

- We want to compute Θ^* exactly, using the observed data and the Minimal Relevant Partition.
- This partition over the space of valuation is such that any two consumers who have valuation in the same set would exhibit the same choice behaviour 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* $(v_1, \dots, v_J), (p_1, \dots, p_J) \in \mathbb{R}^J$, *where* $v \equiv (v_0, v_1, \dots, v_J)$ *and* $p \equiv (p_0, p_1, \dots, p_J)$ *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):*

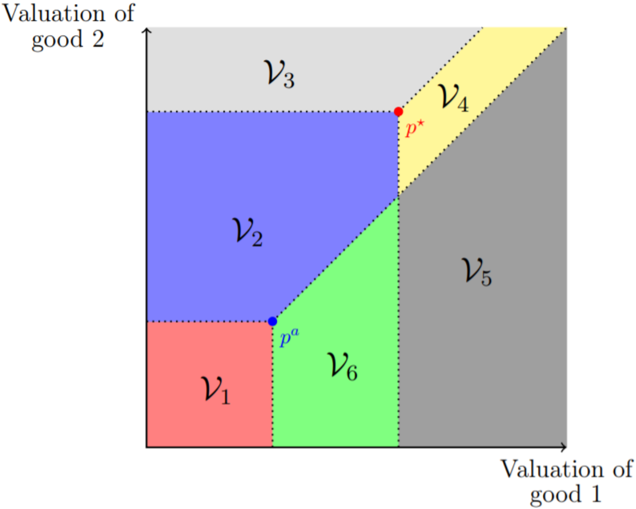
$$v, v' \in \mathcal{V} \text{ for some } \mathcal{V} \in \mathbb{V} \quad \Leftrightarrow \quad Y(v,p) = Y(v',p) \text{ for all } p \in \mathcal{P}. \tag{17}$$

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

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

- Then we know that

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

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

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

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

Few Concluding Remarks

- In practice when you have multiple markets and and $x's$, but you care about the aggregate, what you do is to averge the bounds

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

- How to use this method in the wild? We need
  - Some exogenous variation in costs
  - Discrete choice setting (otherwise, just instrument)
  - visceral hatred for distributional assumptions
- Understand that it can be challenging to get strict bounds, therefore you should have enough variations in costs
- Noise
- Imposing restrictions