Class 3

September 7 2021

Review from first class:

Formal Definition

Definition

$m^* \in M$ is identified in M iff $\forall m \in M$, $m \neq m^*$, $P(m) \neq P(m^*)$.

Definition

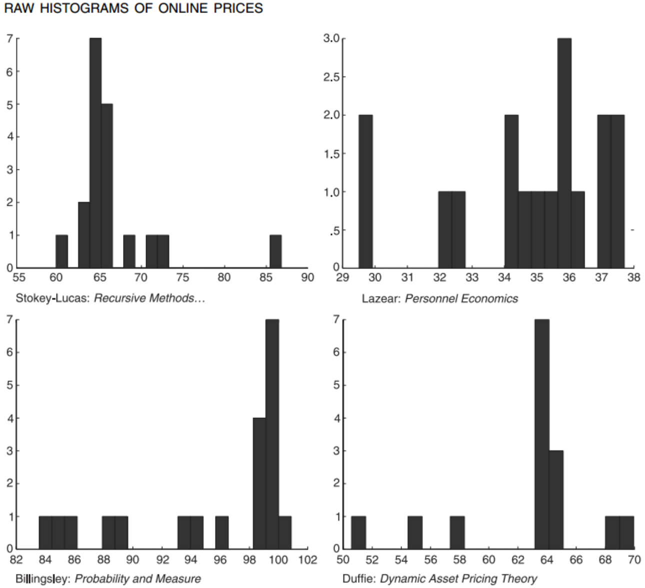
$m^* \in M$ is not identified in M iff $\exists m \in M$, $m \neq m^*$, $P(m) = P(m^*)$.

If for some m and m', we have $P(m) = P(m')$, we say that m and m' are **observationally equivalent**.

Example: Search (Hong and Shum Rand '06)

- Hong and Shum (06) consider both simultaneous and sequential search.
- Here, we consider simultaneous search.
- Based on Burdett and Judd (1983), model of price dispersion with firm mixed strategies. (Do you know this model?)

Hong and Shum, Setting



Model

- Consumer i has search cost C_i.
- Decides how many times to search ex-ante (i.e., simultaneous).
- Primitive: $F_C(\cdot)$. Assume F_C has support on $[\underline{C}, \overline{C}]$, $\underline{C} > 0$.
  - We take as given that $\underline{C}$ and $\overline{C}$ are known.
- Consumer's utility is $\overset{\text{value of good}}{v_i} - \overset{\text{search cost}}{(n-1)C_i} - \overset{\text{price}}{p_i}$
  - n is # of search
- For now, assume away the decision of whether or not to buy (say lower bound of v_i is higher than upper bound of p_i)
- Consider identification of F_C.
- Given distribution of prices F_P (known to consumer), consumer samples $n(C_i)$ prices from the distribution.

- Consumer i's problem

$$n(C_i) = \arg\min_{n \geq 1} (n-1)C_i + \int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p) dp$$

- What is $\int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p) dp$?
- $\int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p) dp = \mathbf{E}[\text{lowest price out of } n \text{ draws}]$

- Consumer i's problem

$$n(C_i) = \arg\min_{n \geq 1} (n-1) C_i + \int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p) dp$$

- What is $\int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p) dp$?
- $\int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p) dp = \mathbf{E}[$lowest price out of n draws$]$
- $n(C_i)$ is nonincreasing in C_i (why?)

- Consumer i's problem

$$n(C_i) = \arg\min_{n \geq 1} (n-1)C_i + \int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p)\,dp$$

- What is $\int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p)\,dp$?
- $\int_{\underline{p}}^{\overline{p}} pn(1 - F_P(p))^{n-1} f_P(p)\,dp = \mathbf{E}[$lowest price out of n draws$]$
- $n(C_i)$ is nonincreasing in C_i (why?)
- Incremental cost: C_i

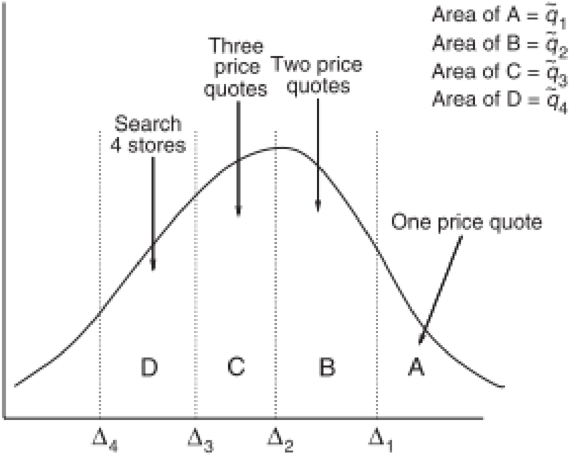
- Consumer i's problem

$$n(C_i) = \arg\min_{n \geq 1} (n-1)C_i + \int_{\underline{p}}^{\overline{p}} pn(1-F_P(p))^{n-1} f_P(p)\,dp$$

- What is $\int_{\underline{p}}^{\overline{p}} pn(1-F_P(p))^{n-1} f_P(p)\,dp$?
- $\int_{\underline{p}}^{\overline{p}} pn(1-F_P(p))^{n-1} f_P(p)\,dp = \mathbf{E}[\text{lowest price out of } n \text{ draws}]$
- $n(C_i)$ is nonincreasing in C_i (why?)
- Incremental cost: C_i
- Incremental Benefit: $\mathbf{E}[\text{lowest price out of } n+1 \text{ draws}] - \mathbf{E}[\text{lowest price out of } n \text{ draws}]$

- Define $\Delta_n = \mathbf{E}[p_{1:n} - p_{1:n+1}]$ where $p_{1:n}$ is the lowest realization from n independent samples from F_P.
  - Δ_n is a function of $P(m)$. Hence Δ_n are all identified.
  - Δ_n: incremental value of an extra search.
  - If $C_i > \Delta_n$, then (incremental cost of search) $>$ (incremental value of search)
  - If $C_i < \Delta_n$, then (incremental cost of search) $<$ (incremental value of search)
  - Δ_n corresponds to the threshold search cost that determines whether consumer searches n or $n+1$ times.
  - Can show $\Delta_{n+1} \leq \Delta_n$ (decreasing marginal benefit), and $\Delta_n \to 0$ as $n \to \infty$.
- Define $q_1 = 1 - F_C(\Delta_1)$: mass of consumers who only take one price draw
- Define $q_K = F_C(\Delta_{K-1}) - F_C(\Delta_K)$ for $K = 2 \cdots \overline{K}$.
  - q_K is the mass of consumers who take K draws.
  - Note that $\underline{C} > 0$ implies $\overline{K}$ is finite.

Graphically...



- We now show that $q_K = F_C(\Delta_{K-1}) - F_C(\Delta_K)$ are identified.
- To do so, use firms' eqm condition in a mixed-strategy price dispersion model based on Burdett and Judd.
- Firm profit from charging price p:

$$\Pi(p) \propto \overbrace{(p-r)}^{\text{Margin}} \sum_{K=1}^{\overline{K}} q_K K \times \overbrace{(1-F_P(p))^{K-1}}^{\text{Prob. } p \text{ is lower than } K-1 \text{ other draws}} .$$

- q_K: mass of consumers who take K price quotes.
- $q_K K$: probability that a given price quote is received by a person who takes K price quotes.
- Needs $\Pi(p) = \Pi(p')$ for all p, p' (b/c mixed strategy).
- In particular, $\Pi(p) = \Pi(p^M)$ where p^M is monopoly price (why?).
  - Observe that p^M should be the upper support of F_P.

- Let $p_1 < p_2 < \cdots < p_T = p^M$ be any price in the support of F_P.

$$(p^M - r)q_1 = (p_i - r)\sum_{K=1}^{\overline{K}} q_K K(1 - F_P(p_i))^{K-1} \text{ for all } i.$$

- As long as $T > \overline{K}$, there is generically a unique solution $(r, q_1, \cdots q_{\overline{K}})$ that solves equation above.
- Hence $(q_1, \cdots q_{\overline{K}})$ are identified.
- This means that F_C is identified on $\overline{K}$ points $(\Delta_1 \cdots \Delta_{\overline{K}})$
  - $q_1 = 1 - F_C(\Delta_1)$, and $q_K = F_C(\Delta_{K-1}) - F_C(\Delta_K)$ for $K = 2 \cdots \overline{K}$.
- F_C is not identified on all other points (partial identification)
- Q: propose an estimator for $F_C(\Delta_1) \cdots F_C(\Delta_{\overline{K}})$.

TABLE 2 **Search-Cost Distribution Estimates for Nonsequential-Search Model**

Product	K^{a}	M^{b}	$\tilde{q}_1^{c}$	$\tilde{q}_2$	$\tilde{q}_3$	Selling Cost r	MEL Value
Parameter estimates and standard errors: nonsequential-search model							
Stokey-Lucas	3	5	.480 (.170)	.288 (.433)		49.52 (12.45)	102.62
Lazear	4	5	.364 (.926)	.351 (.660)	.135 (.692)	27.76 (8.50)	84.70
Billingsley	3	5	.633 (.944)	.309 (.310)		69.73 (68.12)	199.70
Duffie	3	5	.627 (1.248)	.314 (.195)		35.48 (96.30)	109.13

Search-cost distribution estimates

	Δ_1	$F_c(\Delta_1)$	Δ_2	$F_c(\Delta_2)$	Δ_3	$F_c(\Delta_3)$
Stokey-Lucas	2.32	.520	.68	.232		
Lazear	1.31	.636	.83	.285	.57	.150
Billingsley	2.90	.367	2.00	.058		
Duffie	2.41	.373	1.42	.059		

Questions to think about:

- Simultaneous v.s. Sequential
  - How would you proceed for sequential search
- Learning

Identification: General Additive Models

- Consider an additive model (error is additive)

$$\begin{aligned} y &= g^*(\mathbf{X}) + \varepsilon \\ E[\varepsilon|\mathbf{X}] &= 0 \end{aligned}$$

- $\mathbf{X} \in \mathbb{R}^K$, $\varepsilon \in \mathbb{R}^1$, $y \in \mathbb{R}^1$. $m^* = (g^*, F^*_{\varepsilon,\mathbf{X}}) \in M$
- $M = \{\text{cont. fn.}\} \times \{\text{Dist. over } (\varepsilon, \mathbf{X})\}$.
  - Nonparametric model
- Is m^* identified?
- Yes

- Consider $E[y|\mathbf{X}:m]$:

$$E[y|\mathbf{X}:m] = E[g(\mathbf{X})+\varepsilon|\mathbf{X}:m] = g(\mathbf{X}).$$

- $E[y|\mathbf{X}:m]$ can be derived from $P(m)$ (NOTE: $P(m)$ is $F_{y,\mathbf{X}}$.)
- Hence $g^*(\cdot)$ is identified.
- We now show identification of $F_\varepsilon^*(\cdot|\mathbf{X})$.
  - Note that if we identify $F_\varepsilon^*(\cdot|\mathbf{X})$, we identify $F_{\varepsilon,\mathbf{X}}^*$ b.c. $F_{\mathbf{X}}^*$ is identified too.
- Consider conditional cdf of y given $\mathbf{X}$ under m, $F_y(\cdot|\mathbf{X};m)$.

$$\begin{aligned} F_y(\tau|\mathbf{X};m) &= \Pr(g(\mathbf{X})+\varepsilon \leq \tau|\mathbf{X};m) \\ &= \Pr(\varepsilon \leq \tau - g(\mathbf{X})|\mathbf{X};m) \\ &= F_\varepsilon(\tau - g(\mathbf{X})|\mathbf{X};m) \end{aligned}$$

- Rearrange to see that

$$F_y(\tau+g(\mathbf{X})|\mathbf{X};m) = F_\varepsilon(\tau|\mathbf{X};m).$$

- LHS is known b/c g is identified. Hence $F_\varepsilon^*(\tau|\mathbf{X})$ is identified $\forall\tau$.

General Additive Model

- For any y and $\mathbf{X}$, You can always write

$$y = E[y|\mathbf{X}] + (\mathbf{y} - E[y|\mathbf{X}]).$$

  Note that $E(\mathbf{y} - E[y|\mathbf{X}]|\mathbf{X}) = 0$.
- The generalized additive model is capturing exactly this relationship

$$y = \overbrace{E[y|\mathbf{X}]}^{g(\mathbf{X})} + \overbrace{(\mathbf{y} - E[y|\mathbf{X}])}^{\varepsilon}.$$

- $E[y|\mathbf{X}]$ is called the "regression" of y on $\mathbf{X}$.

Simultaneous Equation Systems

- Consider the canonical demand-supply system as follows:

$$\begin{aligned} \text{Demand} &: \quad Q = \alpha_1 P + \beta_1' X + u_1 \\ \text{Supply} &: \quad Q = \alpha_2 P + \beta_2' X + u_2 \end{aligned}$$

  - Interpretation: First equation is the "demand equation"
    - How much is consumed at given price?
  - Second equation is the "supply equation"
    - How much is produced at a given price
- This formulation is called "structural" b/c the eq. has economic interpretation.
- Assume $X \perp U$ ($U = (u_1, u_2)'$)
- Observe (Q, P, X) (the intersection of Supply and Demand given X)
- Not observe U.
- Question: Is α_1, α_2, β_1, β_2, F_U identified?

Simultaneous Equation Systems

- Rewrite Supply-Demand Eq as

$$\begin{pmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_2 \end{pmatrix} \begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} \beta_1' \\ \beta_2' \end{pmatrix} X + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$\Leftrightarrow \quad \begin{pmatrix} Q \\ P \end{pmatrix} = \begin{pmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_2 \end{pmatrix}^{-1} \begin{pmatrix} \beta_1' \\ \beta_2' \end{pmatrix} X + \begin{pmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_2 \end{pmatrix}^{-1} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}.$$

- Define $A = \begin{pmatrix} 1 & -\alpha_1 \\ 1 & -\alpha_2 \end{pmatrix}$, $B = \begin{pmatrix} \beta_1' \\ \beta_2' \end{pmatrix}$, $U = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, $\Pi = A^{-1}B$.
- Then Supply-Demand eq is

$$\begin{pmatrix} Q \\ P \end{pmatrix} = \Pi X + A^{-1} U.$$

- This is called "reduced-form" because Π does not have a "structural" interpretation.
  - Elements of Π are a combination of demand and supply parameters.

Simultaneous Equation Systems

- Supply-Demand eq is eqv. to

$$\begin{pmatrix} Q \\ P \end{pmatrix} = \overbrace{\Pi X}^{g(\mathbf{X})} + \overbrace{A^{-1} U}^{\varepsilon}$$

- Note that this is a special case of the general additive model, b/c
$$E[A^{-1}U|X] = A^{-1}E[U|X] = 0$$
- Hence $\Pi = A^{-1}B$ and distribution of $A^{-1}U$ are identified.
- Note, however, A, B and F_U are not separately identified.
- In particular, Take any $\tilde{A} \neq A$ that is invertible and consider $\tilde{B} = \tilde{A}A^{-1}B$, and take $F_{U'}$ to be the distribution of $\tilde{A}A^{-1}U$.
  - (A, B, F_U) and $(\tilde{A}, \tilde{B}, F_{U'})$ are obs. equiv.
  - This is true for the reduced form. Notice that the reduced form is equiv. to structural.

Simultaneous Equation Systems

- Where is the "under identification"?
- Recall that $\Pi = A^{-1}B$ and distribution of $A^{-1}U$ are identified.
- What is Π?

$$\Pi = A^{-1}B = \frac{1}{\alpha_1 - \alpha_2} \begin{pmatrix} -\alpha_2 & \alpha_1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1' \\ \beta_2' \end{pmatrix}.$$

- Π has $2 \times K$ elements where $K = \dim(X)$.
- RHS has $2 \times K + 2$ elements $(\beta_1, \beta_2, \alpha_1, \alpha_2)$.
- In general, not identified.

Simultaneous Equation Systems

- **Sufficient condition for identification**:
  - one variable excluded from demand but not from supply AND
  - another variable excluded from supply, but not from demand.
- Consider X_{1i} excluded from demand, i.e., $\beta_{1i} = 0$

$$\Pi = \frac{1}{\alpha_1 - \alpha_2} \begin{pmatrix} -\alpha_2 & \alpha_1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \beta_{11}, \cdots, \beta_{1i-1}, 0, \beta_{1i+1}, \cdots, \beta_{1K} \\ \beta_{21}, \cdots, \beta_{2i-1}, \beta_{2i}, \beta_{2i+1}, \cdots, \beta_{2K} \end{pmatrix}.$$

- Consider $i-$th column of Π, Π_i.

$$\Pi_i = \frac{1}{\alpha_1 - \alpha_2} \begin{pmatrix} \alpha_1 \beta_{2i} \\ \beta_{2i} \end{pmatrix}.$$

- Consider Π_{1i}/Π_{2i}:

$$\Pi_{1i}/\Pi_{2i} = \alpha_1.$$

- α_1 (slope of demand) is identified.

Simultaneous Equation Systems

- Intuitively...
- Consider $\Delta Q/\Delta P$ when $X = (X_1 \cdots, X_i, \cdots, X_K)$ changes to $X + e_i = (X_1 \cdots, X_i + 1, \cdots, X_K)$.
- At X,
$$E\left[\begin{pmatrix} Q \\ P \end{pmatrix} \middle| X\right] = \Pi X.$$
- At $X + e_i$,
$$E\left[\begin{pmatrix} Q \\ P \end{pmatrix} \middle| X + e_i\right] = \Pi X + \Pi_i.$$
- Hence, $\Delta Q/\Delta P$:
$$\Pi_{1i}/\Pi_{2i} = \alpha_1.$$
- Graph

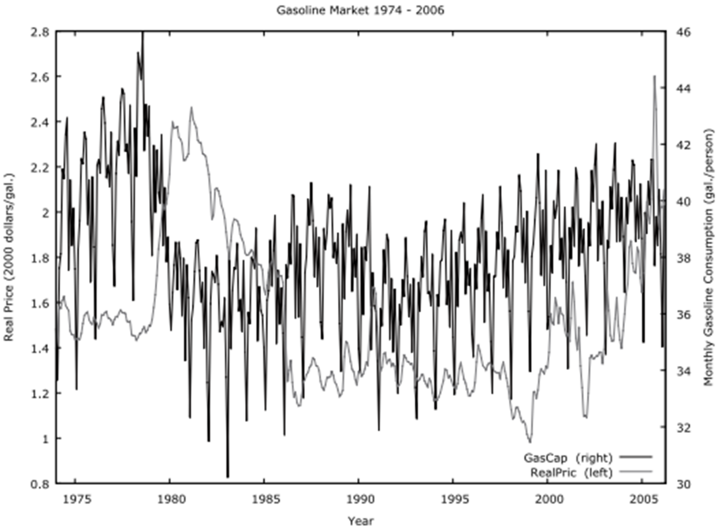
Simultaneous Equation Systems

- People often say “supply side” shocks identify demand.
  - This is what people mean: That there is some X_i which is excluded from demand ($\beta_{1i} = 0$) and not excluded from supply ($\beta_{2i} \neq 0$).
- Similarly, if $\beta_{2j} = 0$, α_2 is identified.
  - This is what people mean when they say “demand side” shocks identify demand.

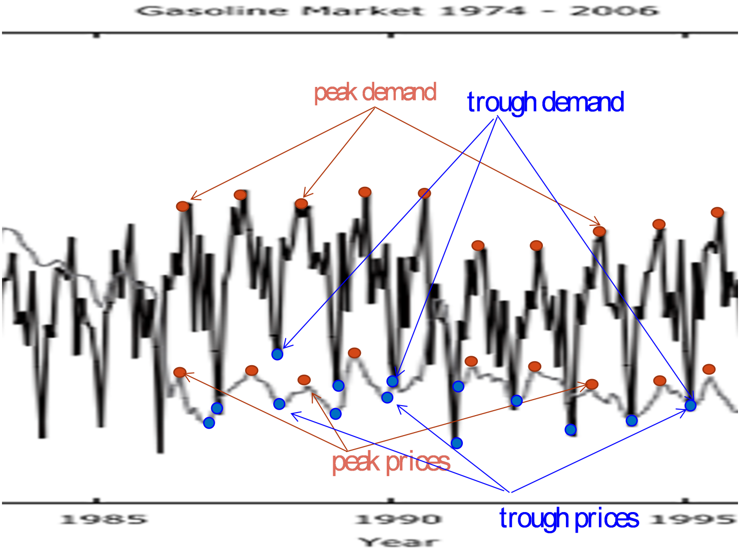
Simultaneous Equation Systems

- Questions to think about:
  - 1. What if we have two variables that are excluded from the first equation?
  - 2. What are some variables that are typically excluded from Demand/Supply Equations?
  - 3. In light of what we just discussed, consider the follwoing (P, Q) pairs for U.S. gas over time

Price and Quantity (P,Q) for U.S. gasoline ('74-'06)



Price & Quantity of Gasoline, 1985-1995



- Question:Think about estimating supply and demand curves. Think through the identification.
- Helpful intermediate steps

- Question:Think about estimating supply and demand curves. Think through the identification.
- Helpful intermediate steps
  - **Question 1** (Focusing on the period between 1985-1995) Explain the periodic dips and spikes in quantity and (to a lesser extent) in prices. Use a supply-demand diagram to explain.

- Question:Think about estimating supply and demand curves. Think through the identification.
- Helpful intermediate steps
  - **Question 1** (Focusing on the period between 1985-1995) Explain the periodic dips and spikes in quantity and (to a lesser extent) in prices. Use a supply-demand diagram to explain.
  - **Question 2** Explain the decline in quantity and increase in prices around 1980 using a supply-demand diagram.

Extensions: Simultaneous Eq Sys w/ N endogenous vars

- Consider
$$AY + BX = U$$
  - A is $N \times N$, Y is $N \times 1$, B is $N \times K$, X is $K \times 1$, and U is $N \times 1$.
  - Y are endogenous and X are exogenous and U are unobservables.
  - Assume A not singular and $U \perp X$.
- Similar as before, $\Pi = -A^{-1}B$ and $F_{A^{-1}U}$ are identified.
- In general, neither F_U, A, nor B are identified
- Need further restrictions.
- See Hausman '83 Handbook of Metrics for various necessary and sufficient conditions.
- See Matzkin for ID of nonparametric simultaneous eq. systems of the form $U = r(Y, X)$.

Identification of Non-additive Models

- Consider
$$y = g^*(X, \varepsilon)$$
    - $X \in \mathbb{R}^K$, $\varepsilon \perp X$.
    - $\varepsilon \in \mathbb{R}^1$ distributed $F^*_\varepsilon(\cdot)$.
    - y and X observed, ε not observed.
- $m^* = (g^*, F^*_\varepsilon)$. We assume that g^* is increasing in ε, that is, for $\varepsilon' > \varepsilon''$, $g^*(X, \varepsilon') > g^*(X, \varepsilon'')$ for all X. We also assume that $g^* \in C$.
- M is the set of pairs (g, F_ε), such that $g(\cdot, \cdot)$ is increasing in the $K + 1$ element.
    - w/o monotinicity requirement on ε, no way $g^*(\cdot, \cdot)$ and F^*_ε can be identified.
    - Consider $m^* = (g^*, F^*_\varepsilon)$ and $m' = (g', F'_\varepsilon)$, s.t. $g'(x, e) = g^*(x, -e)$ and $F'_\varepsilon(\tau) = F^*_{-\varepsilon}(\tau) = 1 - F^*(-\tau)$.

Identification of Non-additive Models

- Given that the non-additive structure is pretty general, it turns out that this model is not identified (even with monotonicity w.r.t. ε).
- Take any str. incr. function s: $\mathbb{R}^1 \to \mathbb{R}^1$ s.t. $s(\mathbb{R}^1) = \mathbb{R}^1$.
- Consider $F'_\varepsilon(\tau) = F^*_\varepsilon \circ s(\tau)$ and $g'(x,\varepsilon) = g^*(x, s(\varepsilon))$.
- Then (g^*, F^*_ε) and (g', F'_ε) are obs. equiv.
  - ex. $s(x) = 2x$

Identification of Non-additive Models

- Why?
- Consider $F_y(\cdot|X;m^*)$:

$$\begin{aligned} F_y(\tau|X;m^*) &= \Pr(y \leq \tau|X;m^*) \\ &= \Pr(g^*(X,\varepsilon) \leq \tau|X;m^*) \\ &= \Pr(\varepsilon \leq g_X^{*-1}(\tau)|X;m^*) \\ &= F_\varepsilon^* \circ g_X^{*-1}(\tau), \end{aligned}$$

  where $g_X^{*-1}(\tau)$ is inverse of $g^*(X,\varepsilon)$ w.r.t ε at X.
- Consider $F_y(\cdot|X;m')$

$$\begin{aligned} F_y(\tau|X;m') &= \Pr(g'(X,\varepsilon) \leq \tau|X;m') \\ &= \Pr(g^*(X,s(\varepsilon)) \leq \tau|X;m') \\ &= F_\varepsilon' \circ s^{-1} \circ g_X^{*-1}(\tau) \\ &= F_\varepsilon^* \circ s \circ s^{-1} \circ g_X^{*-1}(\tau) = F_\varepsilon^* \circ g_X^{*-1}(\tau). \end{aligned}$$

Identification of Non-additive Models

- Hence, $F_y(\cdot|X; m^*) = F_y(\cdot|X; m')$.
- Given that $F_X(\cdot|m') = F_X(\cdot|m^*)$, $F_{y,X}(\cdot|m') = F_{y,X}(\cdot|m^*)$.
  - m' and m^* not identified.
- It turns out that the converse of the preceding claim is also true.

Claim If $m' = (g', F'_\varepsilon)$ and $m^* = (g^*, F^*_\varepsilon)$ are observationally equivalent, then for some incr. function $s(\cdot)$,
$g'(x, s(\varepsilon)) = g^*(x, \varepsilon)$.

Identification of Non-additive Models

- Why?
- If m' and m^* are obs. equiv., then $P(m^*) = P(m')$.
  - In particular, $F_y(\cdot|X;m^*) = F_y(\cdot|X;m')$.
- Consider $F_y(\cdot|X;m^*)$:

$$\begin{aligned} F_y(\tau|X;m^*) &= \Pr(y \le \tau|X;m^*) \\ &= F_\varepsilon^* \circ g_X^{*-1}(\tau). \end{aligned}$$

- Consider $F_y(\cdot|X;m')$:

$$\begin{aligned} F_y(\tau|X;m') &= \Pr(y \le \tau|X;m') \\ &= F_\varepsilon' \circ g_X'^{-1}(\tau). \end{aligned}$$

- $F_y(\cdot|X;m^*) = F_y(\cdot|X;m')$ implies $F_\varepsilon^* \circ g_X^{*-1}(\tau) = F_\varepsilon' \circ g_X'^{-1}(\tau)$.

Identification of Non-additive Models

- Recall
$$F'_\varepsilon \circ g'^{-1}_X(\tau) = F^*_\varepsilon \circ g^{*-1}_X(\tau) \text{ for all } \tau$$
- This implies
$$g'^{-1}_X(\tau) = F'^{-1}_\varepsilon \circ F^*_\varepsilon \circ g^{*-1}_X(\tau).$$
- Hence,
$$g'(X, \varepsilon) = g^*(X, F^{*-1}_\varepsilon \circ F'_\varepsilon(\varepsilon))$$
- Define $s(\varepsilon) = F^{*-1}_\varepsilon \circ F'_\varepsilon(\varepsilon)$. Then $s(\varepsilon)$ is str. incr fn.
- We have
$$g'(X, \varepsilon) = g^*(X, s(\varepsilon)).$$

Identification of Non-additive Models

- Implications:
  - If M only contains fn such that no two functions in M can be expressed as $m(x, \varepsilon) = m'(x, s(\varepsilon))$, m^* is identified.
  - Ex. Class of homothetic fn.
  - See Matzkin for other examples and applications.

Binary Threshold Crossing Models

- Binary threshold crossing

  $$Y = \mathbf{1}\{h(X) - \eta \geq 0\}$$

  where $\mathbf{1}\{E\}$ is an indicator fn.
- Assume $h \in C^0$ and F_η not known. $X \perp \eta$.
  $M = C^0 \times \{\text{Dist.}\}$.
- $P(m^*) = F_{Y,X} = \Pr(Y = 1|X)$.
- Suppose $h(X)$ is utility from doing something
  - Buying a good, bearing a child, taking up employment, starting a new line of business etc.
- Suppose η is the utility of the outside good/option value
  - Not buying, not bearing child, staying out of workforce, waiting for next opportunity
- Binary threshold crossing is a popular model of individual choice.
- Is $h(\cdot)$ identified in M?

Binary Threshold Crossing Models

- consider $\Pr(Y = 1|X)$:

$$\begin{aligned}\Pr(Y &= 1|X) = \Pr(h(X) \geq \eta) \\ &= F_\eta(h(X)) \\ &= F_\eta \circ h(X).\end{aligned}$$

- Hence, only convolution $F_\eta \circ h$ identified.
- Intuitively, consider two models:
- model (A)

$$Y = \mathbf{1}\{h(X) - \eta \geq 0\}$$

- model (B)

$$Y = \mathbf{1}\{G(h(X)) - G(\eta) \geq 0\}$$

- Models (A) and (B) are obs. equiv. if $G(\cdot)$ str. inct. fn.

Binary Threshold Crossing Models

- Suppose we interpret $h(X) - \eta$ as utility.
- Utility fn that are monotone transformation of each other represent the same preference.
- Hence, often it's justified to normalize the distribution of η - say, uniform.
- If F_η known, binary threshold crossing model

$$Y = \mathbf{1}\{h(X) - \eta \geq 0\}$$

  is **identified**.
- This is because

$$\begin{aligned} \Pr(Y &= 1|X) = F_\eta \circ h(X). \\ &\Rightarrow h(X) = F_\eta^{-1}(\Pr(Y = 1|X)). \end{aligned}$$

Binary Threshold Crossing Models

- Alternatively, suppse we set $h(X)$ as linear as $h(X) = \beta X$ and leave F_η unspecified:

$$Y = 1\{\beta' X - \eta \geq 0\}.$$

- Recall

$$\begin{aligned} \Pr(Y &= 1|X) = \Pr(\beta' X \geq \eta) \\ &= F_\eta(\beta' X). \end{aligned}$$

- Consider $\frac{\partial}{\partial X_i} \Pr(Y = 1|X)$, where X_i is the i-th element of X.

$$\begin{aligned} \frac{\partial}{\partial X_i} \Pr(Y &= 1|X) = \frac{\partial}{\partial X_i} F_\eta(\beta' X) \\ &= F'_\eta(\beta' X)\beta_i. \end{aligned}$$

- Taking ratios, we get

$$\frac{\frac{\partial}{\partial X_i} \Pr(Y = 1|X)}{\frac{\partial}{\partial Xj} \Pr(Y = 1|X)} = \frac{\beta_i}{\beta_j}.$$

- If you fix one of the βs to 1, you identify F_η too.

Binary Threshold Crossing Models

- When $h(X)$ has the utility interpretation, it makes sense to normalize one of the βs to 1.
- Then we are measuring utility with respect to the variable whose coefficient is normalized to 1.
- Often people set the coefficient on price equal to -1. Hence the utility of various X is measured in terms of "dollars"
- If one of the βs is normalized to 1, then can identify F_η.

Discrete Choice Models

- Consider the following model

$$\begin{aligned} U_j &= u_j^*(X) + \varepsilon_j \\ Y &= \arg\max_{j \in J}\{U_j\} \end{aligned}$$

- Assume $\varepsilon \perp X$, and observe (Y, X_j).
- ε_j is unobservable.
- First, consider simple case in which $F^*_{\varepsilon_1,\ldots,\varepsilon_J} = \prod F^*_\varepsilon$
- Q: can we identify $(u_1^*(\cdot),\cdots,u_J^*(X),\ F_\varepsilon^*)$?
- No.

Discrete Choice Models

- Cannot identify the scale of u_j^* and F_ε^*. Consider (u_j^*, F_ε^*) and $(C \cdot u_j^*, F_{C\cdot\varepsilon}^*)$, where $F_{C\cdot\varepsilon}^*(\tau) = F_\varepsilon^*(\tau / C)$.
  - Scaling U_j by a constant does not change $P(m)$
- Cannot identify u_j^* up to any additive function of X: Consider (u_j^*, F_ε^*) and $(u_j^* + g(X), F_\varepsilon^*)$, where $g(\cdot)$ does not depend on j.
  - Means that we need normalization such as $u_J^*(X) = 0$.
- These two conditions are natural if U is interpreted as utitlity.
  - scaling of utility is inconsequential
  - only utility differences among choices matter (not levels)

Discrete Choice Models

- Consider the DC model

$$
\begin{aligned}
U_j &= u_j^*(X) + \varepsilon_j \ (1 \le j \le J-1) \\
U_J &= \varepsilon_J \\
Y &= \arg\max_{j \in J}\{U_j\}
\end{aligned}
$$

- Note that $u_J^*(X) = 0$.
- Assume ε_j has c.d.f. type-I extreme, i.e., $F_\varepsilon(\tau) = \exp(-\exp(-\tau))$.
- $m^* = u^*(X)$.
- Is u_j^* $(1 \le j \le J-1)$ identified?

Discrete Choice Models

- Consider the probability of choosing alternative j conditional on X:

$$\Pr(Y = j|X) = \Pr(U_j \geq \max_{j' \in J}(U_{j'})) = \frac{\exp(u_j(X))}{\sum_{j' \in J} \exp(u_{j'}(X))}$$

  - You should derive this if you have never done so
- Taking logs of both sides,

$$\begin{aligned} \log(\Pr(Y &= j|X)) = \log\left(\frac{\exp(u_j(X))}{\sum_{j' \in J} \exp(u_{j'}(X))}\right) \\ &= u_j(X) - \log\left(\sum_{j' \in J} \exp(u_{j'}(X))\right). \end{aligned}$$

- Hence,

$$\begin{aligned} \log(\Pr(Y &= j|X)) - \log(\Pr(Y = J|X)) \\ &= u_j(X) - u_J(X) = u_j(X). \end{aligned}$$

- Hence u_j^* is identified.

Discrete Choice Models

- N.B. A semiparametric estimator consistent:

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

  under appropriate regularity conditions if X is a discrete.
- If X is continuous then replace $\widehat{P_j(x)}$ with a Kernel estimator (e.g. Nadaraya-Watson) or sieves.
- N.B. The Discrete Choice set up with $u_J^* = 0$ is identified if F_ε^* is known. F_ε^* need not be type-I extreme and, more importantly, $\varepsilon_1, \cdots, \varepsilon_J$ can be arbitarily correlated (**pf next class**)