Partial Identification of Structural Models

Francesca Molinari

Presented by Qihong Ruan

Nov 26, 2019

3.1 Discrete Choice in Single Agent Random Utility Model

- $\mathcal{I}$: a population of decision makers
- $\mathcal{Y} = \{c_1, ..., c_{|\mathcal{Y}|}\}$: a set of finite potential alternatives (feasible set)
- $\in^*$: is chosen from

$$\mathbb{P}(c \in^* C) = \mathbb{P}(\pi_i(c) \geq \pi_i(b) \,\forall b \in C)$$

for all $c \in C$, $C \subset Y$, $i \in I$. π_i is a random utility function.

- Example:

$$\pi_{ij} = \beta x_{ij} + \epsilon_{ij}$$

i indexes decision makers, j indexes alternatives, x_{ij} is a vector of observed variables relating to alternative j for person i.

$$\begin{aligned} \mathbb{P}_{ij} = \mathbb{P}(y_{ij} = 1) &= \mathbb{P}(\pi_{ij} \geq \pi_{ik} \,\forall k \neq j) \\ &= \mathbb{P}(\epsilon_{ik} - \epsilon_{ij} \leq \beta x_{ij} - \beta x_{ik} \,\forall k \neq j) \end{aligned}$$

3.1.1 Semiparametric Binary Choice Models with Interval Valued Covariates

Identification Problem 3.1

- Observe (y, x_L, x_U, w) in $\{0,1\} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$, $d < \infty$
- $x \in \mathbb{R}$ is unobservable
- $y = 1(w\theta + \delta x + \epsilon > 0)$, ϵ is continuous conditional on (w, x, x_L, x_U)
- Suppose $\delta > 0$, normalize $\delta = 1$
- R is the joint distribution function of $(y, x, x_L, x_U, w, \epsilon)$
    - $\mathrm{R}(x_L \leq x \leq x_U) = 1$
    - $\mathrm{R}(\epsilon | w, x, x_L, x_U) = \mathrm{R}(\epsilon | w, x)$
    - $\mathrm{R}(\epsilon \leq 0 | w, x) = \alpha$

Problem:
Observe (y, x_L, x_U, w), x is unobservable, $y = 1(w\theta + x + \epsilon > 0)$. What can we learn about θ?

Observe (y, x_L, x_U, w), x is unobservable, $y = 1(w\theta + x + \epsilon > 0)$. What can we learn about θ?

$$y = 1 \Rightarrow \epsilon > -w\theta - x_U$$

$$y = 0 \Rightarrow \epsilon \leq -w\theta - x_L$$

$$-w\theta - x_U \leq -w\theta - x_L$$

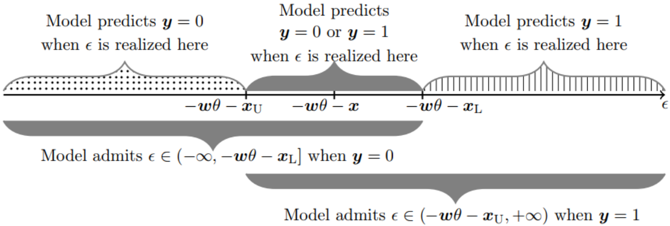


Figure 3.1: Predicted value of y as a function of ϵ, and admissible values of ϵ for each realization of y, in Identification Problem 3.1, conditional on (w, x_L, x_U).

Note: when x is observed, the prediction is unique.

Why does this set-valued prediction hinder point estimation?

$$\begin{aligned}\mathrm{P}(y=1|w,x_L,x_U) &= \int \mathrm{R}(y=1|w,x,x_L,x_U)d\mathrm{R}(x|w,x_L,x_U)\\ &= \int \mathrm{R}(\epsilon > -w\theta - x|w,x)d\mathrm{R}(x|w,x_L,x_U)\end{aligned}$$

The first equation uses the Law of Iterated Expectation; the second equation uses the assumption that $\mathrm{R}(\epsilon|w,x,x_L,x_U) = \mathrm{R}(\epsilon|w,x)$.

- Since $\mathrm{R}(x|w,x_L,x_U)$ is unspecified, we can find multiple values for θ satisfying the assumptions in Identification Problem 3.1 and yielding the observed value of $\mathrm{P}(y=1|w,x_L,x_U)$
- However, not all $\theta \in \Theta$ can be paired with some R
- Thus, θ is partially identified

THEOREM SIR-3.1: The sharp identification region for θ

Under the Assumptions of Identification Problem Problem 3.1, the sharp identification region for θ is

$$\mathcal{H}_{\mathrm{P}}[\theta] = \{\vartheta \in \Theta : \mathrm{P}((w, x_L, x_U) : \{0 \le w\vartheta + x_L \cap \mathrm{P}(y = 1|w, x_L, x_U) \le 1 - \alpha\} \cup \{w\vartheta + x_U \le 0 \cap \mathrm{P}(y = 1|w, x_L, x_U) \ge 1 - \alpha\}) = 0\}. (3.1)$$

Proof. The set of possible values for ϵ given (y, w, x_L, x_U) is

$$\mathcal{E}_\theta(y) \equiv \mathcal{E}_\theta(y, w, x_L, x_U) = \begin{cases} (-\infty, -w\theta - x_L] & \text{if } y = 0, \\ [-w\theta - x_U, +\infty) & \text{if } y = 1. \end{cases}$$

If the model is correctly specified,

$$(\epsilon, w, x_L, x_U) \in (\mathcal{E}_\theta(y), w, x_L, x_U)$$

- Molchanov and Molinari (2018) show that $(\epsilon, w, x_L, x_U) \in (\mathcal{E}_\theta(y), w, x_L, x_U)$ occurs if and only if

$$\mathrm{R}(\epsilon \in C | w, x_L, x_U) \geq \mathrm{P}(\mathcal{E}_\theta(y) \subset C | w, x_L, x_U) \; \forall C \in \mathrm{F},$$

  where F denotes the collection of closed subsets of $\mathbb{R}$.

- Using the Law of Iterated Expectation,

$$\int \mathrm{R}(\epsilon \in C | w, x, x_L, x_U) d\mathrm{R}(x | w, x_L, x_U) \geq \mathrm{P}(\mathcal{E}_\theta(y) \subset C | w, x_L, x_U)$$

- Using $\mathrm{R}(\epsilon | w, x, x_L, x_U) = \mathrm{R}(\epsilon | w, x)$,

$$\int \mathrm{R}(\epsilon \in C | w, x) d\mathrm{R}(x | w, x_L, x_U) \geq \mathrm{P}(\mathcal{E}_\theta(y) \subset C | w, x_L, x_U)$$

- Recall the assumption $\mathrm{R}(\epsilon \leq 0 | w, x) = \alpha$. Let $C = (-\infty, 0]$, then

$$\alpha \geq \mathrm{P}(\mathcal{E}_\theta(y) \subset (-\infty, 0] | w, x_L, x_U)$$

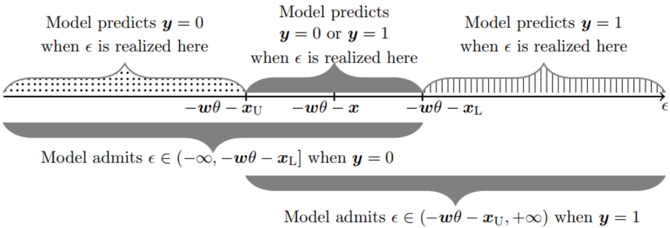


Figure 3.1: Predicted value of $\boldsymbol{y}$ as a function of ϵ, and admissible values of ϵ for each realization of $\boldsymbol{y}$, in Identification Problem 3.1, conditional on $(\boldsymbol{w}, \boldsymbol{x}_{\mathrm{L}}, \boldsymbol{x}_{\mathrm{U}})$.

$$\begin{aligned}\alpha &\geq \mathrm{P}(\mathcal{E}_\theta(y) \subset (-\infty, 0] | w, x_L, x_U) \\ &= \mathrm{P}(y = 0 \cap -w\theta - x_L \leq 0 | w, x_L, x_U) \\ &= \mathrm{P}(y = 0 \cap w\theta + x_L \geq 0 | w, x_L, x_U)\end{aligned}$$

$$\alpha \geq \mathrm{P}(y = 0 | w, x_L, x_U) \quad \forall (w, x_L, x_U) \text{ such that } w\theta + x_L \geq 0$$

$$\begin{aligned}1 - \alpha &\leq 1 - \mathrm{P}(y = 0 | w, x_L, x_U) \; \forall (w, x_L, x_U) \text{ such that } w\theta + x_L \geq 0 \\ &= \mathrm{P}(y = 1 | w, x_L, x_U) \; \forall (w, x_L, x_U) \text{ such that } w\theta + x_L \geq 0\end{aligned}$$

Use $\int R(\epsilon \in C|w,x)dR(x|w,x_L,x_U) \geq P(\mathcal{E}_\theta(y) \subset C|w,x_L,x_U)$ and $R(\epsilon \leq 0|w,x) = \alpha$ again, let $C = [0,+\infty)$, we have

$$\begin{aligned} 1-\alpha &\geq P(\mathcal{E}_\theta(y) \subset [0,+\infty)|w,x_L,x_U) \\ &= P(y=1 \cap -w\theta - x_U \geq 0|w,x_L,x_U) \\ &= P(y=1 \cap w\theta + x_U \leq 0|w,x_L,x_U) \end{aligned}$$

$$\boxed{1-\alpha \geq P(y=1|w,x_L,x_U)\ \forall (w,x_L,x_U) \text{ such that } w\theta + x_U \leq 0. \ (3.2)}$$

$$\boxed{1-\alpha \leq P(y=1|w,x_L,x_U)\ \forall (w,x_L,x_U) \text{ such that } w\theta + x_L \geq 0. \ (3.3)}$$

Any given $\vartheta \in \Theta$, $\vartheta \neq \theta$, violates (3.2) or (3.3) if and only if

$$\begin{aligned} P((w,x_L,x_U) : \{0 \leq w\vartheta + x_L \cap P(y=1|w,x_L,x_U) < 1-\alpha\} \\ \cup \{w\vartheta + x_U \leq 0 \cap P(y=1|w,x_L,x_U) > 1-\alpha\}) > 0 \end{aligned}$$

Notice that when $w\vartheta + x_U > 0$ and $w\vartheta + x_L < 0$, i.e., $-x_U < w\vartheta < -x_L$, ϑ does not violate (3.2) or (3.3), thus ϑ is not distinguishable from θ.

Identification Problem 3.2: Parametric Regression with Interval Covariate Data

- Observe $(y, x_L, x_U, w) \sim \mathrm{P}$ in $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d$, $d < \infty$
- $x \in \mathbb{R}$ is unobservable
- R is the joint distribution of (y, x, x_L, x_U). $\mathrm{R}(x_L \leq x \leq x_U) = 1$; $\mathbb{E}_{\mathrm{R}}(y|w, x, x_L, x_U) = \mathbb{E}_{\mathrm{Q}}(y|w, x)$
- $\mathbb{E}_{\mathrm{Q}}(y|w, x) = f(w, x; \theta)$; f is known and weakly increasing in x

Problem: What can we learn about θ?

The sharp identification region for θ is

$$\mathcal{H}_{\mathrm{P}}[\theta] = \{\vartheta \in \Theta : f(w, x_L; \vartheta) \leq \mathbb{E}_{\mathrm{P}}(y|w, x_L, x_U) \leq f(w, x_U; \vartheta)\} \quad (3.8)$$

- *Proof. (Following the proof of Theorem SIR-2.4)*

$$\begin{aligned} \mathbb{E}_{\mathrm{P}}(y|w, x_L, x_U) &= \int \mathbb{E}_{\mathrm{R}}(y|w, x, x_L, x_U) d\mathrm{R}(x|w, x_L, x_U) \\ &= \int \mathbb{E}_{\mathrm{Q}}(y|w, x) d\mathrm{R}(x|w, x_L, x_U) \\ &= \int f(w, x; \theta) d\mathrm{R}(x|w, x_L, x_U) \end{aligned}$$

  Here we use the Law of Iterated Expectation, $\mathbb{E}_{\mathrm{R}}(y|w, x, x_L, x_U) = \mathbb{E}_{\mathrm{Q}}(y|w, x)$, and $\mathbb{E}_{\mathrm{Q}}(y|w, x) = f(w, x; \theta)$.

- Since f is weakly increasing in x, and $x_L < x < x_U$, $f(w, x_L; \theta) \leq \int f(w, x; \theta) d\mathrm{R}(x|w, x_L, x_U) \leq f(w, x_U; \theta)$

- When $f(w, x_L; \vartheta) \leq \mathbb{E}_{\mathrm{P}}(y|w, x_L, x_U) \leq f(w, x_U; \vartheta)$, ϑ is observationally equivalent to θ

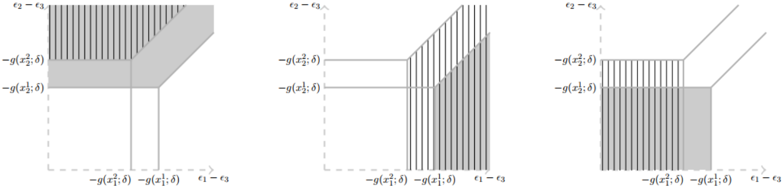
3.1.2 Endogenous Explanatory Variables

Identification Problem 3.3 (Discrete Choice with Endogenous Explanatory Variables)

- Observe random variables $(\mathbf{y}, \mathbf{x}, \mathbf{z}) \sim \mathrm{P}$ in $\mathcal{Y} \times \mathcal{X} \times \mathcal{Z}$
- $v \equiv (\epsilon_{c_1}, ..., \epsilon_{c_{|\mathcal{Y}|}})$, $v \perp\!\!\!\perp \mathbf{z}$, $v \sim \mathrm{Q}$, $\mathrm{Q} \in \mathcal{T}$, $\mathcal{T}$ is a specified family of distributions
- The conditional distribution $\mathrm{S}(v|\mathbf{x}, \mathbf{z})$ is continuous on $(\mathbf{x}, \mathbf{z})$
- The utility function $\pi_i(c) = g(\mathbf{x_c}; \delta) + \epsilon_c$, g is known, $\delta \in \Delta \subset \mathbb{R}^m$, $\forall c \in \mathcal{Y}$
- Normalize $g(\mathbf{x}_{c_{|\mathcal{Y}|}}; \delta) = 0$
- Given $(\mathbf{x}, \mathbf{z}, v)$, suppose $\mathbf{y}$ is the utility maximizing choice in $\mathcal{Y}$, what can we learn about (δ, Q)?

- For any $c \in \mathcal{Y}$ and $x \in \mathcal{X}$, c is chosen if and only if v realizes in the set

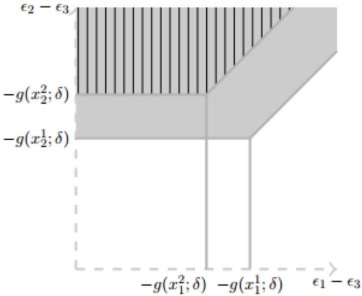
$$\mathcal{E}_\delta(c, x) = \{e \in \mathcal{V} : g(x_c; \delta) + e_c \geq g(x_d; \delta) + e_d \; \forall d \in \mathcal{Y}\}$$



(a) $\mathcal{E}_\delta(\boldsymbol{y}, \boldsymbol{x})$ for $\boldsymbol{y} =$ ~~1~~ 2 (b) $\mathcal{E}_\delta(\boldsymbol{y}, \boldsymbol{x})$ for $\boldsymbol{y} =$ ~~2~~ 1 (c) $\mathcal{E}_\delta(\boldsymbol{y}, \boldsymbol{x})$ for $\boldsymbol{y} = 3$

Figure 3.2: The set $\mathcal{E}_\delta$ in equation (3.9) and the corresponding admissible values for $(\boldsymbol{y}, \boldsymbol{x})$ as a function of $(\epsilon_1 - \epsilon_3, \epsilon_2 - \epsilon_3)$ under the simplifying assumption that $\mathcal{X} = \{x^1, x^2\}$ and $\mathcal{Y} = \{1, 2, 3\}$. The admissible values for $(\boldsymbol{y}, \boldsymbol{x})$ are $\{(c, x^1)\}$ in the gray area, and $\{(c, x^2)\}$ in the area with vertical lines. Because the two areas overlap, the model has set-valued predictions for $(\boldsymbol{y}, \boldsymbol{x})$.

Figure 3.2 plots the set $\mathcal{E}_\delta(\mathbf{y}, \mathbf{x})$ when $\mathcal{Y} = \{1, 2, 3\}$ and $\mathcal{X} = \{x^1, x^2\}$, as a function of $(\epsilon_1 - \epsilon_3, \epsilon_2 - \epsilon_3)$



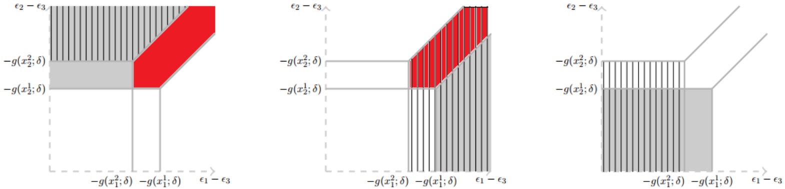
(a) $\mathcal{E}_\delta(\boldsymbol{y}, \boldsymbol{x})$ for $\boldsymbol{y} = 1$ 2

$$\mathbf{y} = 2 \iff g(x_2;\delta) + \epsilon_2 \geq g(x_1;\delta) + \epsilon_1,\ g(x_2;\delta) + \epsilon_2 \geq \epsilon_3$$

$$\iff \epsilon_1 - \epsilon_2 = (\epsilon_1 - \epsilon_3) - (\epsilon_2 - \epsilon_3) \leq g(x_2;\delta) - g(x_1;\delta),\ \epsilon_2 - \epsilon_3 \geq -g(x_2;\delta)$$

$$\iff \epsilon_2 - \epsilon_3 \geq -g(x_2;\delta) + g(x_1;\delta) + (\epsilon_1 - \epsilon_3),\ \epsilon_2 - \epsilon_3 \geq -g(x_2;\delta)$$

- When **x** changes from x^1 to x^2, the region changes. In this case, $g(x_1^2;\delta) > g(x_1^1;\delta)$, $g(x_2^2;\delta) < g(x_2^1;\delta)$, and $g(x_3^1;\delta) = g(x_3^2;\delta) = 0$



(a) $\mathcal{E}_\delta(\boldsymbol{y},\boldsymbol{x})$ for $\boldsymbol{y} =$ ~~1~~ 2 (b) $\mathcal{E}_\delta(\boldsymbol{y},\boldsymbol{x})$ for $\boldsymbol{y} =$ ~~2~~ 1 (c) $\mathcal{E}_\delta(\boldsymbol{y},\boldsymbol{x})$ for $\boldsymbol{y} = 3$

Figure 3.2: The set $\mathcal{E}_\delta$ in equation (3.9) and the corresponding admissible values for $(\boldsymbol{y},\boldsymbol{x})$ as a function of $(\epsilon_1 - \epsilon_3, \epsilon_2 - \epsilon_3)$ under the simplifying assumption that $\mathcal{X} = \{x^1, x^2\}$ and $\mathcal{Y} = \{1, 2, 3\}$. The admissible values for $(\boldsymbol{y},\boldsymbol{x})$ are $\{(c, x^1)\}$ in the gray area, and $\{(c, x^2)\}$ in the area with vertical lines. Because the two areas overlap, the model has set-valued predictions for $(\boldsymbol{y},\boldsymbol{x})$.

- When v is realized at the red area, we have two possible $(\mathbf{x},\mathbf{y})$:

$$(x^1, 2),\ (x^2, 1)$$

- Recall that in Problem 3.1, the model predicts $y = 0$ or $y = 1$ for some ϵ, and hence partial identification results

Compared with Identification Problem 3.1

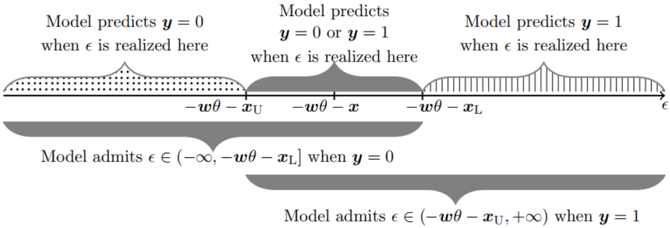


Figure 3.1: Predicted value of y as a function of ϵ, and admissible values of ϵ for each realization of y, in Identification Problem 3.1, conditional on (w, x_L, x_U).

- Problem 3.1: Model predicts multiple values of $\mathbf{y}$ for some ϵ
- Problem 3.3: Model predicts multiple values of $(\mathbf{x}, \mathbf{y})$ for some ϵ
- The set-valued prediction results in partial identification of parameters of interest

The Sharp Identification Region for (δ, Q)

$$\begin{aligned}\mathrm{P}(\mathcal{E}_\delta(\mathbf{y},\mathbf{x}) \subseteq F|\mathbf{z}) &= \sum_{c\in\mathcal{Y}} \mathbf{1}(\mathcal{E}_\delta(c,\mathbf{x}) \subseteq F) \cdot \mathrm{P}(\mathbf{y}=c|\mathbf{z}) \\ &= \int_{x\in\mathcal{X}} \sum_{c\in\mathcal{Y}} \mathbf{1}(\mathcal{E}_\delta(c,x) \subseteq F) \cdot \mathrm{P}(\mathbf{y}=c|\mathbf{x}=x,\mathbf{z}) d\mathrm{P}(x|\mathbf{z})\end{aligned}$$

For given $F \in \mathcal{F}$ and $\delta \in \Delta^m$. $\mathcal{F}$ is the collection of closed subsets of $\mathcal{V}$, and $\mathcal{V}$ is the sample space of $v \equiv (\epsilon_{c_1}, \ldots, \epsilon_{c_{|\mathcal{Y}|}})$.

- This probability can be learned from the observed data $(\mathbf{x}, \mathbf{y}, \mathbf{z})$
- Under the assumptions of Identification Problem Problem 3.3, the sharp identification region for (δ, Q) is

$$\mathcal{H}_{\mathrm{P}}[\delta, \mathrm{Q}] = \{\delta \in \Delta, \mathrm{Q} \in \mathcal{T} : \mathrm{Q}(F) \geq \mathrm{P}(\mathcal{E}_\delta(\mathbf{y},\mathbf{x}) \subseteq F|\mathbf{z}),\ \forall F \in \mathcal{F}\} \tag{3.13}$$

Theorem SIR-3.3 (Discrete Choice with Endogenous Explanatory Variables)

Under the assumptions of Identification Problem Problem 3.3, the sharp identification region for (δ, Q) *is*

$$\mathcal{H}_{\mathrm{P}}[\delta, \mathrm{Q}] = \{\delta \in \Delta, \mathrm{Q} \in \mathcal{T} : \mathrm{Q}(F) \geq \mathrm{P}(\mathcal{E}_\delta(\mathbf{y}, \mathbf{x}) \subseteq F | \mathbf{z}), \; \forall F \in \mathcal{F}\} \tag{3.13}$$

Proof.

Notation: $\mathcal{E}_\delta \equiv \mathcal{E}_\delta(\mathbf{y}, \mathbf{x})$, $(\mathcal{E}_\delta, \mathbf{x}, \mathbf{z}) = \{(\mathbf{e}, \mathbf{x}, \mathbf{z}) : \mathbf{e} \in \mathcal{E}_\delta\}$.

- $(v, \mathbf{x}, \mathbf{z}) \in (\mathcal{E}_\delta, \mathbf{x}, \mathbf{z})$ for the data generating value of (δ, Q) as long as the model is correctly specified
- By Theorem A.1 and Theorem 2.33 in Molchanov and Molinari (2018), this occurs if and only if

$$\mathrm{S}(F | \mathbf{x}, \mathbf{z}) \geq \mathrm{P}(\mathcal{E}_\delta(\mathbf{y}, \mathbf{x}) \subseteq F | \mathbf{x}, \mathbf{z}), \; \forall F \in \mathcal{F}$$

- Integrate $\mathbf{x}$ out at both sides, and use the fact that Q does not depend on $\mathbf{z}$, we have $\mathrm{Q}(F) \geq \mathrm{P}(\mathcal{E}_\delta(\mathbf{y}, \mathbf{x}) \subseteq F | \mathbf{z})$

Why does partial identification result?

$$\mathrm{M}(c|\mathbf{x} \in R_z, \mathbf{z} = z; \delta) = \int_{x \in R_z} \mathrm{S}(\mathcal{E}_\delta(c, \mathbf{x})|\mathbf{x} = x, \mathbf{z} = z) d\mathrm{P}(x|z), \ \forall R_z \subseteq \mathcal{X} \tag{3.10}$$

$$\mathrm{Q}(F) = \int_{x \in \mathcal{X}} \mathrm{S}(F|\mathbf{x} = x, \mathbf{z} = z) d\mathrm{P}(x|z), \ \forall F \subseteq \mathcal{V} \tag{3.11}$$

- The joint distribution of $(\mathbf{x}, v)$ conditional on z is left completely unrestricted (except for (3.11), we can find multiple $(\delta, \mathrm{Q}, \mathrm{S})$ satisfying the maintained assumptions and such that $\mathrm{M}(c|\mathrm{x} \in R_z, \mathrm{z} = z; \delta) = \mathrm{P}(c|\mathbf{x} \in R_z, \mathbf{z} = z) \ \forall c \in \mathcal{Y}$ and $R_z \subset \mathcal{X}$
- McFadden's 1973 conditional logit model yields point identification of δ when $\mathbf{x} \perp\!\!\!\perp v$
- When $\mathbf{x}$ is endogenous, $\mathrm{S}(v|\mathbf{x}, \mathbf{z})$ may change across realizations of $\mathbf{x}$
- For given realization of v, the model admits sets of values for endogenous variables $(\mathbf{y}, \mathbf{x})$, partial identification results

Insights: Models with Endogenous Variables as Incomplete Models

- Chesher, Rosen, and Smolinski (2013) show that one can frame models with endogenous explanatory variables as incomplete models
- Incompleteness here results from the fact that the model does not specify how the endogenous variables **x** are determined
- One can then think of these as models with set-valued predictions for the endogenous variables
- Random set theory can again be leveraged to characterize sharp identification regions

Insights: Point and Partial Identification

- Manski (1985): When (y, w, x) is observed,

$$w\theta + x > 0 \Leftrightarrow \mathrm{P}(y = 1|w, x) > 1 - \alpha$$

using $y \equiv 1(w\theta + x + \epsilon > 0)$ and $\mathrm{R}(\epsilon \leq 0|w, x) \equiv \alpha$.

- Hence, θ is identified relative to $\vartheta \in \Theta$ if

$$\mathrm{P}((w, x) : \{w\theta + x \leq 0 < w\vartheta + x\} \cup \{w\vartheta + x \leq 0 < w\theta + x\}) > 0. \tag{3.4}$$

- Manski and Tamer (2002): When x is unobserved, but $x \in [x_L, x_U]$, the collection of values that cannot be distinguished from θ is

$$\{\vartheta \in \Theta : \mathrm{P}((w, x_L, x_U) : \{w\theta + x_U \leq 0 < w\vartheta + x_L\} \cup \{w\vartheta + x_U \leq 0 < w\theta + x_L\}) = 0\}. \tag{3.5}$$

- The reasoning of point identification can be extended to partial identification

Introduce the Instrument Variable

- Magnac and Maurin (2008) assume that an instrumental variable z is available
  - ϵ is independent of x conditional on (w, z), and $Corr(z, \epsilon) = 0$
  - x is continuous with support $[v_1, v_k]$
  - $\mathbb{P}[x \in [v_i, v_{i+1})|w, z] > 0 \;\forall\; i = 1, ..., k-1$
- If x were observed, follow Lewbel (2000), let

$$\tilde{y} = \frac{y - 1_{x>0}}{f_x(x|w, z)}$$

then

$$\theta = \mathbb{E}_{\mathrm{P}}(zw^T)^{-1}\mathbb{E}_{\mathrm{P}}(z\tilde{y}) \tag{3.6}$$

- If x are interval valued, let x^* takes value $i \in 1, ..., k-1$ if $x \in [v_i, v_{i+1})$, $\delta(x^*) = v_{x^*+1} - v_{x^*}$, $y^* = \frac{\delta(x^*)}{\mathrm{P}(x^*=i|w,z)}y - v_k$, then the sharp identification region for θ is

$$\mathcal{H}_{\mathrm{P}}[\theta] = \mathbb{E}_{\mathrm{P}}(zw^T)^{-1}\mathbb{E}_{\mathrm{P}}(zy^* + zU) \tag{3.7}$$