The Theory of Identification

ECON 31720 Applied Microeconometrics

Francesco Ruggieri

The University of Chicago

October 21, 2020

Framework for the Theory of Identification

- Let $W \in \mathbb{R}^{d_w}$ be an **observed** random vector with **known distribution** G
- $\theta \in \Theta$ is a **parameter** describing a hypothetical **state of the world**
  - θ may be characterized by one or more constants, distributions, or both
- The **parameter space** Θ places **limitations** on the hypothetical states of the world
  - Θ may restrict the support of a constant (e.g. $\beta > 0$)
  - Θ may restrict the set of admissible distributions (e.g. $F \in \mathcal{F} : \mathbb{E}_F [XU] = 0$)

Framework for the Theory of Identification

- A **model** $\{(\theta, G_\theta) : \theta \in \Theta\}$ maps parameters into **potential distributions** of W
  - G_θ is the **implied** distribution of W in the hypothetical state of the world associated with θ
- A **target parameter** $\{(\theta, \pi(\theta)) : \theta \in \Theta\}$ is a function of interest for the researcher
- Importantly, each **model** is a **function** $\mu(\theta) = G_\theta$
  - Each parameter, $\theta \in \Theta$, implies **one and one only** potential distribution G_θ
  - However, it may be the case that $G_{\theta_1} = G_{\theta_2}$ for $\theta_1 \neq \theta_2$
- Analogously, each **target parameter** is a **function** $\pi = \pi(\theta)$
  - Each parameter (state of the world), $\theta \in \Theta$, implies **one and one only** target parameter π
  - However, any given π need **not** be implied by one and one only θ

Framework for the Theory of Identification

Lewbel (2019) considers two additional definitions:

1. The **structure**

$$s(G, \pi) = \{\theta \in \Theta : \mu(\theta) = G, \pi(\theta) = \pi\}$$

is the set of parameters that yield **both** the observed distribution G **and** the target parameter π

2. Two target parameters, π_1 and π_2, are defined to be **observationally equivalent** if

$$\exists G \text{ s.t. } (G, \pi_1) \neq \emptyset \text{ and } s(G, \pi_2) \neq \emptyset$$

Point Identification

- In the lecture notes, a target parameter π is defined to be **point identified** if the set

$$\Pi^* (G) \equiv \{\pi (\theta) : \theta \in \Theta, G_\theta = G\}$$

  includes **a single element**.

- Analogously, Lewbel (2019) defines a target parameter π to be point identified if

$$\nexists \ \tilde{\pi} \neq \pi \quad \text{s.t.} \quad \pi \text{ and } \tilde{\pi} \text{ are observationally equivalent}$$

- In other words, there do **not** exist states of the world (i.e., parameters) that **both**
  1. imply the **observed distribution of the data** via the posited model ($G_\theta = G$), **and**
  2. imply **two distinct target parameters**

Examples of Point Identification: Median

- **Observed data**: a real-valued random variable $W \in \mathbb{R}$ with distribution G
- **Parameter**: the marginal distribution of W, $\theta = F$
- **Parameter space**: all distributions of W with a strictly monotonic distribution function

$$\Theta = \{F \in \mathcal{F} : F(w) > F(w') \; \forall w > w'\}$$

- The **model** is trivially $G_\theta(w) \equiv \mathbb{P}_\theta(W \leq w) = F(w)$

Examples of Point Identification: Median

- **Target parameter**: the median of W, $\pi \equiv \inf \{w \in \mathbb{R} : F(w) \geq 0.5\}$
- In this case, it is not possible for $\pi \neq \tilde{\pi}$ to be **observationally equivalent** because

$$F(\pi) = F(\tilde{\pi}) = 0.5 \implies \pi = \tilde{\pi} \quad \forall F \in \Theta$$

and each possible distribution of W has a **unique distribution function**.

Examples of Point Identification: Linear Supply and Demand

- **Observed data**: a random vector $W \equiv (Q, P, Z)$ with $Q, P \in \mathbb{R}_+$ and $Z \in \mathbb{R}$
- **Parameter**: $\theta = (\alpha, \beta, \gamma, F)$, where
  - α, β, γ are real constants, and F is the joint distribution function for (P, Z, U, V)
- **Parameter space**: all constants and distributions of W such that demand equals supply

$$\Theta = \{\theta \equiv (\alpha, \beta, \gamma, F) : Q = \beta P + \gamma Z + U = \alpha P + V, U \perp\!\!\!\perp Z, V \perp\!\!\!\perp Z\}$$

- The **model** is a standard system of supply and demand equations:

$$G_\theta(q, p, z) \equiv \mathbb{P}_\theta(\beta P + \gamma Z + U \leq q, P \leq p, Z \leq z) = \mathbb{P}_\theta(\alpha P + V \leq q, P \leq p, Z \leq z)$$

- **Target parameter**: the price elasticity of demand, $\pi = \alpha$

Examples of Point Identification: Linear Supply and Demand

- The model **implies** the following **reduced-form** coefficients:

$$Q = \phi_1 Z + R_1 \qquad Q = \phi_2 Z + R_2 \quad \text{with} \quad \phi_1 = \frac{\alpha\gamma}{\alpha - \beta} \text{ and } \phi_2 = \frac{\gamma}{\alpha - \beta}$$

- The **model** can therefore be rewritten as

$$G_\theta(q, z) \equiv \mathbb{P}_\theta(\phi_1 Z + R_1 \leq q, Z \leq z) = \mathbb{P}_\theta(\phi_2 Z + R_2 \leq q, Z \leq z)$$

with ϕ_1, ϕ_2, R_1, R_2 defined as resulting from the simultaneous equations above

- Notice that $\alpha = \frac{\phi_2}{\phi_1}$. Target parameters $\widehat{\alpha}$ and $\widetilde{\alpha}$ will be **observationally equivalent** if

$$\widehat{\alpha} \neq \widetilde{\alpha} \quad \text{but} \quad \phi_1 = \phi_2 = 0$$

which occurs if $\boldsymbol{\gamma = 0}$, i.e., the exogenous supply shifter does not affect quantity supplied

- For α to be point identified, Θ must be updated with the **assumption** that $\gamma \neq 0$

Examples of Point Identification: Latent Error Distribution

- **Observed data**: a random vector $W \equiv (X, Y)$ with $X \in \mathbb{R}$ and $Y \in \{0, 1\}$
- **Parameter**: $\theta = F$, where F is the joint distribution function for (X, U)
- **Parameter space**: all joint distributions such that X and U are independent

$$\Theta = \{F \in \mathcal{F} : X \perp\!\!\!\perp U \text{ and } X \text{ has continuous support}\}$$

- The **model** is $G_\theta (x, y) \equiv \mathbb{P}_\theta (\mathbb{I}[X + U > 0] \leq y, X \leq x)$
- **Target parameter**: the marginal distribution function of U, $H_U (u)$

Examples of Point Identification: Latent Error Distribution

- Notice that, because $Y \in \{0, 1\}$ and $X \perp\!\!\!\perp U$, for any $x \in \mathcal{X}$

$$\mathbb{E}_G[Y|X = x] = \mathbb{P}_F(X + U > 0|X = x) = \mathbb{P}_F(x + U > 0) = 1 - H_U(-x)$$

- The conditional expectation $\mathbb{E}[Y|X]$ is a **feature of the observed data**

- The target parameter can be **point identified** for all u in the support of $-X$, $\mathcal{X}^-$

$$H_U(u) = 1 - \mathbb{E}_G[Y|X = -u] \quad \forall u \in \mathcal{X}^-$$

- Two distinct marginal distributions of U cannot imply equal $\mathbb{E}[Y|X = x]$ for all x

$$\widehat{H}_U(u) \neq \widetilde{H}_U(u) \text{ for some } u \implies \mathbb{E}_{\widehat{F}}[Y|X = x] \neq \mathbb{E}_{\widetilde{F}}[Y|X = x] \text{ for some } x$$

  Thus, no marginal distributions of U can be observationally equivalent

Partial Identification

- In the lecture notes, a target parameter π is defined to be **partially (or set) identified** if

$$\Pi^* (G) \equiv \{\pi(\theta) : \theta \in \Theta, G_\theta = G\}$$

has **more than one element** and is a strict subset of $\mathbb{R}^{d_\pi}$

- Analogously, Lewbel (2019) defines a target parameter π to be partially identified if

$$\exists\ \tilde{\pi} \in \mathbb{R}^{d_\pi} \quad \text{s.t.} \quad \pi \text{ and } \tilde{\pi} \text{ are observationally equivalent}$$

- If **all possible** $\tilde{\pi}$ are observationally equivalent to π, then π is **not** (set) identified

- In other words, it is **not** the case that **all** states of the world (i.e., parameters) **both**
  1. Imply the **observed distribution of the data** via the posited model ($G_\theta = G$), **and**
  2. Imply **two or more distinct target parameters**

Examples of Partial Identification: Expenditure Shares

- **Observed data**: a real-valued random variable W_1 with marginal distribution G
  - Suppose W_1 denotes the fraction of a consumer's budget that is spent on food
- **Parameter**: $\theta = F$, where F is the joint distribution function of W
  - Each W_j in $W \equiv \left(W_1, \ldots, W_{\bar{j}}\right)$ denotes the budget share that is spent on a good
- **Parameter space**: all distributions of W with support $[0,1]^{\bar{j}}$:

$$\Theta = \left\{F \in \mathcal{F} : F \text{ has support } [0,1]^{\bar{j}}\right\}$$

- The **model** is $G_\theta\left(w_1, \ldots, w_{\bar{j}}\right) \equiv \mathbb{P}_\theta\left(W_1 \leq w_1, \ldots, W_{\bar{j}} \leq w_{\bar{j}}, \sum_{j=1}^{\bar{j}} W_j = 1\right)$
- **Target parameter**: the expected fraction of income spent on clothing: $\pi = \mathbb{E}\left[W_2\right]$

Examples of Partial Identification: Expenditure Shares

- Because W_1 is observed, the expected food expenditure share can be trivially identified:

$$\mathbb{E}_F\left[W_1\right] = \mathbb{E}_G\left[W_1\right]$$

- W_2 is unobserved, so the expected clothing expenditure share **cannot** be **point identified**

- However, knowledge of $\mathbb{E}\left[W_1\right]$ can be used to **bound** $\mathbb{E}\left[W_2\right]$
  - Given the restriction on the support of W, the **lower bound** is trivially 0
  - Because $\sum_{j=1}^{\bar{j}} W_j = 1$ by assumption and $\mathbb{E}\left[W_1\right]$ is point identified, the **upper bound** is $1 - \mathbb{E}\left[W_1\right]$, which occurs if food and clothing jointly exhaust consumer budgets (on average)

- Thus, $\mathbb{E}\left[W_2\right]$ is **partially identified** as

$$\pi \equiv \mathbb{E}\left[W_2\right] \in \left[0, 1 - \mathbb{E}_G\left[W_1\right]\right]$$

1. Framework for the Theory of Identification

2. Point Identification
   - Example: Median
   - Example: Linear Supply and Demand
   - Example: Latent Error Distribution

3. Partial Identification
   - Example: Expenditure Shares
   - Example: Demand for Differentiated Products

4. Falsifying Selection on Observables (Heckman, Ichimura, and Todd 1998)

5. Summary

Examples of Partial Identification: Demand for Differentiated Products

- Demand is commonly estimated with **discrete choice** models of product differentiation

- Individual i's **utility** from purchasing and consuming product $j \in \{0, 1, \ldots, \bar{j}\}$ is

$$U_{ij} = \alpha_i P_{ij} + X_{ij}' \beta_i + \eta_j + \varepsilon_{ij}$$

  where α_i and β_i are **random coefficients**, P indicates **prices**, X is a vector of **observed** product characteristics, and η encompasses **unobserved** product features

- The distribution of utility draws, ε, is typically parameterized as a Type-I Extreme Value
  - This modeling choice is referred to as **conditional logit**
  - With this parameterization of the i.i.d. shocks, individual **choice probabilities** are

$$\mathbb{P}(i \to j) = \frac{\exp\left(\alpha_i P_{ij} + X_{ij}' \beta_i + \eta_j\right)}{\sum_{k=0}^{\bar{j}} \exp\left(\alpha_i P_{ik} + X_{ik}' \beta_i + \eta_k\right)}$$

Examples of Partial Identification: Demand for Differentiated Products

- Parameters can be conveniently estimated with **maximum likelihood**
- However, the T1EV parameterization imposes **severe restrictions** on consumer behavior
  - The most salient restriction is the **Independence of Irrelevant Alternatives** property:

$$\frac{\mathbb{P}(i \to j)}{\mathbb{P}(i \to k)} = \frac{\exp\left(\alpha_i P_{ij} + X_{ij}'\beta_i + \eta_j\right)}{\exp\left(\alpha_i P_{ik} + X_{ik}'\beta_i + \eta_k\right)} \quad \forall j, k$$

  Whether a consumer's choice set is expanded (restricted) by the inclusion (exclusion) of a product has no effect on the **relative choice probabilities** of any pair of alternatives

- If one could just **not parameterize** the distribution of latent utility draws...

Examples of Partial Identification: Demand for Differentiated Products

Tebaldi, Torgovitsky, and Yang (2019):

- Does **not parameterize** the distribution of latent utility draws
- **Partially identifies** target parameters associated with consumer demand counterfactuals
- Assumes that consumer valuations and product prices are **additively separable**:

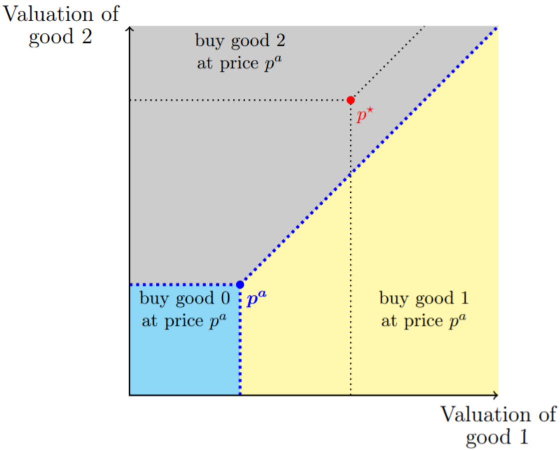
$$Y_i \equiv \arg \max_{j \in \{0,1,\ldots,\bar{j}\}} V_{ij} - P_j$$

  where V denotes valuations and P indicates prices

- Key intuition: given one or more **price vectors**, it is possible to **partition** the space of valuations into sets in which consumer behavior is **observationally equivalent**

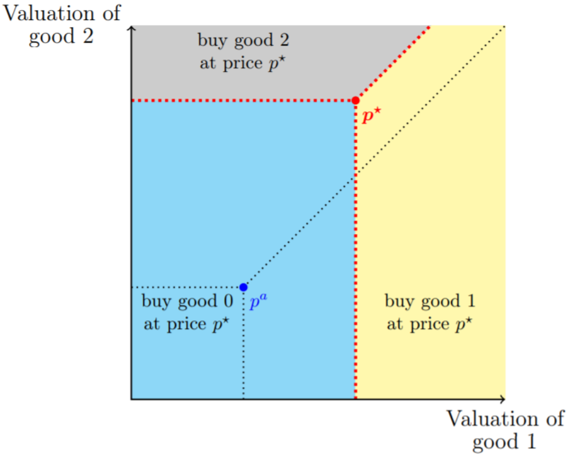
Examples of Partial Identification: Demand for Differentiated Products

Consider the case in which there are $\bar{j} = 2$ products and let p^a be an **observed price vector**.



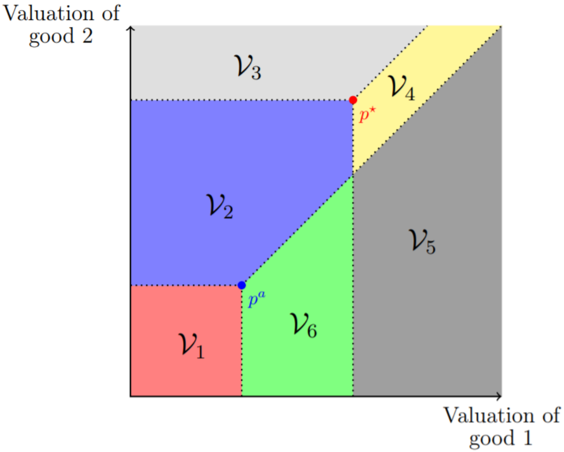
Examples of Partial Identification: Demand for Differentiated Products

Now let p^* be a **counterfactual price vector**.



Examples of Partial Identification: Demand for Differentiated Products

Combine the previous two figures and construct the **Minimal Relevant Partition (MRP)**.



Examples of Partial Identification: Demand for Differentiated Products

- **Consumer valuations differ** within each subset of the space of valuations
- But **consumer choices do not vary**, so those valuations are observationally equivalent
  - E.g. consumers with valuations in $\mathcal{V}_4$ choose good 2 when $p = p^a$ and good 1 when $p = p^*$
- The observed data consist of **consumption choice shares** when $p = p^a$
  - E.g. $P(Y = 0) = 1 - \alpha - \beta$, $P(Y = 1) = \alpha$, $P(Y = 2) = \beta$

Examples of Partial Identification: Demand for Differentiated Products

Suppose the **target parameter** were the share of consumers who buy good 2 if $p = p^*$

- Model assumptions are **not sufficient to point identify** $\phi_3 = \int_{\mathcal{V}_3} f(v)\, dv$
- But observed choice shares can be used to **construct bounds** for the target parameter
- "Worst-case" scenario: consumers who buy good 2 if $p = p^a$ have valuations in $\mathcal{V}_2$ and/or $\mathcal{V}_4$
- "Best-case" scenario: consumers who buy good 2 if $p = p^a$ have valuations in $\mathcal{V}_3$
- The target parameter can be **bounded below** by 0 and **above** by $P(Y = 2) = \beta$

Falsifying Selection on Observables

The following example is based on Heckman, Ichimura, and Todd (1998):

- $D \in \{0, 1\}$ indicates **participation in a job training program**
- $X \in \mathbb{R}^{d_x}$ is a vector of predetermined **observable** characteristics
- $Y \in \mathbb{R}$ denotes **earnings** at some point **after** the job training program
- Consider two **alternative** institutional scenarios:
  1. The job training program **accepts all individuals** who wish to participate
  2. Job training is **randomly offered** to a **subset** of applicants

Falsifying Selection on Observables

Consider the case in which all applicants are accepted into the program:

- **Observed data**: a random vector, $W \equiv (Y, D, X)$, jointly distributed according to G
- **Parameter**: $\theta = F$, where F is the joint distribution function for $(Y(0), Y(1), D, X)$
- **Parameter space**: all distributions such that **selection on observables** holds

$$\Theta = \{\theta \in \Theta : (Y(0), Y(1)) \perp\!\!\!\perp D|X \text{ under } F\}$$

- **Model**: $G_\theta(y, d, x) = \mathbb{P}_\theta(DY(1) + (1 - D)Y(0) \leq y, D \leq d, X \leq x)$
- No focus on a specific target parameter, so the **identified set** is

$$\Theta^*(G) \equiv \{\theta \in \Theta : G_\theta(y, d, x) = G(y, d, x) \ \forall y, d, x\}$$

Falsifying Selection on Observables

- A model is said to be **falsifiable** if there exists a known function $\tau : \mathcal{G} \to \{0, 1\}$ such that
  1. $\tau(G) = 1 \implies \Theta^*(G) = \emptyset$
  2. $\tau(G) = 1$ for at least one $G \in \mathcal{G}$

- **Assume** $\exists\, \tau$ that meets these two conditions. Thus, $\exists\, G$ such that $\Theta^*(G) = \emptyset$

- Argue **by contradiction**. Suppose there existed a state of the world (i.e., a θ) such that

$$\mathbb{P}_\theta\left(Y(0) \leq y_0, Y(1) \leq y_1 | D = d, X = x\right) \equiv \mathbb{P}_G\left(Y \leq y_0 | D = 0, X = x\right) \mathbb{P}_G\left(Y \leq y_1 | D = 1, X = x\right)$$
$$\mathbb{P}_\theta\left(D = d, X = x\right) = \mathbb{P}_G\left(D = d, X = x\right)$$

for all y_0, y_1, d, x, i.e., potential outcomes are independent of D conditional on X.

Falsifying Selection on Observables

- The proposed state of the world (i.e., parameter) θ:

  1. Satisfies **selection on observables**, i.e., $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, because $\mathbb{P}_\theta$ does **not** depend on $d = 0, 1$.

  2. **Implies a distribution** that is "consistent" with the distribution of **observed data**, G:

$$\begin{aligned}
\mathbb{P}_\theta\left(Y \leq y | D = 0, X = x\right) &= \mathbb{P}_\theta\left(DY(1) + (1 - D)Y(0) \leq y | D = 0, X = x\right) \\
&= \mathbb{P}_\theta\left(Y(0) \leq y | D = 0, X = x\right) \\
&= \mathbb{P}_\theta\left(Y(0) \leq y, Y(1) \leq +\infty | D = 0, X = x\right) \\
&= \mathbb{P}_G\left(Y \leq y | D = 0, X = x\right) \mathbb{P}_G\left(Y \leq +\infty | D = 1, X = x\right) \\
&= \mathbb{P}_G\left(Y \leq y | D = 0, X = x\right) \times 1 \\
\mathbb{P}_\theta\left(Y \leq y | D = 1, X = x\right) &= 1 \times \mathbb{P}_G\left(Y \leq y | D = 1, X = x\right)
\end{aligned}$$

  for all $x \in \mathcal{X}$. Thus, $G_\theta = G$, so $\boldsymbol{\theta} \in \boldsymbol{\Theta}^*\left(\boldsymbol{G}\right)$.

- $\theta \in \Theta$ and $\theta \in \Theta^*\left(G\right)$ contradict that $\Theta^*\left(G\right) = \emptyset$. Thus, the model is **not falsifiable**.

Falsifying Selection on Observables

Consider the case in which a **random subset** of the applicants is offered participation:

- $S \in \{0, 1\}$ denotes whether a worker **applied for/selected into** the program
- $R \in \{0, 1\}$ denotes whether a worker **was randomized into** the program
- $D \equiv SR$: workers are "treated" if they apply for **and** are randomized into the program
- In the first scenario, selection on observables entails assuming that $(Y(0), Y(1)) \perp\!\!\!\perp S|X$
  - Non-applicants' earnings are comparable to **potential** untreated earnings of participants
  - As shown above, this model is **untestable** (not falsifiable) because all applicants are treated
- In the second scenario, non applicants' earnings should be comparable to **observed** earnings of applicants who are **not randomized** into the program (and have the same x)

Falsifying Selection on Observables

- **Parameter**: $\theta = F$, where F is the joint distribution function for $(Y(0), Y(1), R, S, X)$

- **Parameter space**: all distributions such that **selection on observables** holds and program participation is **randomly assigned to applicants**

$$\Theta = \{\theta \in \Theta : (Y(0), Y(1)) \perp\!\!\!\perp SR|X \text{ and } (Y(0), Y(1)) \perp\!\!\!\perp R|S = 1 \text{ under } F\}$$

- **Model**: $G_\theta(y, r, s, x) = \mathbb{P}_\theta(SRY(1) + (1 - SR)Y(0) \leq y, R \leq r, S \leq s, X \leq x)$

- Any state of the world **implies** a distribution for "unsuccessful" applicants:

$$\begin{aligned}\mathbb{P}_\theta(Y \leq y|R = 0, S = 1, X = x) &= \mathbb{P}_\theta(Y(0) \leq y|R = 0, S = 1, X = x)\\ &= \mathbb{P}_\theta(Y(0) \leq y|S = 1, X = x)\\ &= \mathbb{P}_\theta(Y(0) \leq y|S = 0, X = x)\\ &= \mathbb{P}_\theta(Y \leq y|S = 0, X = x)\end{aligned}$$

Falsifying Selection on Observables

Thus, a **function τ** can be constructed as follows:

$$\tau(G) = \mathbb{I}\left[\mathbb{P}_G(Y \leq y | R = 0, S = 1, X = x) \neq \mathbb{P}_G(Y \leq y | S = 0, X = x)\right]$$

If $\boldsymbol{\tau}(\boldsymbol{G})$ **were equal to 1** and θ **belonged to the identified set**, then

$$\begin{aligned} \mathbb{P}_\theta(Y \leq y | R = 0, S = 1, X = x) &= \mathbb{P}_G(Y \leq y | R = 0, S = 1, X = x) \\ &\neq \mathbb{P}_G(Y \leq y | S = 0, X = x) \\ &= \mathbb{P}_\theta(Y \leq y | S = 0, X = x) \end{aligned}$$

which **would contradict the equality** proven in the previous slide. This model is **falsifiable**.

Summary

- Given a model, a target parameter is
  - **Point identified**, if **there exists no other** observationally equivalent target parameter
  - **Set identified**, if **there exist other** observationally equivalent target parameters (while at least another is not)
- Without further assumptions, **selection on observables** is **not falsifiable**
  - Heckman, Ichimura, and Todd (1998) discusses a setting in which it is falsifiable