Lecture Note 13: Partial Identification and Bounds

Identification

Let's go back to the general extremum estimator theory of LN2 and 3. Recall that a key condition for consistency of the extremum estimator was condition (i): $Q(\theta)$ is uniquely maximized at θ_0.

We said that this is related to identification of the parameter θ_0.More generally, we can think of identification in the following way:

Suppose we have a random sample of $z_i \overset{\text{iid}}{\sim} P$, where P is a probability distribution in some set of distributions $\mathcal{P}$. In large samples, we can learn P. Does knowledge of P let us know what the value of θ is? If so, we say that θ is (point) identified.

In other words, it is possible to learn θ from observation of z. The condition (i) cannot hold without point identification, but also requires that the criterion function be a "good" one.

You should think of θ in a general way: it could be a parameter of a parametric model, but more generally it could be any feature of the distribution of z that is of particular interest.

Example: recall the probit model without variance normalization:

$$Pr(y_i = 1|x_i) = \Phi(x_i'\beta/\sigma).$$

If we think of $z_i = (x_i, y_i)$, then $P(z_i)$ is the joint distribution of (x_i, y_i), which can be expressed as the marginal distribution of x_i (which we leave unspecified) and the conditional distribution of y_i given x_i is the probit model above. So the set $\mathcal{P}$ consists of all joint distributions satisfying: $P_x(x_i)$ is any distribution, and $P(y_i|x_i)$ is a distribution of the form $Pr(y_i = 1|x_i) = \Phi(x_i'\beta/\sigma)$.

In large samples we can learn the marginal distribution of x_i and the conditional distribution $Pr(y_i = 1|x_i)$. But this does not pin down β and σ, since

$$\Phi(x_i'\beta/\sigma) = \Phi(x_i'\tilde{\beta}/\tilde{\sigma}),$$

where $\tilde{\beta} = \beta \cdot c$ and $\tilde{\sigma} = \sigma \cdot c$.

Usually, only β/σ is of economic interest, so the normalization $\sigma = 1$ is reasonable, and there is no identification problem.

In other cases, however, meaningful economic parameters may fail to be identified by the data.

Partial Identification with Censored Data

This discussion is based on the book *Identification Problems in the Social Sciences* by Charles Manski.

Suppose we are interested in a linear regression model

$$E[y|x] = x'\beta.$$

(Since we are focusing on identification, I am dropping the i subscript.)

However, we do not always observe y. Let $T = 1$ if y is observed, and $T = 0$ if y is not observed. We always observe x and T.

The data can be thought of as $z = (x, T, y \cdot T)$. From a large random sample, we can learn the marginal distribution $P(x, T)$, and the conditional distribution $P(y|x, T = 1)$. From the conditional distribution, we get $E[y|x, T = 1]$.

Define $\epsilon = y - E[y|x] = y - x'\beta$. So

$$y = x'\beta + \epsilon, \quad \text{where } E[\epsilon|x] = 0.$$

So

$$\begin{aligned} E[y|x, T = 1] &= E[x'\beta + \epsilon|x, T = 1] \\ &= E[x'\beta|x, T = 1] + E[\epsilon|x, T = 1] \\ &= x'\beta + E[\epsilon|x, T = 1] \end{aligned}$$

Now, suppose $\epsilon \perp T|x$, where $\perp$ mean "independent of."

Then $E[\epsilon|x, T = 1] = E[\epsilon|x] = 0$, so

$$E[y|x, T = 1] = x'\beta.$$

We can learn the left hand side, so provided there is some variation in x, we can infer β.

However, the assumption that ϵ is independent of T may be unreasonable. Essentially, it says that y and T are conditionally independent.

Suppose instead we make no assumptions about how ϵ and T are related. Then

$$E[y|x, T = 1] = x'\beta + h(x),$$

where $h(x) \equiv E[\epsilon|x, T = 1]$ is an function of x.

Then β is <u>not identified</u>. There may be different pairs $(\beta, h(\cdot))$ that correspond to $E[y|x, T = 1]$.

Sometimes, researchers will make specific parametric assumptions about ϵ and its relationship to T. These assumptions will put more specific structure on the function h (often making it nonlinear), and it may be the case that β is then formally identified.

Typically, however, this type of identification is sensitive to the parametric assumptions on ϵ (which restrict $h(\cdot)$), and the assumed linearity of $E[y|x]$.

Since parametric assumptions are typically made for convenience, rather than being strongly motivated by economic theory, it can be dangerous to rely too heavily on parametric assumptions for identification. It is a good idea to start by trying to establish identification without functional form assumptions. If identification fails without parametric restrictions, then your empirical results may be quite sensitive to the assumptions.

Returning to the censored regression example, let us remove the linearity assumption (that $E[y|x] = x'\beta$). Now, we regard $E[y|x]$ as an unknown function of x, and ask: what can we say about this function given knowledge of the marginal distribution of x, T and the conditional distribution of y given x and $T = 1$?

We can write

$$E[y|x] = E[y|x, T = 1]P(T = 1|x) + E[y|x, T = 0]P(T = 0|x).$$

The data identify $E[y|x, T = 1]$ and the two probabilities $P(T = 1|x)$ and $P(T = 0|x)$. But we cannot learn $E[y|x, T = 0]$, since we never observe y when $T = 0$. So, in the absence of further assumptions or restrictions, we cannot learn $E[y|x]$.

Suppose we assume

$$y \perp T|x.$$

Then $E[y|x, T = 1] = E[y|x, T = 0] = E[y|x]$. So we can learn $E[y|x]$, and this does not require linearity of the expectation. But, similar to the linear case, this requires a strong assumption, that the missing-data indicator is conditionally independent of the outcome.

Drop the independence assumption, but suppose that y is binary. Then $E[y|x, T]$ is bounded below by 0 and bounded above by 1. So

$$E[y|x, T = 1]P(T = 1|x) \leq E[y|x]$$

and

$$E[y|x] \leq E[y|x, T = 1]P(T = 1|x) + P(T = 0|x).$$

So, while $E[y|x]$ is not (point) identified, we can say that it is between two functions:

$$E[y|x, T = 1]P(T = 1|x)$$

and

$$E[y|x, T = 1]P(T = 1|x) + P(T = 0|x).$$

So for any given x, we can identify a *set* of possible values for $E[y|x]$ that are consistent with the observable data. We say that $E[y|x]$ is partially-identified. If $P(T = 0|x)$ is small, then these "bounds" might provide some useful information.

Manski and Tamer (2002): Interval-censored data

In some data sets, some variables are interval-censored. For example, some surveys define ranges of income (e.g. $[0, 20000], [20001, 50000], [50001, 100000], [100001, \infty)$). The survey only records which income category the individual falls into.

Notation: we have random variables y, x, v, v_0, v_1. The variable x is a $k \times 1$ random vector, the others are all scalar. In addition

$$v_0 \leq v \leq v_1.$$

We observe y, x, v_0, v_1, but not v.

Problem: based on knowledge of $P(y, x, v_0, v_1)$, what can we learn about:

$$\begin{array}{ll} (a) & E[y|x, v] \\ (b) & E[v|x] \end{array}$$

Assumptions IMMI:

Interval (I): $P(v_0 \leq v \leq v_1) = 1$.

Monotonicity (M): $E[y|x, v]$ exists and is weakly increasing in v.

Mean Independence (MI): $E[y|x, v, v_0, v_1] = E[y|x, v]$.

Proposition: Let assumption I hold. Then

$$E[v_0|x] \leq E[v|x] \leq E[v_1|x].$$

(Proof is obvious.)

Proposition: Let assumptions IMMI hold. Let $\gamma \in \mathbb{R}$. Then

$$\sup_{v_1 \leq \gamma} E[y|x, v_0, v_1] \leq E[y|x, v = \gamma] \leq \inf_{v_0 \geq \gamma} E[y|x, v_0, v_1].$$

Proof: We'll prove the lower bound (the upper bound argument is analogous). By the law of iterated expectations and assumption MI,

$$\begin{aligned} E[y|x, v_0, v_1] &= \int E[y|x, v, v_0, v_1] dP(v|x, v_0, v_1) \\ &= \int E[y|x, v] dP(v|x, v_0, v_1) \end{aligned}$$

By assumption I and M, for any $\gamma_0 \leq \gamma_1$,

$$E[y|x, v = \gamma_0] \leq \int E[y|x, v] dP(v|x, v_0 = \gamma_0, v_1 = \gamma_1) \leq E[y|x, v = \gamma_1].$$

Therefore,

$$E[y|x, v = \gamma_0] \leq E[y|x, v_0 = \gamma_0, v_1 = \gamma_1] \leq E[y|x, v = \gamma_1].$$

Now, let γ_1 be any value satisfying $\gamma_1 \leq \gamma$. By assumption M,

$$E[y|x, v = \gamma_1] \leq E[y|x, v = \gamma].$$

And by the argument above,

$$E[y|x, v_0 = \gamma_0, v_1 = \gamma_1] \leq E[y|x, v = \gamma].$$

Since this holds for every $\gamma_1 \leq \gamma$, it also holds that

$$\sup_{v_1 \leq \gamma} E[y|x, v_0, v_1] \leq E[y|x, v = \gamma].$$

$\square$

Estimation: Suppose we have estimators for $E[v_0|x]$, $E[v_1|x]$, and $E[y|x, v_0, v_1]$, call then $E_n[v_0|x]$, $E_n[v_1|x]$, and $E_n[y|x, v_0, v_1]$.

For concreteness, suppose that x, v_0, v_1 have discrete distributions. Then natural choices are the empirical conditional averages:

$$E_n[v_0|x] = \frac{\sum_{i=1}^n v_{0i} 1(x_i = x)}{\sum_{i=1}^n 1(x_i = x)},$$

$$E_n[v_1|x] = \frac{\sum_{i=1}^n v_{1i} 1(x_i = x)}{\sum_{i=1}^n 1(x_i = x)},$$

and

$$E_n[y|x, v_0, v_1] = \frac{\sum_{i=1}^n y_i 1(x_i = x, v_{0i} = v_0, v_{1i} = v_1)}{\sum_{i=1}^n 1(x_i = x, v_{0i} = v_0, v_{1i} = v_1)}.$$

If x, v_0, or v_1 have continuous distributions, however, we may need to use more complicated estimators.

To estimate the bound for $E[v|x]$, we can just use

$$\big[E_n[v_0|x], \quad E_n[v_1|x]\big]$$

Provided that $E_n[v_0|x]$ and $E_n[v_1|x]$ are consistent, then the endpoints will be estimated consistently.

For the estimation of $E[y|x,v]$ we need that $E[y|x,v_0,v_1]$ is uniformly consistent over the support of x, v_0, v_1. Then

$$\sup_{v_1 \le \gamma} E_n[y|x,v_0,v_1] \xrightarrow{p} \sup_{v_1 \le \gamma} E[y|x,v_0,v_1],$$

and similarly for the upper bound.

Parametric Regression Model

Let us add a further assumption, that the regression function $E[y|x,v]$ has a parametric form:

$$E[y|x,v] = f(x,v,\theta),$$

where f is a known function and $\theta \in \Theta$, a finite-dimensional parameter space. We keep assumptions I and MI, and replace assumption M with

Parametric Regression (PR): For each x and each $\theta \in \Theta$, $f(x,v,\theta)$ is weakly increasing in v.

Then, in general, it is not possible to point-identify θ. The set of parameter values consistent with the observed data distribution can be defined as follows: let

$$V(\theta) \equiv \{(x,v_0,v_1) : f(x,v_1,\theta) < E[y|x,v_0,v_1] \text{ or } E[y|x,v_0,v_1] < f(x,v_0,\theta)\},$$

$$\Theta^* \equiv \{\theta \in \Theta : P[V(\theta)] = 0\}.$$

In general, the identified set Θ^* is not a singleton, nor is it easy to give a simple characterization of it. Nevertheless, we could still try to construct a set-valued estimator for Θ^*.

Let

$$\eta(x,v_0,v_1) \equiv E[y|x,v_0,v_1],$$

$$g_1(\theta,x,v_0,v_1) \equiv 1(f(x,v_1,\theta) < \eta(x,v_0,v_1)),$$

$$g_0(\theta,x,v_0,v_1) \equiv 1(\eta(x,v_0,v_1) < f(x,v_0,\theta)).$$

Proposition Consider the minimization problem

$$\min_{\theta \in \Theta} Q(\theta,\eta),$$

where

$$Q(\theta,\eta) = \int g_1(\theta,x,v_0,v_1)[f(x,v_1,\theta)-\eta(x,v_0,v_1)]^2 + g_0(\theta,x,v_0,v_1)[f(x,v_0,\theta)-\eta(x,v_0,v_1)]^2 dP(x,v_0,v_1).$$

Every $\theta \in \Theta^*$ solves this problem, and no θ that is not in Θ^* solves this problem.

Proof Let $\theta \in \Theta^*$. Then, for all x, v_0, v_1 in the support of $P(x, v_0, v_1)$, we must have

$$f(x, v_1, \theta) \geq E[y|x, v_0, v_1],$$

implying $g_1(\theta, x, v_0, v_1) = 0$. Likewise, we must have $f(x, v_0, \theta) \leq E[y|x, v_0, v_1]$, so $g_0(\theta, x, v_0, v_1) = 0$. So $Q(\theta, \eta) = 0$ for all $\theta \in \Theta^*$.

Now consider a θ not in Θ^*. Then

$$g_1(\theta, x, v_0, v_1)[f(x, v_1, \theta) - \eta(x, v_0, v_1)]^2 + g_0(\theta, x, v_0, v_1)[f(x, v_0, \theta) - \eta(x, v_0, v_1)]^2 > 0$$

for all $(x, v_0, v_1) \in V(\theta)$. Since $P[V(\theta)] > 0$ for $\theta \notin \Theta^*$, $Q(\theta, \eta) > 0$.
$\square$

This suggests that we can construct an estimator by solving an empirical version of this minimization problem.

Let $\eta_n(x, v_0, v_1)$ be an estimator of $\eta(x, v_0, v_1) = E[y|x, v_0, v_1]$. Let

$$g_{n1}(\theta, x, v_0, v_1) = 1(f(x, v_1, \theta) < \eta_n(x, v_0, v_1)),$$

and similarly for g_{n0}. Let the estimator $\hat{\Theta}$ be the set of all solutions to

$$\min_{\theta \in \Theta} Q_n(\theta, \eta_n),$$

where

$$Q_n(\theta, \eta_n) = \frac{1}{n} \sum_{i=1}^{n} g_{n1}(\theta, x_i, v_{0i}, v_{1i})[f(x_i, v_{1i}, \theta) - \eta_n(x_i, v_{0i}, v_{1i})]^2 + g_{n0}(\theta, x_i, v_{0i}, v_{1i})[f(x_i, v_{0i}, \theta) - \eta_n(x_i, v_{0i}, v_{1i})]^2.$$

Manski and Tamer show that this estimator consistently estimates the identified set Θ^*. However, developing an asymptotic distribution theory for this type of estimator is much more difficult than for a standard point estimator. Some results are given in Chernozhukov, Hong, and Tamer (2004). This is an area still under active development.