Lecture 12. Set Estimation and Inference in Moment Condition Models

- The notes cover the paper "Estimation and Inference on Parameter Sets in Econometric Models" by Chernozhukov, Hong, and Tamer (Econometrica, 2007).

- Consider a population criterion function $Q(\theta) \geqslant 0$. An economic model $\theta \in \Theta \subset \mathbb{R}^k$ passes empirical restrictions if $Q(\theta) = 0$. Denote the set of parameters that pass these restrictions as Θ_I. That is,

$$\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\} = \arg\min_{\theta \in \Theta} Q(\theta).$$

  Θ_I will be called the identified set.

- $Q(\theta)$ typically embodies moment restrictions arising from economic theory and other considerations. In particular, **moment inequality and equality** restrictions lead to objective functions $Q(\theta)$ of GMM type. Concrete examples follow.

- Our goal is to provide set estimators $\widehat{\Theta}_I$ based on $Q_n(\theta)$ that are

  1) consistent,

  2) converge to Θ_I at the fastest rate,

  3) have confidence interval property, and

  4) computationally tractable.

- These results extend the classical theory for the case when Θ_I is a singleton.

Model I: Moment Inequality Problems

The moment restrictions are computed with respect to the probability law P of the data and take the form

$$E_P[m_i(\theta)] \leqslant 0, \tag{1}$$

where $m_i(\theta) = m(\theta, w_i)$ is a vector of moment functions parameterized by θ and determined by a vector of real random variables w_i. Therefore the set of parameters θ that pass the testable restrictions is given by

$$\Theta_I = \{\theta \in \Theta : E_P[m_i(\theta)] \leqslant 0\}.$$

It is interesting to comment on the structure of the set Θ_I in this model. When the moment functions are linear in parameters, the set Θ_I is given by an intersection of linear half-spaces and could be a triangle, trapezoid, or a polyhedron, as in Examples 1 and 2 introduced below. When moment functions are non-linear, the set Θ_I is given by an intersection of nonlinear half-spaces which boundaries are defined by nonlinear manifolds.

The set Θ_I can be characterized as the set of minimizers of the criterion function

$$Q(\theta) := \|E_P[m_i(\theta)]_+ W^{1/2}(\theta)\|^2, ^{\text{a}} \tag{2}$$

where $W(\theta)$ is a continuous and positive definite for each $\theta \in \Theta$. Therefore, the inference on Θ_I may be based on the empirical analog of Q:

$$\begin{aligned} Q_n(\theta) &:= \|E_n[m_i(\theta)]'_+ W_n^{1/2}(\theta)\|^2, \\ E_n[m_i(\theta)] &= \frac{1}{n}\sum_{i=1}^{n} m_i(\theta), \end{aligned} \tag{3}$$

where $W_n(\theta)$ is a uniformly consistent estimate of $W(\theta)$. In applications $W_n(\theta)$ is often taken to be an identity matrix or chosen to weight the individual empirical moments by estimates of inverses of their individual variances.

The modified objective function

$$\widetilde{Q}_n(\theta) = Q_n(\theta) - \inf_{\theta' \in \Theta} Q_n(\theta')$$

is another useful analog of $Q(\theta)$ for inference.

This analog mimics quasi-likelihood ratio statistic more closely, and thus improves power, when $\inf_{\theta' \in \Theta} Q_n(\theta') \neq 0$ in finite samples.

Example 1: Interval Data.

The first simplest example is motivated by missing data problems, where Y is the unobserved real random variable bracketed below by Y_1 and above by Y_2.

Let $\{(Y_{1i}, Y_{2i}), i = 1, ..., n\}$, be an i.i.d. sequence of real random variables with law P on $\mathbb{R}^d$. The parameter of interest $\theta = E_P[Y]$ is known to satisfy the restriction

$$E_P[Y_1] \leqslant \theta \leqslant E_P[Y_2].$$

The identified set is therefore given by an interval

$$\Theta_I = \{\theta : E_P[Y_1] \leqslant \theta \leqslant E_P[Y_2]\}.$$

This example falls in the moment-inequality framework with moment function

$$m_i(\theta) = (Y_{1i} - \theta, \theta - Y_{2i})'.$$

Therefore, set Θ_I can be characterized as the set of minimizers of

$$Q(\theta) = \|E_n[m_i(\theta)]\|_+^2 = (E_P[Y_{1i}]-\theta)_+^2+(E_P[Y_{2i}]-\theta)_-^2,$$

with the sample analog

$$Q_n(\theta) = (E_n[Y_{1i}] - \theta)_+^2 + (E_n[Y_{2i}] - \theta)_-^2.$$

Example 2. Interval Outcomes in Regression Models. A regression generalization of the previous basic example is immediate. Suppose a regressor vector X_i is available, and the conditional mean of unobserved Y_i is modelled using linear function $X_i'\theta$. The parameters of this function can be bounded using inequality

$$E_P[Y_{1i}|X_i] \leqslant X_i'\theta \leqslant E_P[Y_{2i}|X_i].$$

These conditional restrictions can be converted to unconditional ones by considering inequalities

$$E_P[Y_{1i}Z_i] \leqslant \theta' E_P[X_i Z_i] \leqslant E_P[Y_{2i}Z_i],$$

where Z_i is a vector of positive transformations of X_i, for instance, $Z_i = \{1(X_i \in \mathcal{X}_j), j = 1, ..., J\}$, for a suitable collection of regions $\mathcal{X}_j$. The identified set is therefore given by an intersection of linear half-spaces in $\mathbb{R}^d$.

This examples also falls in the moment inequality framework, with moment function given by

$$m_i(\theta) = ((Y_{1i} - \theta' X_i)Z_i', -(Y_{2i} - \theta' X_i)Z_i')'.$$

In auction analysis, the bracketing of the latent response – bidder's valuation – by functions of observed bids is very natural and occurs in a variety of settings, cf. Haile and Tamer (2003). Analogous situations occur in income surveys.

Example 3. Optimal Choice of Economic Agents and Game Interactions

Another application of (1) is in the optimal choice behavior of firms and economic agents. Suppose that a firm can make two choices $D_i = 0$ or $D_i = 1$. Suppose that the profit of the firm from making choice D_i is given by $\pi(W_i, D_i, \theta) + U_i$, where U_i is a disturbance such that $E[U_i|X_i] = 0$, for X_i representing information available to make the decision, and W_i are various determinants of the firm's profit, some of which may be included in X_i. For example, W_i may include the actions of other firms that affect the firm's profit. From a revealed preference principle, the fact that the firm chooses D_i necessarily implies that

$$E[\pi(W_i, D_i, \theta) | X_i] \geqslant E[\pi(W_i, 1 - D_i, \theta) | X_i].$$

Therefore, we can take the moment condition in (1) to be

$$m_i(\theta) = (\pi(W_i, 1 - D_i, \theta) - \pi(W_i, D_i, \theta)) Z_i \leqslant 0,$$

where Z_i is the set of positive instrumental variables defined as positive transformations of X_i, for instance, defined as in the previous example.

This simple example highlights the structure of empirically testable restrictions arising from the optimizing behavior of firms and economic agents.

These testable restrictions are given in the form of moment inequality conditions. It could be noted that this simple example also allows for game-theoretic interactions among economic agents.

The moment inequality conditions of the above kind are ubiquitous and are known to arise in (more realistic) dynamic settings. See Ciliberto and Tamer (2003) and Ryan (2005) for more details.

Similar principles are used in Blundell, Browning, and Crawford (2005) to analyze **bounds on demand functions**. Related ideas also appear in an area of **stochastic revealed preference analysis**, e.g. see Varian (1984) and McFadden (2005).

Example 4. Based on Hansen, Heaton, and Luttmer (1995):

The price vector P_t of securities with payoff S_t satisfy

$$P_t \geqslant E[Q_t(\theta)S_t|X_t],$$

when the short-sale constraints are present, where $Q_t(\theta)$ is the pricing kernel. An asset pricing model θ provides a pricing kernel $Q_t(\theta)$. The model can be tested using moment inequality conditions with moment function

$$m_t(\theta) := -(P_t - Q_t(\theta)S_t)Z_t,$$

for Z_t equal to a collection of positive transforms of X_t. The set of models θ that provide adequate pricing kernels are given by

$$\Theta_I = \{\theta \in \Theta : E_P[m_t(\theta)] \leqslant 0\}.$$

The above illustrates how market frictions introduce moment inequalities in the asset pricing framework.

Hansen, Heaton, and Luttmer (1995) actually use these equations as a starting point for a derivation of a region of feasible (possibly conditional) means and variances for the pricing kernel, subject to volatility and specification error bounds. They also provide a consistent estimator but don't consider inference. The results of this paper allow to construct both estimators and confidence regions for these regions.

Model II: Moment Equalities

Moment equalities are more traditional in empirical analysis. The economic models, indexed by θ, are assumed to satisfy the set of testable restrictions given by moment equalities:

$$E_P[m_i(\theta)] = 0, \text{ that is } \Theta_I = \{\theta \in \Theta : E_P[m_i(\theta)] = 0\}. \tag{4}$$

When the moment functions are linear in parameters, the set Θ_I is either a point or a hyperplane intersected with parameter space Θ. When moment functions are non-linear, the set Θ_I is typically a manifold, which also includes the case of isolated points (a zero-dimensional manifold).

The set Θ_I can be characterized as the set of minimizers of the conventional generalized method of moments function

$$Q(\theta) := \|E_P[m_i(\theta)]' W^{1/2}(\theta)\|^2, \tag{5}$$

where $W(\theta)$ is a continuous and positive-definite matrix for each $\theta \in \Theta$. The inference on Θ_I may be based on the conventional generalized-method-of-moments function

$$\begin{aligned} Q_n(\theta) &:= \|E_n[m_i(\theta)]' W_n^{1/2}(\theta)\|^2, \\ E_n[m_i(\theta)] &= \frac{1}{n}\sum_{i=1}^{n} m_i(\theta), \end{aligned} \quad (6)$$

where $W_n(\theta)$ is a uniformly consistent estimate of $W(\theta)$. In applications, $W_n(\theta)$ can be an identity matrix or an estimate of the inverse of the asymptotic covariance matrix of empirical moment functions.

In many situations, we can also use the modified objective function for inference:

$$\widetilde{Q}_n(\theta) = Q_n(\theta) - \inf_{\theta' \in \Theta} Q_n(\theta').$$

This modification is useful in cases where Q_n does not attain value 0 in finite samples. In such cases, using the modified objective function typically leads to power improvements, as is well-known in point-identified cases.

Example 5. Structural Simultaneous Equations. Consider the structural instrumental variable estimation of returns to schooling. Suppose that we are interested in the following example where potential income Y is related to education E through a flexible, quadratic functional form,

$$Y = \theta_0 + \theta_1 E + \theta_2 E^2 + \epsilon = X'\theta + \epsilon,$$

for

$$\theta = (\theta_0, \theta_1, \theta_2) \text{ and } X = (1, E, E^2)'.$$

Although parsimonious, this simple model is not point identified in the presence of the standard quarter-of-birth instrument suggested in Angrist and Krueger (1992).[a] In the absence of point identification, all of the parameter values θ consistent with the instrumental orthogonality

[a]The instrument is the indicator of the first quarter of birth. Sometimes the indicators of other quarters of birth are used as instruments. However, these instruments are not correlated with education (correlation is extremely small) and thus bring no additional identification information.

restriction $E[(Y - \theta' X)Z] = 0$ are of interest for purposes of economic analysis. Phillips (1989) develops a number of related examples.

Similar partial identification problems arise in nonlinear moment and instrumental variables problems, see e.g. Demidenko (2000) and Chernozhukov and Hansen (2005). In Chernozhukov and Hansen (2005), the parameters θ of the structural quantile functions for returns to schooling satisfy the restrictions:

$$E[(\tau - 1(Y \leqslant X'\theta))Z] = 0,$$

where $\tau \in (0, 1)$ is the quantile of interest. This is an example of a nonlinear instrumental variable model, where the identification region, in the absence of point identification, would generally be given by a nonlinear manifold. Chernozhukov and Hansen (2004) and Chernozhukov, Hansen, and Jansson (2005) analyze an empirical returns-to-schooling example and a structural demand example where such situations arise.

Estimation

A lower contour set $C_n(c)$ of level c of the sample criterion function Q_n is defined as

$$C_n(c) := \{\theta : Q_n(\theta) \leqslant c/n\}.$$

The estimator $\widehat{\Theta}_I$ will take the form:

$$\widehat{\Theta}_I = C_n(\widehat{c}).$$

For estimation purposes, we will need that the level $c = \widehat{c}$, which could be data-dependent, diverges very slowly to infinity; for concreteness, we could set $\widehat{c} = \ln n$. Divergence would generally be needed to cover general cases.

In Examples like 1-3 no growth condition would typically be needed, and for estimation purposes – though not for inference purposes – one could even set $\widehat{c} = 0$ or to any other positive constant. For example, in Example 1, $\widehat{c} = 0$ would give us the estimate $C_n(0) = [E_n[Y_1], E_n[Y_2]]$.

The analysis of the rates of convergence and consistency will make use of the Hausdorff distance between sets, which is defined as

$$d_H(A, B) := \max \left[\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \right],$$
$$d(b, A) := \inf_{a \in A} \|a - b\|,$$

and $d_H(A, B) := \infty$ if either A or B is empty.

The general consistency (**Theorem 3.1**) result

$$d_H(C_n(\widehat{c}), \Theta_I) \to_p 0,$$

follows from mild assumptions on Q_n.

Specifically, we require the uniform convergence of the sample function Q_n to the limit continuous function Q over the compact parameter space Θ, where the rate of convergence over set Θ_I is $1/n$.

The rate of convergence (**Theorems 3.1 and**

3.2) is:

$$d_H(C_n(\widehat{c}), \Theta_I) = O_p(\sqrt{\max(\widehat{c}, 1)/n}),$$

which is very close to $1/\sqrt{n}$ rate of convergence (and is exactly $1/\sqrt{n}$ in many moment inequality examples, e.g. Example 1, 2, and 3).

The assumption required for this is an approximately quadratic behavior of $Q_n(\theta)$ over suitable neighborhoods of Θ_I. To get the sharp rate we need degenerate interior asymptotics – Q_n must vanish on appropriate contractions of the identified set Θ_I.

Confidence Region for Θ_I:

If we want a confidence region for Θ_I, simply set

$$CR = C_n(\widehat{c}) = \{\theta \in \Theta : nQ_n(\theta) \leqslant \widehat{c}\}$$

where $\widehat{c}$ is an estimate of $c(\alpha)$ – the α-quantile of

$$\mathcal{C}_n = \sup_{\theta \in \Theta_I} nQ_n(\theta)$$

Clearly, event $\mathcal{C}_n \leqslant c(\alpha)$ is equivalent to $\Theta_I \subset C_n(c(\alpha))$, provided Θ_I is compact. Then under conditions of **Theorem 3.3**

$$\liminf_n P(\Theta_I \subset CR) \geqslant \alpha.$$

The above confidence region is simply the inversion of the quasi-likelihood ratio test of the null hypothesis that $\sup_{\theta \in \Theta_I} Q(\theta) = 0$.

Critical value $\widehat{c}$ can be obtained by

(a) a generic subsampling method (problem independent) applied to a feasible version of $\mathcal{C}_n$ (with Θ_I replaced by an estimate),

(b) by simulating the limit distribution (better, but problem specific).

Subsampling is constructed in **Theorem 3.4**. Limit distributions are constructed in **Theorem 4.4** for moment condition problems.

It should be mentioned that naive (canonical) bootstrap does not work for this problem.

Confidence Region for θ_0:

Suppose there is a true parameter θ_0 and we want to construct a confidence region for this parameter,[a] not the confidence region for Θ_I. If so, simply set

$$CR = \{\theta \in \Theta : nQ_n(\theta) \leqslant \widehat{c}(\theta)\},$$

where $\widehat{c}(\theta)$ is an estimate of $c(\alpha, \theta)$ – the α-quantile of

$$\mathcal{C}_n(\theta) = nQ_n(\theta).$$

Then under conditions of **Theorem 5.1**

$$\lim_n P(\theta_0 \in CR) \geqslant \alpha.$$

The above confidence region is simply the inversion of the quasi-likelihood ratio test of the null hypothesis that $Q(\theta) = 0$. We simply collect

[a] In economic modelling, it is often hard to make the case that there is the true model (parameter), since economic models are typically highly simplified constructs that aim to explain or fit certain (but not all!) features of the real world. We therefore might prefer confidence regions for Θ_I.

all values θ where this hypothesis cannot be rejected, and you get CR above.

Critical value $\widehat{c}(\theta)$ can be obtained by

(a) a generic subsampling method (problem independent),

(b) by simulating the limit distribution (better, but problem specific).

Limit distributions are constructed for moment condition problems.

It should be mentioned that naive (canonical) bootstrap does not work for this problem.

Illustration with Moment Inequalities. Suppose that

$$\sqrt{n}\Big(E_n[m_i(\theta)] - E_P[m_i(\theta)]\Big) \Rightarrow \Delta(\theta),$$

where $\Delta(\theta)$ is a mean zero Gaussian process with a.s. continuous paths. Then we show that (1)

$$\mathcal{C}_n(\theta) \to_d \mathcal{C}(\theta) := \|(\Delta(\theta) + \xi(\theta))' W^{1/2}(\theta)\|_+^2,$$

where $\xi(\theta) := (\xi_j(\theta), j \leqslant J)$ with

$$\xi_j(\theta) = -\infty \text{ if } E_P[m_{ij}(\theta)] < 0$$

and

$$\xi_j(\theta) = 0 \text{ if } E_P[m_{ij}(\theta)] = 0.$$

(2)

$$\mathcal{C}_n \to_d \mathcal{C} = \sup_{\theta \in \Theta_I} \mathcal{C}(\theta).$$

If the data are i.i.d., we can estimate the quantiles of $\mathcal{C}(\theta)$ and $\mathcal{C}$ by simulating the distribution of the variables

$$\mathcal{C}_n^*(\theta) := \|(\Delta_n^*(\theta) + \widehat{\xi}(\theta))' W_n^{1/2}(\theta)\|_+^2,$$

and

$$\mathcal{C}_n^* := \sup_{\theta \in \widehat{\Theta}_I} \mathcal{C}_n^*(\theta),$$

where $\widehat{\xi}(\theta) := (\widehat{\xi}_j(\theta), j = 1, ..., J)'$ with

$$\widehat{\xi}_j(\theta) := -\infty \text{ if } E_n[m_{ij}(\theta)] \leqslant -c_j \sqrt{\log n/n},$$

and

$$\widehat{\xi}_j(\theta) := 0 \text{ if } E_n[m_{ij}(\theta)] > -c_j \sqrt{\log n/n},$$

for some positive constants $c_j > 0$. We simulate $\Delta_n^*(\theta)$ as

$$\Delta_n^*(\theta) := n^{-1/2} \sum_{i=1}^{n} [m_i(\theta) z_i^*],$$

and $(z_i^*, i \leqslant n)$ is a n-vector of i.i.d. $N(0,1)$ variables. Similarly, we can simulate $\Delta_n^*(\theta)$ using "the" bootstrap. For this purpose, we take $z_i^* = k_i^* - 1$ for each i, and $(k_i^*, i \leqslant n)$ is an n-vector of variables following the multinomial distribution with success probabilities $1/n$, defined over n-trials.

Interval Censored Regression: Empirical Monte-Carlo

- Y is wages and salaries (complete),
- X is education.
- Estimate the linear model:

$$E_\theta[Y|X] = \theta_0 + \theta_1 X$$

When Y is complete, the linear model is identified:

Figure 1: Point Identified Case: Subsampling vs χ^2 Ellipses

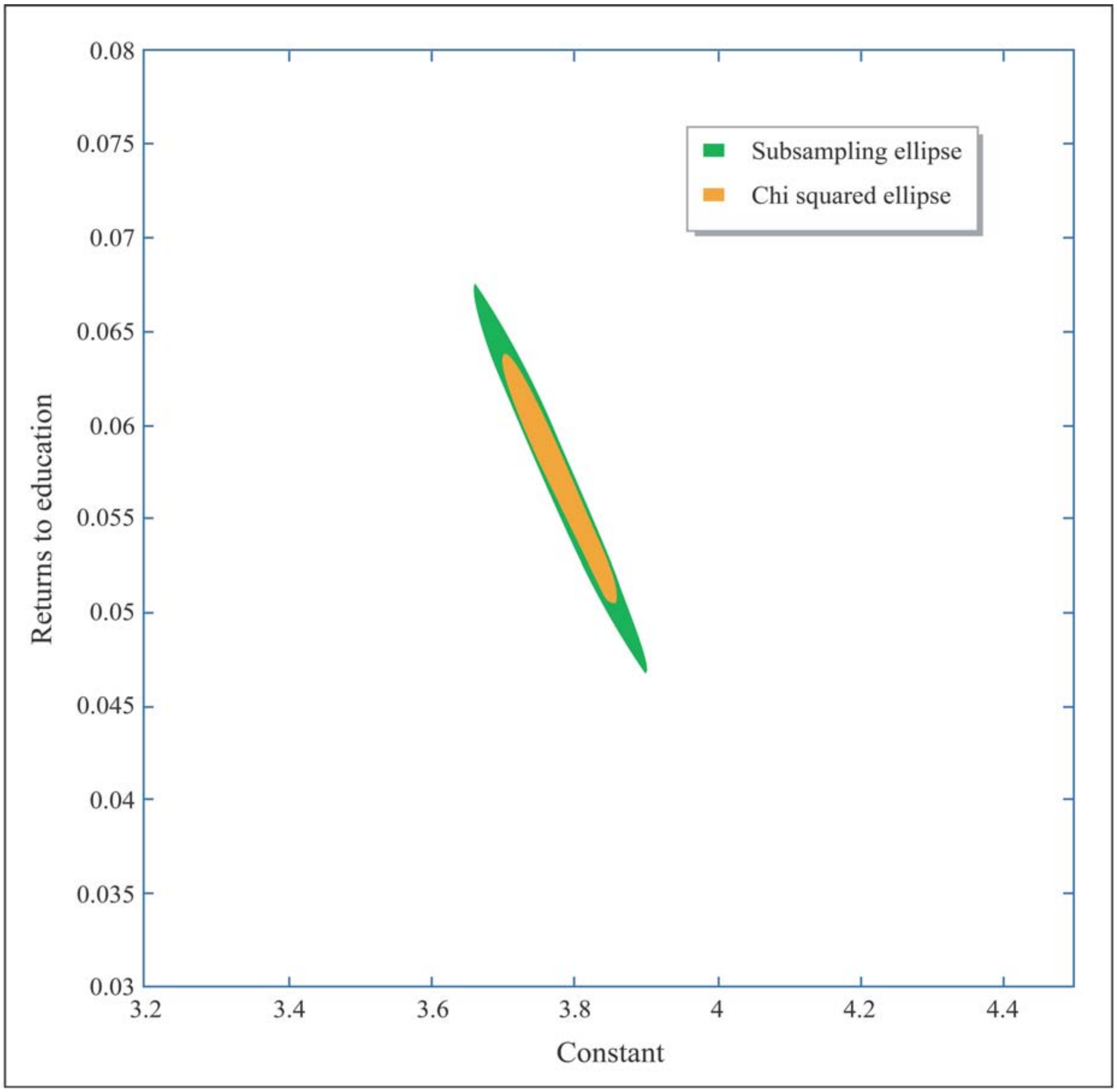

Image by MIT OpenCourseWare.

Now bracket Y into categories (in thousands):

```
[0,5], [5,7.5], [7.5,10], [10,12,5], [12.5,15],
[15,20], [20,25], [25,30], [30,35], [35,40],
[40,50], [50,60], [60,75], [75,100],
[100,150],[150,100000]
```

Figure 2: Θ_I vs $C_n(\widehat{c}_{.95})$. n=600, b=150.

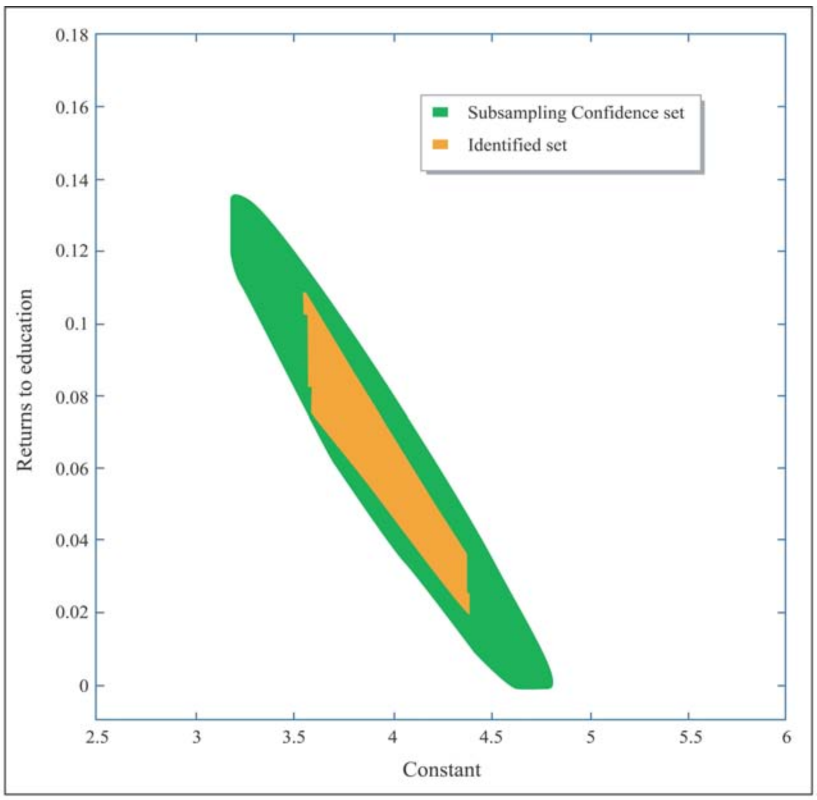

Image by MIT OpenCourseWare.

Figure 3: Θ_I vs $C_n(\widehat{c}_{.95})$. n=10000, b=2000.

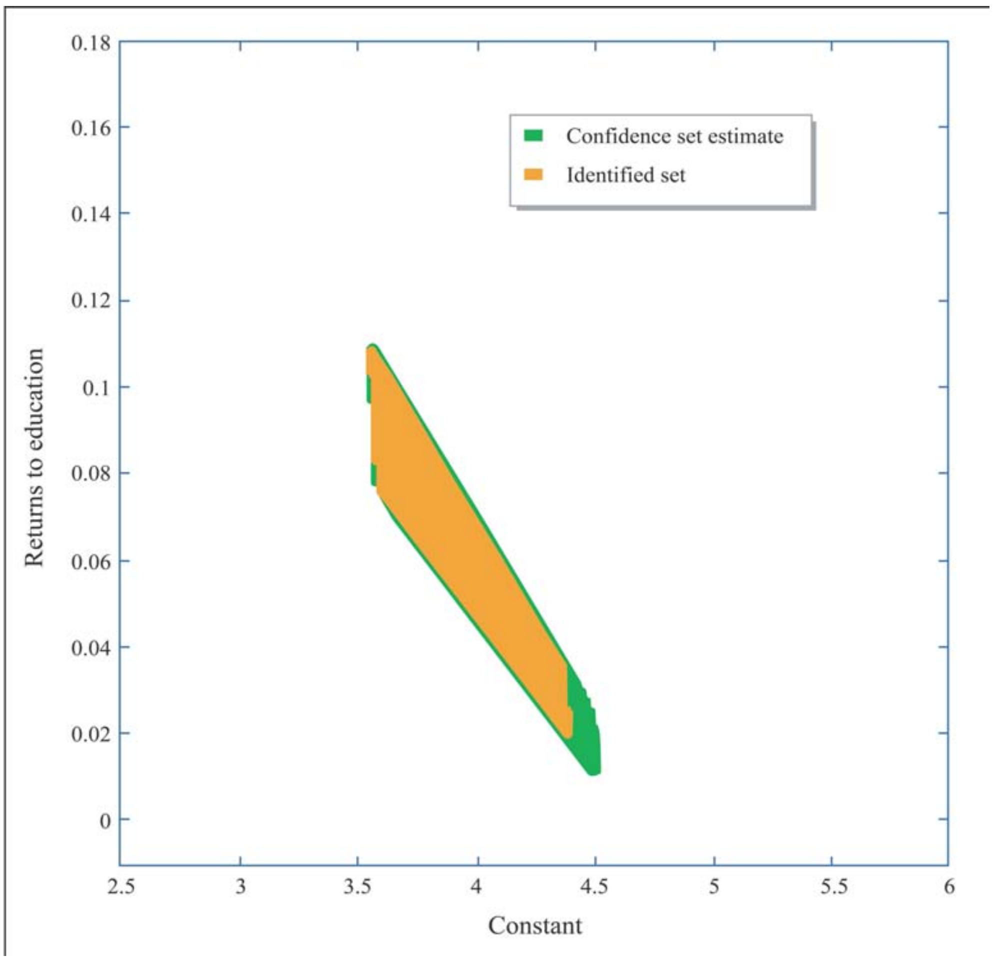

Image by MIT OpenCourseWare.

References

Angrist, J., and A. Krueger (1992): "Age at School Entry," *Journal of the American Statistical Associtation*, 57, 11–25.

Blundell, R., M. Browning, and I. Crawford (2005): "Best Nonparametric Bounds on Demand Responses," Working Paper, University College London.

Chernozhukov, V., and C. Hansen (2004): "Instrumental Quantile Regression," Working Paper, MIT.

——— (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 81, 806–839.

Chernozhukov, V., C. Hansen, and M. Jansson (2005): "Finite-Sample Inference for Quantile Regression Models," Working Paper, MIT.

Ciliberto, F., and E. Tamer (2003): "Market Structure and Multiple Equilibria in Airline Markets," Working Paper, Northwestern University.

Demidenko, E. (2000): "Is this the least squares estimate?," *Biometrika*, 87(2), 437–452.

Haile, P., and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1–51.

Hansen, L., J. Heaton, and E. G. Luttmer (1995): "Econometric Evaluation of Asset Pricing Models," *The Review of Finacial Studies*, 8, 237–274.

McFadden, D. L. (2005): "Revealed stochastic preference: a synthesis," *Economic Theory*, 26, 245–264.

Phillips, P. C. B. (1989): "Partially identified econometric models," *Econometric Theory*, 5(2), 181–240.

Ryan, S. P. (2005): "The Costs of Environmental Regulation in a Concentrated Industry," Working Paper, MIT.

Varian, H. R. (1984): "The nonparametric approach to production analysis," *Econometrica*, 52(3), 579–597.