Part 8: Partial Identification

Chris Conlon

Microeconometrics

June 28, 2017

Overview

This Lecture will cover (roughly) the following papers:
Theory:

- Parts of Manski's book
- Chernozhukov Hong and Tamer (2007)

Empirics:

- Haile and Tamer

Motivation

- Most of this course has been about what is the minimal set of assumptions we can place on our data in order to identify parameters of interest.
- Now we consider: suppose we are willing to place even fewer assumptions on our data but are willing to accept a set of θ that satisfy our restrictions rather than a single estimate $\hat{\theta}$.
- Reasons this might be a good idea:
  - Maybe we don't want to impose 'fully rational' behavior on agents.
  - Maybe we want to eliminate certain assumptions about functional forms, etc.
  - Maybe full solution approaches are computationally infeasible (such as complicated censoring, or dynamic games).

Censored Data

- Suppose that Y is subject to censoring such that it is not always observed.
- (Y, X, T) where T is a binary variable obtained by random sample.
- Y is only observed if $T = 1$, X, T always observed. (suppose $T = 1$ is being employed).
- In a large sample I can learn $P(X, T)$ and $P(Y|X, T = 1)$ where $P(\cdot)$ is the joint distribution.

$$\begin{aligned} P(Y|X) = P(Y|X, T = 1)P(T = 1|X) + \\ P(Y|X, T = 0)P(T = 0|X) \end{aligned}$$

- Challenge is that $P(Y|X, T = 0)$ is unobserved and we cannot learn it without more assumptions.

Some Ideas

Missing at Random:

- Assume that $Y \perp T | X$.

$$P(Y|X, T = 1) = P(Y|X, T = 0)$$

- $P(Y|X, T = 1)P(T = 1|X) = P(Y|X, T = 0)P(T = 1|X)$ so everything is known and $P(Y|X)$ is identified.

Suppose we know nothing...

Then $Y \in (-\infty, +\infty)$ so the conditional expectation is unbounded!

$E[Y|X] = E[Y|X, T = 1]P(T = 1|X) + E[Y|X, T = 0]P(T = 0|X)$

Now consider whether $Y \in B$ (some set).

$$\begin{aligned} E[Y \in B|X] = &E[Y \in B|X, T = 1]P(T = 1|X) \\ &+E[Y \in B|X, T = 0]P(T = 0|X) \end{aligned}$$

Consider instead the probability of being in the set which is always $\in [0, 1]$.

$$\begin{aligned} P[Y \in B|X] = &P[Y \in B|X, T = 1]P(T = 1|X) \\ &+P[Y \in B|X, T = 0]P(T = 0|X) \end{aligned}$$

$P[Y \in B|X, T = 0]$ is unknown, but it must be in $[0, 1]$

Suppose we know nothing...

We plug in $P[Y \in B|X, T = 0] = 0$ and $P[Y \in B|X, T = 0] = 1$:

$$P[Y \in B|X] \in (P[Y \in B|X, T = 1]P(T = 1|X),$$
$$P[Y \in B|X, T = 1]P(T = 1|X) + P(T = 0|X))$$

- The width of the interval is $P(T = 0|X)$ so the more data are missing, the less we know.
- We say that $P(Y \in B|X)$ is partially identified.
- This is the best we can do without further information bounds are sharp.

Add an Exclusion Restriction

- Break up X into (W, V) so that $P(Y|W,V) = P(Y|W)$ (it does not depend on V).
- This is like having V be an instrumental variable.
- Assume V takes on v_1, v_2 to make algebra easy.
- Now we have two restrictions:

$$
\begin{aligned}
P(Y \in B|W, V = v_1, T = 1)P(T = 1|W, V = v_1) &\leq \color{red}{P(Y \in B|W)} \\
P(Y \in B|W, V = v_2, T = 1)P(T = 1|W, V = v_2) &\leq \color{red}{P(Y \in B|W)} \\
\color{red}{P(Y \in B|W)} \leq P(Y \in B|W, V = v_1, T = 1)&P(T = 1|W, V = v_1) \\
&+P(T = 0|W, V = V_1) \\
\color{red}{P(Y \in B|W)} \leq P(Y \in B|W, V = v_2, T = 1)&P(T = 1|W, V = v_2) \\
&+P(T = 0|W, V = V_2)
\end{aligned}
$$

Add an Exclusion Restriction

Construct the greatest lower bound:

$$\max_j P(Y \in B|W, V = v_j, T = 1)P(T = 1|W, V = v_j)$$

Construct the least upper bound:

$$\min_j P(Y \in B|W, V = v_j, T = 1)P(T = 1|W, V = v_j) \\ + P(T = 0|W, V = v_j)$$

With a number of levels of the instrument this can reduce the width of the interval substantially.

Treatment Effects

- Recall for treatment effects we observed $Y(1)$ only for individuals for whom $T = 1$ and $Y(0)$ only for individuals for whom $T = 0$.
- $Y = Y(1) * T + Y(0) * (1 - T)$.
- We can learn $P(X, T)$ and $P(T|X)$ as well as $P(Y(1)|X, T = 1)$ and $P(Y(0)|X, T = 0)$.
- We do not learn the counterfactuals: $P(Y(1)|X, T = 0)$ and $P(Y(0)|X, T = 1)$.
- We are interested in $ATE(X) = E[Y(1)|X] - E[Y(0)|X]$. Write the following:

$$\begin{aligned} E[Y(1)|X] = E[Y(1)|X, T = 1]P(T = 1|X) + \\ {\color{red}E[Y(1)|X, T = 0]}P(T = 0|X) \end{aligned}$$

Can we put restrictions on $E[Y(1)|X, T = 0]$ other than $(-\infty, \infty)$?

Binary Outcomes

- If $Y(1), Y(0)$ are binary outcomes then $E[Y(1)|X, T=0] \in [0,1]$ and $E[Y(0)|X, T=1] \in [0,1]$.
- Upper Bound on $ATE(X)$

$$\begin{aligned} E[Y(1)|X,T=1]P(T=1|X) &+& P(T=0|X) \\ -E[Y(0)|X,T=0]P(T=0|X) && \end{aligned}$$

- Lower Bound on $ATE(X)$

$$\begin{aligned} E[Y(1)|X,T=1]P(T=1|X) &-& P(T=1|X) \\ -E[Y(0)|X,T=0]P(T=0|X) && \end{aligned}$$

- So width is $P(T=0|X) + P(T=1|X)$.
- We can also allow $P(Y(1) \in B|X) - P(Y(0) \in B|X)$ using the same argument.

More Assumptions/Tighter Bounds

By making additional assumptions we can improve the bounds:

- Unconfounded treatment assignment: $T \perp Y(1), Y(0)|X$.
  - $E[Y(1)|X, T = 1] = E[Y(1)|X, T = 0]$ and $E[Y(0)|X, T = 1] = E[Y(0)|X, T = 0]$
  - Leads to point identification of $ATE(x)$ even if Y unbounded.
- Ordered outcomes: $Y(1) \geq Y(0)$ for all individuals
- Roy Model: Individuals choose T leading to highest outcome
  - $Y(1) > Y(0) \rightarrow T = 1$
  - $Y(1) \leq Y(0) \rightarrow T = 0$

The Mixing Problem

- Drop X to make life easy.
- Recall that a randomized experiment identifies both $P(Y(0))$ and $P(Y(1))$. In large samples we can learn about the marginal distributions of $Y(0)$ and $Y(1)$.
- The randomized experiment tells us what happens if everyone gets $T = 0$ or everyone gets $T = 1$.
- Often we are interested in something else.
- Program is available to everyone but voluntary.
- We make treatment available only to some people, and doctors/caseworkers decide whether to give the treatment.
- This is known as the mixing problem.

Haile and Tamer (2003)

Ascending (English) Auctions

- Often modeled as button auction.
- Good for theory, not as good empirically.
- Jump bidding, non-bidding, not bidding your value, etc.
- Haile and Tamer investigate a bounds approach, which is elegant and likely to have application in other assymetric information problems.

Assumptions

- A1: Bidders do not bid above their valuation $b_{it} \leq u_{it}\ \forall i$.
- A2: Bidders do not let someone else win at a price they are willing to beat.

Note: this allows for jump bidding, bids $\neq$ valuations, etc.

Haile and Tamer (2003)

The assumptions imply that $b^{(i;n)} \leq u^{(i;n)}$

$$G_B^{(i;n)}(u) \geq F_U^{(i;n)}(u) \quad \forall i, u, n$$

We use the properties of order statistics for an iid sample of size n from distribution F.

$$F_{(i:n)}(s) = \frac{n!}{(n-i)!(i-1)!} \int_0^{F(s)} t^{i-1}(1-t)^{n-1} dt$$

This is increasing in $F(\cdot) \in [0,1]$ so that $F^{(i:n)}(s)$ uniquely determines values for $F(s) \forall s$.
Then we can define an implicit mapping where $\phi : F^{(i:n)} \rightarrow F(s)$

$$\begin{aligned} H &= \frac{n!}{(n-i)!(i-1)!} \int_0^{\phi} t^{i-1}(1-t)^{n-1} dt \quad H \in [0,1] \\ F_U(u) &= \phi(F_U^{i:n}(u); i, n) \end{aligned}$$

Haile and Tamer (2003)

Upper Bound

Since $\phi : [0,1] \to [0,1]$ is strictly increasing then:

$$\phi(G_B^{i:n}(u); i, n) \geq F_U(u)$$

So given an estimate of $G_B^{i:n}(u)$ we can get an upper bound on $F_U(x)$. The most informative (least) upper bound is

$$F_U^+(u) = \min_{i,n} \phi(G_B^{i:n}(u); i, n)$$

Lower Bound

[A2] tells us that losing bidders have valuations less than $b^{n:n} + \Delta$, where Δ is the min bid increment.

$$u^{n-1:n} < b^{n:n} + \Delta$$

$$G_\Delta^{n:n}(u) \leq F_U^{n-1:n}(u) \quad \forall n, u$$

Look for greatest of $|\{\underline{n}, \ldots, \overline{n}\}|$ lower bounds.

Haile and Tamer (2003)

Nonparametric Estimator

Easy to construct non-parametric estimators

$$\hat{G}_B^{i:n}(b) = \frac{1}{T_n}\sum_{t=1}^{T}\mathbb{I}[n_t = n, b^{i:n_t} \leq b]$$

$$\hat{G}_B^{n:n}(b) = \frac{1}{T_n}\sum_{t=1}^{T}\mathbb{I}[n_t = n, b^{n:n_t} + \Delta_t \leq b]$$

Plug in to get bounds. Trick is in asymptotics (bootstrap goes through).

Potential Problem

In finite sample bounds may cross $\rightarrow$ work with weighted averages.

Haile and Tamer (2003)

TABLE 2
Gaps Between First- and Second-Highest Bids

Quantiles	High Bid	Gap	Minimum Increment	Gap ÷ Increment
10%	9,151	30	4.1	1.2
25%	22,041	92	10.1	6.9
50%	55,623	309	23.4	14.8
75%	127,475	858	52.1	20.0
90%	292,846	2,048	110.5	76.4

Haile and Tamer (2003)

TABLE 3
SUMMARY STATISTICS

	Mean	Standard Deviation	Minimum	Maximum
Number of bidders	5.7	3.0	2	12
Year	1985.2	2.6	1982	1990
Species concentration	.68	.23	.24	1.0
Manufacturing costs	190.3	43.0	56.7	286.5
Selling value	415.4	61.4	202.2	746.8
Harvesting cost	120.2	34.1	51.1	283.1
Six-month inventory*	1,364.4	376.5	286.4	2,084.3
Zone 2 dummy	.88		0	1

* In millions of board feet.

Haile and Tamer (2003)

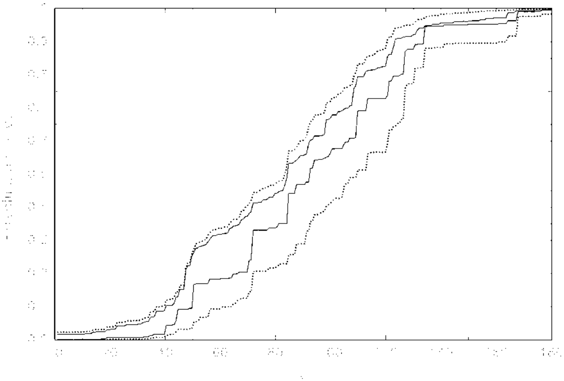


Fig. 10.—U.S. Forest Service timber auctions. Solid curves are estimated bounds, and dotted curves are bootstrap confidence bands.

Haile and Tamer (2003)

TABLE 4
Simulated Outcomes with Alternative Reserve Prices

	Reserve Price					
	p_L		$(p_L + p_U)/2$		p_U	
	Distribution of Valuations					
	F_L	F_U	F_L	F_U	F_L	F_U
Reserve price when $v_0 = \$20$	62.40		86.02		109.65	
Change in profit	6.96	−2.78	6.67	−2.74	1.74	−18.57
Pr(no bids)	.00	.02	.07	.12	.19	.41
Reserve price when $v_0 = \$40$	74.93		92.29		109.65	
Change in profit	7.64	−.61	7.61	−1.14	6.30	−10.04
Pr(no bids)	.03	.05	.11	.18	.19	.41
Reserve price when $v_0 = \$60$	85.67		103.39		121.11	
Change in profit	9.23	1.92	12.04	3.14	7.21	−6.05
Pr(no bids)	.07	.12	.15	.28	.35	.58
Reserve price when $v_0 = \$80$	98.20		112.34		126.48	
Change in profit	13.65	7.63	15.03	6.82	10.44	.96
Pr(no bids)	.13	.24	.28	.46	.46	.72
Reserve price when $v_0 = \$100$	111.09		122.54		134.00	
Change in profit	20.09	15.94	21.65	16.87	17.00	14.30
Pr(no bids)	.28	.45	.45	.60	.67	.80
Reserve price when $v_0 = \$120$	144.74		156.01		167.29	
Change in profit	32.06	31.31	33.72	31.64	31.56	28.87
Pr(no bids)	.84	.86	.84	.89	.88	.97

Note.—Profit and reserve price figures are given in 1983 dollars per MBF. See text for additional details.

Moment Equalities

We are already familiar with the idea of moment equalities of the form: $E[g(z_i, \theta)] = 0$:

- We can estimate them by forming $G_n(\theta) = \frac{1}{N} g_i(z_i, \theta)$ which is a $q \times 1$ vector.
- Form the objective function $Q(\theta) = G_n(\theta)' * W * G_n(\theta)$ where W is the $q \times q$ weighting-matrix.
- Find $\hat{\theta}_{GMM} = \arg\max_{\theta} Q(\theta)$ where θ is $k \times 1$.
- There is often a discussion about over or under or just identification depending on whether $k > q$ or $k < q$ or $k = q$.
  - When we are overidentified: Often $Q(\hat{\theta}_{GMM}) > 0$ even if $E[g(z_i, \theta_0)] = 0$ at the true population value θ_0.
  - If we are underidentified then $Q(\hat{\theta}_{GMM}) = 0$ at multiple values of θ.

Moment Inequalities: Chernozhukov, Hong, and Tamer (2007)

Suppose we could make such a strong assumption, suppose instead that we were willing to assume $E[g(z_i, \theta)] \geq 0$.

- Instead of a single $\hat{\theta}$ we want to characterize a set $\Theta_I \subset \Theta$ that satisfies the moment inequality restrictions.
- We can define $[x]_+$ to be the non-negative portion of x: $x_j = \max\{x_j, 0\}$.
- We can define $[x]_-$ to be the non-positive portion of x: $x_j = \min\{x_j, 0\}$.
- Now we can define the objective function: $Q(\theta) = E[g(z_i, \theta)]'_- \cdot W \cdot E[g(z_i, \theta)]_-$.
- Or in finite sample $Q(\theta) = G_n(\theta)'_- * W * G_n(\theta)_-$
- We want to find a θ to minimize $Q(\theta)$ but we only care about moment conditions that are violated from below (not from above).

Moment Inequalities: CHT (2007)

What is identified?

- Asymptotically we have that $Q(\theta) = E[g(z_i, \theta)]'_{-} \cdot W \cdot E[g(z_i, \theta)]_{-} = 0$.
- What is identified is a set. We might be tempted to define $\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$ and treat that as the identified set.
- What if we did that for the usual moment equality case (such as overidentified GMM)?
  - Remember it can be that: $Q(\hat{\theta}_{GMM}) > 0$
  - Defining $\Theta_I = \{\theta \in \Theta : Q(\theta) = 0\}$ might be a problem in finite sample.
- Instead let $\Theta_I = \{\theta \in \Theta : Q(\theta) = a_N\}$ where a_N is some small positive number that gets smaller as N gets larger.
  - Most of the time $a_n = c/n$ (but you can cook up problems where this is not true!).

Moment Inequalities: Inference

Inference can be somewhat complicated. We are now trying to construct a confidence set rather than a confidence interval

1. Should the confidence set contain each element of the identified set with a fixed probability (95%)?
2. Should the confidence set the entire identified set some probability (95%)?

Most people work on (1).

Moment Inequalities: Inference

- Many estimated confidence sets tend to be conservative that is they tend to be larger than they need to be.
- For example, for each element of θ we could construct $[\theta_k^{LB}, \theta_k^{UB}]$, but this assumes that the confidence set is a hyperrectangle but it might be some ellipse wholly within that hyperrectangle.
- Often we have some parameters that are point identified while one one or two parameters are partially identified within the same model.
  - e.g.: We might know the determinants of the variable profits, but we might only be able to recover bounds on fixed costs or entry costs.

Moment Inequalities: Inference

There is a big (and fast growing) literature on how to construct confidence sets:

- Imbens and Manski (2004)
- Romano and Shaikh (multiple papers)
- Andrews and Soares, Andrews Berry Jia (2004).
- Pakes, Porter, Ho, and Ishii (2015)

Example: 2×2 Entry Game

		BK IN	BK OUT
McD	IN	$(\alpha_m + \delta_b + \epsilon_m, \alpha_b + \delta_m + \epsilon_b)$	$(\alpha_m + \epsilon_m, 0)$
	OUT	$(0, \alpha_b + \epsilon_b)$	$(0, 0)$

- Two players: Burger King and McDonald's.
- Each player can play (IN, OUT).
- In market i we can write the profits of j as $\pi_{ij} = \alpha_j + \delta_k \cdot d_{ik} + X_i\beta + \epsilon_{ij}$.
- $d_{ij} = 1$ if IN, $d_{ij} = 0$ if OUT.
- Competitive effect $\delta_k < 0$.
- Firms observe everything (including ϵ)

Example: 2×2 Entry Game

Ignore covariates βX_i, suppose we have lots of data from independent plays of the game.

- $d_{ij} = 1[\pi_{ij} \geq 0]$ (Firms enter when profits are positive).
- Econometrician only observes entry status: $d_{ij} = 1[\pi_{ij} \geq 0]$ for B, M.
- Can we recover $\theta = [\delta_b, \delta_m, \alpha_a, \alpha_m]$?
- What if we assume that $\epsilon_{ij} \sim N(0,1)$ and is IID?

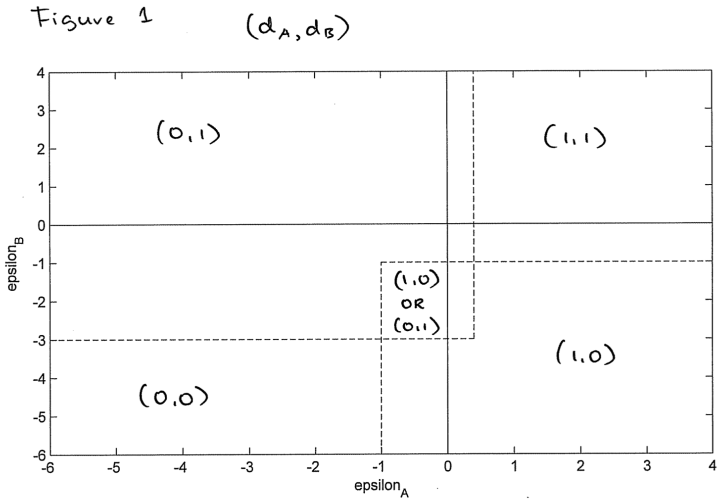
Example: 2×2 Entry Game

What if:

$$-\alpha_B < \epsilon_B \leq -\alpha_B - \delta_M$$
$$-\alpha_M < \epsilon_M \leq -\alpha_M - \delta_B$$

- Then $(d_b, d_m) = (1, 0)$ and $(d_b, d_m) = (0, 1)$ both satisfy the profit conditions!
- This shouldn't be too surprising: the original game has multiple equilibria.
- We don't observe the selection rule which determines behind the scenes which equilibria gets played.
- Even if we knew the form of the errors (ϵ_b, ϵ_m) we still can't map the data to a likelihood.
- You cannot write $Pr((d_b, d_m) = (1, 0))$ as a function of parameters!
- We can make assumptions on the selection rule (Berry 1992) such as "more profitable player moves first".

Entry Game



$\alpha_A = 1 \qquad \delta_A = -1.4$

$\alpha_B = 3 \qquad \delta_B = -2$

Deriving Bounds

$$H_{L,01}(\theta) \leq Pr((d_{i,b}, d_{i,m}) = (0,1)) \leq H_{U,01}(\theta).$$

Where

$$\begin{aligned} H_{L,01}(\theta) &= Pr(\epsilon_b < -\alpha_b, -\alpha_m < \epsilon_m) + \\ & \quad Pr(-\alpha_b \leq \epsilon_b < -\alpha_b - \delta_b, -\alpha_m - \delta_m < \epsilon_m) \end{aligned}$$

and

$$\begin{aligned} H_{U,01}(\theta) &= Pr(\epsilon_b < -\alpha_b, \alpha_m < \epsilon_m) + \\ & \quad Pr(-\alpha_b \leq \epsilon_b < -\alpha_b - \delta_b, -\alpha_m - \delta_m < \epsilon_m) + \\ & \quad Pr(-\alpha_b \leq \epsilon_b < -\alpha_b - \delta_b, -\alpha_m < \epsilon_m < -\alpha_m - \delta_m) \end{aligned}$$

Entry Game

Similar expressions can be derived for the probability $\Pr((d_{Am}, d_{Bm}) = (1,0))$. Thus in general we can write the information about the parameters in large samples as

$$\begin{pmatrix} H_{L,00}(\theta) \\ H_{L,01}(\theta) \\ H_{L,10}(\theta) \\ H_{L,11}(\theta) \end{pmatrix} \leq \begin{pmatrix} \Pr((d_{Am}, d_{Bm}) = (0,0)) \\ \Pr((d_{Am}, d_{Bm}) = (0,1)) \\ \Pr((d_{Am}, d_{Bm}) = (1,0)) \\ \Pr((d_{Am}, d_{Bm}) = (1,1)) \end{pmatrix} \leq \begin{pmatrix} H_{U,00}(\theta) \\ H_{U,01}(\theta) \\ H_{U,11}(\theta) \\ H_{U,11}(\theta) \end{pmatrix}.$$

Note: For $d = (1,1)$ and $d = (0,0)$ the upper and lower bounds coincide.

Generalized Inequality Restrictions

We can write this in the GIR form by defining

$$\psi(d_A, d_B|\alpha_A, \alpha_B, \delta_A, \delta_B) = \begin{pmatrix} H_{U,00}(\theta) - (1-d_A)\cdot(1-d_B) \\ (1-d_A)\cdot(1-d_B) - H_{L,00}(\theta) \\ H_{U,01}(\theta) - (1-d_A)\cdot d_B \\ (1-d_A)\cdot d_B - H_{L,01}(\theta) \\ H_{U,10}(\theta) - d_A\cdot(1-d_B) \\ d_A\cdot(1-d_B) - H_{L,10}(\theta) \\ H_{U,11}(\theta) - d_A\cdot d_B \\ d_A\cdot d_B - H_{L,11}(\theta) \end{pmatrix},$$

so that the model implies that at the true values of the parameters

$$\mathbb{E}\left[\psi(d_A, d_B|\alpha_A, \alpha_B, \delta_A, \delta_B)\right] \geq 0.$$

Generalized Inequality Restrictions

We can write this in the GIR form by defining

$$\psi(d_A, d_B|\alpha_A, \alpha_B, \delta_A, \delta_B) = \begin{pmatrix} H_{U,00}(\theta) - (1-d_A)\cdot(1-d_B) \\ (1-d_A)\cdot(1-d_B) - H_{L,00}(\theta) \\ H_{U,01}(\theta) - (1-d_A)\cdot d_B \\ (1-d_A)\cdot d_B - H_{L,01}(\theta) \\ H_{U,10}(\theta) - d_A\cdot(1-d_B) \\ d_A\cdot(1-d_B) - H_{L,10}(\theta) \\ H_{U,11}(\theta) - d_A\cdot d_B \\ d_A\cdot d_B - H_{L,11}(\theta) \end{pmatrix},$$

so that the model implies that at the true values of the parameters

$$\mathbb{E}\left[\psi(d_A, d_B|\alpha_A, \alpha_B, \delta_A, \delta_B)\right] \geq 0.$$

Homework

You can actually estimate this without too much difficulty ...