Lecture 3: AEA Continuing Education.
Discrete Choice and Moment Inequalities.

Lecturer: Ariel Pakes, Harvard University.

Consider two problems in discrete choice
• ν_1 errors caused by mis-measurement or mis-specification of functional form.
• Choice and group specific fixed effects.

But handling them requires additional assumptions. What this lecture shows you is that there is a simple non-parametric way of controlling for:

1. ν_1 errors on the "right hand side", caused by either errors in variables or miss-specification. Here we can handle fixed effects that are group specific (that is there are at least two observations with the same value of ν_2) but no additional ν_2 error.

2. A free distribution of ν_2 errors in the presence of choice and group specific fixed effects if there is no ν_1 error.

3. With additional assumptions on the ν_2 error we can add state dependence to the second case.

I do not know of non-parametric ways of controlling simultaneously for ν_1 and ν_2 errors. The interpretation to (1) then is that the group specific fixed effects together with the included observed covariates captures all of the determinants of choice while the interpretation of (2) is that the group and choice specific fixed effects capture all of the sources of measurement and specification error. So some thought in which set of assumptions best describes the data and model you are estimating is required before deciding on your estimator.

I first take on the non-parametric ν_1 case. I am going to do this through an empirical example as it will also show how we come to problems such as these and how the techniques enable us to get answers to important policy issues.

Hospital Choices, Hospital Prices and Financial Incentives to Physicians by, Kate Ho and Ariel Pakes. AER, 2014

Questions We Are Trying To Address.

- Do hospital referrals respond to the price paid by the insurer to the hospital, and how do price responses differ with the incentive structure built into insurance contracts with physician groups?

- What are the trade-offs between cost, "quality", and convenience factors implicit in the hospital choice function, and how do they relate to the incentives built into insurance contract?

Relationship to US Health Reforms.

• PPACA: ACOs are groups of providers who share responsibility for managing a large part (often all) of the health care needs of a group of Medicare patients (and private sector ACOs are forming in parallel; often same ACO).

• Cost control incentives for ACOs and California's physician groups are similar:

(i) both based on costs incurred by the group as a whole (no rules on how to pass down to particular providers), and
(ii) both either bear financial risk for hospital payments

or benefit from hospital savings relative to a benchmark (for ACOs it depends on a quality adjustment, but $\approx$ 50% of saving).

Large and important prior literatures which: (i) estimate models of hospital choice from discharge records, and (ii) establish that HMOs generate cost savings.

One part of our analysis focuses on a part of the allocation process that is not studied elsewhere:
(i) referral responses to insurer-paid hospital prices,
(ii) dependence of response to incentives in provider's insurer contracts.

Our main question. *What impact do shared savings programs have on the cost and quality of care?*

Institutional Setting: The California Medical Care Market 2003

• Focus on HMOs (53% of employed pop.)

• 7 largest HMOs had 87% of HMO market: we consider all but Kaiser (about 1/3 of these; they do not report prices).

• Physician contracts: HMOs have non-exclusive contracts with large physician groups.

• Two payment mechanisms for physician groups:
(1.) Capitation; fixed payment per patient ($\approx$ 75% of hospitals payments in our data).
(2.) Fee-for-service contracts.

Data for Our Analysis.

We utilize hospital discharge data for California in 2003; only women in labor (largest patient group). Census of hospital discharges of private HMO enrollees.

- Patient characteristics: insurer name, hospital name, detailed diagnoses, procedures, age, gender, zip code, list price, outcome measures.

- Hospital characteristics: average discount, address, teaching status, number of beds, services, annual profits.

Data. Does not identify patients' physician groups or details of compensation schemes. Do observe each patient's insurer and percent of each insurer's payments for primary services that are capitated. Considerable dispersion in this across insurers:

- Blue Cross: 38% capitated payments.

- Pacificare: 97% capitated payments.

Overview of the Model

Physician and patient jointly make the hospital choice within a choice set determined by the insurer. $W_{i,\pi,h}$ provides observed part of the plan and severity specific ordering of hospitals that this generates.

$$W_{i,\pi,h} = \theta_{p,\pi} p_{i,\pi,h} + g_{\pi}(q_h(s), s_i) + \theta_{d,\pi} d(l_i, l_h)$$

- $p_{i,\pi,h}$ = the expected price at hospital h for treating patient i in plan π.

- $d(l_i, l_h)$ = distance between hospital and patient's (home) location

- s_i = measure of patient severity

- $q_h(s, s_i)$ = a different quality for each sickness level in each hospital.

- $g_\pi(\cdot)$ = plan-specific non-parametric function of $q_h(s)$ and s_i which allows hospital quality rankings to differ by severity and plan.

Note:

- plan trade-offs between "quality", price, and convenience can differ across severities,

- different plans can make these trade-offs differently.

Our questions in the context of the model.

- The price coefficient
  - is it negative?
  - more negative when insurer gives physicians incentives to control costs?

- Do the plans which are more averse to price send patients to lower quality or to more distant hospitals?

Previous Hospital Choice Literature and Price Effects.

Prior Literature. Utility as a function of distance, hospital quality, hospital-patient interactions.

- Largely multinomial logit models with no price term (logic: neither the doctor nor the patient pays)*..

- The hospital quality terms and the patient-hospital quality interactions are a particular parameterization of our $g_\pi(q_h(s), s)$ terms.

*Gaynor and Vogt construct and use a single price for each hospital.

The Price Variable

- Agents do not know price when decision is made: need an expected price.

- Assume expected price per entering diagnosis/hospital is the average list price per diagnosis/hospital multiplied by a discount.

- Have data on average list price per entering diagnosis (like hotel "rack rate") and average discount at hospital level.

- Define price = expected list price*(1-average discount)

Major problems with price measure: the ν_1 error.

- Expectational and/or measurement error: likely more of a problem for severities with a small number of patients.

- Discounts are negotiated separately with each plan and discounts vary widely.

Notes on the Data.

- Discharge Descriptive Statistics. Note the average discount is 67.5

- Adverse Outcomes. Note: Discharge not home = patients; discharged to acute care or special nursing facility, deaths, and discharge against medical advice.

- Plan characteristics (capitation, premium/month).

- Prices and Outcomes by Patient Age and Charlson Score. Charlson score (Charlson et al, 1987, *Journal of Chronic Diseases*): clinical index that assigns weights to comorbidities other than principal diagnosis where higher weight indicates higher severity. Values 0-6 observed in data.

Discharge Descriptive Stats	Mean	(Std Dev)
Number of patients	88,157	
Number of hospitals	195	
Teaching hospital	0.27	
Ave. Dist. to Feasible Hosp.	24.6	(25.6)
Dist. to Chosen Hosp.	6.7	(10.3)
List price ($)	$13,312	($13,213)
List price*(1-discount)	$4,317	($4,596)
Length of Stay	2.54	(2.39)
Outcome Measures		
Infant Readmission	9.42%	(.1%)
Mother Readmission	2.39%	(.1%)
Discharge Not Home	6.60%	(.1%)
Plan Characteristics	capit.	pm/pm
Pacificare (FP)	.97	149.9
Aetna (FP)	.91	152.4
Health Net (FP)	.80	184.9
Cigna (FP)	.75	n.a.
Blue Shield (NFP)	.57	146.3
Blue Cross (FP)	.38	186.9

Prices and Outcomes By Patient Type

	N	P(1-δ)	Percentage M-Read	I-Read	N-Home
Age					
≤ 40	71073	4259	2.4	9.4	6.5
≥ 40	2044	5420	3.5	9.6	9.9
Charlson					
0	71803	4256	2.3	9.4	6.5
≥ 0	1314	6227	5.8	12.3	10.5

- Price and outcome measures vary in the expected direction with age and Charlson score.

- Most women are under 40, and most women come in with a zero Charlson score (severity, however, also differs by diagnosis).

Mimic Prior Literature: Multinomial Logits

$$W_{i,\pi,h} =$$

$$\theta_{p,\pi}(\delta_h lp(c_i,h)) \; + \; g_\pi(q_h(s),s_i) \; + \; \theta_d d(l_i,l_h) \; + \; \varepsilon_{i,\pi,h}$$

where

- $\epsilon_{i,\pi,h}$ is added extreme value disturbance
- δ_h = 1 - discount, and $lp(c_i,h)$ = average list price for type c_i at h (so there is no measurement error in price).
- s_i has age groups (4), principal diagnosis (21), Charlson score (6) and diagnosis generating Charlson score.

Multinomials restrictions

• No expectational error in price. Particularly problematic when small number of patients in group.

• Need to limit controls for $g_\pi(q_h(s), s)$ due to sample size per (s, h); #h = 195, #s= 105 (omitting zeros ≈ 16,000 fixed effects).

$$g_\pi(q_h(s), s_i) = q_h + \beta z_h x(s_i)$$

where

• q_h= hospital fixed effects,

• z_h= hospital characteristics (teaching hospitals, FP hospitals, hospitals with transplant services, a measure of quality of labor and birth services)

• $x(s_i)$ = P(adverse outcomes — age, principal diagnosis, Charlson score).

If it is **incorrect**

$$g_\pi(q_h(s), s_i) = q_h + \beta z_h x(s_i) + e_\pi(q_h(s), s_i).$$

Cases:

• Fix $s_i = s^*$, $e_\pi(q_h(s), s^*)$ is absorbed in q_h.

• Let s_i vary. Then residual will be positively correlated with price if more severely ill women go to higher quality hospitals and higher quality hospitals are more highly priced.

Logit Results: 1

	All	Least sick	Sick
Price	.010**	-0.017*	.012**
	(.002)	(.009)	.002
Distance	-.215**	-.215**	-.217**
	(.001)	(.002)	(.002)
Distance^2	.001**	.001**	0.001**
	(.000)	(.000)	(.000)
$z_h x(s_i)$	Y	Y	Y
(15 coeffts)			
Hosp. F.E.s	Y	Y	Y
(194 coeffts)			
N	88,157	43,742	44,059

Notes:

- Least sick patients are aged 20-39 with zero Charlson scores and all diagnoses "routine".
- Price is average list price by age group (4) $\times$ principal diagnosis

(21) $\times$ Charlson Score (6) $\times$ diagnosis Charlson score was in where possible.

Results. • Positive coefficient consistent with omitted variable bias.

• When we control for hospital and consider a more homogeneous severity group get a negative coefficient.

• When we look at the group with a lot of variance in severity the coefficient is even more positive.

Conclude: It is likely that there is an omitted variable and it is related to severity-hospital interactions.

Results: Logit Analysis 2

	% capit	Discharges	Least sick patients Estimates
Spec 1: Price x			
constant			.069**(.014)
% capit			-.127** (.016)
Spec 2: Price x			
Pac-care	0.97	7,633	-0.077** (0.01)
Aetna	0.91	3,173	-0.011 (0.016)
HN	0.80	8,182	-0.038** (0.01)
Cigna	0.75	4,001	-0.021 (0.014)
BS	0.57	7,992	0.018 (0.011)
BC	0.38	12,761	0.008 (0.011)
Distance			-0.215** (0.002)
Distance squared			0.001** (0.000)
$z_h x(s_i)$ controls			Y
Hospital F.E.s			Y
N			43,742

Conclusions.

- **Capitation.** Matters for price effect (Capitation coefficient for sick patients was $-.025(.006)$.)

- **Magnitudes.** From logits, even for least sick patients, questionable

-*Distance.* Average impact of a 1 mile increase in distance to hospital h (all else fixed) on P_{ih} is a 13.7% reduction. Comparable to prior estimates.

-*Price.* A \$1000 price increase for Pacificare enrollees (97% capitated) in hospital h a 5.2% reduction in P_{ih}. $\approx$ \$2,600 per mile for Pacificare (average price $\approx \$3,400$), and higher for all others.

Inequalities Analysis

$$W_{i,\pi,h} = \theta_{p,\pi}(\delta_h lp(c_i, h)) + g_\pi(q_h(s), s_i) + \theta_{d,\pi} d(l_i, l_h)$$

- Analysis done separately by plan.

- Normalize $\theta_{d,\pi} = 1$. s_i determines severity group and c_i determines price group. Both far more detailed than logit analysis.

- s_i groups: age $\times$ principal diagnosis $\times$ Charlson score $\times$ diagnosis generating Charlson score $\times$ rank of most serious comorbidity.

- Price groups. Severity $\times$ actual comorbidity $\times$ # comorbidities.

- Groupings done at suggestion of Columbia Presbyterian obstetricians: rank of most serious comorbidities determines hospital choice, but not the number of comorbidities (which do determine hospital costs).

- $g_\pi(q_h(s), s_i)$ freely interacts s_i groups ($\approx$ 105) with q_h (hospital F.E.). We assume it absorbs all unobserved quality variation that affects hospital choice.

Assumptions. For price start with

$$p_i^o \equiv \delta_h^o lp^o(c_i^o, h) = \delta_h lp(c_i, h) + \varepsilon_{i,\pi,h},$$

and in robustness analysis allow for within hospital variation (including plan effects), and

$$g_\pi(q_h(s), s_i^o) = g_\pi(q_h(s), s_i) + \epsilon_{s_i,\pi,h},$$

where for all ϵ

$$E[\varepsilon_{i,\cdot,h} | z_{i,\pi,h} \neq p^o(c_i, h), h] = 0.$$

Inequalities Analysis: Developing Estimator.

Use Revealed Preference: If h' was feasible for i_h

$$W_{i_h,\pi,h} \geq W_{i_h,\pi,h'}.$$

Procedure: Find all pairs of **same** π and s but **different** c patients, say i_h, $i_{h'}$ s.t.:

- $i_{\pi,h}$ visited h and had alternative h'
- $i_{\pi,h'}$ visited h' and had alternative h.

Then sum the two revealed preference inequalities, and then average over such couples.

- Equal and opposite $g_\pi(\cdot)$ terms drop out. , I.e.the ν_2

errors drop out.

- Terms that are errors in price average out.

Formalities. For any x let $x(i_h, h, h') \equiv x_{i_h,h} - x_{i_h,h'}$. If $s_{i_h} = s_{i_{h'}}$, $h' \in C(i_h)$, $h \in C(i_{h'})$ then

$$0 \leq W(i_h, h, h') = \theta_p p^o(i_h, h, h')$$

$$+ [g(q_h, s^o) - g(q_{h'}, s^o)] - d(i_h, h, h') + \varepsilon(i_h, h, h')$$

and

$$0 \leq W(i_{h'}, h', h) = \theta_p p^o(i_{h'}, h', h)$$

$$+ [g(q_{h'}, s^o) - g(q_h, s^o)] - d(i_{h'}, h', h) + \varepsilon(i_{h'}, h', h).$$

So

$$0 \leq W(i_h, h, h') + W(i_{h'}, h', h) =$$

$$\theta_p(p^o(i_h,h,h') + p^o(i_{h'},h',h)) - (d(i_h,h,h') + d(i_{h'},h',h))$$

$$+ \varepsilon(i_h,h,h') + \varepsilon(i_{h'},h',h).$$

For each $x(i_h,h,h')$ let

$$\overline{x}(h,h') = \frac{1}{N_{h,h'}} \sum_{i_h : h' \in C(i_h)} x(i_h,h,h').$$

The sum of the revealed preference inequalities over $\{i_h : h' \in C(i_h)\}$ and $\{i_{h'} : h \in C(i_h)\}$ is

$$\frac{-1}{N_{h,h'}N_{h',h}}\Big(\theta_p[\overline{p}^o(h,h') + \overline{p}^o(h',h)] + \overline{d}(h,h') + \overline{d}(h',h)\Big)$$

$$\to_P \quad \kappa \geq 0, \text{ at } \theta_p = \theta_0.$$

Can interact this with any positive function and it should still be positive at $\theta = \theta_0$.

Moments. Interact the original inequalities with "instruments" ($E[\varepsilon|z] = 0$) of the same sign, proceeding as above, and then summing over $h' > h$. Instruments

- a constant term

- the positive and negative parts of distance differences for h and for h' (paper has a robustness test for errors in this below).

Limitation of this methodology

- Unobservables causing selection absorbed in $g_\pi(.)$; i.e. there is no idiosyncratic structural (or ν_2) error.

- Assume price variable has no “systematic” error (i.e. it averages out for each plan). Assume the distance measure is exact. (Paper has robustness test for this).

Current limitations analogous to those for “matching estimators”, only here we have inequalities from revealed preference for each matched agent.

Benefits:

- Differencing out the $g_{\pi}(.)$ terms makes detailed hospital-quality/patient-severity/plan controls possible,

- Averaging over patients addresses measurement error problems in price variable (and only require mean independence, and not a distributional assumption, of the error and we do use average discounts).

Price and Severity Groups.

- 106 populated severity groups; 272 populated price groups (157 hospitals with over 1000 switches).

- Columns of table: severity groups aggregated (over age, principal diagnosis,...) into max rank.

- Rows give us the average of the price group in that severity group. Prices increase with the number of comorbidities in each rank.

- Given a severity, within rank differences in the price group relative to distance differences give us the distance-price trade-off for a given severity.

Table 5: Prices: Aggregated Price and Severity Groups.

Number diags of max rank	Max rank 1 Pats	Max rank 1 Price ($)	Max rank 1 SD	Max rank 2 Pats	Max rank 2 Price ($)	Max rank 2 SD
1	23029	3431 (15)	1612	13128	4968 (42)	2476
2	11757	4145 (28)	2180	4196	6019 (88)	2785
3	4077	4682 (60)	2356	1274	7428 (212)	3609
4	1179	5505 (149)	2590	380	8602 (462)	5283
5	331	6189 (254)	3123	110	10186 (1002)	6084
≥ 5	95	7663 (936)	4896	55	13365 (1596)	8880
Total	40468	3857 (15)		19143	5488 (40)	

Preparatory Analysis and Results.

- 73 - 283 moments per insurer.

- Can accept that the variance of price groups within severity group does not effect our measures of adverse effects (standard χ^2 tests).

- Price variation: Moving from severity to price groupings explains an additional 12% of variance in price (from 50% to 62% of total variance).

- Calculate t-statistic for each of our 977 moments at $\theta = \theta_0$. 6.1% negative, $\approx$ .7% with $t \leq -2$.

- Do analysis with and without moments with $t < -2$, not much difference. Health Net sometimes varies with whether you include the extra 2 moments and so I include both but use those that drop the $t < -2$.

Baseline Price Coefficients.

	% capit.	$\hat{\theta}_{p,\pi}$	$[CI_{LB},$	$CI_{UB}]$
Using observed discount d_h				
Pac/care	0.97	-1.50	[-1.68,	-1.34]
Aetna	0.91	-0.92	[-0.95,	-0.86]
(Hnet	.80	-.17	[-0.27,	-0.13])
HNet	drop t $\leq$ -2	-0.78	[-0.80,	-0.44]
Cigna	0.75	-0.35	[-0.40,	-0.33]
BS	0.57	-0.06	[-0.15,	0.23]
(BC	0.38	-0.10	[-0.24,	-.01])
BC	drop t $\leq$ -2	-0.29	[-0.31,	-0.25]

Note. (Except Blue Shield; A Not For Profit.)

- All negative.
- Ordered by capitation rates.
- Confidence intervals do not overlap.

Magnitude of Results

- $\eta^{d,p}$ = percent distance reduction needed to compensate for a 1% price increase (using $\widehat{d}$):

	P-coeff=	Logits	Inequalities
	patients=	less-sick	all
HMO	% cap	$\eta^{d,p}$	$\eta^{d,p}$
Pac-care	0.97	0.33	11.10
Aetna	0.91	0.10	11.47
Health Net	0.80	0.15	06.52
Cigna	0.75	0.10	02.49
BC	0.38	-0.03	03.24

Note: Price elasticities.

- $\geq$ order of magnitude larger than logits.
- vary a great deal with capitation rates.

Plan-Specific Trade-offs: Quality, Cost, & Distance.

Need plan-specific estimates of hospital "quality" ($\forall s$). Revealed preference implies that for each (π, s, h, h')

$$q_h^\pi(s) - q_{h'}^\pi(s) \ \geq \ g^\pi(s; h, h') \ = \theta_p^\pi \overline{p}(s; h, h') + \overline{d}(s; h, h')$$

For each couple of hospitals in the same market we have two such inequalities:

• One for those who chose h over h', and one for those who chose h' over h.

• # inequalities per market $= H_m(H_m - 1)$, where $H_m =$ # of hospitals in market.

Reliability of Ordering: Transitivity.

Note that regardless of θ_p

$$\overline{p}(h,h') \geq 0 \text{ and } \overline{d}(h,h') \geq 0, \Rightarrow q_h \succ q_{h'};$$

This defines a partial order which *is independent of our estimate*. However, that partial order need not obey transitivity. I.e. we could have

$$A \succ B,\ B \succ C, \text{ but } C \succ A; \text{ or just; } A \succ B, \ \& \ B \succ A.$$

• No non-transitive cycle‘s non-parametrically. Very few parametrically and **all** associated with differences in means which were not significantly different from zero.

Quality Bounds.

Quality estimates depend on θ_p, but our results change very little as we vary the θ_p's over their c.i.'s. Here I use point estimates of θ_p.

For a given (π, s) the quality of hospital h relative to a (market-specific) base hospital (H), is bounded by

$$\overline{q}(h) \equiv \min_{h' \neq h} E[-\widehat{q}(H, h') - \widehat{q}(h', h)] \geq q_h$$

$$\geq \max_{h' \neq h} E[\widehat{q}(h', H,) + \widehat{q}(h, h')] \equiv \underline{q}(h).$$

We stack these inequalities for each hospital, weight each by its estimated standard error, and then find the (set) estimator that minimizes the squared inequality violations (drop all comparisons with less than 5 switches)

Constraining the Quality Estimates.

• Too many $q_h^\pi(s)$ estimates (and sample sizes too small). Aggregate to five "Super-severity" groups: four determined by obstetrician (ordered by size) and a "remainder". Use five biggest markets (Bay Area, Inland Empire, LA, Orange County, San Diego); they contain almost all data with 5 or more switches.

Question. Are the insurers orderings affine transforms of one another? Or can we accept

$$q_h^\pi(s) = \alpha_{\pi,m,s} + \beta_{\pi,m,s} q_{s,h}.$$

m is for market. Note that the estimates of $\{q_h^\pi(s)\}$ are independent across plans. So if they are alike it is not because of the way we construct the estimates.

Implications of Accepting.

Preferences are linear functions of price, distance and a quality measure which is *common across insurers*. Can then compare trade-offs across insurers.

If in addition

$$\beta_{\pi,m,s} = \beta_{\pi},$$

we can omit the (m, s) indices and divide $W_{i,\pi,h}$ by β_{π} to obtain

$$E[W_{i,\pi,h}|c_i, l_i, l_h] \propto -\Big(\frac{\theta_{\pi}}{\beta_{\pi}}\Big)p(c_i, h, \pi) - \Big(\frac{1}{\beta_{\pi}}\Big)d(l_i, l_h) + q_{h,s_i}.$$

This makes differences in plans' trade-offs between cost, perceived quality, and distance transparent.

Details.

- Normalize BC coeff.s = 1.

- First equality takes 1,078 estimates to 452 parameters. $R^2 = .982$ (see picture).

- Both equalities give us 380 parameters $R^2 = .957$ (see picture) Fall in R^2 is .00009 per added constraint.

- Results when we aggregate all markets are incredibly precise, but they are a lot more variant when we disaggregate by market.

Trade-Offs for the Different Plans.

Insurer	P-care	Aetna	Hnet	Cigna	BC
% cap	0.97	0.91	0.80	0.75	0.38
$-\theta_\pi^p$	1.50	0.92	0.78	0.35	0.29
$-\beta_\pi$	5.13	3.12	2.63	1.20	1.00
θ_π^p/β_π	.293	.295	.297	.291	.290
$-1/\beta_\pi$	0.20	0.32	0.38	0.83	1.00
Upper and Lower Bounds on C.I. θ_p/β_π^*					
Upper	0.34	0.35	0.34	0.35	0.31
Lower	0.26	0.25	0.26	0.25	0.25.

*Calculated as lower bound (upper bound) θ_p divided by upper bound (lower bound) β_π.

• Just as the price coefficient increases montonically with capitation rate, so does the quality coefficient.

• The implication is that their ratio, which is the trade-off between price and our quality measure, is virtually identical across plans.

• More highly capitated plans send their patients further to obtain the same quality – but do not sacrifice quality.

Results consistent with (independent) outcome data.

χ^2 tests of differences of four outcomes* conditional on our five severities. Tested each couple of the five insurers for each severity and none significant.

*Mother and infant readmission within twelve months, mother and infant not discharged to home.

A Simple Counterfactual Analysis

- Findings indicate that patients in insurers using capitation incentives are referred to lower-priced hospitals. Assume the introduction of capitation would prompt low-capitation insurers to “act like” high-capitation insurer.

- Ask: under this assumption, how much would be saved if the use of capitation contracts was increased?

- Consider patients of Blue Cross (lowest-capitation insurer). Assume that increasing percent capitation to Pacificare level would change BC utility equation to that of Pacificare. Simulate BC patients' hospital referrals under Pacificare utility equation.

- Finding: 4.8% reduction in average price paid, with 6 mile average increase in distance, very little quality change. Savings in other diagnosis groups could be higher - since procedures and costs might vary more across hospitals

Potential implications for structure of ACOs

- Analysis assumes physicians free to choose hospitals within the existing networks

- Of approx 430 ACOs established by 1/13, around 50% integrated with a hospital system, the rest sponsored by physician group

- If ACOs integrated with hospital(s), with physician incentives to use own hospitals, cost reductions could be more limited.

- Could offset the benefits of integration (e.g. improved information flow, care coordination).

Conclusions.

- Hospital referrals are sensitive to price.

- The sensitivity of the hospital referrals to price depends on the extent to which the contracts the plan signs with physician groups are capitated: a 60% difference in extent of capitation triples coefficients and more than triples elasticities.

- At least in an environment where hospitals and other providers are separate, the higher capitation insurers substitute convenience for cost, but **do not** sacrifice quality for cost. They simply send their patients further to get the same quality of care.

I now move to the non-parametric case where there are only ν_2 errors and choice specific fixed effects (or dummy variables). This is based on the paper

Moment Inequalities for Multinomial Choice with Fixed Eects, by Ariel Pakes and Jack Porter.

Setup.

Observations (i, t) make choices $d \in D$, where $\#D = \mathcal{D}$. i indexes "group", of observations, t indexes

within-group. Formalization is the random utility model with choice specific effects

$$U_{d,i,t} = g_d(x_{i,t}, \theta_0) + f_d(\lambda_{d,i}, \varepsilon_{d,i,t}), \quad (1)$$

(often $f_d(\cdot) = \lambda_{d,i} + \varepsilon_{d,i,t}$). Do need $f(\cdot)$ additively separable from $g_d(\cdot)$.

The observed choice is

$$y_{i,t} \in \underset{d \in D}{\text{argmax}}\, U_{d,i,t}.$$

Motivation.

Familiar models.

• Panel data: Choice specific individual effects in panel data. E.g: Conditional logit (Chamberlain, 1980).

• Market demand: Consumers (t) within markets (i) chose products (d) with unobserved attributes ($\lambda_{i,d}$) (BLP/MicroBLP,1995/2004, Pure Characteristics,2007).

• *"Descriptive" form for discrete choice*. Provides the discrete choice analogue of the "within" analysis used in continuous choice problems. Examples.

Assumption.

$\epsilon_{i,t} \equiv [\epsilon_{i,1,t}, \ldots \epsilon_{i,\mathcal{D},t}]$; $\epsilon_i \equiv [\epsilon_{i,1}, \ldots \epsilon_{i,T}]$; $\lambda_i \equiv [\lambda_{i,1}, \ldots, \lambda_{i,\mathcal{D}}]$.

Assumption 1 (main stochastic assumption).
The conditional distributions of $\varepsilon_{i,s}$ and $\varepsilon_{i,t}$ conditional on $(x_{i,s}, x_{i,t}, \lambda_i)$, for any $s \neq t$, are the same, or

$$\varepsilon_{i,s}|x_{i,s}, x_{i,t}, \lambda_i \ \sim \ \varepsilon_{i,t}|x_{i,s}, x_{i,t}, \lambda_i.$$

- No other restriction on ε_i (either over t, or across $d \in D$).

Relationship to Prior Literature.

• Assumption used in a *binary* choice linear panel data problem by Manski (1987) who gives conditions for point identification, and in non-linear single index problems where the outcome determined by fixed effects, a disturbance, and is monotone in an index of covariates and parameters, (Honore, 1992; Abrevaya, 2000).

• Neither approach used above generalizes to multinomial choice.

• Panel Data Model: nests conditional logit (Chamberlain, 1980), and more generally the "strict exogeneity" assumptions (Chamberlain, 1980)

$$\varepsilon_{i,s}|x_{i,1},\ldots x_{i,T},\lambda_i \sim \varepsilon_{i,t}|x_{i,1},\ldots,x_{i,T},\lambda_i.$$

• Demand Systems: nests "BLP/MicroBLP" and the "Pure Characteristics" demand models (choice specific fixed effects plus random coefficients).

Advantages/Disadvantages of Approach.

General Panel Data.

• $\varepsilon_{i,s}$ freely related to $\varepsilon_{i,t}$ (the assumption only constrains the marginal distribution of the vector $\varepsilon_{i,s}$).

• $F_{i,\varepsilon}(\cdot)$ can vary over i (only need $T_i \geq 2$).

• Circumvents incidental parameter problem.

Demand Models (with panels).

• $\varepsilon_{i,d,t}$ arbitrarily related to $\varepsilon_{i,s,t}$ (no vestiges of IIA or limits on substitutability), and

• the marginal distribution of $\varepsilon_{i,d,t}$ can vary with d.

Weaknesses.

• In general we only get set identification.

• Might be relatively uninformative when $\mathcal{D}$ is large.

Parametric ε_i Distribution.

• No other assumption (including the homogeneity assumption above). Should work for large $\mathcal{D}$. This is revealed preference; but this time with disturbances (though parametric).

• A "truly" dynamic choice model with switching cost model and unobserved heterogeneity: with no need to specify agent's perceptions of the future or to calculate the value function).

Logic Underlying Estimation.

The stochastic assumption restricts the marginal distribution of $\epsilon_{i,t}$ to be the same for different t. This plus the fact that the $\lambda_{d,i}$ do not vary over t implies that the relative response probabilities of observations (i, s) and (i, t) are determined solely by comparing

$$g_d(x_{i,s}, \theta) \;\; to \;\; g_d(x_{i,t}, \theta).$$

Non-parametric results come from this fact.

Structure of Talk.

- Begin with a single conditional moment inequality (makes role of assumptions and the logic transparent).

- Add information; provide additional conditional moment inequalities.

- Endogenous r.h.s. variables; heterogeneity and state dependence;.

- Parametric Information on Disturbance Distribution:
(i) Static multinomial choice (set valued generalization of conditional logit to any parametric distribution),
(ii) Dynamic discrete choice with unobserved heterogeneity and switching costs.

Simple Conditional Moment Inequality

Fix (and omit) the index i, the group index, and let $D = \#\mathcal{D}$. Fix θ and for that θ compare the index functions for period s to those for t and find that choice that maximizes the difference in them, i.e.

$$d^D(x_s, x_t, \theta) = argmax_{c \in \mathcal{D}}[g_c(x_s, \theta) - g_c(x_t, \theta)].$$

Notes:

• $d^D(x_s, x_t, \theta)$ does not depend on either ϵ_s or ϵ_t, so when we consider the probability that s choses d^D and t does not there is no selection problem.

• $d^D(x_s, x_t, \theta)$ can be computed without knowing the $\lambda_{i,d}$ (which difference out).

- **Note.** If $d \equiv d^D(x_s, x_t, \theta_0)$, then **for all** $c \neq d$

$$[g_c(x_s, \theta_0) + \lambda_c - (g_{d=D}(x_s, \theta_0) + \lambda_{d=D})] \leq$$

$$[g_c(x_t, \theta_0) + \lambda_c - (g_{d=D}(x_t, \theta_0) + \lambda_{d=D})].$$

$\Rightarrow$

$$\max_{c \neq d} [g_c(x_s, \theta_0) + \lambda_c - (g_d(x_s, \theta_0) + \lambda_d)]$$

$$\leq \max_{c \neq d} [g_c(x_t, \theta_0) + \lambda_c - (g_d(x_t, \theta_0) + \lambda_d)]$$

- $\Rightarrow$ a conditional moment inequality for the difference in indicator functions. For now assume no ties. Then if $\Omega_{s,t} = (x_s, x_t, \lambda)$,

$$\mathrm{Pr}(y_t = d|\Omega_{s,t}) =$$

$$P(\varepsilon_{d,t} \geq \max_{c \neq d} [g_c(x_t, \theta_0) + \lambda_c - g_d(x_t, \theta_0) - \lambda_d] + \varepsilon_{c,t}|\Omega_{s,t}),$$

$$\leq \text{ (by our inequality)}$$

$$P(\varepsilon_{d,t} \geq \max_{c \neq d} [g_c(x_s, \theta_0) + \lambda_c - g_d(x_s, \theta_0) - \lambda_d] + \varepsilon_{c,t}|\Omega_{s,t}),$$

$$= \text{ (by equivalence of marginals)}$$

$$P(\varepsilon_{d,s} \geq \max_{c \neq d} [g_c(x_s, \theta_0) + \lambda_c - g_d(x_s, \theta_0) - \lambda_d] + \varepsilon_{c,s}|\Omega_{s,t}).$$

$$= \mathrm{Pr}(y_s = d|\Omega_{s,t}). \ \spadesuit$$

Additional Inequalities.

Fix θ and rank the differences this induces. So

$$d^D(x_s, x_t, \theta) = argmax_{c \in \mathcal{D}}[g_c(x_s, \theta) - g_c(x_t, \theta)],$$

$$d^{D-1}(x_s, x_t, \theta) = argmax_{c \in \mathcal{D}, c \neq d^D}[g_c(x_s, \theta) - g_c(x_t, \theta)],$$

$$d^{D-2}(x_s, x_t, \theta) = argmax_{c \in \mathcal{D}, c \notin \{d^D, d^{D-1}\}}[g_c(x_s, \theta) - g_c(x_t, \theta)],$$

$$\ldots$$

$$d^1(x_s, x_t, \theta) = argmin_{c \in \mathcal{D}}[g_c(x_s, \theta) - g_c(x_t, \theta)].$$

• For $\theta = \theta_0$ the probability of choosing one of the w-highest ranked differences is

$$P(\{y_s = d^D\} \cup \{y_s = d^{D-1}\} \cup \cdots \cup \{y_s = d^{D-w}\} | \Omega_{s,t}).$$

• We develop inequalities for differences in the probabilities of these choices between s and t for $w \in (1 \ldots D]$.

Proof of Inequality.

$$P(\{y_s = d^D\} \cup \{y_s = d^{D-1}\} \cup \cdots \cup \{y_s = d^{D-w}\} | \Omega_{s,t}) =$$

$$P\Big\{\epsilon_s : \bigcup_{r=0}^{w} \{\varepsilon_{d^{D-r}} \geq \max_{c \neq d^{D-r}} g_{c,s}(\cdot) + \lambda_c - g_{d^{D-r},s}(\cdot) - \lambda_{d^{D-r}} + \varepsilon_c \Big| \Omega_{s,t}\}$$

which by equivalence of sets

$$P\{\epsilon_s : \bigcup_{r=0}^{w} \{\varepsilon_{d^{D-r}} \geq \max_{c \notin \{d^D,\ldots,d^{D-w}\}} g_{c,s}(\cdot) + \lambda_c - g_{d^{D-r},s}(\cdot) - \lambda_{d^{D-r}} + \varepsilon_c | \Omega_{s,t}\}$$

Note that for $r \leq w$,

$$\max_{c \notin \{d^D,\ldots,d^{D-w}\}} g_{c,s}(\cdot) + \lambda_c - g_{d^{D-r},s}(\cdot) - \lambda_{d^{D-r}} + \varepsilon_c | \Omega_{s,t}\} \leq$$

$$\max_{c \notin \{d^D,\ldots,d^{D-w}\}} g_{c,t}(\cdot) + \lambda_c - g_{d^{D-r},t}(\cdot) - \lambda_{d^{D-r}} + \varepsilon_c | \Omega_{s,t}\}, \Rightarrow .$$

So

$$P\{\epsilon_s : \bigcup_{r=0}^{w} \{\varepsilon_{d^{D-r}} \geq \max_{c \notin \{d^D,\ldots,d^{D-w}\}} g_{c,s}(\cdot) + \lambda_c - g_{d^{D-r},s}(\cdot) - \lambda_{d^{D-r}} + \varepsilon_c | \Omega_{s,t}\}$$

$\geq$ (from the inequality above)

$$P\{\epsilon_s : \bigcup_{r=0}^{w} \{\varepsilon_{d^{D-r}} \geq \max_{c \notin \{d^D,\ldots,d^{D-w}\}} g_{c,t}(\cdot) + \lambda_c - g_{d^{D-r},t}(\cdot) - \lambda_{d^{D-r}} + \varepsilon_c | \Omega_{s,t}\}$$

which from our stochastic assumption, $=$

$$P\{\epsilon_t : \bigcup_{r=0}^{w} \{\varepsilon_{d^{D-r}} \geq \max_{c \notin \{d^D,\ldots,d^{D-w}\}} g_{c,t}(\cdot) + \lambda_c - g_{d^{D-r},t}(\cdot) - \lambda_{d^{D-r}} + \varepsilon_c | \Omega_{s,t}\}$$

(which from the set equivalence argument above) $=$

$$= \Pr(\{y_t = d_{s,t}^{(D)}\} \cup \{y_t = d_{s,t}^{(D-1)}\} \cup \cdots \cup \{y_t = d_{s,t}^{(D-w)}\} \mid \Omega_{s,t}). \spadesuit$$

So our first proposition concerns

$$m_w(y_s, y_t, x_s, x_t, \theta) =$$

$$\mathbf{1}\left\{y_t \in \cup_{r=0}^{w}\{d_{s,t}^{D-r}(\theta)\}\right\} - \mathbf{1}\left\{y_s \in \cup_{r=0}^{w}\{d_{s,t}^{D-r}(\theta)\}\right\}.$$

Proposition. $\forall w \leq D-1$

$$0 \leq E_\epsilon[m_w(y_{i,s}, y_{i,t}, x_{i,s}, x_{i,t}, \theta_0) \mid x_{i,s}, x_{i,t}], \quad \spadesuit.$$

I.e. for any $h(x_{i,s}, x_{i,t}) \geq 0$

$$0 \leq E_\epsilon[m_w(y_{i,s}, y_{i,t}, x_{i,s}, x_{i,t}, \theta_0) h(x_{i,s}, x_{i,t})].$$

• **Potential content of inequalities.** Let r rank differences. If the realizations of

$$y_s = d^r_{s,t}(\theta) \text{ and of } y_t = d^{r-k}_{s,t}(\theta),$$

then the (s,t) comparison can contribute to the empirical analogues of restrictions from at most k conditional moments for any fixed θ.

• The conditional expectations of

$$\left\{\mathbf{1}\{y_s \in \cup^D_{q=d^r_{s,t}(\theta)-j}\} - \mathbf{1}\{y_t \in \cup^D_{q=d^r_{s,t}(\theta)-j}\}\right\}^{k-1}_{j=0},$$

should be positive at $\theta = \theta_0$.

Information in comparisons. If “rank” = the rank of the difference in index functions, and

- $rank[y_{i,s}(\theta)] \geq rank[y_{i,t}(\theta)]$ then the comparison does not provide evidence against that value of θ.

- However if $rank[y_{i,s}(\theta)] < rank[y_{i,t}(\theta)]$, i.e. if $k < 0$, then there is evidence against θ and that evidence is stronger the larger (in absolute value) is $|k|$ and the values of $\{g_d(x_{i,t}, \theta) - g_d(x_{i,s}, \theta)\}_{d=y_{i,s}}^{y_{i,s}+k}$.

• **Importance of using all the choices**. Say we would have only used first ranked choice. Then we only have $m^D_{s,t}(\cdot,\theta)$, and

$$y_{i,s} \neq D_{s,t}(\theta) \text{ and } y_{i,t} \neq D_{t,s}(\theta) \Rightarrow m^D_{s,t}(\cdot,\theta) = 0.$$

If D is large $P\{y_{i,s} = D_{s,t}(\theta)\}$ is typically small (even at $\theta = \theta_0$) $\Rightarrow$ we only use a small fraction of data (even if $y_{i,s} = D_{s,t}(\theta) - 1$ *and* the difference in index functions is negligible).

Asymptotics.

• Can let either n, T or both grow large. If n grows large holding T fixed the number of parameters grows with the number of observations, but we have circumvented the incidental parameter problem in the estimates of θ.

• More detail on the asymptotics requires assumptions on the dependence structure of the observations. Then one can use the recent literature on estimating parameters when the model generates conditional moment inequalities (Andrews and Shi, 2013; Chetverikov 2013; Armstrong 2014; Chernozhukov, Lee and Rosen, 2013, ...).

Monte Carlo Example: Two Period Entry Model*

The Need for a "Within" Estimator (here i indices markets and t firms in market). Goal: provide a summary of how the values of active firms depend on the number and type of competitors **conditional on** market "profitability": i.e. general demand and supply conditions which are, in part, unobserved.

• Early work that did not "endogenize" the disturbances in the entry equation typically found that the value of active firms was positively related to the number of competitors; more profitable markets had more firms

*From Pakes, A.: "Behavioral and Descriptive Forms of Choice Models", *IER*, 2014.

and we could not fully condition on market profitability. This starts the two period entry literature (Bresnahan and Reiss 1990, Berry 1992, Tamer 2003, and Ciliberto and Tamer 2009).

• We propose to treat the profitability of the market as an "unobservable" and do a "within-market" analysis. Discrete analogue to group specific fixed effects in continuous problems which lead to "within" analysis (or difference in differences).

How do we proceed? i indexes market, and t firms.

• Compute Markov Perfect equilibria to a set of infinite horizon entry games with different market sizes.

• Use equilibrium policies to simulate market structures.

- Regress the value function of firms in the simulated market structures on market size and the number and type of competitors.
- This summary is computed in a way which endows the coefficients with both a meaning and comparability across studies of different markets (elsewhere I have called it a "descriptive form").
- The error is now mean independent of observables by construction, and it can include approximation and measurement errors (unlike prior two-period models).
- Finally assume market size is unobserved, so it becomes the "choice-specific" fixed effect, and use our estimation technique to try to recover the descriptive form that conditions on market size.

DGP for Monte Carlo. Firms differ in their: (i) location (l; east or west), (ii) quality of their product (q), and their market (differentiated by size, our λ, with sixty values).

- Consumers differ in their location and their sensitivity to price.
- Six firms can participate in each market. In each period some are potential entrants who can enter either east or west or not enter.
- Incumbents decide whether to exit and if they continue they can invest in quality.
- The outcome of the investment is stochastic, and there are random draws on the sunk cost of entry and exit values in every period.
- There is also a fixed cost which differs with the quality level and by market.

Equilibria. Static equilibrium conditions on the characteristics of active firms and is Nash in prices. Dynamic choices are Markov Perfect in entry, exit, and investment policies (an Ericson-Pakes, 1995, model computed with the Pakes-McGuire,1994 algorithm).

Notation:
Q divides firms into high and low quality,
$n_{Q,l}$ is the number competitors with the same (Q, l)
n_Q is the number of competitors with the same Q.
n_l is the number of competitors with the same l.
M is market size.

Goal of Analysis. The algorithm delivers a value functions for each active firm and potential entrant, for an incumbent that would be $V(l, q, s)$ where s is the vector (counting measure) which tells us how many competitors at each (l, q). We regress

$$V(l, q, s) = (n_{Q,l}, n_Q, n_l, Q, M)\beta + \varepsilon$$

gives us a minimum mean square predictor of the value of being active conditional on market size (our "descriptive form"). We then try and predict the descriptive form from the entry and exit decisions assuming M is unknown.

Table 1: True Descriptive Forms.

	Expected Discounted Value			
θ	With Mkt.Size		No Mkt.Size.	
$n_{l,q}$	-0.73	(0.01)	-0.48	(0.01)
n_{-l}	-0.58	(0.01)	-0.16	(0.02)
n_{-q}	-0.07	(0.01)	0.18	(0.01)
q=H	1.59	(0.01)	1.69	(0.02)
R^2	0.80	n.r.	0.22	n.r.

• Easy to see the effects of market size. The coefficients in the two columns differ markedly and the R^2 drops to one quarter of its value. There are good reasons to endogenize entry decisions.

• Even after allowing for the free $\{\lambda\}_i$ twenty per cent of the variance is not accounted for. This variance is orthogonal to x by construction: it is the approximation error from the descriptive form.

• Approximation error is ignored in prior models.

Estimation Notes: Using the P-P Algorithm.

• Use entry-exit decisions which allow for choice specific FE but no M to try to replicate descriptive form.

• The residuals here do not necessarily satisfy that within a market

$$\epsilon|_{x_s, x_t, \lambda} \sim \epsilon|_{x_r, x_s, \lambda}.$$

• The potential entrants have a trinomial choice (don't enter, enter west, or enter east). The incumbents can only chose to exit or continue.

Estimation Questions.

- Does the restrictive assumption on heteroscedasticity matter?

- Does the addition of 2nd choice data help?

- Does it matter if we only compare agents with the same choice set (prior models lump potential entrants and incumbents together)?

• Always cover the true parameter, but use of only the highest ranked difference leads to wide confidence intervals.

• Adding the inequality for the second highest index decreases the width of the confidence interval by a factor of about 3.

• When we compare across choice sets we actually do a bit better, but the difference is small.

Table 2. P-P Estimators of Descriptive Form.

Estimator	$n_{l,q}$	n_{-l}	n_{-q}
Descriptive Form			
1. OLS with λ	-0.73 (.03)	-0.25 (.04)	-0.10 (.01)
Use Only Largest Difference in Index: All Choice Sets.			
2. Point Estimate	-0.53	-0.42	-0.02
3. Confidence Interval	[-.96,-.19]	[-.94,.01]	[-.34,.02]
Use 1^{st} & 2^{nd} Differences: Compare Across Choice Sets.			
4. Point Estimates	-0.61	-0.28	-0.13
5. Confidence Interval	[-.77 -.49]	[-.36,-.19]	[-.16,-.09]
Use 1^{st} & 2^{nd} Differences: Only Within Choice Sets			
6. Point Estimates	-0.60	-0.26	-0.14
7. Confidence Interval	[-.74 -.44]	[-.39,-.22]	[-.22,-.07]

Adding Information: $\varepsilon \sim F_i(\cdot|\beta)$.

Case 1: Static Multinomial Choice. Likely important for large choice sets. Focus on the panel data case: i.e. limits in N (limits in T is covered by the demand literature). For this case Chamberlain (1980) assumed a logistic distribution and provided a conditional mle. We generalize to any joint distribution.

Assumption 3. $F_i(\varepsilon|\Sigma(\beta))$, for $\beta \in \mathcal{R}^k$.

Notes. $E[\varepsilon_i|x_i, \lambda_i] = 0$ since λ_i picks up means. No other assumption on the distribution of ε_i: i.e. no need for the group homogeneity assumption. $\Rightarrow$ if $F_i(\cdot)$ has the same marginals over t prior results generate a non-parametric test of whether $F_i(\cdot)$ could be right.

Logic. If $y_s = d$ but $y_t = q$ then by revealed preference

$$
\begin{aligned}
g_d(x_s, \theta_0) - g_q(x_s, \theta_0) + \lambda_d - \lambda_q + \varepsilon_{d,s} - \varepsilon_{q,s} \quad &\geq \quad 0 \\
\geq \quad g_d(x_t, \theta_0) - g_q(x_t, \theta_0) + \lambda_d - \lambda_q + \varepsilon_{d,t} - \varepsilon_{q,t}.
\end{aligned}
$$

So if $\Delta^c_{s,t}(\theta) \equiv g_c(x_s, \theta) - g_c(x_t, \theta)$,

$$\varepsilon_{d,s} - \varepsilon_{q,s} - (\varepsilon_{d,t} - \varepsilon_{q,t}) \geq \Delta^q_{s,t}(\theta_0) - \Delta^d_{s,t}(\theta_0).$$

It follows that

$$Pr\Big(\varepsilon_{d,s} - \varepsilon_{q,s} - (\varepsilon_{d,t} - \varepsilon_{q,t}) - (\Delta^q_{s,t}(\theta_0) - \Delta^d_{s,t}(\theta_0)) \geq 0 | x_s, x_t, \lambda, \beta_0\Big)$$

$$\geq Pr(y_s = d, y_t = q | x_s, x_t, \lambda, \theta_0, \beta_0). \quad \spadesuit$$

To obtain our proposition for this case define

$$m_F(y_s, y_t, x_s, x_t, \theta) \equiv$$

$$1\{\varepsilon_{d,s} - \varepsilon_{q,s} - (\varepsilon_{d,t} - \varepsilon_{q,t}) - (\Delta^q_{s,t}(\theta) - \Delta^d_{s,t}(\theta)) \geq 0\} - 1\{y_s = d, y_t = q\}.$$

Proposition 4 Assumption (3) implies that for all (s, t) and for all $(d, q) \in \{(d, q) : d \neq q, (d, q) \in \mathcal{D}^2\}$

$$E[m_F(y_s, y_t, x_s, x_t, \theta_0)|x_s, x_t, \beta_0] \geq 0. \spadesuit$$

For each (s, t) couple this gives us $2\mathcal{D}(\mathcal{D} - 1)$ inequalities; one each for $\{y_s = d, y_t = q\}$ for $d \neq q, (d, q) \in \mathcal{D}^2$, and the analogous inequalities for $\{y_t = d, y_s = q\}$; and there are $T_i(T_i - 1)$ such couples for each i.

References

- Abrevaya, J. "Rank estimation of a generalized fixed-effects regression model," *Journal of Econometrics*, 2000, 95(1): 1-23.

- Andrews, D., and X. Shi, "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, 81(2): 609-666.

- Andrews, D. and G. Soares "Inference for Models Defined by Moment Inequalities Using Generalized Moment Selection Procedures," *Econometrica*, 2010, 78(1): 119-157.

- Armstrong, T., “A Note on Minimax Testing and Confidence Intervals in Moment Inequality Models, Yale working paper, 2014.

- Berry, B. “Estimation of a Model of Entry in the Airline Industry,” *Econometrica*, 1992, 60(4): 889-917.

- Berry, S., J. Levinsohn, and A. Pakes, “Automobile Prices in Market Equilibrium,” 1995, *Econometrica*, 63(4): 841-890.

- Berry, S., J. Levinsohn, and A. Pakes, “Estimating Differentiated Product Demand Systems from a

Combination of Micro and Macro Data: The Market for New Vehicles,'' 2004, *Journal of Political Economy*, 112(1): 68-105.

- Berry, S. and A. Pakes, ''The Pure Characteristics Model of Demand,'' 2007, *International Economic Review* 48 (4), 1193-1225 (special issue in honor of Dan McFadden).

- Bresnahan, T., and P. Reiss, ''Entry in Monopoly Markets,'' *The Review of Economic Studies*,'' 1990, 57(4) 531-553.

- Chamberlain, G., "Analysis of Covariance with Qualitative Data," *Review of Economics Studies*, 1980, XLVII, 225-238.

- Charlson, M., P. Pompei, K. Ales, and C.R. MacKenzie, "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation., *Journal of Chronic Diseases*, 1987, 40(5): 373383.

- Chernozhukov, V., S. Lee, and A. Rosen, "Intersection Bounds: Estimation and Inference," *Econometrica*, 2013, 81(2): 667-737.

- Chetverikov, D., “Adaptive Tests of Conditional Moment Inequalities,” UCLA working paper, 2013.

- Ciliberto, F. and E. Tamer, “Market Structure and Multiple Equilibria in Airline Markets, *Econometrica*, 2009, 77(6): 1791-1828.

- Ericson, R., and A. Pakes, “Markov Perfect Industry Dynamics: A Framework for Empirical Work, *Review of Economic Studies*, 1995, 62(1): 53-82.

- Gaynor, M., and W. Vogt. “Antitrust and Competition in Health Care Markets. In *Handbook of*

Health Economics. Vol. 1B, edited by Anthony J. Culyer and Joseph P. Newhouse, 2000, 14051487. Amsterdam: Elsevier B.V.

- Ho, K. and A. Pakes, “Hospital Choices, Hospital Prices, and Financial Incentives to Physicians,” *American Economic Review*, 2014, 104(12): 3841-3884.

- Honore, B., “Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects,” *Econometrica*, 1992, 60(3): 533-565.

- Manski, C., “Semiparametric Analysis of Random Effects Linear Models From Binary Response Data, *Econometrica*, 1987, 55(2): 357-362.

- Manski, C., and E. Tamer, “Inference on Regressions with Interval Data on a Regressor or Outcome, *Econometrica*, 2002, 70(2): 519-546.

- Melichar, L., “The Effect of Reimbursement on Medical Decision Making: Do Physicians Alter Treatment in Response to a Managed Care Incentive? *Journal of Health Economics,* 2009, 28(4): 902907.

- Pakes, A., “Behavioral and Descriptive Forms of Choice Models,” *International Economic Review*, 2014, 55(3): 603-624.

- Pakes, A., and P. McGuire, “Computing Markov Perfect Nash Equilibrium: Numerical Implications of a Dynamic Dierentiated Product Model, *RAND Journal of Economics*, 1994, 25(4): 555-589.

- Pakes, A., and J. Porter, “Moment Inequalities for Multinomial Choice with Fixed Eects,” 2014, Harvard working paper.

- Tamer, E., “Incomplete Simultaneous Discrete Response Model with Multiple Equilibria, *Review of Economic Studies*, 2003, 70(1): 147-167.