Structural Behavioral Economics
NH Summer School

Stefano DellaVigna

June 28, 2018

Section 1

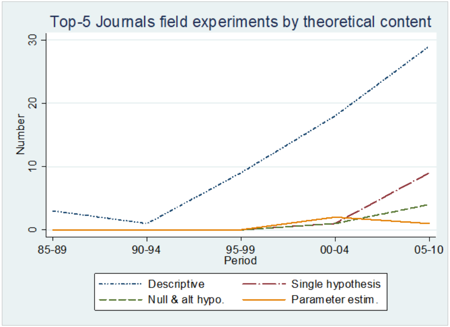
Overview

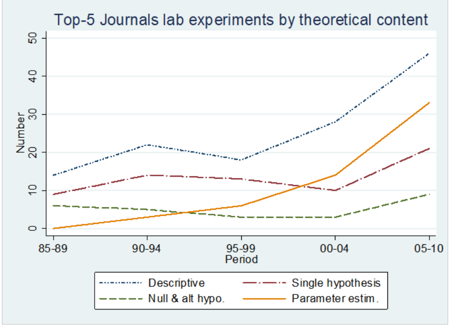
Overview

- Structural estimation in behavioral economics
  - Use model for estimation
  - Estimate key model parameters
- What do we mean by structural?
  "Estimation of a model on data that recovers parameter estimates (and c.i.s) for some key model parameters"
- Chapter in preparation for first *Handbook of Behavioral Economics*

Overview

- Chapter Focused on applications to field evidence. Structural lab evidence is way ahead (**Card, DellaVigna, Malmendier, JEP 2012**)





Overview Advantages

- Six advantages to Structural Behavioral Econ:
  1. (Calibration) It builds on, and expands, great behavioral tradition of calibrating models: Are magnitudes right?
  2. (Model and Assumptions) Helps to better understand the model and clarifies implicit assumptions
  3. (Stability) Helps to understand whether key behavioral parameters are stable, including out of sample
  4. (Out of Sample) Allows for out of sample predictions which can be tested
  5. (Design) Can lead to better experimental design
  6. (Welfare and Policy) It allows for welfare evaluation and policy counterfactuals

Overview Limitations

- Three limitations to Structural Behavioral Econ:
  1. (Not the Right Tool) Not all questions lend themselves obviously to parameter estimation
  2. (Complexity and Time Costs) It will, generally, take long, and there is higher possibility of errors
  3. (Robustness) Need extra work to make sure estimates are robust, and which assumptions are driving them

Overview Nuts and Bolts

- See the chapter for Nuts and Bolts:
  1. Common methods of estimation
  2. Where the error term comes from
  3. Estimating some common models
  4. Getting Started

Section 2

Advantages

Calibration

- Importance of calibrating models is lesson ONE from behavioral economics
- Example 1: Inertia in retirement savings
  - Standard model can explain *qualitative* pattern given switching costs k
  - But magnitudes? Costs would need to be ridiculous (**O'Donoghue and Rabin, 1998**)
  - Instead, procrastination plausible for naïve β-δ model even with β very close to 1 (O'Donoghue and Rabin, 1999; 2001)

Calibration

- Importance of calibrating models is lesson ONE from behavioral economics
- Example 1: Inertia in retirement savings
  - Standard model can explain *qualitative* pattern given switching costs k
  - But magnitudes? Costs would need to be ridiculous (**O'Donoghue and Rabin, 1998**)
  - Instead, procrastination plausible for naïve β-δ model even with β very close to 1 (O'Donoghue and Rabin, 1999; 2001)
  - Extend O'Donoghue and Rabin calibration with more realistic assumptions (stochastic k)

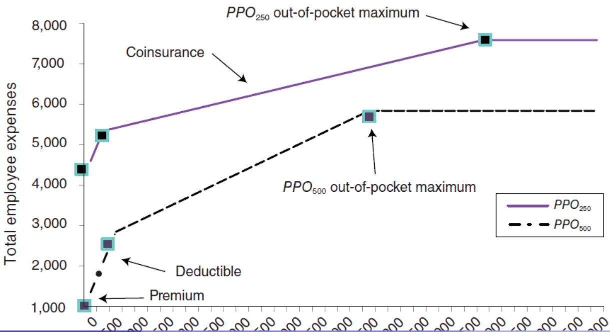
Calibration

- Importance of calibrating models is lesson ONE from behavioral economics
- Example 1: Inertia in retirement savings
  - Standard model can explain *qualitative* pattern given switching costs k
  - But magnitudes? Costs would need to be ridiculous (**O'Donoghue and Rabin, 1998**)
  - Instead, procrastination plausible for naïve β-δ model even with β very close to 1 (O'Donoghue and Rabin, 1999; 2001)
  - Extend O'Donoghue and Rabin calibration with more realistic assumptions (stochastic k)
- Example 2: **Rabin (EMA 2000)** calibration theorem on risk

Calibration

- Inertia in health insurance: **Handel (AER 2013)**
  - Year *t*: firm introduces *new plans* and require *active choice*
  - Year *t+1*: some plans change, choice by default
  - Some plans: year-*t* plan dominated in year *t+1* (but not in t)
  - Do employees still choose it at *t+1*? 80% do!

Panel B. PPO health insurance plan characteristics, t_1 low-income family



Calibration

- Structural estimation
  - value for insurance based on previous risk at *t-1*
  - Models switching cost as cost k to pay when switch (no cost in year t when active choice)

Empirical Model Results					
Parameter	Primary	Base	MH Robust	γ Robust	ϵ Robust
Switching Cost - Single, η_0	1729 (28)	1779 (72)	1859 (107)	2430 (116)	1944 (150)
Switching Cost - Family, $\eta_0 + \eta_2$	2480 (26)	2354 (62)	2355 (113)	3006 (94)	2365 (34)

- Maximum likelihood estimation: \$2,000 clearly unlikely to capture administrative costs
  - $\rightarrow$More likely captures procrastination or inattention
  - (Precise) estimate of \$2,000 drives home the point also to non-behavioral economists

Calibration

- Ref.-dep. job search (**DellaVigna et al, QJE 2017)**
- Fit of ref. dep. Model similar with β model and δ model

Calibration

- Ref.-dep. job search (**DellaVigna et al, QJE 2017)**
- Fit of ref. dep. Model similar with β model and δ model

- BUT δ model has 15-day δ=.9 (implausible impatience)
- β model instead has β =.6 (in range of other estimates)

Model and Assumptions

- Point 1. Strengthens evidence-theory debate
  - Sketch of model will not suffice
  - Full specification in order to do estimation, forced to work out details
  - Run countless simulations at different parameter values
  - $\rightarrow$Leads to understanding model better
  - Does model really do what you thought it would do?
  - $\rightarrow$Can even lead to theoretical break-throughs
  - **Barseghyan, Molinari, O'Donoghue, Teitelbaum** in working out AER 2013 estimation of insurance choice got result on non-identification of loss aversion with KR preferences

Model and Assumptions

- Point 2. Better empirical test
- Example 1: **Genesove and Mayer (QJE)** is pioneering application of reference dependence to housing
  - Individuals hate to sell house at loss relative to purchase price
    - Yet, GM did not work out a model of reference dependence
    - If one writes one, one does not get the GM specification
    - One does get, though, the prediction of bunching at the last house price sale (which they did not test)
  - Example 2: Similar issues (as well known now) for **Camerer et al. (1997)** cab drivers paper

Model and Assumptions

- Point 3. Clarify needed assumptions
- Real-effort experiment (e.g., **Gneezy et al., 2003; Gill et al., 2016**)

$$Effort_{n,s,r} = \beta T + \gamma \boldsymbol{X_{n,s}} + \varepsilon_{n,s,r}$$

- What assumption are behind such specification?

Model and Assumptions

- Point 3. Clarify needed assumptions
- Real-effort experiment (e.g., **Gneezy et al., 2003; Gill et al., 2016**)

$$Effort_{n,s,r} = \beta T + \gamma \boldsymbol{X_{n,s}} + \varepsilon_{n,s,r}$$

- What assumption are behind such specification?
    - Effort is outcome of maximization decision, $max_{e_i} s\left(T_i\right) e_i - \frac{\exp(\gamma e_i)}{\gamma}\eta_i$
    - Assume η log-normal, $ln(\eta_i) \sim N(\gamma k\left(X_i\right), \gamma^2\sigma^2)$.
    - Taking first order conditions,

$$e_i = \frac{1}{\gamma} log\left[s\left(T_i\right)\right] - k\left(X_i\right) + \epsilon_i.$$

- Can estimate with Non-linear Least Squares, almost like OLS (nl in Stata) – Easy!

Model and Assumptions

- Do you buy the needed assumptions?
- Can assume alternative functional form assumptions
- **Power Cost Function:**

$$c(e) = \frac{e^{1+s}}{1+s}$$

- Implied expression for effort is **(DellaVigna, List, Malmendier, and Rao, 2016)**

$$log(e_i) = \frac{1}{\gamma} log\left[s\left(T_i\right)\right] - k\left(X_i\right) + \epsilon_i$$

Stability

- Behavioral economics has convergence on some parsimonious models:
    - Beta-delta model
    - Reference-dependence
- But is there reasonable agreement in parameters across settings?
- Key debate with psychology: can we make any *quantitative* predictions?

Stability

- Behavioral economics has convergence on some parsimonious models:
  - Beta-delta model
  - Reference-dependence
- But is there reasonable agreement in parameters across settings?
- Key debate with psychology: can we make any *quantitative* predictions?
- Case 1. Social Preferences. (Matthew's point)
  - Inequity aversion explains sharing 50-50 in dictator sharing
  - For same parameters, would give half of what is in your wallet to each homeless

Stability

- Case 2. Beta-delta model
  - Evidence from the field of present bias
  - BUT Laboratory evidence (**Andreoni and Sprenger AER**) estimated β close to 1
  - Design with effort choice a la **Augenblick, Niederle, and Sprenger, QJE** appears to solve the puzzle: $\beta = 0.9$ over effort, not on money since timing of money is fungible

Stability

- Case 3. Reference-dependent model
- Most of the focus is on loss aversion parameter λ and on reference point r.
- But probability weighting $\pi(p)$ plays an important role
    - overweighting of small probabilities
    - **Sydnor (2012) and Barseghyan et al:** helps explain home insurance purchases
    - **Barberis (2018):** can explain preferences for IPOs which are skewed
    - Evidence from laboratory lottery choices is strong: $\pi(0.01) = 0.06$

Stability

- Inspired by this, simple design: compare effect of two incentives
  - A. Piece rate of p
  - B. Piece rate of $100p$, paid with probability 0.01
  - With probability weighting, B should be more effective

Table 2b. Evidence for Overweighting of Small Probabilities, Studies with Probabilistic Incentives

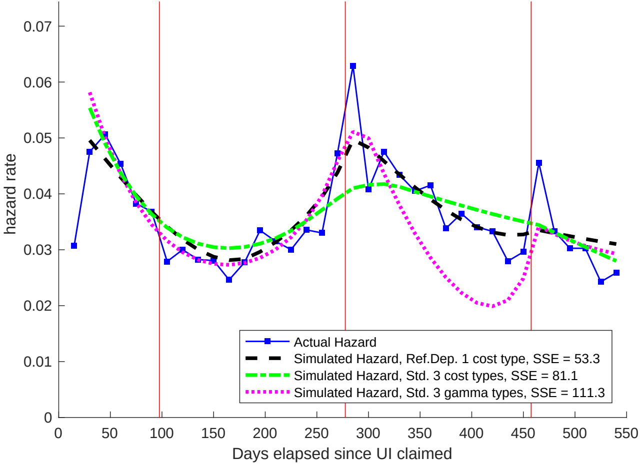
Paper	Subjects	Effort Task	Sample Size	Treatments (Certain Reward vs. Probabilistic Reward with low p)	Effort with Certain Reward, Mean (S.D.)	Effort with Probabilistic Reward, Mean (S.D.)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel B. Field Studies Comparing Certain Reward to Probabilistic Reward						
DellaVigna and Pope (forthcoming)	Mturk	Button Presses	555 (P), 558 (C)	1% chance of winning US$1 (P) vs. fixed payment of US$0.01 (F) per 100 presses	2029 (27.47)	1896 (28.44)
Halpern et al. (2011)	Resident Physicians in a US Database	Survey Response	358 (P), 400 (C)	0.4% chance of winning US$2500 (P) vs. fixed payment of US$10 (F) for response	0.558 (0.497)	0.511 (0.500)
Thirumurthy et al. (2016)	Men aged 21 to 39 years old in Kenya	Uptake of Circumcision	302 (P), 308 (C)	Mixed lottery with expected retail value of US$12.50 (P) vs. food voucher worth US$12.50 (F)	0.084 (0.278)	0.033 (0.179)
Diamond and Loewy (1991)	Undergraduates in State University	Recycling	78 (P), 113 (C)	5% chance of winning $5 and 1% chance of winning $25 (P) vs. $0.50 voucher for campus store (F)	0.212 (0.409)	0.308 (0.462)
Dolan and Rudisill (2014)	16 to 24 year olds in England	Return Test Kit via Mail	247 (P), 549 (C)	10% chance of a 50 GBP Tesco voucher (P) vs. 5 GBP Tesco voucher (F)	0.732 (0.443)	0.706 (0.455)

- Do not find evidence of probability weighting!

Out of Sample

- Key advantage of structural estimates is to allow for out of sample predictions (McFadden et al. 1977; Todd and Wolpin, 2006)
- Some examples in behavioral economics
- Example. Job Search. **DellaVigna et al. (2017 QJE)**
  - Estimate reference dependent model has good fit, even with just 1 type
  - A specific "standard" model (3-type heterogeneity in elasticity) can also fit well
  - Examine out of sample, consider smaller reform 2 years prior
  - Also consider group with lower earnings which had different reform

Out of Sample



(a) Out-of-sample predictions of models for unemployment system 2 years prior to reform and empirical hazard

Experimental Design

- Structural estimation is most often used on observational data (e.g. consumption/savings)
  - In fact, sometimes used as substitute for clear identification (not recommended)
- BUT estimation is *perfect* for experiments
- Advantage 1. Get best of both worlds
  1. Get reduced-form results / treatment effects
  2. PLUS estimate parameters based on experimental results

Experimental Design

- Advantage 2: estimation informs the exp. design
- Idea:
  1. set up model + estimation before running experiment
  2. Create simulated data set (possibly using pilot data)
  3. Attempt to estimate on this data
- Often you will realize that
  - you need an extra treatment in design...
  - or more sample in one treatment...
  - or you are badly underpowered
  - $\rightarrow$Change design! (Cannot do this in obs. studies)
- Different from reduced-form power studies, as this is about estimating the parameters

Experimental Design

- Example 1: Time preference experiments a la **Andreoni and Sprenger** or **Augenblick et al.**
  - Designed so as to estimate time preferences
  - Also, design to estimate confounding parameters (curvature of utility and cost of effort function)

Experimental Design

- Example 1: Time preference experiments a la **Andreoni and Sprenger** or **Augenblick et al.**
  - Designed so as to estimate time preferences
  - Also, design to estimate confounding parameters (curvature of utility and cost of effort function)
- Example 2: Claw-back design **(Hossain and List MS)** with 2 arms
  1. Gain X if reach threshold T -> Return to effort = X
  2. Lose X if do not reach threshold T Return to effort = λX

  - $\rightarrow$Cannot identify λ because do not know cost

Experimental Design

- Example 1: Time preference experiments a la **Andreoni and Sprenger** or **Augenblick et al.**
    - Designed so as to estimate time preferences
    - Also, design to estimate confounding parameters (curvature of utility and cost of effort function)
- Example 2: Claw-back design **(Hossain and List MS)** with 2 arms
    1. Gain X if reach threshold T -> Return to effort = X
    2. Lose X if do not reach threshold T Return to effort = λX

    - $\rightarrow$Cannot identify λ because do not know cost
- **DellaVigna and Pope (RES 2017)** – add third treatment
    3. Gain 2X if reach threshold Return to effort = 2X

    - Now just compare effort in (2) and (3) to see whether λ>2

Experimental Design

- What design tricks facilitate estimation?
- Trick 1. "Price out" behavioral parameters comparing to anintervention in \$
  - **DellaVigna, List, Malmendier (2012)**– survey treatment where advertise \$ payment
  - **Andreoni and Sprenger (AER 2011); Augenblick, Niederle, and Sprenger (QJE 2014)** –variation in interest rate
  - **DellaVigna, List, Malmendier, and Rao, 2016:** Real Effort experiments and identification of cost of effort
- Trick 2. Within-subject experiment
  - Often allows to extract more information, BUT trade-off with simplicity of design
  - **Allcott and Taubinsky (AER 2016)** on energy inattention
  - **Taubinsky and Rees-Jones (RES forthc.)** on limited attention to taxes

Welfare and Policy

- Advantage of estimating model is... you can use it!
  - Compute welfare of setting versus counterfactuals
  - Estimate effect of potential policies
- Some examples:
  - **DellaVigna, List, and Malmendier (QJE 2012)** on charity
  - **Handel (AER 2013)** on health insurance
  - **DellaVigna, List, Malmendier, Rao (REStud 2017)** on voting
  - **Bernheim, Fradkin, Popov (AER 2015)** on retirement saving
  - **Allcott and Taubinsky (AER 2015)** on energy

Welfare and Policy: Voting

- **DellaVigna, List, Malmendier, and Rao (RES 2017)** considers effects of a get-out-the-vote intervention
- Many studies of the impact of GOTV on turnout, not on welfare

	Voters and Non-Voters Have Different Auxiliary Parameters		Voters and Non-Voters Hav Same Auxiliary Parameter	
Implications for GOTV	Voter	Non-Voter	Voter	Non-Vote
Utility from being Asked Once Whether One Voted	-2.8 (1.2)	-5.9 (1.5)	-3.7 (1.6)	-10.6 (2.6)
Implied GOTV Effect from Asking One More Time (N+1)	+0.003 (0.0005)		+0.005 (0.0007)	
Implied Number of Subjects Targeted with GOTV Intervention to Get One Additional Vote (N+1)	295 (84.9)		206 (69.5)	
Utility Cost to Get One Additional Vote (N+1)	1189 (2684.4)		1326 (449.6)	

Section 3

Limitation of Structural Behavioral Economics

Overview Limitations

- Three limitations to Structural Behavioral Econon:
  1. (Not the Right Tool) Not all questions lend themselves obviously to parameter estimation
  2. (Complexity and Time Costs) It will, generally, take long, and there is higher possibility of errors
  3. (Robustness) Need extra work to make sure estimates are robust, and which assumptions are driving them

Limitation 1: Not Right Tool

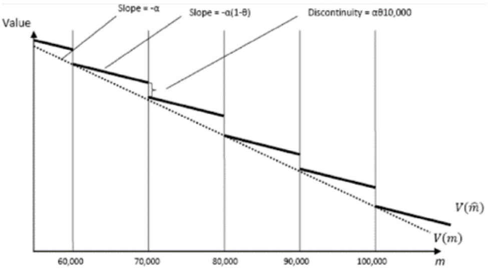
- Not all questions lend themselves obviously to parameter estimation
  - Exploratory question on which we do not have a good model
    - Example: **Dana, Weber, and Kuang (2007)** on moral wriggleroom
    - Example: Framing effects (eg **Benartzi and Thaler, 2002**)
  - Many reduced form/policy question
    - **Bhargava and Manoli AER**
  - Models and Axioms
    - Models can provide comparative statics test of different models, or axioms
    - Can do so without structural estimation, do not need to specify all assumptions

Limitation 2: Complexity and Time Costs

- Estimation complex, takes a long time, need to learn computational tools
- And... there will be bugs in your code
  - Test very extensively, have other people peer review the code
  - Use simulations: simulate and estimate
- Keep in mind the objective: Complexity of the model and estimation is not the aim, it is a necessary evil
- BUT structural model can be simple
  - If rich data provides necessary variation (*sufficient statistic* approach)
  - OR if data collection / experiment is set up to make *estimation simple*

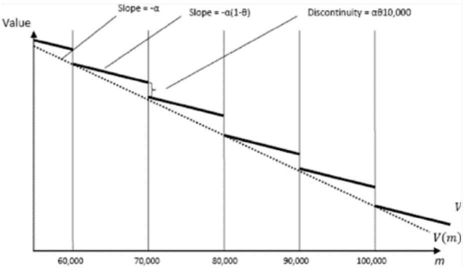
Limitation 2: Complexity and Time Costs

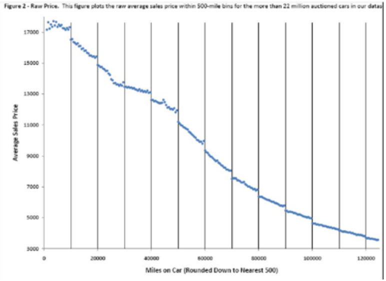
- **Lacetera, Pope, and Sydnor (AER 2012)** – Limited attention θ to left-digit bias for odometer readings



Limitation 2: Complexity and Time Costs

- **Lacetera, Pope, and Sydnor (AER 2012)** – Limited attention θ to left-digit bias for odometer readings





- Using 22 million (!) used car auction transactions
- Can estimate structural parameter θ by... OLS!
  - Delta method to get $\Theta = 0.31$ (s.e. 0.01)

Limitation 3: Robustness to Assumptions

- **Warning 1.** Estimation and implications are only as good as assumptions going into it
  - Test much robustness
- **Warning 2.** Standard errors do not acknowledge model mis-specification → Point estimates are likely too precise
- **Warning 3.** Do you do as well out of sample as in sample?

Limitation 3: Robustness to Assumptions

- Disclose key assumptions: **Allcott and Taubinsky (AER 2016)**
- Consumers make choices between incandescent and CFL

Choice A
Philips 60-Watt-Equivalent Compact Fluorescent Light Bulb, 1-Pack

Choice B
Philips 60-Watt Incandescent Light Bulbs, 4-Pack

Decision Number	Choice A 60-Watt-Equivalent Compact Fluorescent Light Bulb, 1-Pack	Choice B 60-Watt Incandescent Light Bulbs, 4-Pack
	Purchase **Choice A for free**	Purchase **Choice B for $10**
1)	○	○
	Purchase **Choice A for free**	Purchase **Choice B for $8**
2)	○	○
	Purchase **Choice A for free**	Purchase **Choice B for $6**
3)	○	○

- One group receives information on energy savings, other not

Limitation 3: Robustness to Assumptions

- Moderate Shift in demand curve due to information
- What if not sure about some key assumptions?

Row	Scenario	(1) Optimal Subsidy ($/pckg)	(2) Welfare Effect of Ban ($/pckg)	(3) Effect of Ban (Percent of "Perceived Surplus")
1	Base	3	-0.44	-41
	If censored, assume ...			
2	WTP={$12,-$12}	3	-0.34	-36
3	WTP={$20,-$20}	3	-0.60	-47
4	self-reported hypothetical WTP	3	-0.61	-43
	Scale average marginal bias to match ...			
5	consumers who pass review "quiz"	3	-0.41	-38
6	consumers w/ "correct" endline beliefs	3	-0.13	-12
7	Balanced Treatment group	3	-0.48	-45
8	10 percent confidence bound	1	-0.92	-86
9	90 percent confidence bound	(Ban)	0.05	4
	Additional Distortion Computed from Excess Mass Test			
10	Excess mass consumers have $v = 7.66$	8	1.22	114

Section 4

Nuts and Bolts

Common Methods of Estimation

1. Minimum Distance Estimator / Method of Moments
2. Maximum Likelihood
3. Others (OLS / NLLS / "Bunching")

Minimum Distance

- Minimum Distance Steps
  1. Pick the observed empirical moments to match, $\hat{m}$
  2. Solve/ Simulate the model at a given set of parameters θ and generate the same moments, $m(\theta)$
  3. Find the set of parameters that minimize the distance between the empirical and model-generated moments
     $\hat{\theta} = argmin(m(\theta) - \hat{m})' W (m(\theta) - \hat{m})$

  - Example: **Laibson, Maxted, Repetto, and Tobacman (2015)**
    - Study consumption-savings of households
    - Step 1: Many possible moments, pick important ones
    - Step 2 (hardest for them): Write full consumption-savings problem as function of model parameters
    - Step 3 (easy): Grid search of parameters such that would have good fit

Empirical Moments used in Method of Simulated Moments

% Visa 21-30	0.815
% Visa 31-40	0.782
% Visa 41-50	0.749
% Visa 51-60	0.659
mean Visa 21-30	0.199
mean Visa 31-40	0.187
mean Visa 41-50	0.261
mean Visa 51-60	0.276
wealth 21-30	1.23
wealth 31-40	1.86
wealth 41-50	3.24
wealth 51-60	5.34

Structural Estimation Results

	Present Biased	Exponential	Data
Parameter estimates			
	0.5054	**1**	-
	-0.1481	-	-
	0.9872	**0.8926**	-
	-0.0089	-0.0083	-
CRRA	**1.2551**	**1.0047**	-
	-0.1564	-0.2857	-
Second-stage moments			
% Visa 21-30	0.598	0.704	0.815
% Visa 31-40	0.607	0.693	0.782
% Visa 41-50	0.588	0.654	0.749
% Visa 51-60	0.569	0.601	0.659
mean Visa 21-30	0.232	0.204	0.199
mean Visa 31-40	0.237	0.225	0.187
mean Visa 41-50	0.217	0.21	0.261
mean Visa 51-60	0.196	0.193	0.276
wealth 21-30	1.299	0.441	1.23
wealth 31-40	1.819	0.015	1.86
wealth 41-50	2.925	-0.047	3.24
wealth 51-60	[illegible]	[illegible]	[illegible]

Minimum Distance

- Advantages of Minimum Distance:
  - Transparent in what identifies: moments you pick
  - Perfect for experiments:
    - Moment 1 is X in control, Moment 2 is X in treatment
  - Also, allows you to post the moments even if data is confidential
- Disadvantages of Minimum Distance:
  - Does not use all the information in the data, only what is contained in moments
  - Sensitive to choice of moments

Maximum Likelihood

- Maximum Likelihood Steps
  1. Specify fully the statistical model, deriving likelihood $L(x|\theta)$, where x is data
  2. Maximize likelihood given the data, picking parameters $\hat{\theta}$
- **Augenblick and Rabin (RES, forthc.)**
  - Model implies that optimal effort is given by

$$e^* = \arg\max_e \quad \delta^{T-k} \cdot (e \cdot w) - \frac{1}{\beta^{\mathbf{1}(k=t)}} \cdot \frac{1}{\beta_h^{\mathbf{1}(p=1)}} \cdot \delta^{t-k} \cdot \frac{1}{\varphi \cdot \gamma}(e+10)^{\gamma}.$$

  - Leads to first order condition

$$e^* = \left(\frac{\delta^{T-k} \cdot \varphi \cdot w}{\frac{1}{\beta^{\mathbf{1}(k=t)}} \cdot \frac{1}{\beta_h^{\mathbf{1}(p=1)}} \cdot \delta^{t-k}}\right)^{\frac{1}{\gamma-1}} - 10.$$

  - Assuming additive noise in observed effort yields likelihood

$$L(e_j) = \phi\left(\frac{e_j^* - e_j}{\sigma}\right)$$

Maximum Likelihood

- Advantages of Maximum Likelihood:
  - Uses all the information in the data
- Disadvantages of Minimum Distance:
  - Identification is less transparent
  - Observations with a very low likelihood could be driving the results

II. Randomness

- Consider a typical model solution, eg., Nash eq., or Market Equilibrium
- Solution often is one equilibrium action, or one price
- Yet, in reality we always a distribution of outcomes
- Where does the randomness come from?
- That will be key step to take model into econometrics.
- Three broad categories
  1. Random utility (McFadden logit)
  2. Random coefficients
  3. Implementation error

Section 5

Conclusion

Conclusion

- Summary: Behavioral economics is normal science
- As such, we should see a variety of approaches used in empirical work, including structural estimates
- Much more in the chapter!