Econ 219B
Psychology and Economics: Applications
(Lecture 13)

Stefano DellaVigna

April 18, 2018

Outline

1. Behavioral Political Economy II
2. Methodology: Structural Behavioral Economics
3. Behavioral Finance
4. Behavioral Corporate Finance
5. Behavioral Labor

Section 1

Behavioral Political Economy II

What explains political participation?

- Olson (1965): Public good problem: Even if think participation is right, individually better off staying at home
  - Example 1: Riots and protests
  - Example 2: Voter turnout at the polls $\rightarrow$ Probability of being pivotal very small
- Series of papers introduce variants of social preferences to explain participation in political activities

- **Passarelli and Tabellini (2013):**
  - Focus on protests
  - Assume negative reciprocity and role of emotions
  - Individuals treated poorly by government get glow from protesting

Model

- For individual i:
  - Cost of participating to protest ε_i
  - Psychological benefit of participation to protest a_i
  - Benefit a_i depends on aggrievement:

$$a_i = \begin{cases} 0 & \text{if} \quad V_i \geq \hat{V} \\ \omega \left(V - \hat{V} \right)^2 & \text{if} \quad V_i < \hat{V} \end{cases}$$

  - V_i is welfare of individual i with given policy
  - $\hat{V}$ is what individual thinks appropriate (can be self-biased)
  - Ad-hoc form of reference dependence
  - When aggrieved, individual willing to incur cost of participation because of glow from participation

DellaVigna, List, Malmendier, Rao (REStud 2017)

- Related idea: Explain voter turnout with social preferences
- Tie to social interactions
- Identify using field experiment design

Determinants of Voting

Four determinants of voting

1. Pivotality pV
   - p = subjective probability of being pivotal
   - V = value of deciding the election
2. Warm glow g
3. Cost of voting c
   - cost of voting
4. Social Image utility

   Focus of this paper

   - s_V = utility from saying one voted
   - s_N = utility from saying one did not vote   **social image**
   - L = psychological cost of lying   **dishonesty**

- Non-voters lie about voting if $s_V - L > s_N \leftrightarrow s_V - s_N > L$
- Voters lie if $s_N - L > s_V$

(Net) Expected Utility from Voting

Voting iff

$$\underbrace{pV + g - c}_{= \, \varepsilon} + \underbrace{N \left[\max\left(s_V, s_N - L\right) - \max\left(s_N, s_V - L\right)\right]}_{\substack{= \textit{ net utility gain from having voted,} \\ \textit{due to being asked once}}} \geq 0$$

Can rewrite as: $N\Phi\left(s_v - s_L, L\right) + \varepsilon \geq 0$

where $\quad \Phi\left(s_v - s_L, L\right) = \begin{cases} \min(s_V - s_N, L) & \text{if } s_V - s_N \geq 0 \\ \max(s_V - s_N, L) & \text{if } s_V - s_N < 0 \end{cases}$

Experimental Design

- Field experiment: door-to-door survey
  - Match households to voting records
  - Identify all-voter and all-non-voter households

- Cross-randomize
  1. Whether individuals receive advance notice of survey.
     - Individuals can avoid (or seek) surveyor at a cost.
  2. Vary payment and length of survey to estimate elasticity
  3. Incentives to lie / tell truth about voting.

- Get-Out-The-Vote experiment related to model
  - Inform some people that we will visit them after the election to ask whether they voted

Field Experiment - Implementation

- Single-family homes in towns around Chicago
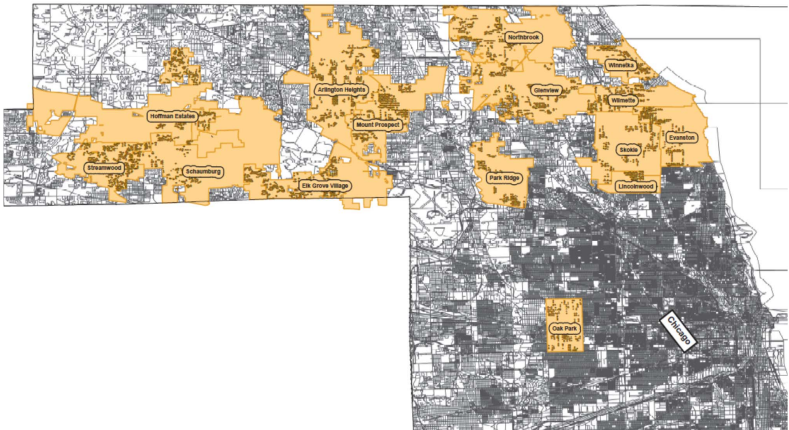
Exp 1: Announcing Content of Survey


Control: Unannounced Visit

Model Predictions

- **Prop. 1.** With pride in voting ($s_V > 0$), voters should be more likely to be at home and answer the door if informed of election survey

- **Prop. 2.** With stigma from not voting ($s_N < 0$), non-voters should be less likely to be at home and answer the door if informed of election survey

- **Prop. 3.** The probability of lying about voting should increase in the incentive to do so

- **Prop. 4.** The probability of voting should increase in the number of times asked N

- **Sorting in Response to Election Survey -- Voters**
- Voters -> No evidence of sorting in, some evidence of sorting
- No evidence of pride in voting on average
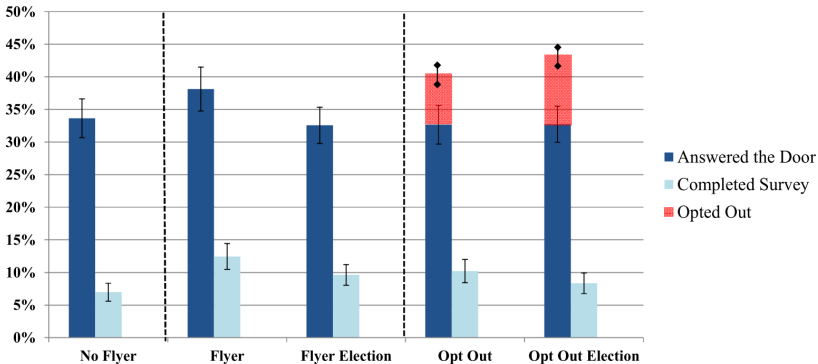

Voters (N=6,873)
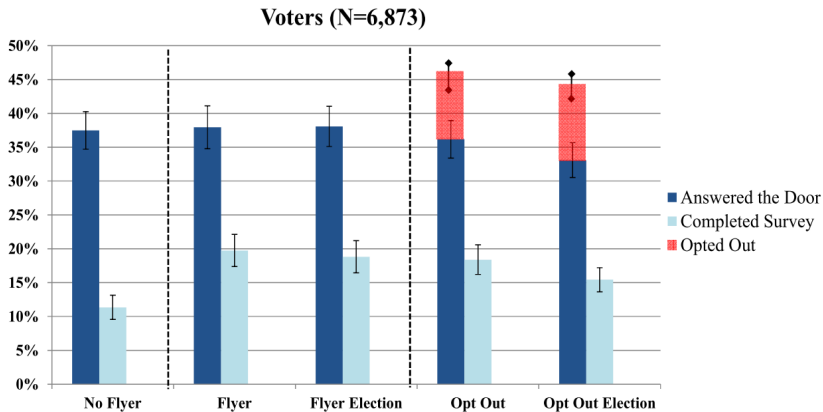
- **Sorting in Response to Election Survey -- Voters**
- However, 2010 election was low point for democratic voters
- 2/3 of registered voters in towns we reached are Democrats
- What if we split by voting record in primaries?
- Evidence of sorting in for Republicans

Republican Voters (N=1,918)   **Democratic Voters (N=3,018)**

■ Answered the Door   □ Completed Survey
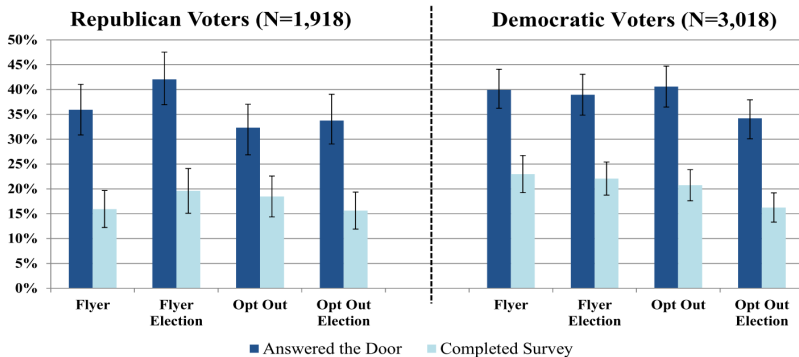
- **Sorting in Response to Election Survey – Non-Voters**
- Non-voters-> Strong evidence of sorting out
- Evidence of stigma from not voting and lying costs

Non-Voters (N=6,324)

Legend:
- Answered the Door
- Completed Survey
- Opted Out

Categories: No Flyer, Flyer, Flyer Election, Opt Out, Opt Out Election

Exp 2: Varying payment and length of svy

THE UNIVERSITY OF
CHICAGO

University of
Chicago Study

Researchers will
visit this address
tomorrow (/)
between   and
to conduct a
5 minute survey.

THE UNIVERSITY OF
CHICAGO

University of
Chicago Study

Researchers will visit this
address tomorrow (/)
between   and   to
conduct a **10 minute
survey**.

You will be paid **$10 in
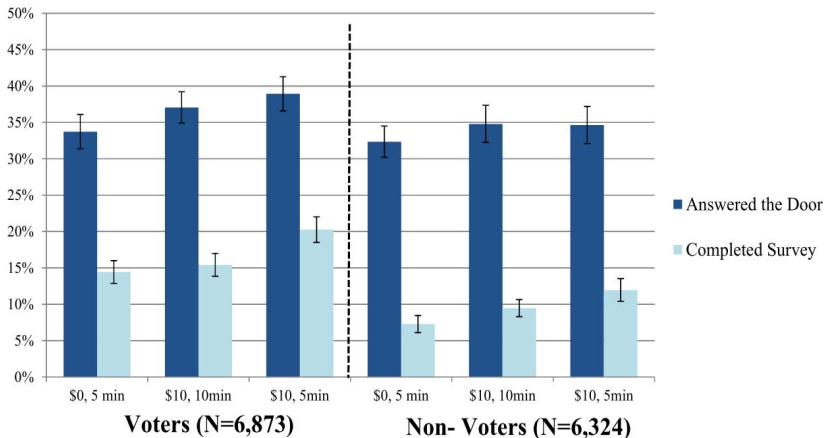cash** for your participation.

THE UNIVERSITY OF
CHICAGO

University of
Chicago Study

Researchers will visit this
address tomorrow (/)
between   and   to
conduct a **5 minute
survey**.

You will be paid **$10 in
cash** for your participation.

- **Response to Incentives**
- Response to payment and duration
- Election warning effect on non-voters ~ **$10** decrease in pay

Exp. 3: Lying Incentives

- Crossed treatment: Incentive to lie in 10-minute survey
- *No Incentive.* Just ask whether voted in 2010 election
- *8-Minute Incentive.* (<u>8 minute incentive to say 'did not vote'</u>)
  - "**We have 10 minutes of questions about your voter participation in the 2010 congressional election, but if you say that you did not vote then we only have 2 minutes of questions. Either way you answer you will be paid $10.** [Show the end of the survey if answer to #2 is NO]
  Did you vote in the 2010 congressional election?"

  - For voters it is incentive to <u>lie</u>
  - For non-voters this is incentive to tell <u>truth</u>

  - Novel survey instrument → Use to estimate *counterfactual* utility

Lying Incentives

- In 5-minute surveys:
- *No Incentive.* Just ask whether voted in 2010 election
- *$5 Incentive.* (<u>$5 incentive to say did not vote</u>)
  - "**We have 5 minutes of questions about your participation in the 2010 congressional election, but if you say that you did not vote then we would like to ask you an extra 1 minute of questions and we will pay you an extra $5 for answering these additional questions [*IF PAID*: for a total of $15]. If you say that you voted then we will just ask you the original 5 minutes of questions. [*IF PAID*: Either way you answer you will be paid $10.] Did you vote in the 2010 congressional election?"**

  - Incentive to lie for voters, to tell the truth for non-voters

- **Response to Incentives to Say 'Did Not Vote'**
- Small impact on voters: 2 percentage points increase in lying
  → Strong social image utility and/or lying cost
- Sizeable impact on non-voters: 12 percentage point decrease in lying → Non-voters are closer to indifference
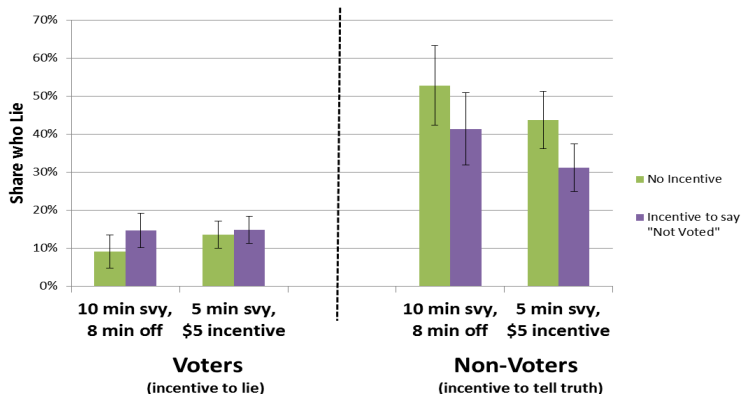
Structural Estimation

- **Structural estimates (Minimum-distance estimator)**

- Minimize distance between predicted moments $m\left(\vartheta\right)$ and observed ones $\hat{m}$

$$\min_{\vartheta} \left(m\left(\vartheta\right) - \hat{m}\right)' W \left(m\left(\vartheta\right) - \hat{m}\right)$$

- Moments $m\left(\vartheta\right)$:
  1. Probability of opening door to surveyor $\left(P\left(H\right)_j^S\right)$
  2. Probability of filling survey $\left(P\left(S\right)_j^S\right)$
  3. Probability of checking the opt-out box
  4. Probability of lying about voting

- All moments $\hat{P}$ are probabilities, straight from Figures

Election Field Experiment - Estimation

- What is ϑ?

- Main parameters
  - mean and s.d. of s_i^V – signalling utility of saying one voted
  - mean and s.d. of s_i^N – signalling utility of saying one did not vote
  - $L_i \geq 0$ – lying cost

- Auxiliary parameters:
  - Willingness to do survey
  - Value of time
  - Cost of avoiding suveyor

Estimation with Selection

- Estimation approach: Incorporate selection into V/NV

- Parameters (sV, sN) *predict* becoming voter or non-voter

$$pV + g - c + N\left[\max\left(s_V, s_N - L\right) - \max\left(s_N, s_V - L\right)\right] \geq 0$$

$$= \varepsilon$$

  - Assume epsilon Normal
  - Voters and non-voters drawn from same population
  - Draw parameters, determine selection into voters or non-voters
  - Match to moments using simulations
  - Assume number of times asked N from survey
  - Additional moment: baseline turnout rate (60 percent)

- Total value of voting depends on *N*
- Survey: *How often have you been asked whether you voted?*
  - *9 times* for 2008 presidential election

Estimation with Selection

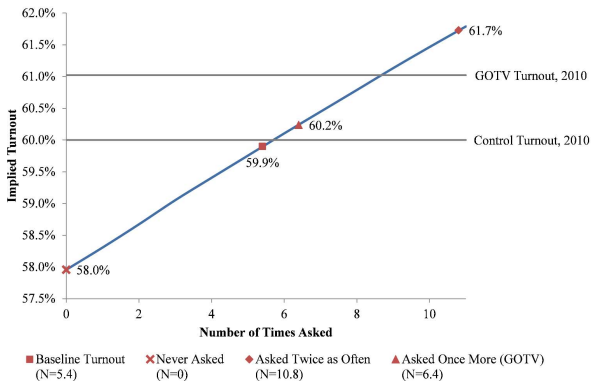**Table 3. Simulated Minimum-Distance Estimates, Benchmark Results**

	Voters and Non-Voters Have Same Auxiliary Parameters	Voters and Non-Voters Have Different Auxiliary Parameters
Voting Parameters		
Mean Social Image Value of Saying Voted (s_V)	-6.3	-3.9
	(2.07)	(1.47)
Mean Social Image Value of Saying Did Not Vote (s_N)	-21.7	-11.3
	(3.19)	(1.77)
Std. Dev. of s_V and s_N	19.7	9.5
	(2.83)	(1.29)
Lying Cost L (in $)	16.4	7.6
	(2.82)	(1.21)
Mean Value of Other Reasons to Vote (ε)	95.0	64.1
	(114.33)	(167.90)
Std. Dev. of Other Reasons to Vote (ε)	490.6	318.7
	(454.75)	(691.37)

- Lying cost L estimated

Estimation with Selection

- Implications: estimate impact on voting if
  - No one asked
  - Twice as many people asked
  - Also impact of being asked one more time (next)

Baseline Turnout (N=5.4) | Never Asked (N=0) | Asked Twice as Often (N=10.8) | Asked Once More (GOTV) (N=6.4)

Estimation with Selection

- Other implications of estimates

Table 4. Implied Value of Voting and Welfare Effects of GOTV

	Voters and Non-Voters Have Same Auxiliary Parameters		Voters and Non-Voters Have Different Auxiliary Parameters	
Implications for Value of Voting to Tell Others	Voter	Non-Voter	Voter	Non-Voter
Implied Value of Voting "To Tell Others" (N=5.4)	41.4	26.1	18.3	13.3
	(5.6)	(10.2)	(4.6)	(3.3)
Baseline Turnout	0.604		0.599	
	(0.011)		(0.011)	
Implied Change in Turnout if Never Asked About Voting	-0.027		-0.019	
	(0.0153)		(0.0031)	
Implied Change in Turnout if Asked About Voting Twice as Often	+0.025		+0.018	
	(0.0081)		(0.0079)	
Implications for GOTV	Voter	Non-Voter	Voter	Non-Voter
Utility from being Asked about Voting Once	-3.7	-10.6	-2.8	-5.9
	(1.6)	(2.6)	(1.2)	(1.5)
Implied GOTV Effect (N+1)	+0.005		+0.003	
	(0.0007)		(0.0005)	
Implied Number of GOTV Subjects to Get One Additional Vote (N+1)	206		295	
	(69.5)		(84.9)	
Disutility Cost of Getting One Additional Vote (N+1)	-1326		-1189	
	(449.6)		(2684.4)	

Prospective Election Field Experiment

- If estimates are correct, being asked one more time increases the value of voting by $1.50-$3.00

- Experiment in week before elections in 2010 and 2012
  - Control (C) group: No contact
  - Control Flyer (CF) group: Flyer reminds households to vote
  - Treatment Flyer (TF) group: Flyer reminds households to vote, AND announces that a surveyor will come by to ask whether they voted in one of the following three weeks

- Comparison of turnout rate in TF group versus CF group provides evidence on impact of social image motive on voting

Prospective Election Field Experiment

THE UNIVERSITY OF
CHICAGO

University of Chicago
Study

**Don't forget to vote in the
2012 Presidential Election.**

**Election Day is Tuesday,
November 6, 2012.**

- Control
  Flyer

  ←

- Treatment
  Flyer

  →

THE UNIVERSITY OF
CHICAGO

University of Chicago
Study

**Researchers will contact
you within three weeks of
the election (between 11/7
and 11/27) to conduct a
survey on your
voter participation.**

**Don't forget to vote in the
2012 Presidential Election.**

**Election Day is Tuesday,
November 6, 2012.**

Prospective Election Experiment

Table 7. Results for Get-Out-The-Vote Treatments

Specification:	OLS Regressions			
Dependent Variable:	Indicator for Voting in Election in Year t			
Election:	Congressional Elections in Nov. 2010		Presidential Elections in Nov. 2012	
	(1)	(2)	(3)	(4)
Constant	0.6000***		0.7312***	
	(0.0109)		(0.0033)	
Flyer with Voting Reminder	-0.0020	-0.0031	0.0060	0.0046
	(0.0152)	(0.0083)	(0.0056)	(0.0034)
Flyer with Announcement Will Ask About Voting	0.0120	0.0102	0.0023	0.0056
	(0.0157)	(0.0084)	(0.0056)	(0.0034)
Omitted Treatment	No Flyer		No Flyer	
Control for past Voting since 2004		X		X
Difference (Flyer Will Ask - Flyer Reminder)	0.0140	0.0133	-0.0037	0.0010
p-value for test of equality, 2-sided	p=0.365	p=0.120	p=0.561	p=0.811
p-value for test of equality, 1-sided	p=0.182	p=0.060*		p=0.405
R2	0.0001	0.4024	0.0000	0.3251
N	N = 31,306	N = 31,304	N = 93,805	N = 93,805

- 1.3pp. effect in 2010 (marg. Significant 1-sided)
- 0.1pp. Effect in 2012 (highly competitive election)
- Estimates consistent with predicted small effect from model

Section 2

Methodology: Structural Behavioral Economics

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**)
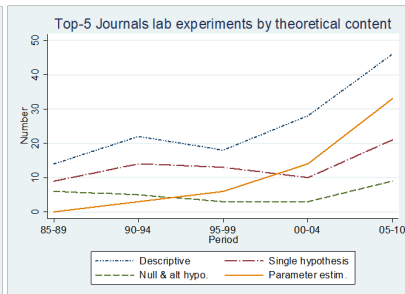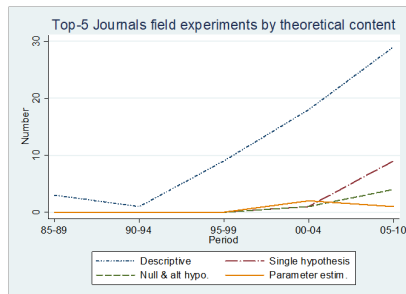
Overview Advantages

- Five 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

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)
  - Problem set 1: Extend O'Donoghue and Rabin calibration with more realistic assumptions (stochastic k)
- Example 2: **Rabin (EMA 2000)** calibration theorem on risk

Calibration

- The lesson of calibration is that it is important to check for reasonable values of the parameters
- Example. **DellaVigna et al., 2017** on job search
    - Exp disconting and $\beta\delta$ model have about same fit
    - BUT δ model has 15-day δ=.9 (implausible impatience)
    - β model instead has β =.6 (in range of other estimates)
    - estimated values for exp discounting outside the range of plausible, not so for $\beta\delta$ model

Model and Assumptions

- Point 1. <u>Strengthens evidence-theory debate</u>
    - 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
    - →Leads to understanding model better
    - Does model really do what you thought it would do?
    - →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. <u>Better empirical test</u>
- 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. <u>Clarify needed assumptions</u>
- 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 of time preference (Laibson, 1997; O'Donoghue and Rabin, 1999)
    - Reference-dependence model of risk preferences (Kahneman and Tversky, 1979; Koszegi and Rabin, 2006
- For these models, is there reasonable agreement in parameters across settings?
- Case 1. Beta-delta model
    - Evidence from the field provided evidence 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 2. 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 Overwighting 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

- Structural estimates allow for out of sample predictions (McFadden et al. 1977; Todd and Wolpin, 2006)
- Some examples in behavioral economics
- Example 1. Social Image. **DellaVigna et al. (2017 RES)**.
    - Estimate impact of get-out-the-vote intervention based on flyer experiment
- Example 2. 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

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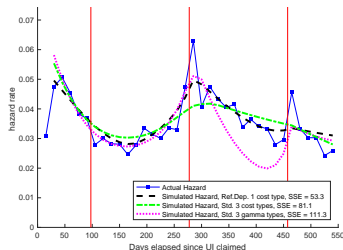
(a) Out-of-sample predictions of models for unemployment system 2 years prior to reform and empirical hazard
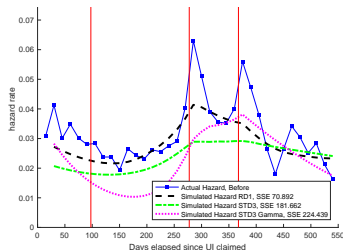
Experimental Design

- 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)
- Example 2: Charity and Social Pressure Experiment **(DLM QJE)**
  - Add entire set of experiments to estimate cost of sorting in/out of the home
- Example 3. Limited attention and taxation
  - **Allcott and Taubinsky (AER 2016)** on energy inattention
  - **Taubinsky and Rees-Jones (2016)** on limited attention to taxes

Experimental Design

- What design tricks facilitate estimation?
- Trick 1. "Price out" behavioral parameters comparing to an intervention in $
  - **DLM (2012)**– survey treatment where advertise $ payment
  - **Andreoni and Sprenger (AER 2011); Augenblick, Niederle, and Sprenger (QJE 2014)** –variation in interest rate
- 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 (2016)** on limited attention to taxes
  - Real Effort experiments and identification of cost of effort **(DellaVigna, List, Malmendier, and Rao, 2016)**

Welfare and Policy

- Advantage of estimating model is. . . you can use it!
    - Compute <u>welfare</u> of setting versus counterfactuals
    - Estimate effect of potential <u>policies</u>
- 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

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

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 can be relatively easy, but it typically is not
- You will learn a lot about optimization methods!
- And... there will be bugs in your code
    - Test very extensively, have other people peer review the code
    - Use simulations: simulate and estimate
    - Example of error: **Apesteguia and Ballester (JPE forthcoming)** point flaw in earlier Harrison paper
- 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 3: Robustness to Assumptions

- What about pitfalls?
- Assume you fully specify model and estimate it
    - You spend years of life doing it
    - Model implies welfare and policy implications
    - Now you sell the implications for policy and welfare
- **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 $\rightarrow$ Point estimates are likely too precise
- **Warning 3.** Do you do as well out of sample as in sample?

Limitation 3: Robustness to Assumptions

- How to avoid pitfall: **Allcott and Taubinsky (AER 2016)**
- Consumers make choices between incandescent and CFL

- 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

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, Repetto, Maxted, 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

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

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

  - 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 3

Behavioral Finance

Behavioral Finance Anomalies

- How do 'smart' investors respond to investors with biases?

- First, brief overview of anomalies in Asset Pricing (from Barberis and Thaler, 2004)

1 **Underdiversification.**

   1 Too few companies.

- Investors hold an average of 4-6 stocks in portfolio.
- Improvement with mutual funds

   2 Too few countries.

- Investors heavily invested in own country.
- Own country equity: 94% (US), 98% (Japan), 82% (UK)
- Own area: own local Bells (Huberman, 2001)

   3 Own company

- In companies offering own stock in 401(k) plan, substantial investment in employer stock

② **Naive diversification.**

- Investors tend to distribute wealth 'equally' among alternatives in 401(k) plan (Benartzi and Thaler, 2001; Huberman and Jiang, 2005)

③ **Excessive Trading.**

- Trade too much given transaction costs (Odean, 2001)

④ **Disposition Effect in selling.**

- Investors more likely to sell winners than losers

⑤ **Attention Effects in buying.**

  - Stocks with extreme price or volume movements attract attention (Odean, 2003)

⑥ **Inattention to Fees.**

⑤ Attention Effects in buying.

- Stocks with extreme price or volume movements attract attention (Odean, 2003)

⑥ Inattention to Fees.

Should market forces and arbitrage eliminate these phenomena?

Arbitrage

- By assumption:
    - Individuals attempt to maximize individual wealth
    - They take advantage of opportunities for free lunches

- Implications of arbitrage: 'Strange' preferences do not affect pricing
- Implication: For prices of assets, no need to worry about behavioral stories

- Is it true?

Fictitious Example

- Asset A returns \$1 tomorrow with $p = .5$
- Asset B returns \$1 tomorrow with $p = .5$

- Arbitrage $\rightarrow$ Price of A has to equal price of B
- If $p_A > p_B$,
    - sell A and buy B
    - keep selling and buying until $p_A = p_B$
- Viceversa if $p_A < p_B$

But...

- Problem: Arbitrage is limited (de Long et al., 1991; Shleifer, 2001)
- In Example: can buy/sell A or B and tomorrow get fundamental value
- In Real world: prices can diverge from fundamental value

- Real world example. Royal Dutch and Shell
    - Companies merged financially in 1907
    - Royal Dutch shares: claim to 60% of total cash flow
    - Shell shares: claim to 40% of total cash flow
    - Shares are nothing but claims to cash flow
    - Price of Royal Dutch should be $60/40=3/2$ price of Shell

Royal Dutch and Shell

- p_{RD}/p_S differs substantially from 1.5 (Fig. 1)
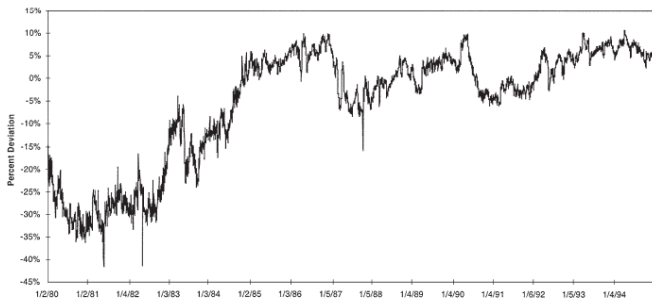
Fig. 1. Log deviations from Royal Dutch/Shell parity. Source: Froot and Dabora (1999).

- Plenty of other examples (Palm/3Com)

What is the problem?

- Noise trader risk, investors with correlated valuations that diverge from fundamental value
  - Example: Naive Investors keep persistently bidding down price of Shell
- In the long run, convergence to cash-flow value
- In the short-run, divergence can even increase
  - Example: Price of Shell may be bid down even more

Noise Traders

- DeLong, Shleifer, Summers, Waldman (*JPE* 1990)
- Shleifer, *Inefficient Markets*, 2000
- Fundamental question: What happens to prices if:
    - (Limited) arbitrage
    - Some irrational investors with correlated (wrong) beliefs
- First paper on Market Reaction to Biases
- *The* key paper in Behavioral Finance

Model Assumptions

A1: arbitrageurs risk averse and short horizon

$\longrightarrow$ Justification?

- Short-selling constraints
  (per-period fee if borrowing cash/securities)
- Evaluation of Fund managers.
- Principal-Agent problem for fund managers.

Model Assumptions

A2: noise traders (Kyle 1985; Black 1986)

- misperceive future expected price at t by $\rho_t \overset{i.i.d.}{\sim} \mathcal{N}(\rho^*, \sigma_\rho^2)$
- misperception *correlated* across noise traders ($\rho^* \neq 0$)

$\longrightarrow$ Justification?

- fads and bubbles (Internet stocks, biotechs)
- pseudo-signals (advice broker, financial guru)
- behavioral biases / misperception riskiness

What else?

- μ noise traders, $(1 - \mu)$ arbitrageurs
- OLG model
    - Period 1: initial endowment, trade
    - Period 2: consumption
- Two assets with identical dividend r
    - safe asset: perfectly elastic supply
      $\implies$ price=1 (numeraire)
    - unsafe asset: inelastic supply (1 unit)
      $\implies$ price?
- Demand for unsafe asset: λ^a and λ^n, with $\lambda^n \mu + \lambda^a (1 - \mu) = 1$.
- CARA: $U(w) = -e^{-2\gamma w}$ (w wealth when old)

$$
\begin{aligned}
E\left[U(w)\right] &= \int_{-\infty}^{\infty} -e^{-2\gamma w} \cdot \frac{1}{\sqrt{2\pi\sigma_w^2}} \cdot e^{-\frac{1}{2\sigma^2}(w-\bar{w})^2}\, dw \\
&= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_w^2}} \cdot e^{-\frac{4\gamma w \sigma_w^2 + w^2 + \bar{w}^2 - 2w\bar{w}}{2\sigma_w^2}}\, dw \\
&= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_w^2}} \cdot e^{-\frac{(w-[\bar{w}-2\gamma\sigma_w^2])^2 - 4\gamma^2\sigma_w^4 + 4\gamma\sigma_w^2\bar{w}}{2\sigma_w^2}}\, dw \\
&= -e^{-\frac{4\gamma\sigma_w^2\bar{w} - 4\gamma^2\sigma_w^4}{2\sigma_w^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_w^2}} \cdot e^{-\frac{(w-[\bar{w}-2\gamma\sigma_w^2])^2}{2\sigma_w^2}}\, dw \\
&= -e^{-\frac{4\gamma\sigma_w^2\bar{w} - 4\gamma^2\sigma_w^4}{2\sigma_w^2}} = -e^{-2\gamma\left(\bar{w}-\gamma\sigma_w^2\right)} \\
&\Downarrow \\
&\max E\left[U(w)\right] \quad \underset{\text{pos. mon. transf.}}{\curvearrowright} \quad \max \bar{w} - \gamma\sigma_w^2
\end{aligned}
$$

Arbitrageurs:

$$\max(w_t - \lambda_t^a p_t)(1 + r)$$
$$+\lambda_t^a(E_t[p_{t+1}] + r)$$
$$-\gamma \left(\lambda_t^a\right)^2 Var_t(p_{t+1})$$

Noise traders:

$$\max(w_t - \lambda_t^n p_t)(1 + r)$$
$$+\lambda_t^n(E_t[p_{t+1}] + \rho_t + r)$$
$$-\gamma \left(\lambda_t^n\right)^2 Var_t(p_{t+1})$$

(Note: Noise traders know how to factor the effect of future price volatility into their calculations of values.)

f.o.c.

Arbitrageurs: $\frac{\partial E[U]}{\partial \lambda_t^a} \stackrel{!}{=} 0$

$$\lambda_t^a = \frac{r + E_t[p_{t+1}] - (1+r)p_t}{2\gamma \cdot Var_t(p_{t+1})}$$

Noise traders: $\frac{\partial E[U]}{\partial \lambda_t^n} \stackrel{!}{=} 0$

$$\lambda_t^n = \frac{r + E_t[p_{t+1}] - (1+r)p_t}{2\gamma \cdot Var_t(p_{t+1})}$$
$$+ \quad \frac{\rho_t}{2\gamma \cdot Var_t(p_{t+1})}$$

Interpretation

- Demand for unsafe asset function of:
  - $(+)$ expected return $(r + E_t[p_{t+1}] - (1 + r)p_t)$
  - $(-)$ risk aversion (γ)
  - $(-)$ variance of return $(Var_t(p_{t+1}))$

  - $(+)$ overestimation of return ρ_t (noise traders)

- Notice: noise traders hold more risky asset than arb. if $\rho > 0$ (and viceversa)
- Notice: Variance of prices come from noise trader risk. "Price when old" depends on uncertain belief of next periods' noise traders.

- Impose general equilibrium: $\lambda^n \mu + \lambda^a (1 - \mu) = 1$ to obtain

$$
\begin{aligned}
1 &= \frac{r + E_t[p_{t+1}] - (1+r)p_t}{2\gamma \cdot Var_t(p_{t+1})} + \mu \frac{\rho_t}{2\gamma \cdot Var_t(p_{t+1})} \text{ or} \\
p_t &= \frac{1}{1+r} \left[r + E_t[p_{t+1}] - 2\gamma \cdot Var_t(p_{t+1}) + \mu \rho_t \right]
\end{aligned}
$$

- To solve for p_t, we need to solve for $E_t[p_{t+1}] = E[p]$ and $Var_t(p_{t+1})$

$$
\begin{aligned}
E[p] &= \frac{1}{1+r} \left[r + E_t[p] - 2\gamma \cdot Var_t(p_{t+1}) + \mu E[\rho_t] \right] \\
E[p] &= 1 + \frac{-2\gamma \cdot Var_t(p_{t+1}) + \mu \rho^*}{r}
\end{aligned}
$$

- Rewrite p_t plugging in

$$p_t = 1 - \frac{2\gamma \cdot Var_t(p_{t+1})}{r} + \frac{\mu\rho^*}{r(1+r)} + \frac{\mu\rho_t}{1+r}$$

$$Var[p_t] = Var\left[\frac{\mu\rho_t}{1+r}\right] = \frac{\mu^2}{(1+r)^2}Var(\rho_t) = \frac{\mu^2}{(1+r)^2}\sigma_\rho^2$$

- Rewrite p_t

$$p_t = 1 + \frac{\mu\rho^*}{r} + \frac{\mu(\rho_t - \rho^*)}{1+r} - 2\frac{\gamma\mu^2\sigma_\rho^2}{r(1+r)^2}$$

- Noise traders affect prices!
- Term 1: Variation in noise trader (mis-)perception
- Term 2: Average misperception of noise traders
- Term 3: Compensation for noise trader risk

Relative returns of noise traders

- Compare returns to noise traders R^n to returns for arbitrageurs R_a:

$$\Delta R = R^n - R^a = (\lambda_t^n - \lambda_t^a)\left[r + p_{t+1} - p_t(1+r)\right]$$

$$E(\Delta R | \rho_t) = \rho_t - \frac{(1+r)^2 \rho_t^2}{2\gamma\mu\sigma_\rho^2}$$

$$E(\Delta R) = \rho^* - \frac{(1+r)^2 (\rho^*)^2 + (1+r)^2 \sigma_\rho^2}{2\gamma\mu\sigma_\rho^2}$$

- Noise traders hold more risky asset if $\rho^* > 0$
- Return of noise traders can be higher if $\rho^* > 0$ (and not too positive)
- Noise traders therefore may outperform arbitrageurs if optimistic!
- (Reason is that they are taking more risk)

Welfare

- Sophisticated investors have higher utility
- Noise traders have lower utility than they expect
- Noise traders may have higher returns (if $\rho^* > 0$)

- Noise traders do not necessarily disappear over time

- Three fundamental assumptions
  1. OLG: no last period; short horizon
  2. Fixed supply unsafe asset (*a* cannot convert safe into unsafe)
  3. Noise trader risk systematic

- Noise trader models imply that biases affect asset prices:
  - Reference Dependence
  - Attention
  - Persuasion

Section 4

Behavioral Corporate Finance

Behavioral Corporate Finance

Baker, Ruback, and Wurgler (2005)

- Behavioral corporate finance:
    - biased investors (overvalue or undervalue company)
    - smart managers
    - (Converse: biased (overconfident) managers and rational investors)

- Firm has to decide how to finance investment project:
    1. internal funds (cash flow/retained earnings)
    2. bonds
    3. stocks

Findings

- Fluctuation of equity prices due to noise traders
- Managers believe that the market is inefficient
    - Issue equity when stock price exceeds perceived fundamental value
    - Delay equity issue when stock price below perceived fundamental value
- Consistent with
    - Survey Evidence of 392 CFO's (Graham and Harvey 2001): 67% say under/overvaluation is a factor in issuance decision
    - Insider trading

- Go over quickly two examples

Long-run performance of equity issuers

- Market Timing prediction: Companies issuing equity underperform later
- **Loughran-Ritter (1995)**: Compare matching samples of
  - companies doing IPOs
  - companies not doing IPOs but have similar market cap.
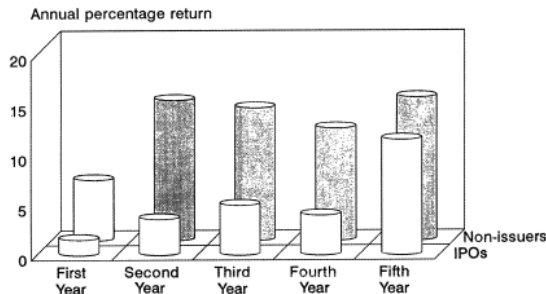
Similar finding with SEOs
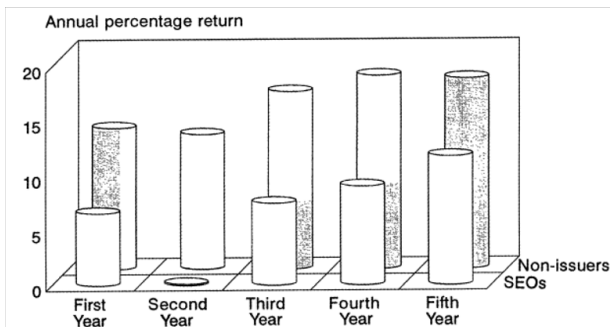
Figure 2. The average annual raw returns for 4,753 initial public offerings (IPOs), and their matching nonissuing firms (top), and the average annual raw returns for 3,702 seasoned equity offerings (SEOs), and their matching nonissuing firms (bottom), during the five years after the issue. The equity issues are from 1970 to 1990. Using the first closing postissue market price, the equally weighted average buy-and-hold return for the year after the issue is calculated for the issuing firms and for their matching firms (firms with the same market capitalization that have not issued equity during the prior five years). On each anniversary of the issue date, the equally weighted average buy-and-hold return during the next year for all of the surviving issuers and their matching firms is calculated. For matching firms that get delisted (or issue equity) while the issuer is still trading, the proceeds from the sale on the delisting date are reinvested in a new matching firm for the remainder of that year (or until the issuer is delisted). The numbers graphed above are reported in Table III.

Section 5

Market Reaction to Biases: Employers

Kahneman, Knetsch and Thaler (1986)

- Telephone surveys in Canada in 1984 and 1985 → Ask questions on fairness

  > Question 4A. A company is making a small profit. It is located in a community experiencing a recession with substantial unemployment but no inflation. There are many workers anxious to work at the company. The company decides to decrease wages and salaries 7% this year.
  >
  >   ($N = 125$)   Acceptable 38%   Unfair 62%
  >
  > Question 4B. . . . with substantial unemployment and inflation of 12%. . . The company decides to increase salaries only 5% this year.
  >
  >   ($N = 129$)   Acceptable 78%   Unfair 22%

  - A real and nominal wage cut is not fair (Question 4A)
  - A real (but not nominal) wage cut is fair (Question 4B)

- If this is true, expect employers to minimize cases of $w_t - w_{t-1} < 0$

Card and Hyslop (1997)

- Examine discontinuity around 0 of nominal wage changes
- Prediction of theory:
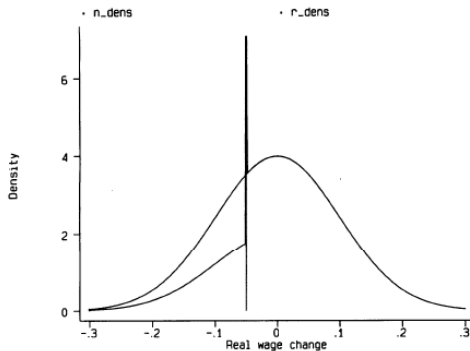
Data and Methodology

- Data sources:
    - 1979-1993 CPS.
        - Rolling 2-year panel
        - Restrict to paid by the hour and to same 2-digit industry in the two years
        - Restrict to non-minimum wage workers
    - PSID 4-year panels 1976-79 and 1985-88
- Use Log Wage changes: $\log w_t - \log w_{t-1}$
- Issue with measurement error and heaping at $\log w_t - \log w_{t-1} = 0$
- Construct counterfactual density of LogWage changes
    - Assume symmetry
    - Positive log wage changes would not be affected

Methodology

- Plots using kernel estimates of density (local smoother)
- Compare the actual distribution and the predicted one
- Evidence from the CPS year-by-year
- Problem more severe in years with lower inflation

- Large effect of nominal rigidities
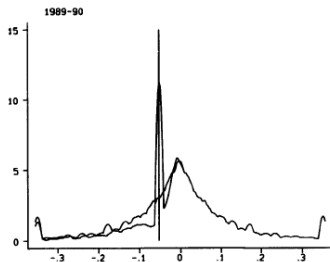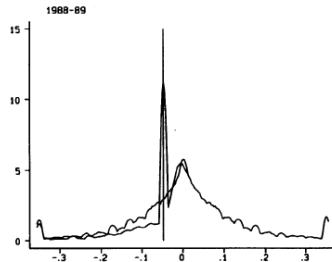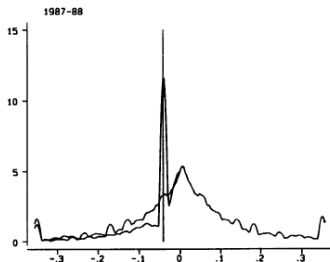- Effect on firings?

Real Wage Changes, 1979-80 to 1982-83
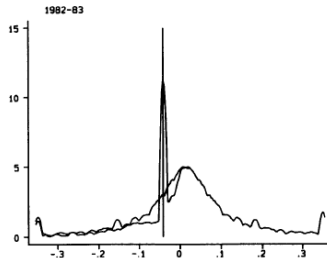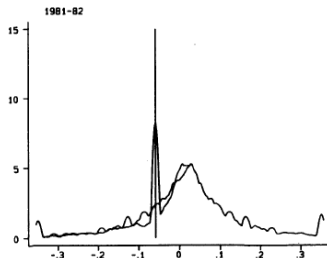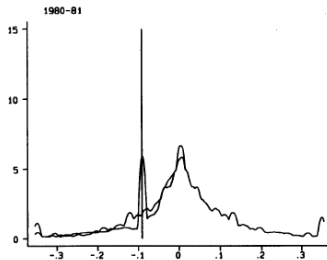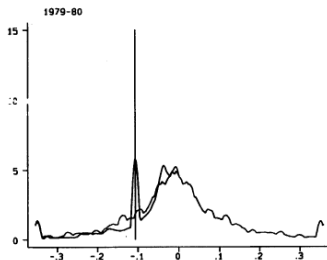
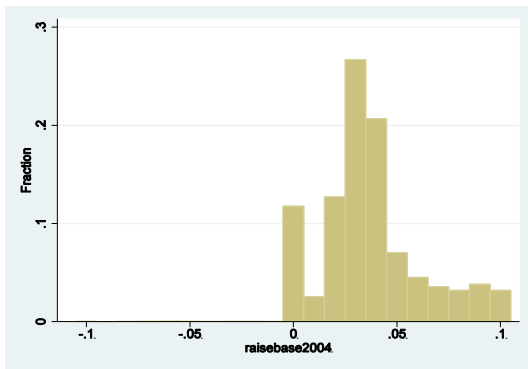# Real Wage Changes, 1983-84 to 1986-87
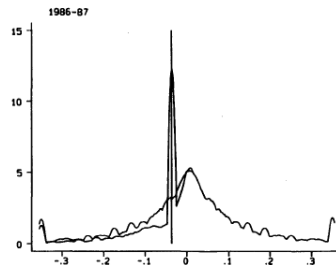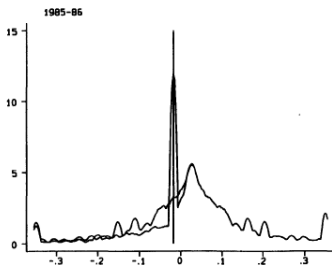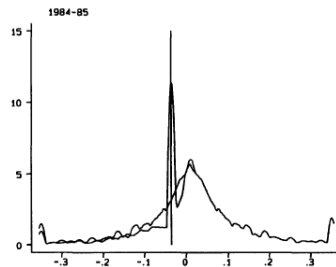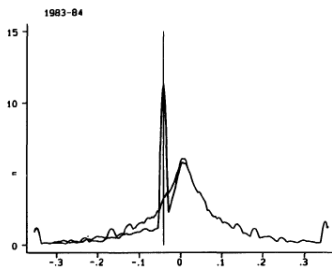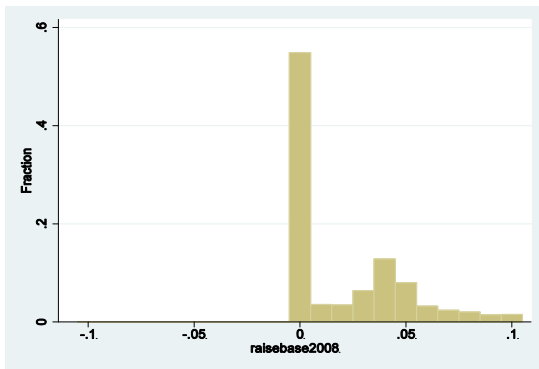
Real Wage Changes, 1987-88 to 1998-91
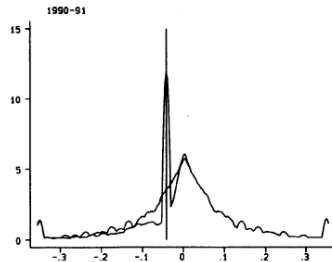
Hipsman (2011): 2003 & 2004

- Administrative data from several firms
  - Base pay % increase among those employed in 2003 and 2004
  - 58 (0.34%) cuts, 1,964 (10.18%) freezes, 15,091 (88.18%) raises

2007 & 2008

- Base pay % increase among those employed in 2007 and 2008
- 46 (0.36%) pay cuts, 6,913 (54.58%) pay freezes, 5,707 (45.06%) pay raises

Conclusions

- Card and Hyslop had *underestimated* the degree of nominal rigidity

- Important implications for labor markets when low inflation
    - If no pay cut, what margin of adjustment?
    - Firing?
    - Less hiring?

- Key under-researched topic in behavioral macro

Section 6

Next Lecture

Next Lecture

- Behavioral Public
- Teaching Evaulations