Econ 219B
Psychology and Economics: Applications
(Lecture 10)

Stefano DellaVigna

April 7, 2010

Outline

1. Attention: Taxes
2. Attention: Small Numbers
3. Attention: Financial Markets
4. Methodology I: Portfolio Methodology
5. Framing
6. Menu Effects: Introduction
7. Menu Effects: Excess Diversification
8. Methodology II: Clustering Standard Errors

1 Attention: Taxes

- **Chetty et al. (2007):** Taxes not featured in price likely to be ignored

- Use data on the demand for items in a grocery store.

- Demand D is a function of:
  - visible part of the value v, including the price p
  - less visible part o (state tax $-tp$)
  - $D = D\left[v - (1-\theta)\,tp\right]$

- Variation: Make tax fully salient ($s = 1$)

- Linearization: change in log-demand

$$\begin{aligned}\Delta \log D &= \log D\left[v - tp\right] - \log D\left[v - (1-\theta)\, tp\right] = \\ &= -\theta tp * D'\left[v - (1-\theta)\, tp\right] / D\left[v - (1-\theta)\, tp\right] \\ &= -\theta t * \eta_{D,p}\end{aligned}$$

  - $\eta_{D,p}$ is the price elasticity of demand

  - $\Delta \log D = 0$ for fully attentive consumers ($\theta = 0$)

  - This implies $\theta = -\Delta \log D / (t * \eta_{D,p})$

- **Chetty et al. (2007) Part I: field experiment**
  - Three-week period: price tags of certain items make salient after-tax price (in addition to pre-tax price).

- Compare sales D to:
  - previous-week sales for the same item
  - sales for items for which tax was not made salient
  - sales in control stores
  - Hence, D-D-D design (pre-post, by-item, by-store)

- Result: average quantity sold decreases (significantly) by 2.20 units relative to a baseline level of 25, an 8.8 percent decline

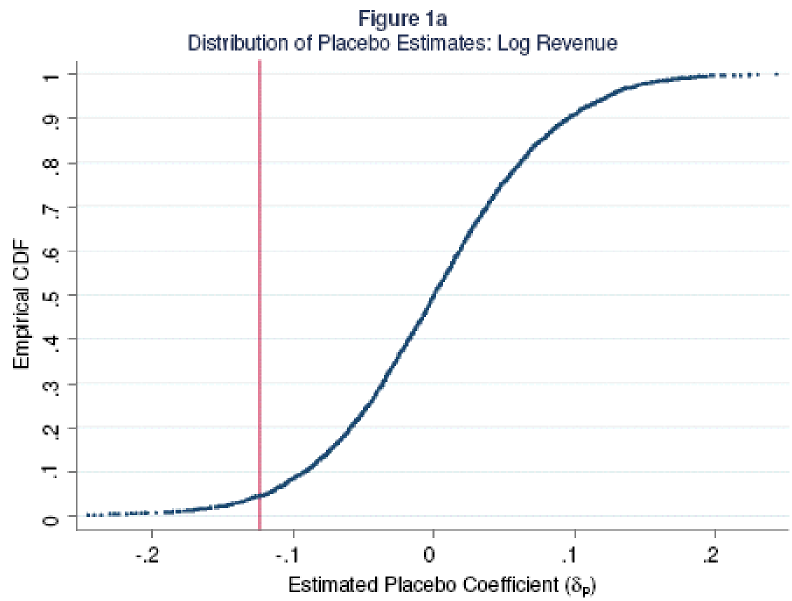
TABLE 3

DDD Analysis of Means: Weekly Quantity by Category

TREATMENT STORE			
Period	Control Categories	Treated Categories	Difference
Baseline (2005:1- 2006:6)	26.48 (0.22) [5510]	25.17 (0.37) [754]	-1.31 (0.43) [6264]
Experiment (2006: 8- 2006:10)	27.32 (0.87) [285]	23.87 (1.02) [39]	-3.45 (0.64) [324]
Difference over time	0.84 (0.75) [5795]	-1.30 (0.92) [793]	**DD_{TS} = -2.14** (0.64) [6588]
CONTROL STORES			
Period	Control Categories	Treated Categories	Difference
Baseline (2005:1- 2006:6)	30.57 (0.24) [11020]	27.94 (0.30) [1508]	-2.63 (0.32) [12528]
Experiment (2006: 8- 2006:10)	30.76 (0.72) [570]	28.19 (1.06) [78]	-2.57 (1.09) [648]
Difference over time	0.19 (0.64) [11590]	0.25 (0.92) [1586]	**DD_{CS} = 0.06** (0.90) [13176]
		DDD Estimate	***-2.20*** (0.58) [19764]

Notes: Each cell shows mean number of units sold per category per week, for various subsets of the sample. Standard errors (clustered by week) in parentheses, number of observations in square

- Compute inattention:
  - Estimates of price elasticity $\eta_{D,p}$: -1.59
  - Tax is .07375
  - $\hat{\theta} = -(-.088)/(-1.59 * .07375) \approx .75$
- Additional check of randomization: Generate placebo changes over time in sales
- Compare to observed differences
- Use Log Revenue and Log Quantity



- Non-parametric p-value of about 5 percent

- **Chetty et al. (2007) Part II: Panel Variation**
  - Compare more and less salient tax on beer consumption
  - Excise tax included in the price
  - Sales tax is added at the register
  - Panel identification: across States and over time
  - Indeed, elasticity to excise taxes substantially larger –> estimate of the inattention parameter of $\hat{\theta} = .94$
- Substantial consumer inattention to non-transparent taxes

TABLE 7

Effect of Excise and Sales Taxes on Beer Consumption

Dependent Variable: Change in Log(per capita beer consumption)	Baseline (1)	Bus Cycle (2)	Bus Cycle Lags (3)	Alc Regulations (4)
ΔLog(1+Excise Tax Rate)	**-0.87**	**-0.91**	**-0.86**	**-0.89**
	(0.17)***	(0.17)***	(0.17)***	(0.17)***
ΔLog(1+Sales Tax Rate)	**-0.20**	**-0.00**	**0.03**	**-0.02**
	(0.30)	(0.30)	(0.30)	(0.30)
ΔLog(Population)	0.03	-0.07	0.05	-0.07
	(0.06)	(0.07)	(0.19)	(0.07)
ΔLog(Income per Capita)		0.22	0.18	0.22
		(0.05)***	(0.05)***	(0.05)***
ΔLog(Unemployment Rate)		-0.01	-0.01	-0.01
		(0.01)**	(0.01)	(0.01)**
Lag Bus. Cycle Controls			x	
Alcohol Regulation Controls				x
Year Fixed Effects	x	x	x	x
F-Test for Equality of Tax Variables (Prob>F)	0.05	0.01	0.01	0.01
Sample Size	1607	1487	1440	1487

Notes: Standard errors, clustered by state, in parentheses: * significant at 10%; ** significant at 5%; *** significant at 1%. All specifications include year fixed effects and log state population. Column 2 controls for log state personal income per capita and log state unemployment rate (unavailable in some states in the early 1970s). Column 3 adds one year lags of personal income per capita and unemployment rate variables. Column 4 controls for changes in alcohol policy by including three separate indicators for whether the state implemented per se drunk driving standards, administrative license revocation laws, or zero tolerance youth drunk driving laws, and the change in the minimum drinking age (measured in years).

2 Attention: Small Numbers

- Are consumers paying attention to full numbers, or only to more salient digits?

- Classical example: $X =$\$5.99 vs. $Y =$\$6.00

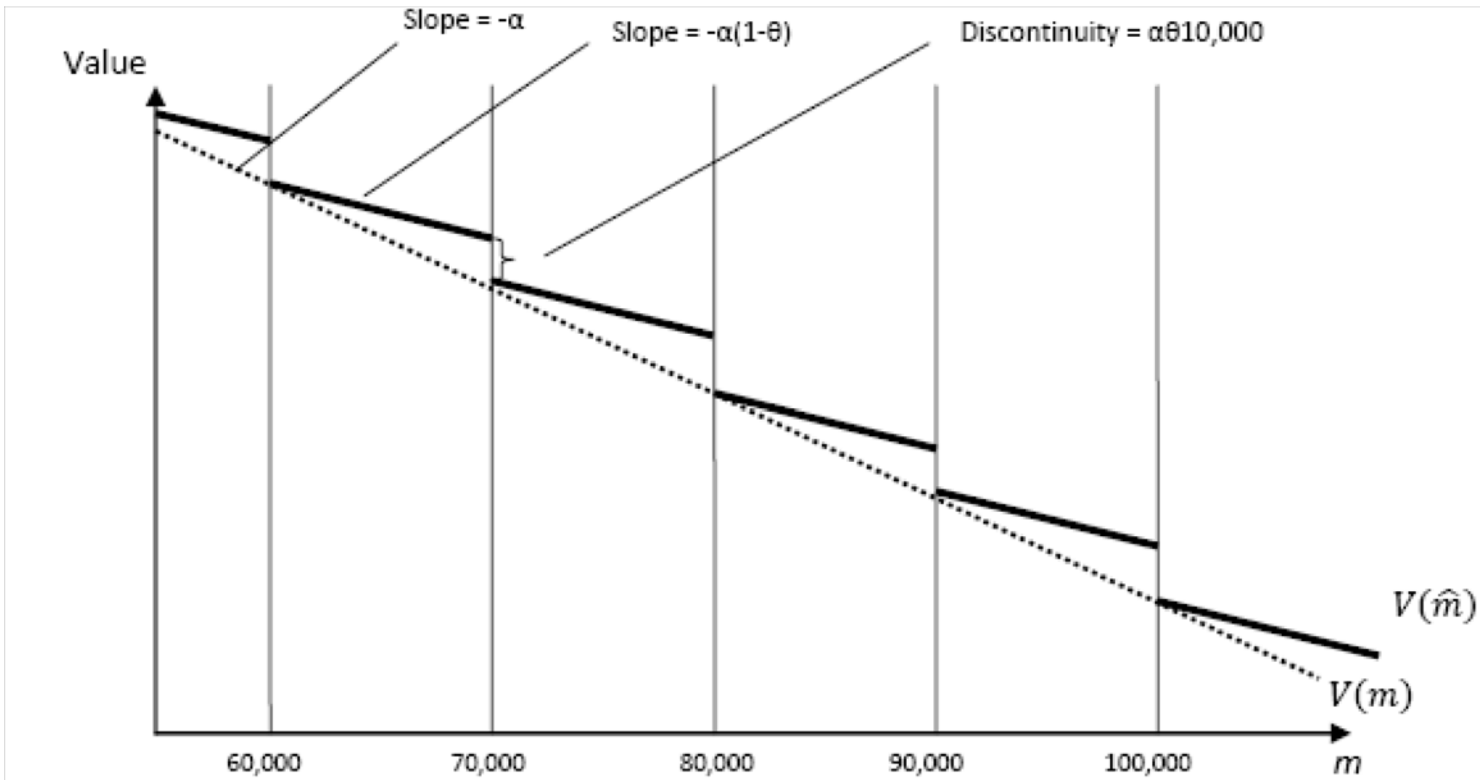
- Consumer inattentive to digits other than first, perceive

$$\begin{aligned} X &= 5 + (1-\theta)\,.99 \\ Y &= 6 \\ Y - X &= .01 + .\theta 99 \end{aligned}$$

- Indeed, evidence of 99 cents effect in pricing at stores

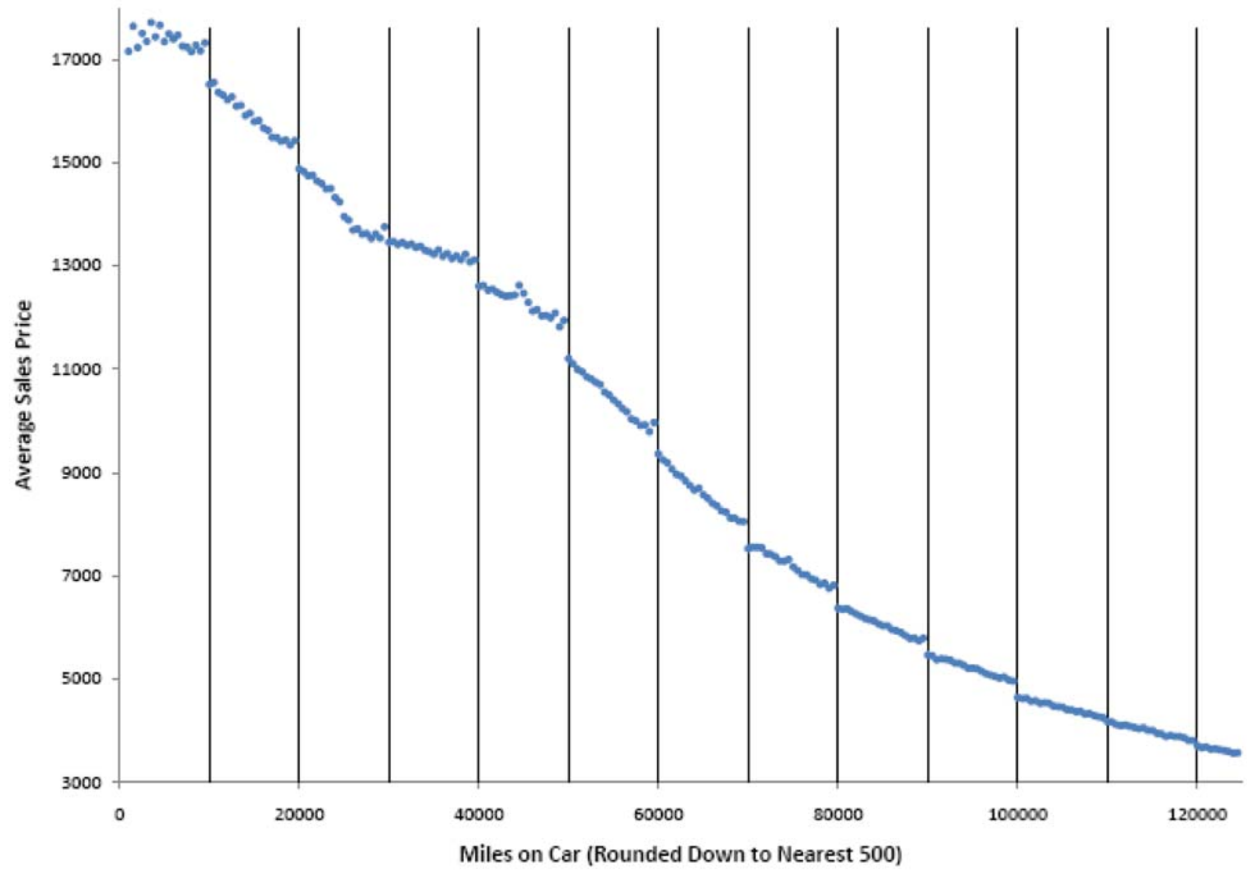
- However, can argue – stakes small for consumers

- **Lacetera, Pope, and Sydnor (2009). Inattention in Car Sales**
- Sales of used cars –Odometer is important measure of value of car



- Data set with 22 million wholesale used car transactions

Figure 2 - Raw Price. This figure plots the raw average sales price within 500-mile bins for the more than 22 million auctioned cars in our datas



- Remarkable precision in the estimates of the discontinuity

- Can estimate $\theta = 0.33$

- Consistent estimate broadly with other evidence

- However: Who des this inattention refer to?

- Data is from sales to car dealers, who are presumably incorporating preferences of buyers

3 Attention: Financial Markets I

- Is inattention limited to consumers?

- Finance: examine response of asset prices to release of quarterly earnings news

- Setting:
  - Announcement a time t
  - v is known information about cash-flows of the company
  - o is new information in earnings announcement
  - Day $t-1$: company price is $P_{t-1} = v$
  - Day t:

    * company value is $v + o$
    * Inattentive investors: asset price P_t responds only partially to the new information: $P_t = v + (1 - \theta)\, o$.
  - Day $t + 60$: Over time,price incorporates full value: $P_{t+60} = v + o$

- Implication about returns:
  - Short-run stock return r_{SR} equals $r_{SR} = (1 - \theta)\, o/v$
  - Long-run stock return r_{LR}, instead, equals $r_{LR} = o/v$
  - Measure of investor attention: $(\partial r_{SR}/\partial o)/(\partial r_{LR}/\partial o) = (1 - \theta)$ –> Test: Is this smaller than 1?
  - (Similar results after allowing for uncertainty and arbitrage, as long as limits to arbitrage — see final lectures)

- Indeed: Post-earnings announcement drift (**Bernard-Thomas, 1989**): Stock price keeps moving after initial signal

- Inattention leads to delayed absorption of information.

- **DellaVigna-Pollet (forthcoming)**
  - Estimate $(\partial r_{SR}/\partial o)/(\partial r_{LR}/\partial o)$ using the response of returns r to the earnings surprise o
  - r_{SR}: returns in 2 days surrounding an announcement
  - r_{LR}: returns over 75 trading days from an announcement

- Measure earnings news o_t:

$$o_t = \frac{e_t - \hat{e}_t}{p_{t-1}}$$

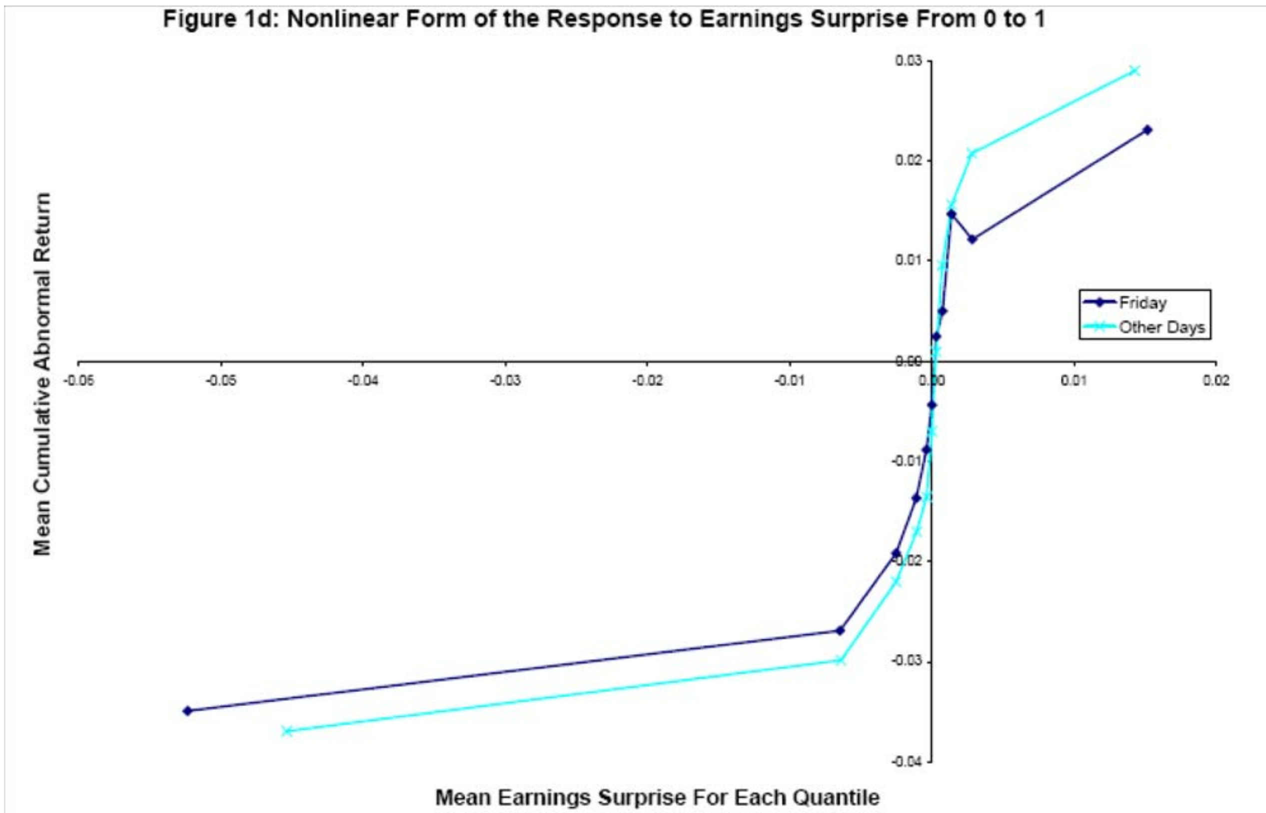
- Difference between earnings announcement e_t and consensus earnings forecast by analysts in 30 previous days
- Divide by (lagged) price p_{t-1} to renormalize

- Next step: estimate $\partial r_{SR}/\partial o$

- Problem: Response of stock returns r to information o is highly non-linear

- How to evaluate derivative?

4 Methodology: Portfolio Methodology

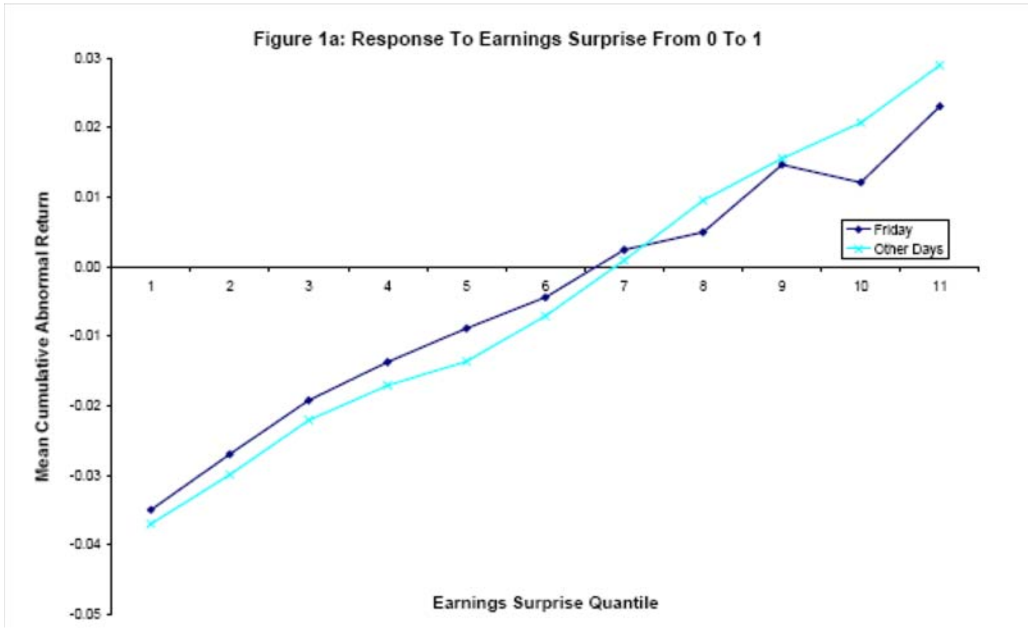


- Economists' approach:
  - Make assumptions about functional form –> Arctan for example
  - Do non-parametric estimate –> kernel regressions

- Finance: Use of quantiles and portfolios (explained in the context of DellaVigna-Pollet (forthcoming))

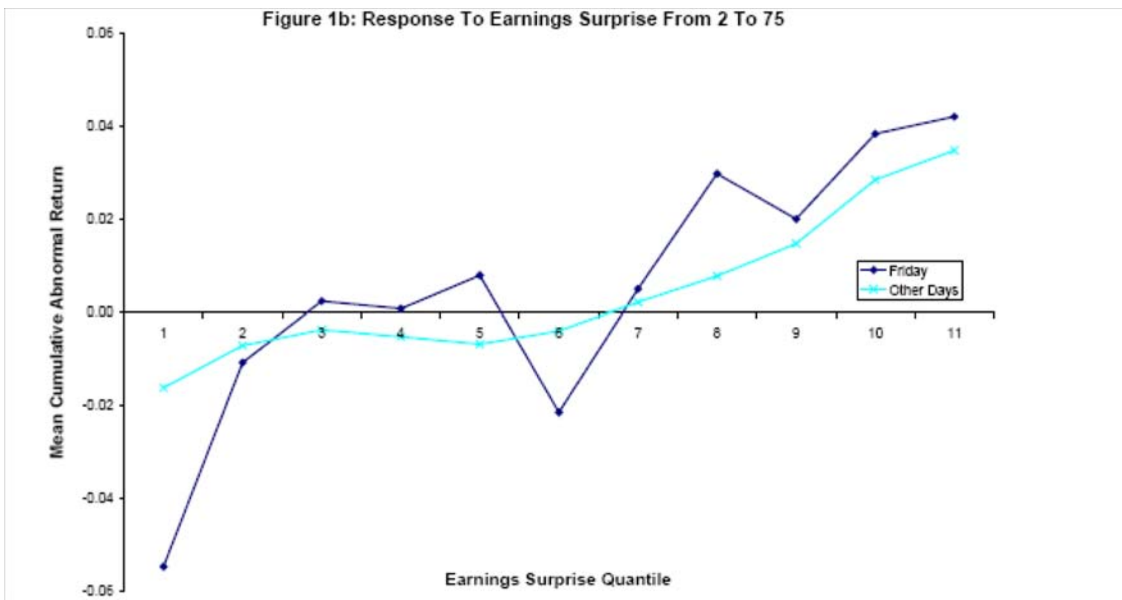
- First methodology: *Quantiles*
  - Sort data using underlying variable (in this case earnings surprise o_t)
  - Divide data into n equal-spaced quantiles: $n = 10$ (deciles), $n = 5$ (quintiles), etc
  - Evaluate difference in returns between top quantiles and bottom quantiles: $Er_n - Er_1$

- This paper:
  - Quantiles 7-11. Divide all positive surprises
  - Quantiles 6. Zero surprise (15-20 percent of sample)
  - Quantiles 1-5. Divide all negative surprise



- Notice: Use of quantiles "linearizes" the function

- Delayed response $r_{LR} - r_{SR}$ (post-earnings announcement drift)



- Inattention:
  - To compute $\partial r_{SR}/\partial o$, use $Er^{11}_{SR} - Er^{1}_{SR} = 0.0659$ (on non-Fridays)
  - To compute $\partial r_{LR}/\partial o$, use $Er^{11}_{LR} - Er^{1}_{LR} = 0.1210$ (on non-Fridays)
  - Implied investor inattention: $(\partial r_{SR}/\partial o)/(\partial r_{LR}/\partial o) = (1-\theta) = .544$ –> Inattention $\theta = .456$
- Is inattention larger when more distraction?
- Weekend as proxy of investor distraction.
  - Announcements made on Friday: $(\partial r_{SR}/\partial o)/(\partial r_{LR}/\partial o)$ is 41 percent –> $\hat{\theta} \approx .59$

- Second methodology: *Portfolios*
  - Instead of using individual data, pool all data for a given time period t into a 'portfolio'
  - Compute average return r_t^P for portfolio t over time
  - Control for Fama-French 'factors':
    - Market return r_t^m
    - Size r_r^S
    - Book-to-Market r_t^{BM}
    - Momentum r_t^M

- $*$ (Download all of these from Kenneth French's website)

– Regression:

$$r_t^P = \alpha + BR_t^{Factors} + \varepsilon_t$$

– Test: Is α significantly different from zero?

- Example in DellaVigna-Pollet (forthcoming)

  – Each month t portfolio formed as follows: $(r_F^{11} - r_F^1) - (r_{Non-F}^{11} - r_{Non-F}^1)$

  – Use returns r_{Drift} (3-75)

  – Differential drift between Fridays and non-Fridays

- Test for significance

	Dependent Variable: Monthly Return on the Zero-Investment Portfolio					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.0384 (0.0134)***	0.0462 (0.0139)***	0.0584 (0.0220)***	0.0218 (0.0079)***	0.0232 (0.0086)***	0.0277 (0.0091)***
VW Index Excess Return (VWRF)	-0.2742 (0.3090)	-0.6419 (0.2778)**	-0.0968 (0.4262)	-0.1842 (0.1865)	-0.1068 (0.2301)	-0.4580 (0.1937)**
Size Factor Return (SMB)		0.2344 (0.4195)	0.5644 (0.6227)	-0.0390 (0.2464)	0.0701 (0.2930)	-0.0137 (0.2438)
Value Factor Return (HML)		-0.4607 (0.6143)	-1.5556 (0.7277)**	0.0762 (0.3329)	-0.3264 (0.2840)	-0.2094 (0.3820)
Momentum Factor Return (UMD)		-0.3994 (0.2632)	-1.1817 (0.6559)*	-0.0696 (0.1740)	-0.0410 (0.2206)	-0.3454 (0.1940)*
One month holding period	X	X	X	X		X
Two month holding period					X	
Top minus bottom quantile	X	X	X		X	
Matched sample			X			
Top two minus bottom two quantiles				X		
Top minus bottom decile						X
R^2	0.0073	0.0385	0.1736	0.0152	0.0153	0.0398
N	N = 125	N = 125	N = 124	N = 130	N = 138	N = 127

* significant at 10%; ** significant at 5%; *** significant at 1%

- Intercept $\hat{\alpha} = .0384$ implies monthly returns of 3.84 percent of pursuing this strategy

5 Attention: Financial Markets II

- **Cohen-Frazzini (forthcoming) – Inattention to subtle links**

- Suppose that you are a investor following company A

- Are you missing more subtle news about Company A?

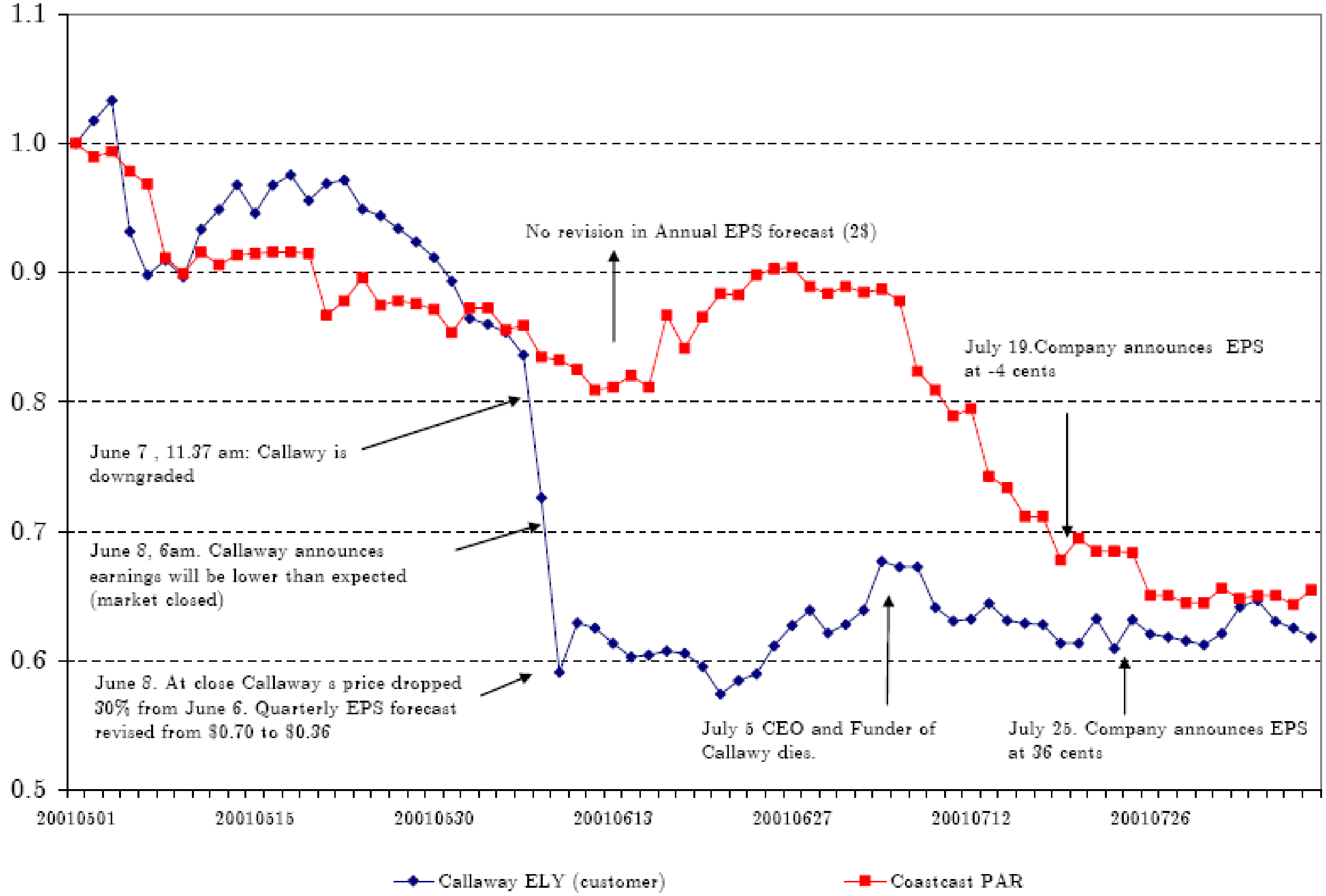
- Example: Huberman and Regev (2001) – Missing the *Science* article

- Cohen-Frazzini (forthcoming) – Missing the news about your main customer

- Example:
  - Coastcoast Co. is leading manufacturer of golf club heads
  - Callaway Golf Co. is leading retail company for golf equipment
  - What happens after shock to Callaway Co.?

Figure 1: Coastcast Corporation and Callaway Golf Corporation

This figure plots the stock prices of Coastcast Corporation (ticker = PAR) and Callaway Golf Corporation (ticker = ELY) between May and August 2001. Prices are normalized (05/01/2001 = 1).



- Data:
  - Customer- Supplier network – Compustat Segment files (Regulation SFAS 131)
  - 11,484 supplier-customer relationships over 1980-2004
- Preliminary test:
  - Are returns correlated between suppliers and customers?
  - Correlation 0.122 at monthly level

- Computation of long-short returns
  - Sort into 5 quintiles by returns in month t of principal customers, r_t^C
  - By quintile, compute average return in month $t+1$ for portfolio of suppliers r_{t+1}^S: $r_{1,t+1}^S, r_{2,t+1}^S, r_{3,t+1}^S, r_{4,t+1}^S, r_{5,t+1}^S$
  - By quintile q, run regression

$$r_{q,t+1}^S = \alpha_q + \beta_q X_{t+1} + \varepsilon_{q,t+1}$$

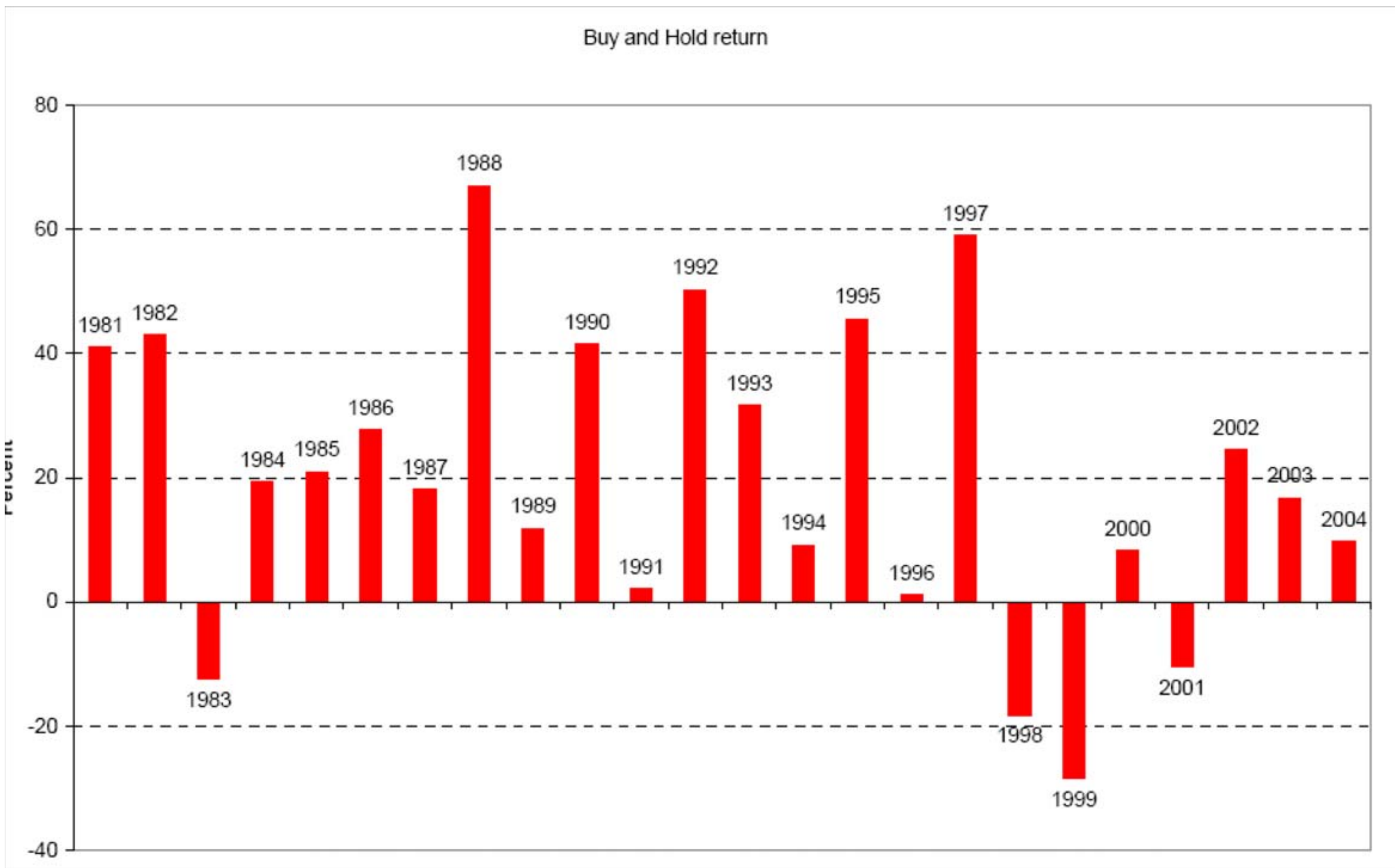
  - X_{t+1} are the so-called factors: market return, size, book-to-market, and momentum (Fama-French Factors)
  - Estimate $\hat{\alpha}_q$ gives the monthly average performance of a portfolio in quintile q
  - Long-Short portfolio: $\hat{\alpha}_5 - \hat{\alpha}_1$

- Results in Table III: *Monthly* abnormal returns of 1.2-1.5 percent (huge)

Panel A: value weights	Q1(low)	Q2	Q3	Q4	Q5(high)	L/S
Excess returns	-0.596	-0.157	0.125	0.313	**0.982**	**1.578**
	[-1.42]	[-0.41]	[0.32]	[0.79]	[2.14]	[3.79]
3-factor alpha	**-1.062**	**-0.796**	**-0.541**	-0.227	**0.493**	**1.555**
	[-3.78]	[-3.61]	[-2.15]	[-0.87]	[1.98]	[3.60]
4-factor alpha	**-0.821**	**-0.741**	-0.488	-0.193	**0.556**	**1.376**
	[-2.93]	[-3.28]	[-1.89]	[-0.72]	[1.99]	[3.13]
5-factor alpha	**-0.797**	**-0.737**	-0.493	-0.019	0.440	**1.237**
	[-2.87]	[-3.04]	[-1.94]	[-0.07]	[1.60]	[2.99]

- Information contained in the customer returns not fully incorporated into supplier returns

- Returns of this strategy are remarkably stable over time

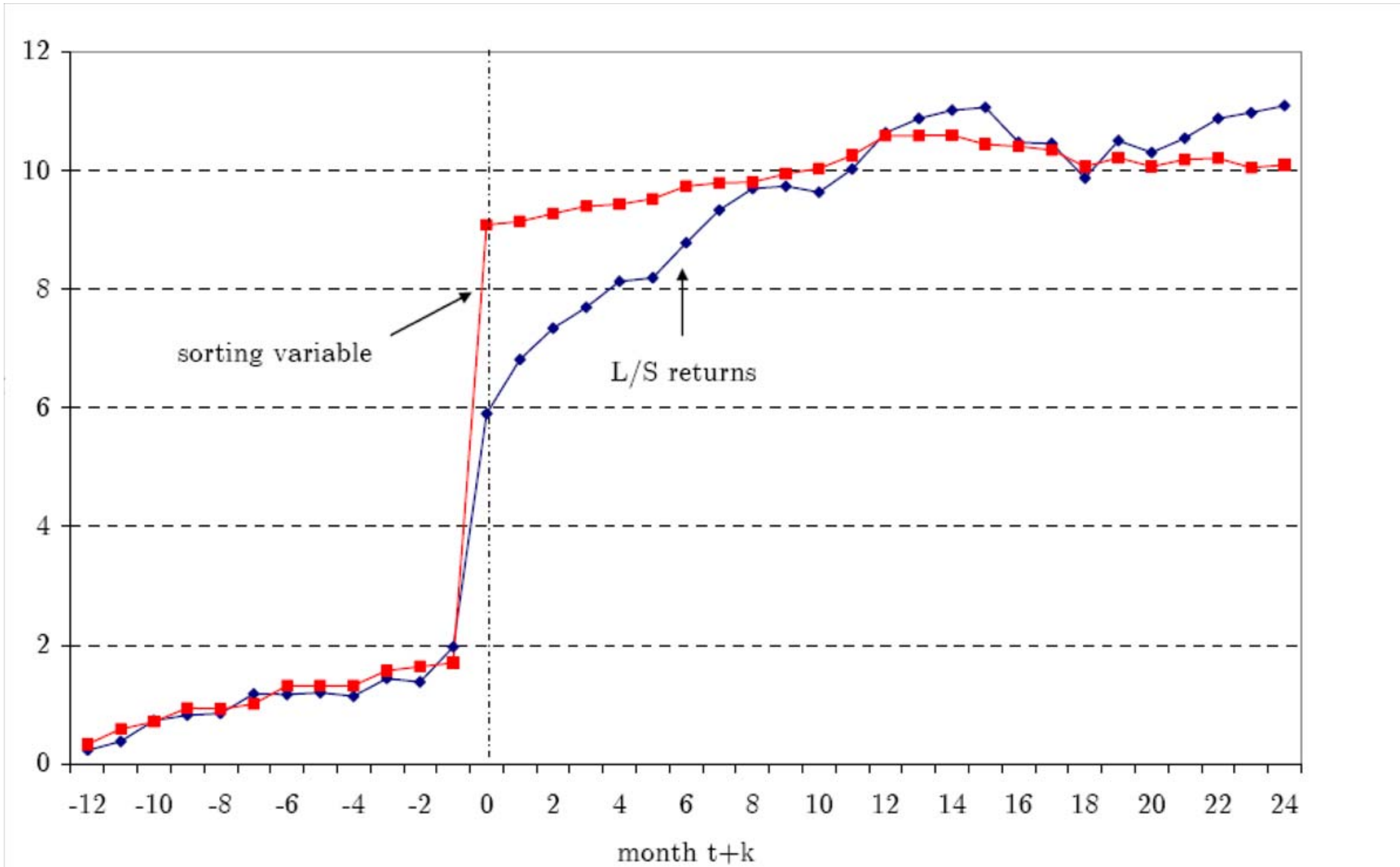


- Can run similar regression to test how quickly the information is incorporated
  - Sort into 5 quintiles by returns in month t of principal customers, r_t^C
  - Compute cumulative return up to month k ahead, that is, $r^S_{q,t->t+k}$
  - By quintile q, run regression of returns of Supplier:

$$r^S_{q,t->t+k} = \alpha_q + \beta_q X_{t+k} + \varepsilon_{q,t+1}$$

  - For comparison, run regression of returns of Customer:

$$r^C_{q,t->t+k} = \alpha_q + \beta_q X_{t+k} + \varepsilon_{q,t+1}$$



12
10
8
6
4
2
0
sorting variable
L/S returns
-12
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
14
16
18
20
22
24
month t+k

- For further test of inattention, examine cases where inattention is more likely

- Measure what share of mutual funds own both companies: COMOWN

- Median Split into High and Low COMOWN (Table IX)

	All stocks		At least 20 mutual funds holding the stock							
			All stocks		At least 10 common funds		Larger firms (CRSP median)		Larger firms (NYSE median)	
Weight	EW	VW	EW	VW	EW	VW	EW	VW	EW	VW
Low COMOWN	**1.653**	**2.301**	**1.659**	**2.306**	1.469	**1.889**	**1.572**	**2.288**	**2.703**	**2.852**
Lower percent of common ownership	[5.46]	[5.24]	[2.96]	[3.64]	[1.75]	[2.08]	[2.82]	[3.60]	[3.49]	[3.55]
High COMOWN	**0.750**	**1.098**	0.528	0.736	0.532	0.835	0.407	0.732	0.611	**1.278**
Higher percent of common ownership	[1.97]	[2.17]	[0.98]	[1.23]	[0.85]	[1.21]	[0.75]	[1.22]	[1.05]	[2.11]
High-Low	**-0.903**	**-1.203**	-1.131	**-1.571**	-0.937	-1.054	-1.165	**-1.557**	**-2.093**	-1.575
	[-2.08]	[-1.99]	[-1.60]	[-1.98]	[-0.92]	[-0.95]	[-1.66]	[-1.96]	[-2.42]	[-1.71]

- Supporting evidence from other similar papers

- **Hong-Torous-Valkanov (2002)**
  - Stock returns in an industry in month t predict returns in another industry in month $t + 1$
  - Investors not good at handling indirect links –> Indirect effects of industry-specific shocks neglected
  - Example: forecasted increase in price of oil
  - Oil industry reacts immediately, Other industries with delay

- **Pollet (2002)**
  - Scandinavian stock market (oil extraction) predicts US stock market (negatively) one month ahead
  - Oil industry predicts several industries one month ahead (again negatively)

- **DellaVigna-Pollet (2007) – Inattention to distant future**
- Another way to simplify decisions is to neglect distant futures when making forecasts
- Identify this using forecastable demographic shifts
- Substantial cohort size fluctuations over the 20th century
- Consumers at different ages purchase different goods
- Changes in cohort size $\Longrightarrow$ predictable changes in profits for different goods
- How do investors react to these forecastable shifts?

- **Example.** Large cohort born in 2004

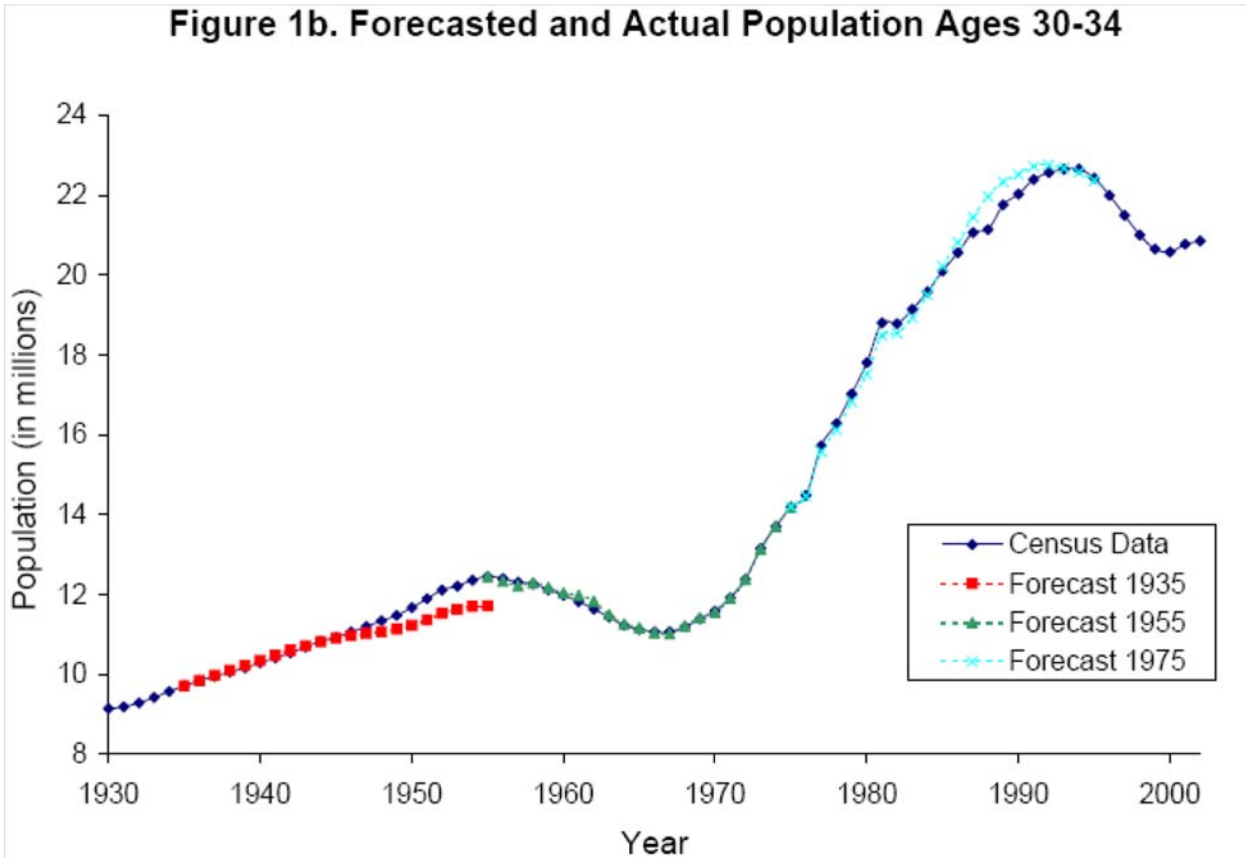
- Positive demand shift for school buses in 2010 $\Longrightarrow$ Revenue increases in 2010

- Profits (earnings) for bus manufacturers?
  - Perfect Competition. Abnormal profits do not change in 2010
  - Imperfect Competition. Increased earnings in 2010

- How do investors react?

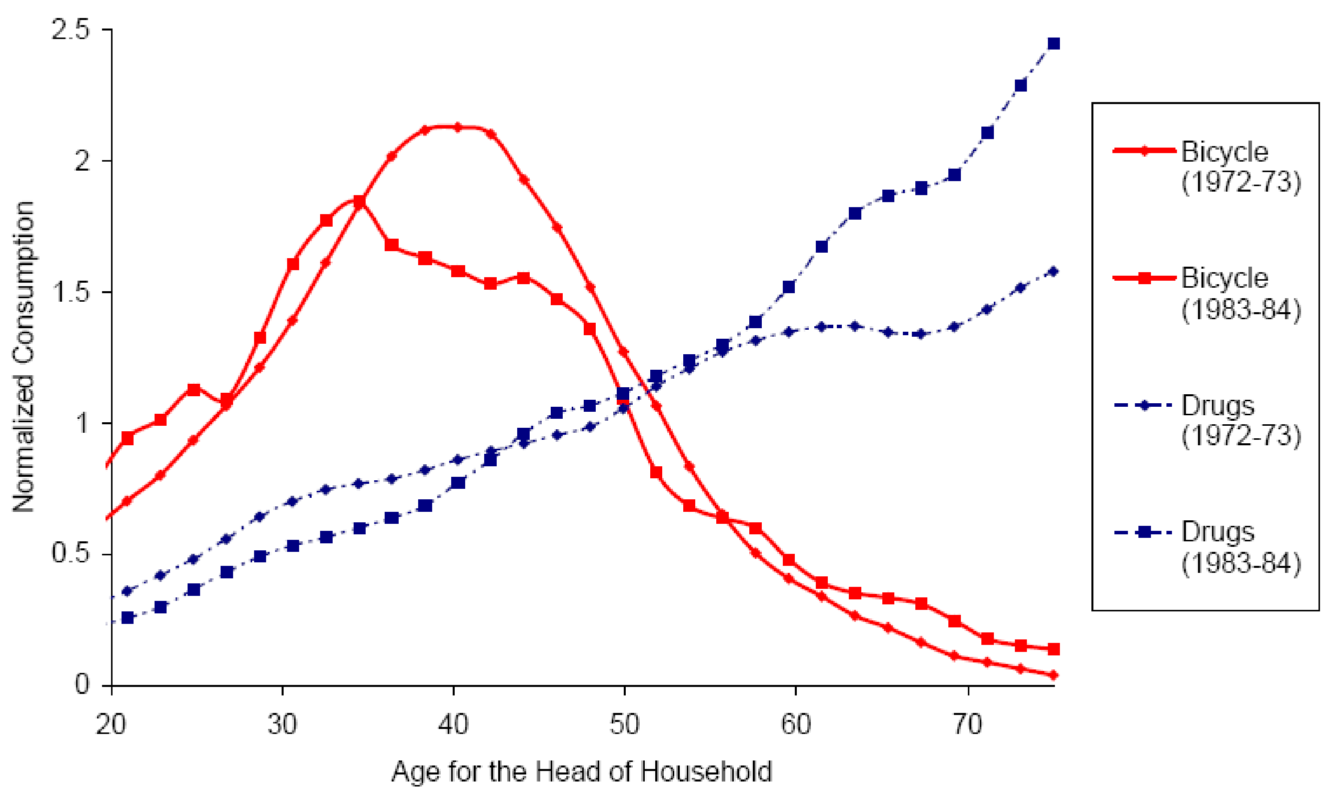
  1. Attentive investors:
     - Stock prices adjust in 2004
     - No forecastability of returns using demographic shifts
  2. Investors inattentive to future shifts:
     - Price does not adjust until 2010
     - Predictable stock returns using contemporaneous demand growth
  3. Investors attentive up to 5 years
     - Price does not adjust until 2005
     - Predictable stock returns using consumption growth 5 years ahead

- **Step 1.** Forecast future cohort sizes using current demographic data



- **Step 2.** Estimate consumption of 48 different goods by age groups (CEX data)

Figure 1. Age Profile of Bicycle and Drugs Consumption



- **Step 3.** Compute forecasted growth demand due to demographics into the future:

  - Demand increase in the short-term: $\hat{c}_{i,t+5} - \hat{c}_{i,t}$

  - Demand increase in the long-term: $\hat{c}_{i,t+10} - \hat{c}_{i,t+5}$

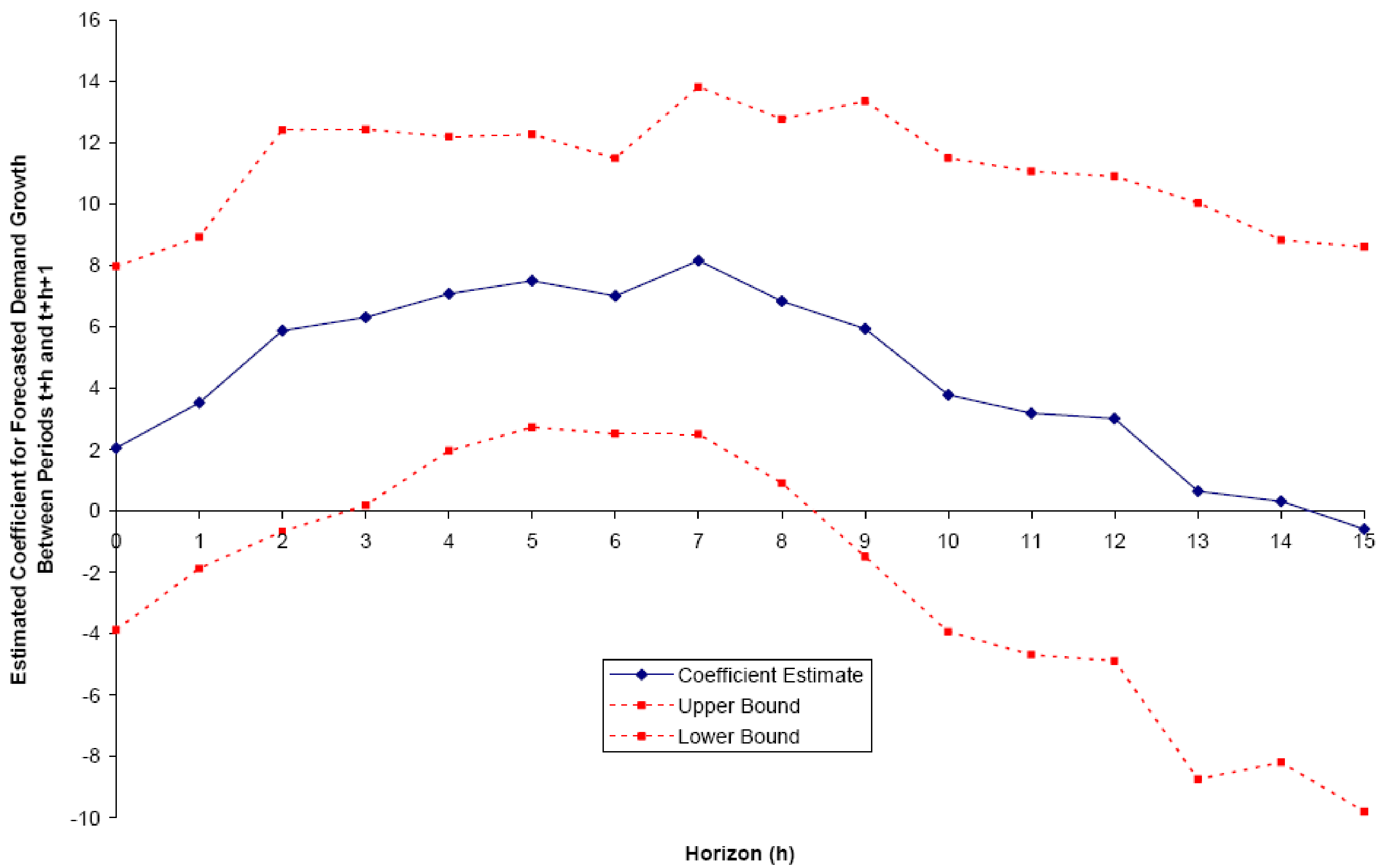
- Does this demand forecast returns? Regression of annual abnormal returns $ar_{i,t+1}$

$$\alpha r_{i,t+1} = \gamma + \delta_0 \left[\hat{c}_{i,t+5} - \hat{c}_{i,t}\right]/5 + \delta_1 \left[\hat{c}_{i,t+10} - \hat{c}_{i,t+5}\right]/5 + \varepsilon_{i,t+1}$$

Table 6. Predictability of Stock Returns Using Demographic Changes

	Dependent Variable: Annual Beta-Adjusted Log Industry Stock Return at t+1								
Sample	**Demographic Industries**						**All Industries**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	-0.0967	0.1004	0.3571	-0.0507	-0.0498	0.0606	-0.0774	-0.0672	0.1213
	(0.05560)*	(0.1122)	(0.0858)***	(0.0332)	(0.0444)	(0.0406)	(0.0472)	(0.0607)	(0.0668)*
Forecasted annualized demand growth between *t* and *t+5*	-0.4484	-0.5726	-2.2113	-1.5509	-1.7362	-2.7576	-1.8485	-1.2779	-2.1448
	(4.3929)	(4.2358)	(3.4036)	(2.7948)	(2.9935)	(2.8176)	(4.2901)	(4.7931)	(3.2678)
Forecasted annualized demand growth between *t+5* and *t+10*	8.7203	11.0365	6.8243	5.3723	5.8355	5.2183	8.3035	10.4185	5.8045
	(4.2206)**	(3.9489)***	(3.5568)*	(3.3562)	(3.3223)*	(2.7478)*	(3.6389)**	(4.2698)**	(3.8659)
Industry Fixed Effects		X	X		X	X		X	X
Year Fixed Effects			X			X			X
Sample: 1974 to 2003	X	X	X				X	X	X
Sample: 1939 to 2003				X	X	X			
R^2	0.0233	0.1121	0.3202	0.0089	0.0676	0.3162	0.0129	0.0484	0.1923
N	*N* = 566	*N* = 566	*N* = 566	*N* = 917	*N* = 917	*N* = 917	*N* = 1387	*N* = 1387	*N* = 1387

Figure 4: Return Predictability Coefficient for Demand Growth Forecasts at Different Horizons



Notes: The estimated coefficient for each horizon is from a univariate OLS regression of abnormal returns at *t*+1 on forecasted consumption growth between *t*+*h* and *t*+*h*+1 for the subsample of *Demographic Industries* over the period 1974-2003. The confidence intervals are constructed using robust standard errors clustered by year and then scaled by a function of the autocorrelation coefficient estimated from the sample orthogonality conditions.

- Results:
  1. Demographic shifts 5 to 10 years ahead can forecast industry-level stock returns
  2. Yearly portfolio returns of 5 to 10 percent
  3. Inattention of investors to information beyond approx. 5 years
  4. Evidence on analyst horizon: Earning forecasts beyond 3 years exist for only 10% of companies (IBES)

- Where else long-term future matters?
  - Job choices
  - Construction of new plant...

6 Framing

- Tenet of psychology: context and framing matter

- Classical example (Tversky and Kahneman, 1981 in version of Rabin and Weizsäcker, forthcoming): Subjects asked to consider a pair of '*concurrent decisions. [...]*

  - **Decision 1**. *Choose between: A. a sure gain of €2.40 and B. a 25% chance to gain €10.00 and a 75% chance to gain €0.00.*

  - **Decision 2**. *Choose between: C. a sure loss of €7.50 and D. a 75% chance to lose €10.00 and a 25% chance to lose €0.00.*'

  - Of 53 participants playing for money, 49 percent chooses A over B and 68 percent chooses D over C

- 28 percent of the subjects chooses the combination of A and D
  * This lottery is a 75% chance to lose ₤7.60 and a 25% chance to gain ₤2.40
  * Dominated by combined lottery of B and C: 75% chance to lose ₤7.50 and a 25% chance to gain ₤2.50
- Separate group of 45 subjects presented same choice in broad framing (they are shown the distribution of outcomes induced by the four options)
  * None of these subjects chooses the A and D combination

- Interpret this with reference-dependent utility function with narrow framing.
  - Approximately risk-neutral over gains –> 49 percent choosing A over B
  - Risk-seeking over losses –> 68 percent choosing D over C.
  - Key point: Individuals accept the framing induced by the experimenter and do not aggregate the lotteries

- General feature of human decisions:
  - judgments are comparative
  - changes in the framing can affect a decision if they change the nature of the comparison

- Presentation format can affect preferences even aside from reference points

- **Benartzi and Thaler (2002): Impact on savings plan choices:**
  - Survey 157 UCLA employees participating in a 403(b) plan
  - Ask them to rate three plans (labelled plans A, B, and C):
    * Their own portfolio
    * Average portfolio
    * Median portfolio
  - For each portfolio, employees see the 5th, 50th, and 95th percentile of the projected retirement income from the portfolio (using Financial Engines retirement calculator)

  - Revealed preferences –> expect individuals on average to prefer their own plan to the other plans

- Results:
  - Own portfolio rating (3.07)
  - Average portfolio rating (3.05)
  - Median portfolio rating (3.86)
  - 62 percent of employees give higher rating to median portfolio than to own portfolio

- Key component: Re-framing the decision in terms of ultimate outcomes affects preferences substantially

- Alternative interpretation: Employees never considered the median portfolio in their retirement savings decision –> would have chosen it had it been offered

- Survey 351 participants in a different retirement plan
  - These employees were explicitly offered a customized portfolio and actively opted out of it
  - Rate:
    - Own portfolio
    - Average portfolio
    - Customized portfolio
  - Portofolios re-framed in terms of ultimate income

- 61 percent of employees prefers customized portfolio to own portfolio
- Choice of retirement savings depends on format of the choices presented
- Open question: Why this *particular* framing effect?
- Presumably because of fees:
  - Consumers put too little weight on factors that determine ultimate returns, such as fees –> Unless they are shown the ultimate projected returns
  - Or consumers do not appreciate the riskiness of their investments –> Unless they are shown returns

- Framing also can focus attention on different aspects of the options

- **Duflo, Gale, Liebman, Orszag, and Saez (2006):** Fied Experiment with H&R Block
  - Examine participation in IRAs for low- and middle-income households
  - Estimate impact of a match

- Field experiment:
  - Random sub-sample of H&R Block customers are offered one of 3 options:
    * No match
    * 20 percent match
    * 50 percent match

- Match refers to first \$1,000 contributed to an IRA

- Effect on take-up rate:

  * No match (2.9 percent)
  * 20 percent match (7.7 percent)
  * 50 percent match (14.0 percent)

- Match rates have substantial impact

- Framing aspect: Compare response to explicit match to response to a comparable match induced by tax credits in the Saver's Tax Credit program
  - Effective match rate for IRA contributions decreases from 100 percent to 25 percent at the $30,000 household income threshold
  - Compare IRA participation for
    * Households slightly below the threshold ($27,500-$30,000)
    * Households slight above the threshold ($30,000-$32,500)
  - Estimate difference-in-difference relative to households in the same income groups that are ineligible for program
  - Result: Difference in match rate lowers contributions by only 1.3 percentage points –> Much smaller than in H&R Block field experiment
- Why framing difference? Simplicity of H&R Block match –> Attention
- Implication: Consider behavioral factors in design of public policy

7 Menu Effects: Introduction

- Summary of Limited Attention:

  - Too little weight on opaque dimension (*Science* article, shipping cost, posted price, news to customers. indirect link, distant future)

  - Too much weight on salient dimension (*NYT* article, auction price, recent returns or volume)

- Any other examples?

- We now consider a specific context: **Choice from Menu** N (typically, **with large** N)
  - Health insurance plans
  - Savings plans
  - Politicians on a ballot
  - Stocks or mutual funds
  - Type of Contract (Ex: no. of minutes per month for cell phones)
  - Classes
  - Charities
  - ...

- We explore 4 +1 (non-rational) heuristics
    1. Excess Diversification
    2. Choice Avoidance
    3. Preference for Familiar
    4. Preference for Salient
    5. Confusion
- Heuristics 1-4 deal with difficulty of choice in menu
    - Related to bounded rationality: Cannot process complex choice –> Find heuristic solution
- Heuristic 5 (next lecture) – Random confusion in choice from menu

8 Menu Effects: Excess Diversification

- First heuristic: **Excess Diversification or 1/n Heuristics**
  - Facing a menu of choices, if possible allocate
  - (Notice: Not possible for example for health insurance plan)

- Example: Experiment of Simonson (1990)
  - Subjects have to pick one snack out of six (cannot pick >1) in 3 different weeks
  - Sequential choice: only 9 percent picks three different snacks
  - Simultaneous choice ex ante: 64 percent chooses three different snacks

- **Benartzi-Thaler (AER, 2001)**
- Study 401(k) plan choices
- Data:
  - 1996 plan assets for 162 companies
  - Aggregate allocations, no individual data
- Average of 6.8 plan options per company
- Lacking individual data, cannot estimate if allocation is truly 1/n
- Proxy: Is there more investment in stocks where more stocks are offered?

- They estimate the relationship

$$\%Invested\ In\ Equity = \alpha + .36\ (.04) * \%Equity\ Options + \beta X$$

TABLE 7—THE RELATIVE NUMBER OF EQUITY-TYPE INVESTMENT OPTIONS AND ASSET ALLOCATION: A REGRESSION ANALYSIS (DEPENDENT VARIABLE: THE PERCENTAGE OF PLAN ASSETS INVESTED IN EQUITIES)

WLS regression model	Intercept	Relative number of equity options	Indicator whether the plan offers company stock	Log of the plan assets in thousands	Adjusted R^2
Panel A: No Industry Indicators (N = 162)					
1	22.09	63.14			34.61 percent
	(4.94)	(9.28)			
2	29.72	36.75	15.05		43.45 percent
	(6.73)	(4.49)	(5.10)		
3	10.57	36.77	14.78	1.40	44.16 percent
	(0.89)	(4.52)	(5.03)	(1.74)	
Panel B: Including Industry Indicators Based on 2-Digit SIC Codes (N = 142)					
4		58.68			55.12 percent
		(8.29)			
5		43.90	12.93		58.91 percent
		(5.39)	(3.26)		
6		47.07	9.09	4.13	61.79 percent
		(5.93)	(2.25)	(2.96)	

Notes: The initial sample consists of the June 1996 MMD sample of 401(k) plans. Eight plans with less than four investment options were excluded, resulting in a sample of 162 plans. When we include industry indicators, the sample is further reduced to 142 plans due to missing industry information. The table reports WLS regression estimates with plan assets as weights (t-statistics are in parentheses).

- For every ten percent additional offering in stocks, the percent invested in stocks increases by 3.6 percent
- Notice: availability of company stocks is a key determinant of holdings in stocks
- Issues of endogeneity:
  - Companies offer more stock when more demand for it
  - Partial response: Industry controls
- Additional evidence based on a survey
  - Ask people to allocate between Fund A and Fund B
  - Vary Fund A and B to see if people respond in allocation

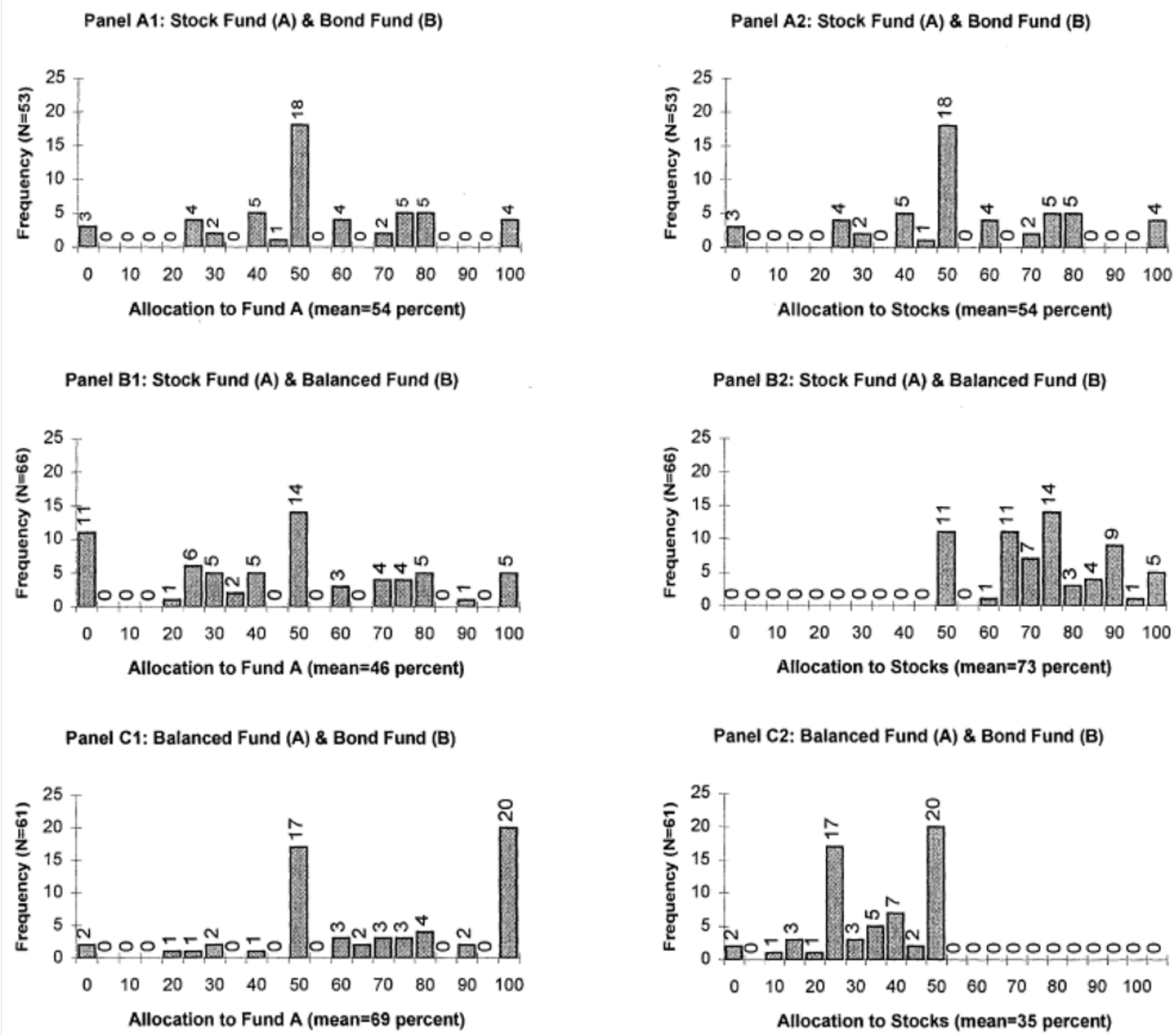
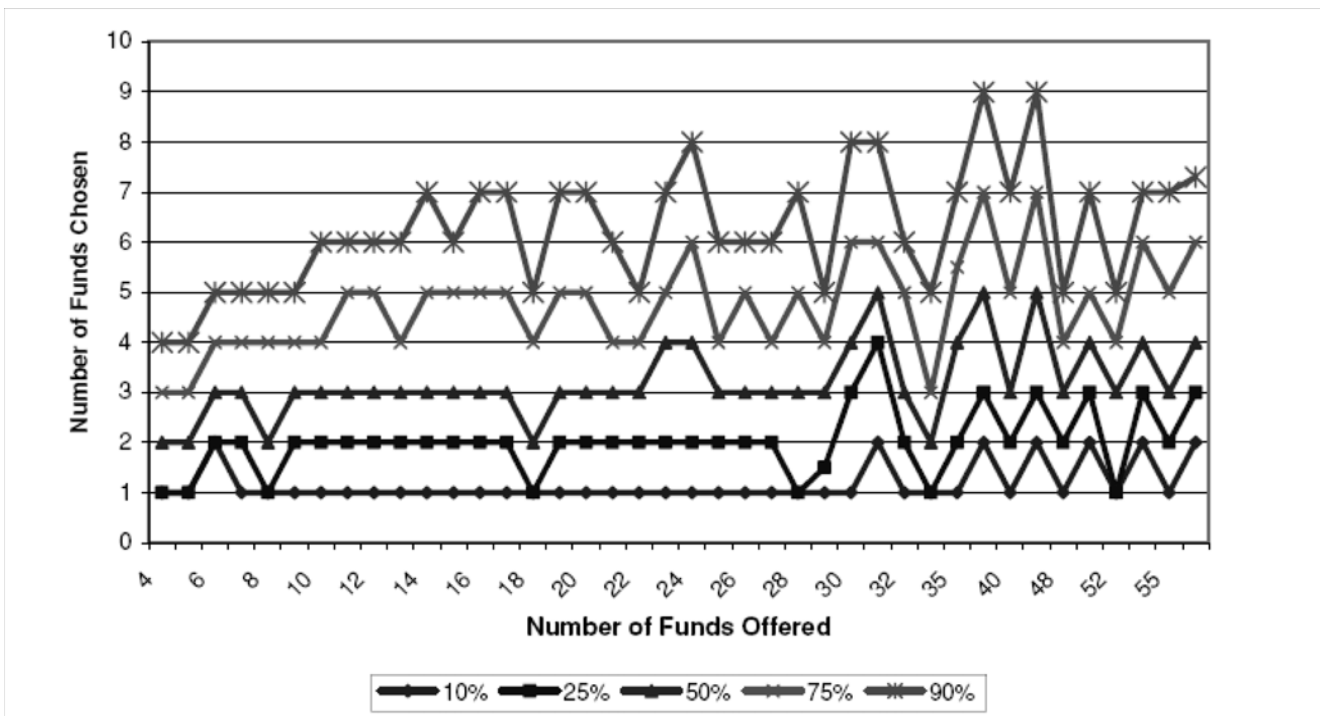


FIGURE 1. VERBAL SAVINGS QUESTIONNAIRE: HISTOGRAMS OF THE ALLOCATION TO FUND A AND THE RESULTING ALLOCATION TO STOCKS

- People respond to changes in content of Fund A and B, but incompletely
- Issues:
  - Not for real payoff
  - Low response rate (12%)
  - People dislike extreme in responses

- **Huberman-Jiang (JF, 2006)**
- Data:
  - Vanguard data to test BT (2001)
  - Data on individual choices of participants
  - Half a million 401(k) participants
  - 647 Defined Contribution plans in year 2001
  - Average participation rate 71 percent
- Summary Statistics:
  - 3.48 plans choices on average
  - 13.66 plans available on average

- **Finding 1**. People do not literally do 1/n, definitely not for n large
  - Flat relationship between $\#Chosen$ and $\#Offered$ for $\#Offered > 10$
  - BT (2001): could not estimate this + $\#Offered$ rarely above 15



- Regressions specification:

$$\#Chosen = \alpha + \beta * \#Offered + \beta X$$

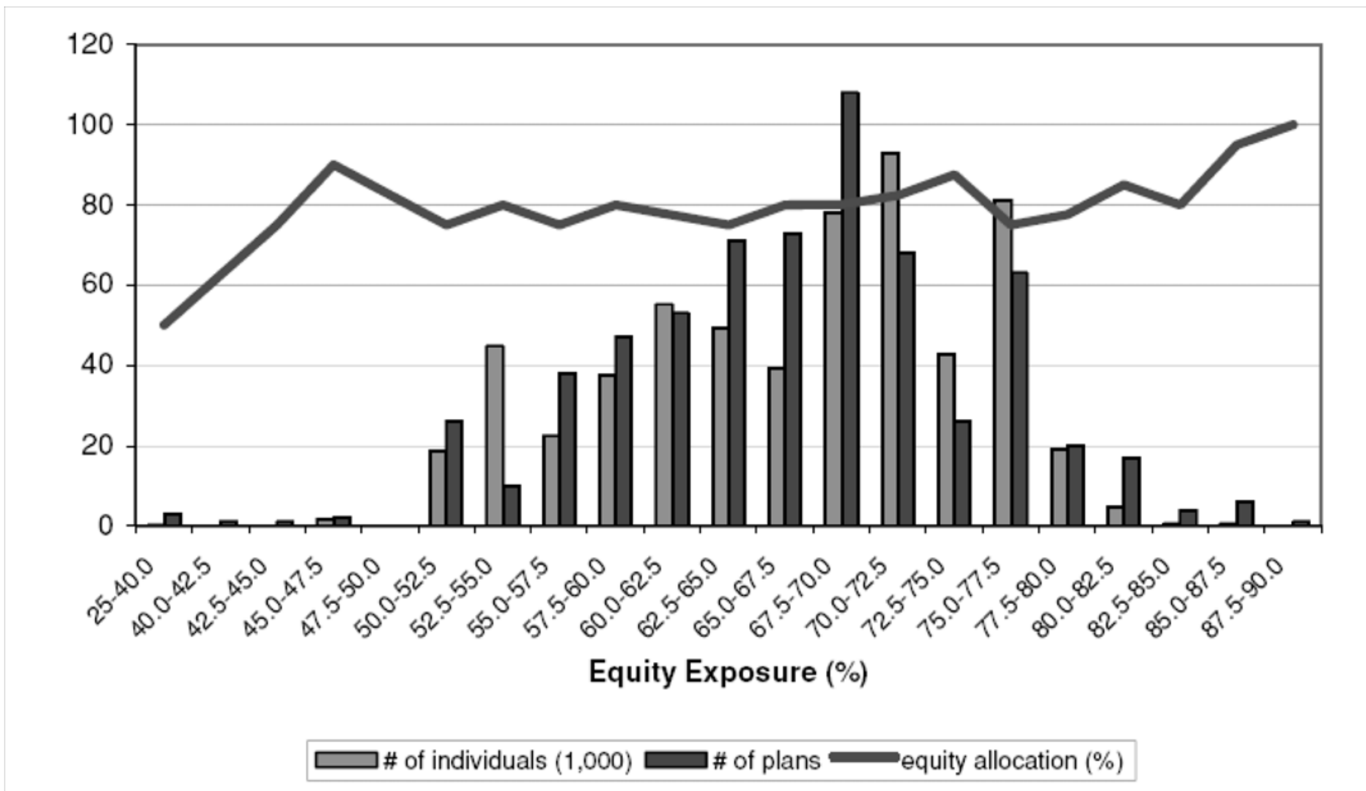
	All Participants			
	NCHOSEN			
	(1)		(2)	
	COEF × 100	SE × 100	COEF × 100	SE × 100
NCHOICE	0.95	0.70	1.03	0.70
CONTRIBUTION	10.54*	0.56	—	—
COMP	−0.02	2.30	33.05*	2.87
WEALTH	1.20*	0.51	3.90*	0.55
FEMALE	14.51*	1.97	14.84*	1.95
AGE	−1.66*	0.10	−1.35*	0.09
TENURE	0.88*	0.26	0.95*	0.26
MATCH	0.00	0.24	0.00	0.23
COMPSTK	70.67*	12.72	67.16*	12.68
DB	−6.31	15.35	−6.06	15.21
WEB	1.17	0.71	1.39	0.71
NEMPLOY	−10.28*	4.79	−9.25*	4.73
Intercept	1036.95	284.44	664.25	290.06
No. of individuals and plans	572,157	641	572,157	641
R^2	0.075		0.060	

- **Finding 2.** Employees do 1/n on the *chosen* funds if
  - number n is small
  - $1/n$ is round number

No. of Funds Chosen (1)	New Entrants (%) (2)	$\underline{H}$ (3)	$\bar{H}$ (4)	$Freq_1$ (%) (5)	$Freq_1 / \max_{j \neq 1}(Freq_j)$ (6)
1	38.6	1.0000	1.0000	–	–
2	17.5	0.5000	0.5050	64.0	12.81*
3	15.6	0.3333	0.3356	17.9	1.78*
4	13.2	0.2500	0.2513	37.4	8.89*
5	7.3	0.2000	0.2008	26.6	8.19*
6	3.5	0.1667	0.1672	1.3	0.25
7	1.8	0.1429	0.1433	1.0	0.19
8	1.1	0.1250	0.1253	3.9	1.14
9	0.6	0.1111	0.1114	5.1	1.20
10	0.4	0.1000	0.1002	53.3	13.50*

- **Finding 3.** Equity choice (most similar to BT (2001))

- In aggregate very mild relationship between $\%Equity$ and $\%EquityOffered$



- Split by $\#Offered$:
  1. For $\#Offered \leq 10$, BT finding replicates:

$$\%Equity = \alpha + .292 * \%EquityOffered$$
$$(.063)$$

  2. For $\#Offered > 10$, no effect:

$$\%Equity = \alpha + .058 * \%EquityOffered$$
$$(.068)$$

	(1)		(2)		(3)		(4)	
	All *NFunds*				*NFunds* ≤ 10		*NFunds* > 10	
	COEF	SE	COEF	SE	COEF	SE	COEF	SE
Panel A: Full Sample—Uniform Sensitivity								
%EQOffered	0.175	0.274	0.177*	0.088	0.292*	0.107	0.058	0.09
R^2	0.000		0.061		0.063		0.068	

- Psychologically plausible:
  - Small menu set guides choices –> Approximate 1/n in weaker form
  - Larger menu set does not

- BT-HJ debate: Interesting case
  - Heated debate at beginning
  - At the end, reasonable convergence: we really understand better the phenomenon
  - Convergence largely due to better data

9 Methodology: Clustering Standard Errors

- Econometric issue: Errors correlated across groups of observations

- Example 1–Huberman and Jiang (2006):
  - Errors correlated within a plan over time
  - Cluster at the plan level

- Example 2–Conlin, O'Donoghue, and Vogelsang (2007)
  - Correlations within day due to shock (TV ad)—>Cluster by day
  - Correlation within household over time —>Cluster by household

- Example 3. Earnings announcement panel
  1. Persistent shock to Company over time (Autocorrelation)
  2. Correlation in shocks across companies within date (Cross-Sectional correlation)

- OLS standard errors assume i.i.d. cross-sectionally and over time

- Clustered standard errors can take care of Issue 1 or 2 — not both:

  1. Cluster by State (Company):
     - Assume independence across States (companies)
     - Allow for any correlation over time within State (company)
  2. Cluster by year (date)
     - Assume independence across years (dates)
     - Allow for any correlation within a year (date) across States (companies)

- How does this work?

- Assume simple univariate regression:

$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$$

- OLS estimator:

$$\hat{\beta} = \beta + \left(x'x\right)^{-1} x'\varepsilon = \beta + \frac{Cov\left(x, \varepsilon\right)}{Var\left(x\right)}$$

- $Var\left(\hat{\beta}\right)$ under i.i.d. assumptions (with $\hat{\sigma}^2 = \sum_{it} \hat{\varepsilon}_{it}^2 / NT$):

$$Var\left(\hat{\beta}\right)_{OLS} = \left(x'x\right)^{-1} \sum_{i,t} \left(x_{it}\hat{\varepsilon}_{it}\right)\left(\hat{\varepsilon}_{it}x_{it}\right)\left(x'x\right)^{-1} = \frac{\hat{\sigma}^2}{\sum x_{it}^2}$$

- White-heteroskedastic:

$$Var\left(\hat{\beta}\right)_{Het} = \frac{1}{\sum_{it} x_{it}^2} \sum_{it} \frac{x_{it}^2 \hat{\varepsilon}_{it}^2}{\sum x_{it}^2}$$

- White-heteroskedastic:

$$Var\left(\hat{\beta}\right)_{Het} = \frac{1}{\sum_{it} x_{it}^2} \sum_{it} \frac{\left(x_{it}\hat{\varepsilon}_{it}\right)^2}{\sum x_{it}^2}$$

  - Notice: Second sum is weighted average of $\hat{\varepsilon}_{it}^2$, with more weight given to observations with higher x_{it}^2
  - If high x_{it}^2 is associated with high $\hat{\varepsilon}_{it}^2$, $Var\left(\hat{\beta}\right)_{Het} > Var\left(\hat{\beta}\right)_{OLS}$

- Standard Errors Clustered by I (allow for autocorrelation):

$$Var\left(\hat{\beta}\right)_{Clust} = \frac{1}{\sum_{it} x_{it}^2} \sum_{i} \frac{\left(\sum_t x_{it}\hat{\varepsilon}_{it}\right)^2}{\sum x_{it}^2}$$

  - First sum all the covariances $x_{it}\hat{\varepsilon}_{it}$ within a cluster
  - Then square up and add across the clusters
  - Notice: This is as if one cluster (one i) was one observation

- That is, this form of clustering allows

$$E(u_{it}u_{it'}|X_{it}X_{it'}) \neq 0$$

  - Correlation within cluster i

- Requires

$$E(u_{it}u_{i't'}|X_{it}X_{i't'}) = 0$$

  for $i \neq i'$

  - No correlation across clusters

- When is $Var\left(\hat{\beta}\right)_{Clust} > Var\left(\hat{\beta}\right)_{Het}$?

- Example: Assume $I = 2$, $T = 2$

$$Var\left(\hat{\beta}\right)_{Het} = \frac{1}{\sum_{it} x_{it}^2} \frac{(x_{11}\hat{\varepsilon}_{11})^2 + (x_{12}\hat{\varepsilon}_{12})^2 + (x_{21}\hat{\varepsilon}_{21})^2 + (x_{22}\hat{\varepsilon}_{22})^2}{\sum x_{it}^2}$$

- Compare to

$$\begin{aligned} Var\left(\hat{\beta}\right)_{Clust} &= \frac{1}{\sum_{it} x_{it}^2} \frac{(x_{11}\hat{\varepsilon}_{11} + x_{12}\hat{\varepsilon}_{12})^2 + (x_{21}\hat{\varepsilon}_{21} + x_{22}\hat{\varepsilon}_{22})^2}{\sum x_{it}^2} = \\ &= Var\left(\hat{\beta}\right)_{Het} + \frac{1}{\sum_{it} x_{it}^2} \frac{2x_{11}\hat{\varepsilon}_{11}\hat{\varepsilon}_{12}x_{12} + 2x_{21}\hat{\varepsilon}_{21}\hat{\varepsilon}_{22}x_{22}}{\sum x_{it}^2} \end{aligned}$$

  - Hence, $Var\left(\hat{\beta}\right)_{Clust} > Var\left(\hat{\beta}\right)_{Het}$ if $Ex_{i1}x_{i2} > 0$ and $E\hat{\varepsilon}_{i1}\hat{\varepsilon}_{i2} > 0$ –>Positive correlation within cluster (that is, over time) among x variables and ε

- Positive correlation –> Standard errors understated if no clustering

- Notice that instead this does not capture correlation across clusters, that is, $E\hat{\varepsilon}_{1t}\hat{\varepsilon}_{2t} = 0$ and $Ex_{1t}x_{2t} > 0$

- Assume now that we cluster by T instead (allow for cross-sectional correlation):

$$Var\left(\hat{\beta}\right)_{Clust} = Var\left(\hat{\beta}\right)_{Het} + \frac{1}{\sum_{it} x_{it}^2} \frac{2x_{11}\hat{\varepsilon}_{11}\hat{\varepsilon}_{21}x_{21} + 2x_{12}\hat{\varepsilon}_{12}\hat{\varepsilon}_{22}x_{22}}{\sum x_{it}^2}$$

- Hence, $Var\left(\hat{\beta}\right)_{Clust} > Var\left(\hat{\beta}\right)_{Het}$ if $Ex_{1t}x_{2t} > 0$ and $E\hat{\varepsilon}_{1t}\hat{\varepsilon}_{2t} > 0$ –>Positive correlation within a time period across the observations among x variables and ε

- Calculation of Adjustment of Standard Errors due to Clustering
  - T observations within cluster
  - Within-cluster correlation of x_s: ρ_x
  - Within-cluster correlation of ε: ρ_ε

- Compare $Var\left(\hat{\beta}\right)_{Clust}$ and $Var\left(\hat{\beta}\right)_{OLS}$:

$$Var\left(\hat{\beta}\right)_{Clust} = Var\left(\hat{\beta}\right)_{OLS} * \left(1 + (T-1)\,\rho_x \varrho_\varepsilon\right)$$

  - Standard errors downward biased with OLS if $\rho_x \varrho_\varepsilon > 0$, or positive correlations (as above)
  - No bias if no correlation in *either* x or ε
  - Bias larger the larger is T
  - Illustrative case: Suppose all observations within cluster identical ($\rho_x = \rho_\varepsilon = 1$) –> Bias $= T$

- Issues with clustering:

- Issue 1. **Number of clusters**
  - Convergence with speed I $->$ Need a large number of clusters I to apply LLN
  - Beware of papers that apply clustering with $<$20 clusters
  - **Cameron-Gelbach-Miller (2008):** Test with good finite sample properties even for $I \approx 10$

- Issue 2. **Cluster in only one dimension**
  - Clustering by I controls for autocorrelation
  - Clustering by T controls for cross-sectional correlation
  - How can control for both? **Cameron-Gelbach-Miller (2006):** Two-way clustering, can do so

- **Cameron-Gelbach-Miller (2006).** Double-clustered standard errors with respect to I and T

- Procedure:

  1. Compute standard errors clustering by I –> Compute $V\left(\hat{\beta}\right)_{Cl-I}$
  2. Compute standard errors clustering by T–> Compute $V\left(\hat{\beta}\right)_{Cl-T}$
  3. Compute standard errors clustering by $T*I$ (this typically means s.e.s not clustered, just robust)–> Compute $V\left(\hat{\beta}\right)_{Cl-T*I}$
  4. Final variance and covariance matrix is

$$V\left(\hat{\beta}\right)_{DoubleCl} = V\left(\hat{\beta}\right)_{Cl-I} + V\left(\hat{\beta}\right)_{Cl-T} - V\left(\hat{\beta}\right)_{Cl-T*I}$$

- Intuition: It's variance obtained clustering along one dimension (say, I), plus the additional piece of variance along the other dimension that goes beyond the robust s.e.s

- Readings on clustered standard errors:
  - **Stata Manual** –> basic, intuitive
  - **Bertrand-Duflo-Mullainathan (QJE, 2004)** –> Excellent discussion of practical issues with autocorrelation in diff-in-diff papers, good intuition
  - **Peterson (2007)** –> Fairly intuitive, applied to finance
  - **Cameron-Trivedi (2006) and Wooldridge (2003)** –> More serious treatment
  - **Colin Cameron (Davis)'s website** –> Updates

10 Next Lecture

- Menu Effects II
- Persuasion
- Social Pressure