Observational Empirics in Behavioral Economics

Summer Institution in Behavioral Economics
July 2018

Kelly Shue (Yale and NBER)

How to do behavioral research using observational data

Disclaimer: These are my personal thoughts; Others may disagree and there are many paths to research success.

1. Why use observational data?
2. Coming up with research ideas
3. Obtaining data
4. Two templates for behavioral observational papers: "**Facts First**" and "**Hypothesis First**"
5. Example of a "fact-first" observational paper

Why use observational data?

We know of many behavioral biases that are plausible in theory and manifest in controlled laboratory experiments

How much do these biases still matter in settings with

- High monetary stakes and/or large welfare consequences?
- Experienced professionals making decisions?
- The disciplining presence of market competition and arbitrage?
- Repeated interactions and the potential for learning?

Observational data helps address these concerns

Why use observational data?

What about field experiments?

A large scale field experiment is superior to observational data!

- But, large-scale field experiments are very expensive, and approval from the IRB or the government is difficult to obtain
- Also ethical considerations

Observational data offers large sample sizes and is easier to collect

- But, you have to be more creative in finding credible random variation

Downsides of observational data

Scope for randomization and experimentation is limited

- Have to rely on discontinuities or IV strategies to establish causation, which measure local average treatment effects
- Difficult to test for the exact decision-making error that leads to the behavior, so papers often focus on the reduced-form effect and then speculate about channels

However, most readers (I hope) should be Bayesians

- Even if a paper can't prove something with 100% certainty, the paper can make a contribution by introducing a new idea or offering evidence that is ***convincing enough*** to help update priors

Coming up with research ideas

It is hard! Some ideas come from…

Thinking about the existing theory and body of research

- Work through a list of biases (loss aversion, prospect theory, belief in the law of small numbers, confirmation bias, etc.)
- Potential downside: Might be viewed as a minor extension of a known phenomenon or your advisor's research agenda

Introspection and thinking about where you see mistakes

- Reflect upon your previous work experience in industry or as an RA

Coming up with research ideas

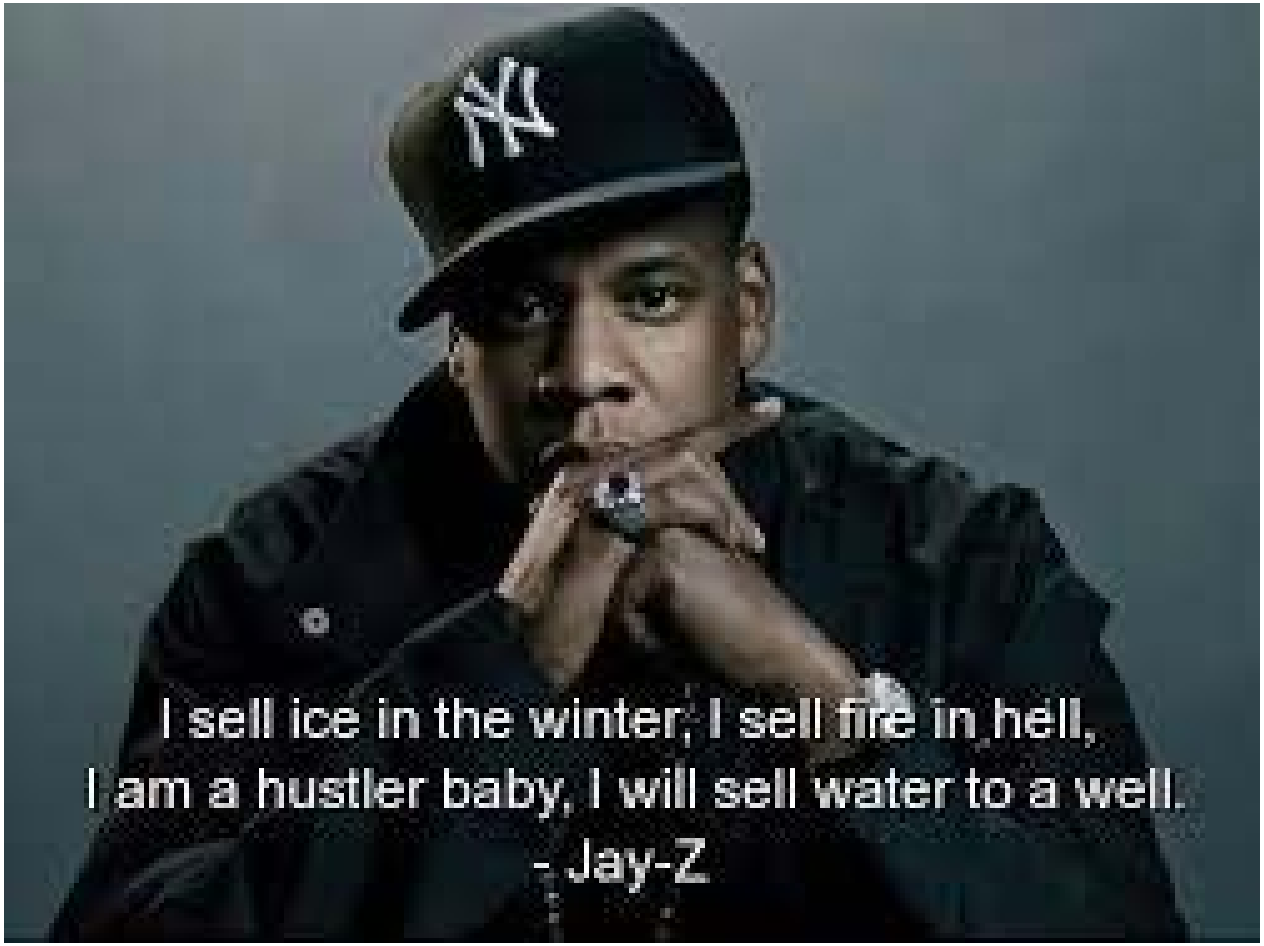
Some research ideas come after the data

- While research ideally starts hypothesis first, sometimes we get lucky and a dataset lands in our lap
- Run a lot of summary stats and graph the data to generate research ideas
- Last resort: make a “stone soup”

Be engaged with both academic and everyday life, because a research idea could come from anywhere

- Attend seminars, do referee reports, talk with friends and family including those outside of economics, read books, read blogs, read the news, watch TV
- Exploit any interdisciplinary advantages
- **Keep an idea list!**

Obtaining data



Obtaining data

"Your comparative advantage as a graduate student is getting data" – David Card

1. Use data that is available for free or accessible through a university subscription (WRDS, CRSP, NLSY, CPS, NBER, PSID, HRS, etc.)

Take advantage of your current status as a graduate student

- Download everything available while you are affiliated with a school with a generous data subscription policy
- After you graduate, your data access is approximately the union of all the data that is directly accessible to those in your PhD social network (subject to legal limitations)

Obtaining data

2. Apply for access to restricted data

- Apply early and be aware of the risks! E.g., census data access can take years and is difficult to access continuously, which is very bad for successful navigation of R&R's.
- For Census, you can try to get access as a research assistant rather than directly
- For Fed data, graduate students are at an advantage. You can access restricted data through the Fed summer PhD program
- Some companies / research groups will let you use their data if you work for them for a year, but be prepared to "sell your soul"

Obtaining data

3. Approach other academics that have data (sometimes with the option of coauthorship)

- I've asked for data from other people's field experiments to test a different behavioral research question
  - Positive side effect: Nobody can accuse you of setting up the experiment to get your desired results
- Some data from previously published papers are posted on Journal websites
- Authors may say "no," but they often suggest alternative data sources or alternative ways to access the same data

Obtaining data

4. Ask companies for data

- Some executives are flattered by the request
  - E.g., my coauthor Alan Benson was the TA for an executive education class at Sloan. He successfully asked his student (a C-level executive) for valuable company records
- Don't limit yourself to only top-level executives
  - E.g., I obtained access to HBS student records going back to 1949 because I approached an administrator at Alumni Relations instead of going through the Dean's Office
- Don't be afraid of rejection: "Once a week, I would ask a company for their data" – Devin Pope

Obtaining data

5. Scrape data from online websites

- Massive amounts of data available
- No need to know Python (e.g. freelancer.com, or hire a CS undergraduate)
- Create data from text using machine learning methods and textual analysis
- Public records of hospital visits, pollution, weather, agriculture, oil and gas, real estate transactions, etc.
- Explore recent data leaks, e.g., Panama Papers (again, subject to legal limitations)

Obtaining data

6. Request data from non-standard sources

FOIA (Freedom of Information Act) requests can be used to obtain detailed records from any public institution

- E.g. voting records, schedule of FDA approval process, judicial decisions, etc.

Other sources:

- Sports data (high quality, but less important for welfare)
- Transportation data (taxis, flights, private planes and boats)
- Non-US countries may have more generous data sharing policies and be less concerned about privacy

Obtaining data

7. Don’t give up on the old, well-mined, data

Sometimes, it is sufficient to look at the old data with fresh eyes

Examples of visualizing old data in new ways:

- Asset pricing in the frequency domain (Dew-Becker and Giglio 2018)
- The Speed of Price Discovery: Trade Time vs Clock Time (Santosh 2018)
- I’ll give another example at the end of this lecture

Template #1 for a behavioral observational paper: **Facts First**

1. Establish an interesting empirical fact
   - Start with the simplest exposition of the data
   - Rely heavily on graphical evidence
   - Slowly add technical complexity to assess mechanisms, heterogeneity, and robustness

2. Provide a discussion of your preferred psychological theory that explains the empirical fact
   - Consider and attempt to reject alternative explanations
   - If the empirical fact is important enough, this section can be more speculative – the main takeaway is the fact itself

Also known (disparagingly?) as a "puzzle" or "anomaly" paper in Finance

Template #2 for a behavioral observational paper: **Hypothesis First**

1. Motivate your hypothesis
   - If the hypothesis is new, motivate with related evidence
   - If the hypothesis is old, discuss the existing theory and empirical research (make sure to include psychology research), and explain why more observational evidence is needed

2. Test the hypothesis in observational data
   - What are the benefits of your particular empirical setting?
   - Test the main and secondary predictions of the theory
   - Present placebo tests, heterogeneity tests, robustness tests

3. Consider and attempt to reject alternative explanations

Tradeoffs

Fact First

- Pro: The fact stands by itself, regardless of the mechanism
- Pro: More acceptable to referees who aren't friendly toward behavioral explanations
- Mixed: Not a pure "behavioral" paper, more like, e.g., a public policy paper with a possible behavioral explanation
- Con: High bar for the interestingness of the fact
- Con: You might get accused of data snooping

Hypothesis First

- Pro: Less likely to be accused of data snooping
- Con: Even higher bar for identification, although you *might* be able to lower the identification bar if the hypothesis is novel or the paper's main contribution is theoretical

Examples of these two templates

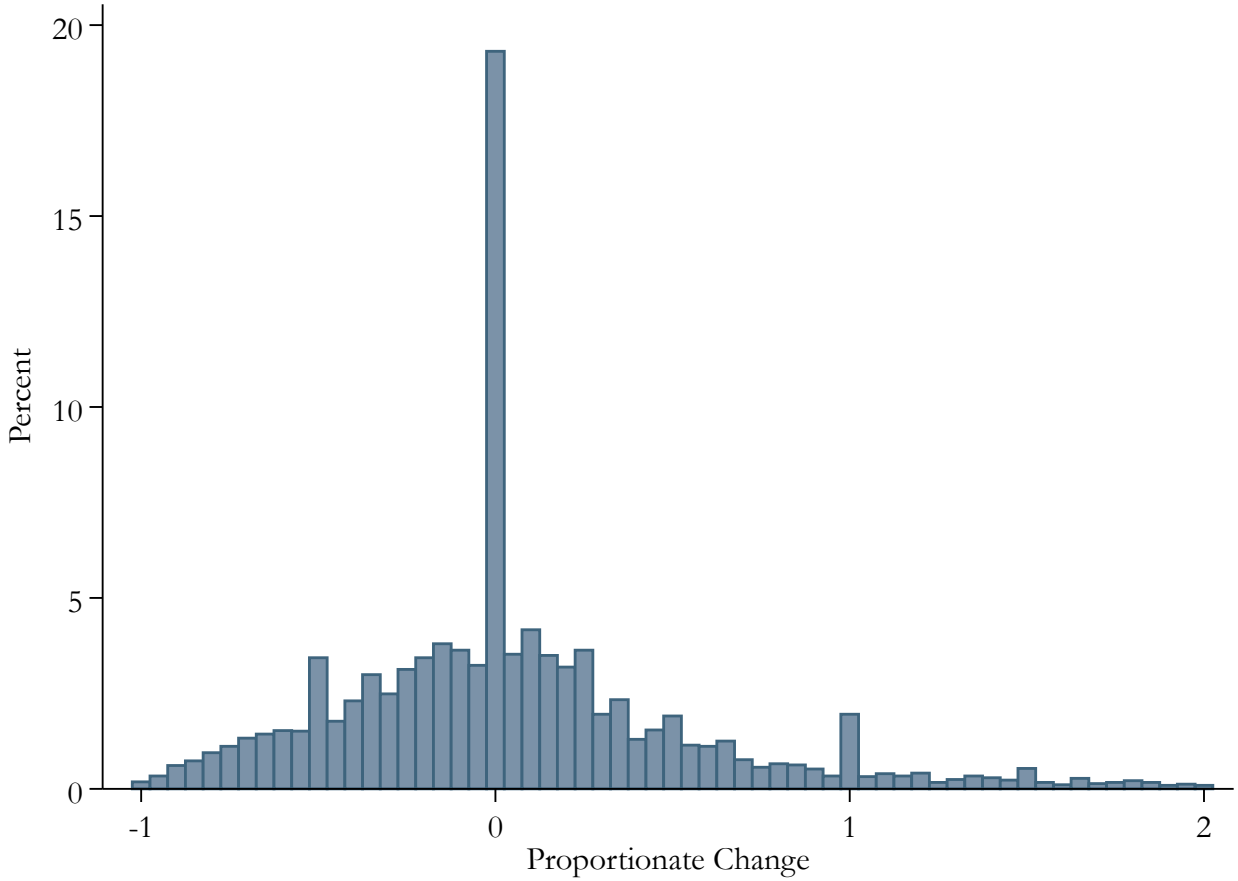
Hypothesis First

- "Consistent Good News and Inconsistent Bad News"
- "Can the Market Multiply and Divide? Non-proportional Thinking in Financial Markets"

Facts First

- "Growth through Rigidity: An Explanation for the Rise in CEO Pay"

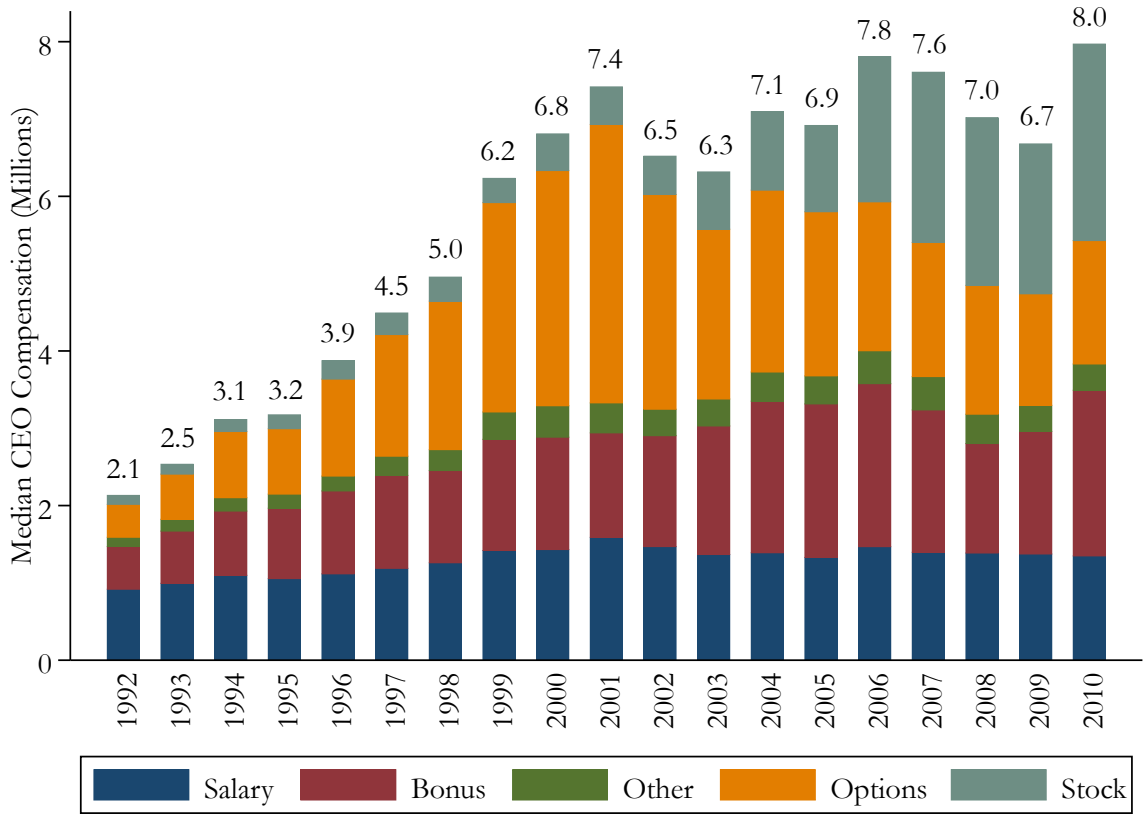
Options grants are rigid in number (S&P 500 CEOs)



Histogram: Proportional change in number of options granted this year relative to last year

- Tend to grant **same** or **round multiple** of last year's number

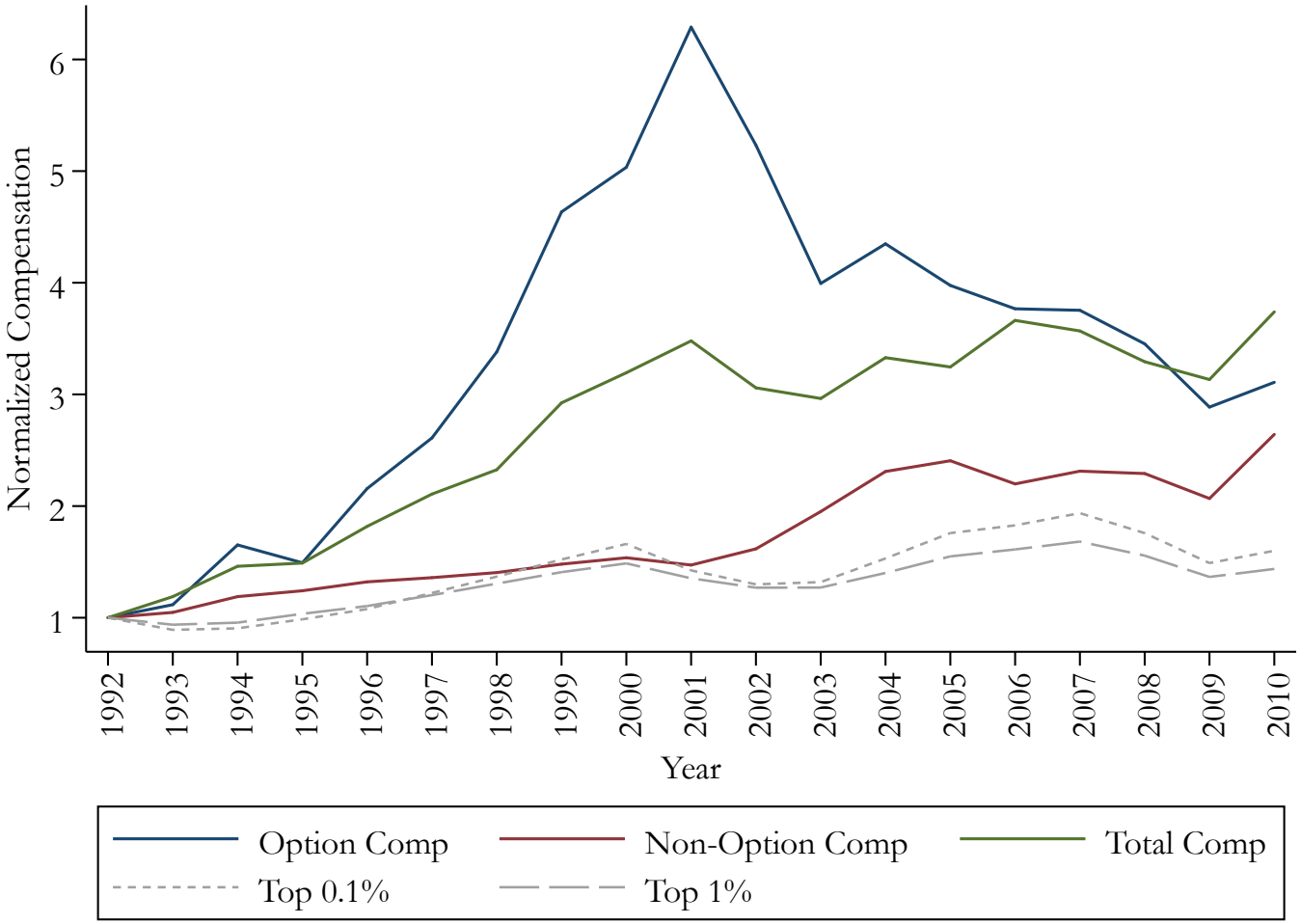
S&P 500 CEO compensation



Dramatic growth in CEO pay during Tech Boom: 1990s and early 2000s

- Off the long run trend
- Growth driven by options grants (grant date value)

Growth relative to other high income groups



Executive compensation outpaced growth among other high income professions during the Tech Boom, slowed in mid 2000s

Other trends in executive compensation

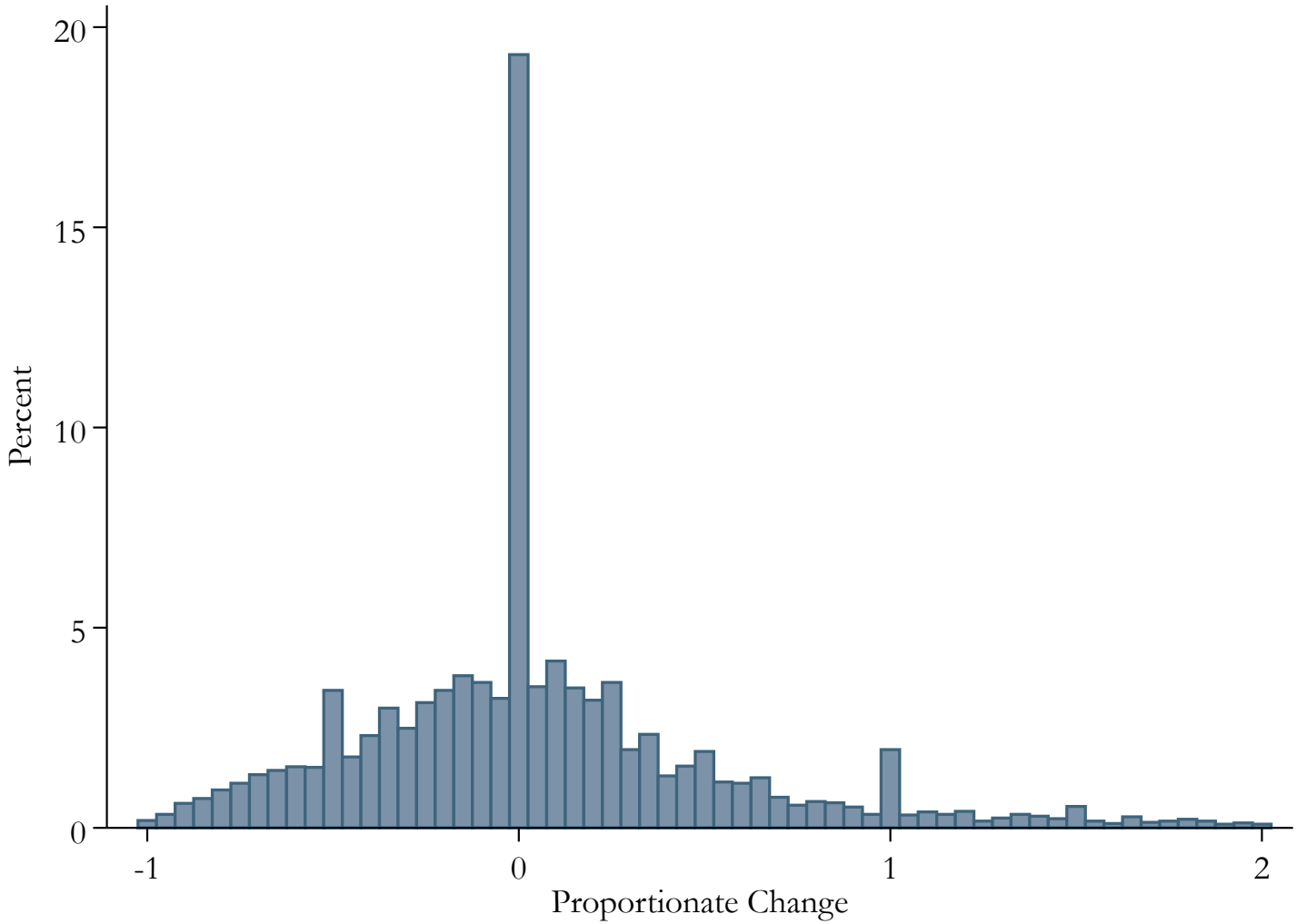
Dispersion in compensation across firms rose dramatically in 1990s

US CEO pay exceeds pay in other countries by $> 2X$ (controlling for industry and firm size) and more equity-based

Near perfect correlation between average executive pay and the S&P 500 index in the 1990s and early 2000s.

- Murphy (2012): Consistent with compensation committees focusing on number rather than value

How can rigidity help explain these trends?



How can rigidity help explain these trends?

		Stock price		
		Year 1 Grant	**Year 2 Grant**	**Year 3 Grant**
Plan		**100**	**120**	**144**
Fixed Value	Value of Options	$1,000,000	$1,000,000	$1,000,000
	Number of Options	28,128	23,440	18,752
Fixed Number	Value of Options	$1,000,000	$1,200,000	$1,440,000
	Number of Options	28,128	28,128	28,128

- Number rigidity $\Rightarrow$ Grant date value moves proportionally with share price
- We focus on options (bulk of growth in pay) but same logic applies for number-rigidity in restricted stock grants

How can rigidity help explain these trends?

- Many firms start granting options in early '90s
- Conditional on paying options, **tendency toward number-rigidity**
- Number-rigidity, combined with high equity returns during the Tech Boom could account for option-driven
  - Increase in level of total compensation
  - Increase in dispersion and correlation with market returns
  - "Excess" pay compared to other countries
- Salary and bonus are **downward rigid in value** and did not offset gains in options
- Number-rigidity declined in 2000s with regulatory changes requiring disclosure/expensing of grant date fair value of options

Two broad classes of theories

#1: All parties involved in the compensation-setting process understand option valuation, and number-rigidity arises from optimal contracting

- Keeps ownership % roughly constant or growing at a fixed rate
- Steep pay-for-performance incentives, including pay for luck

May be optimal, but

- Potential mismatch between many theories and downward rigidity in other subcomponents of pay

Two broad classes of theories

#2: At least one party in the compensation-setting process is unsophisticated about how to value options

Option grants were new, many did understand or trust valuation methods such as Black-Scholes

- Used number as a proxy for value
- Akin to ***money illusion***: Thinking about real money in nominal units

Lack of sophistication may be on the part of the CEO, board, shareholders (outrage constraint), law makers, and/or lower level employees

Towers Perrin CompScan Survey (1998)

Last year, you personally received options to purchase 1,000 shares ... at the stock's then current price of $50. This year, the share price is up to $70. How many options should you get?

A) 1,500
B) 1,000
C) 715

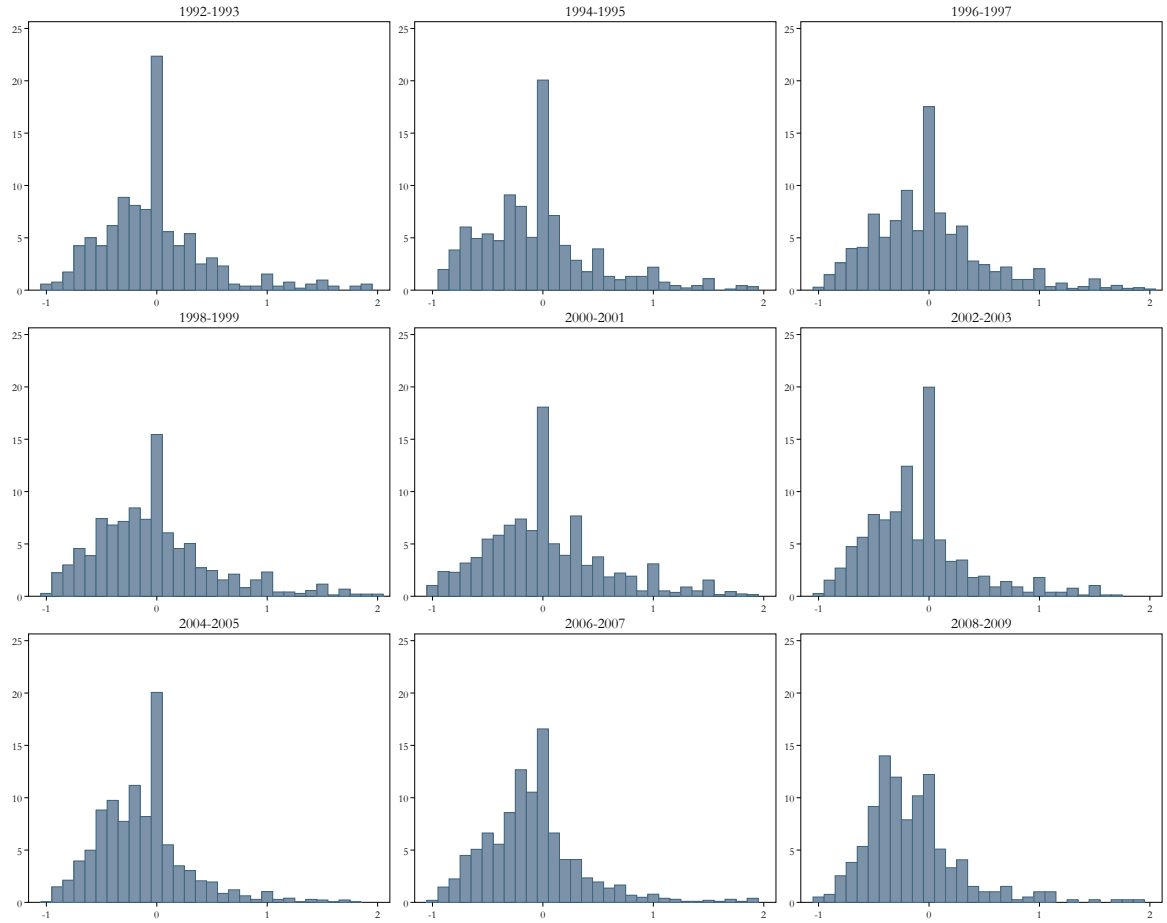
If you chose A, you're in the vast majority [>50% of survey respondents] of option recipients who think they should get more, not fewer, options when the price goes up.

If you selected B ... you're not expecting a bigger grant, more than the 1,000 options you received last year, but you also can't see why the size should be cut back when the stock has performed well.

Sensitivity to reporting requirements

- FASB and SEC began requiring that firms expense/disclose options at the grant date fair value
- In comparison, stock grants are easier to value and the grant date value was always disclosed and expensed
- Sensitivity to reporting requirements doesn't mesh with simple explanations of rigidity as an optimal choice
- After the change in regulations in 2006, number rigidity and the growth of compensation stopped!
- Firms that voluntarily expensed the value of options prior to 2006 did not grant fixed numbers of options and had lower compensation growth

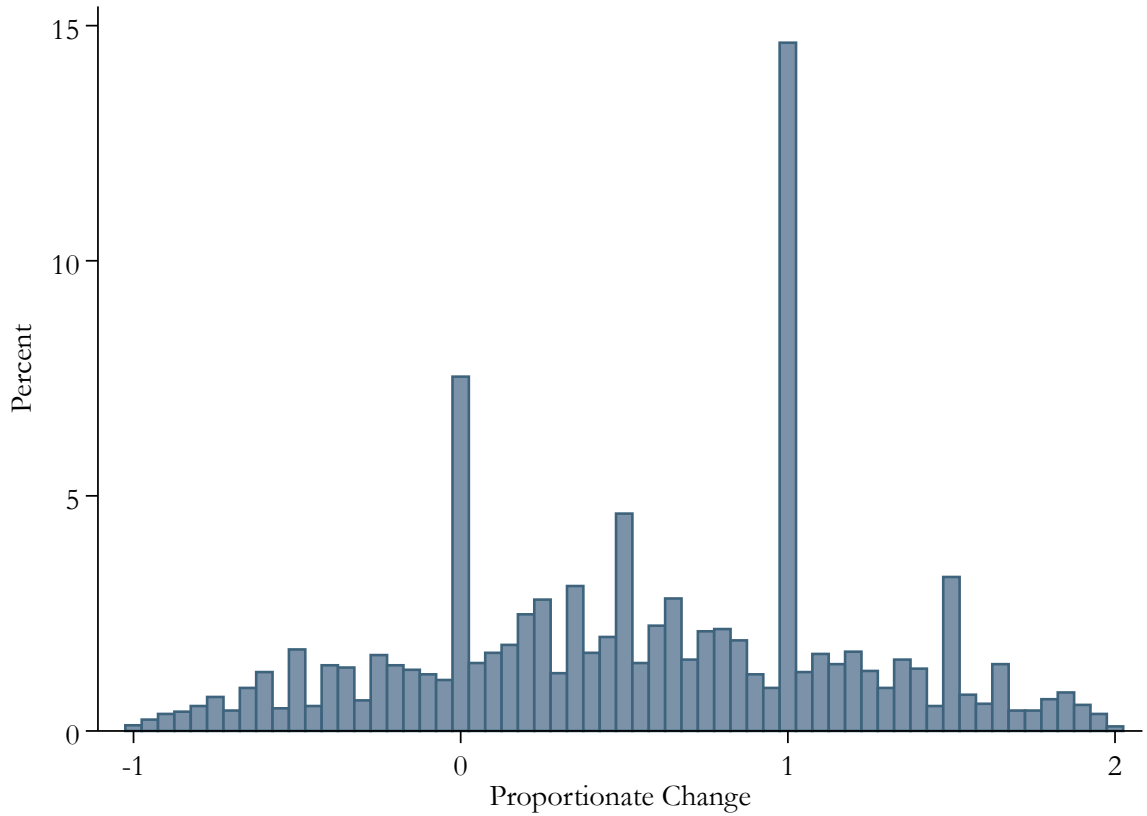
Changes over time (firm return $\geq 25\%$ sample)



For firms with returns $\geq 25\%$, we should see mass to the left of 0 if we expect the value of pay to grow by less than 25%: This is increasingly true over time

Number-rigidity after stock splits

"Stock splits also offer an opportunity to readjust grant levels, moving back toward more competitive levels, without jolting employees' perceptions or expectations quite so drastically."



Testing the two theories for sources of number-rigidity

1. Expensing/disclosure of the value of options $\Longrightarrow$ less likely to grant number rigid options
   - If number-rigidity is optimal, the contract should remain optimal regardless of whether number or value is directly disclosed/expensed
2. Better governed firms $\Longrightarrow$ less likely to grant number rigid options
3. Did sophisticated CEOs conceal high pay from naive shareholders? Maybe, but some CEOs may have been unsophisticated / lost money
   - Financially sophisticated CEOs $\Longrightarrow$ number rigidity less likely
   - Some CEOs receive same number of options after a stock split

Concluding thoughts

Benefits of observational data

- More and more available
- Provides enormous statistical power
- Can show evidence of where psychology is actually having an impact on people's lives

Keep it **simple**

- Research is a much easier and pleasant process if you are documenting an empirical fact that is jumping out of the data (no need to p-hack)
- Striking facts stand on their own
- Sophisticated empirical tests are the "icing on the cake"