Class 12 - Design I - (Quasi)-Experimental Data

Agenda

- Quasi-experimental designs: what, why, when, how (30 minutes)
- Application paper discussion (30 minutes)
- Replication presentation (Group 15 minutes)
- Skills corner - Class walkthrough in R (25 minutes)
- General discussion (10 minutes)
- *Break* then Class 13

Quasi-experimental designs

Preamble: Curve-fitting v. causality

Two different regimes:

- Curve-fitting: Statistically identifying the parameters of a mathematical relationship between two or more variables based on the observed data, and making a law-like statement that summarizes empirical regularities
- Causality: Imposing further structure to impose an if-then logical sequencing; where a change in one variable will necessarily have a causal impact on the other. Usually couched in terms of counterfactuals or "but for" reasoning: if not for the change in X, we would have observed a different Y

What is a "quasi-experiment"?

A quasi-experiment is a study that takes place in a field setting and involves a change in a key independent variable of interest but relaxes one or both of the defining criteria of laboratory and field experiments: random assignment to treatment conditions and controlled manipulation of the independent variable (Grant and Wall 2009, 655)

What is a "quasi-experiment"?

"Natural experiments" as a form of quasi-experiment:

> *Quasi-experiments also include changes to an independent variable that are naturally occurring rather than manipulated (Grant and Wall 2009, 655)*

What is a "quasi-experiment"?

What about "regression discontinuity designs"?

> *Regression discontinuity designs (RDDs) are sometimes also called cutoff-based designs. They assign units to conditions based on a cut- off score on an ordered assignment variable, with units that fall on one side of the cutoff receiving treatment and those on the other side receiving the comparison condition. (Shadish and Cook 2009, 615)*

What is a "quasi-experiment"?

What about "interrupted time series" or "structural breaks"?

> *Similar to RDD, an effect is measured as a change in the slope or intercept of the time series at the point of treatment introduction. (Shadish and Cook 2009, 617)*

Why are quasi-experiemental designs useful?

Per Grant and Wall (2009), the benefits include but are not limited to:

1. strengthening causal inference when random assignment and controlled manipulation are not possible or ethical; and
2. building better theories of time and temporal progression.

Employing a quasi-experiement

What are the common threads among many of these designs?

Two groups: treatment and control

- The amount of “randomness” differs across the designs
- The distinction between treatment and control also varies

What are the common threads among many of these designs?

Two+ time periods or comparison points: pre-/post or with/without treatment

- The period of followup before and after differs
- The “fuzziness” of the treatment can also vary

What are the common threads among many of these designs?

Accounting for covariates

- Random assignment is used to “wash out” differences between the groups
- Propensity scores and control variables used to account for residual differences

The aim: The experimental ideal

1. Experimenter intervention in selection of indepedent variable (treatment)
2. Random assignment to treatment and control group
3. Creation of a stark counterfactual: what would happen **but for** the treatment?

Assuming complications like "compliance" and other experimental design concerns are addressed, this allows for a clean determination of counterfactuals

The aim: The experimental ideal

Per Rubin's Potential Outcomes Framework (Angrist and Pischke 2008):

$$E[y_{1i}|d_i = 1] - E[y_{0i}|d_i = 0] =$$
$$\boldsymbol{E[y_{1i} - y_{0i}|d_i = 1]} + (E[y_{0i}|d_i = 1] - E[y_{0i}|d_i = 0])$$

- The LHS (top) is the difference between the treatment and control group in the population
- The bolded term is the "average treatment effect on the treated" - the causal effect that occurs for the treated group
- The RHS term in parentheses is the selection bias

The aim: The experimental ideal

$$E[y_{1i}|d_i = 1] - E[y_{0i}|d_i = 0] =$$
$$\boldsymbol{E[y_{1i} - y_{0i}|d_i = 1]} + (E[y_{0i}|d_i = 1] - E[y_{0i}|d_i = 0])$$

The appeal of the experimental ideal is twofold:

1. We have a treatment and control group to estimate the LHS.
2. We use random assignment to set the selection bias term to 0.

Taken together, this allows us to **identify** the causal effect of interest.

How can you design a quasi-experiement?

A potential second-best option: Difference-in-differences

Perhaps we can't get to the experimental ideal.

But let's say we do have a treatment and a control group that have been selected via a "natural experiment" (e.g., close elections, change in regulations, exogenous shocks) that helps us mitigate selection bias.

We can track the change in outcomes over time for both groups and get a difference before the treatment t_1 (pre-treatment) and after the time of treatment t_2 (post-treatment).

We can then look at the difference-in-differences between these two groups to estimate the causal effect.

Estimating difference-in-differences

D-i-D as a comparison of means:
$DiD = (Y_{post,treated} - Y_{pre,treated}) - (Y_{post,control} - Y_{pre,control})$

D-i-D in a regression framework: $Y = \beta_0 + \beta_1 T + \beta_2 \delta + \beta_3 \delta T$

where $\delta = 0$ indicates pre-treatment and $\delta = 1$ indicates post-treatment and T = 1 indicates in the treatment group while T = 0 indicates the control group

In this setup, the estimated coefficient β_3 is the D-i-D estimator

Assumptions

There are many assumptions to make the interpretation of a difference-in-differences casual:

Assumptions

In order to estimate any causal effect, three assumptions must hold: exchangeability, positivity, and Stable Unit Treatment Value Assumption (SUTVA)1
. DID estimation also requires that:

- Intervention unrelated to outcome at baseline (allocation of intervention was not determined by outcome)
- Treatment/intervention and control groups have Parallel Trends in outcome (see below for details)
- Composition of intervention and comparison groups is stable for repeated cross-sectional design (part of SUTVA)
- No spillover effects (part of SUTVA)

Selected comparisons to difference-in-differences

RDD: Does not require temporal differences, identification of effect through discontinuity in a continuous variable that determines placement in treatment group

Structural breaks: Does not have a contemporaneous control group, the assumption is that the existing trend would continue but for the structural break

Randomized field experiments: Closer to experimental ideal in principle but can be logistically challenging and faces hurdles of attrition, incomplete treatment , and Hawthorne effects

Lab experiments (e.g.,Shu et al. (2012) and countless OB lab studies): Closer to experimental ideal but there could be significant issues with external validity / generalizability beyond the lab setting

Applications

Application readings

Let's level-set people's familiarity with these pieces.

- Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. Proceedings of the National Academy of Sciences, 109(38), 15197–15200. doi:10.1073/pnas.1209746109 (see also https://datacolada.org/109)
- Penrosian capacity as a constraint on entrepreneurial growth: An exploratory study employing the dot-com bubble (working paper)

Shu et al (2012)

- What was this paper about?
- What were the findings?
- What was the method?
- What makes sense? What was confusing?

Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end

Lisa L. Shu[a], Nina Mazar[b,1], Francesca Gino[c], Dan Ariely[d], and Max H. Bazerman[c]

[a]Kellogg School of Management, Northwestern University, Evanston, IL 60208; [b]Rotman School of Management, University of Toronto, Toronto, ON, Canada M5S 3E6; [c]Harvard Business School, Harvard University, Boston, MA 02163; and [d]Fuqua School of Business, Duke University, Durham, NC 27708

Edited* by Daniel Kahneman, Princeton University, Princeton, NJ, and approved July 23, 2012 (received for review June 11, 2012)

Many written forms required by businesses and governments rely on honest reporting. Proof of honest intent is typically provided through signature at the end of, e.g., tax returns or insurance policy forms. Still, people sometimes cheat to advance their financial self-interests—at great costs to society. We test an easy-to-implement method to discourage dishonesty: signing at the beginning rather than at the end of a self-report, thereby reversing the order of the current practice. Using laboratory and field experiments, we find that signing before–rather than after–the opportunity to cheat makes ethics salient when they are needed most and significantly reduces dishonesty.

morality | nudge | policy-making | fraud

The annual tax gap between actual and claimed taxes due in the United States amounts to roughly $345 billion. The Internal Revenue Service estimates more than half this amount is due to individuals misrepresenting their income and deductions (1). Insurance is another domain burdened by the staggering cost of individual dishonesty; the Coalition Against Insurance Fraud estimated that the overall magnitude of insurance fraud in the United States totaled $80 billion in 2006 (2). The problem with curbing dishonesty in behaviors such as filing tax returns, submitting insurance claims, claiming business expenses or reporting billable hours is that they primarily rely on self-monitoring in lieu of external policing. The current paper proposes and tests an efficient and simple measure to reduce such dishonesty.

Whereas recent findings have successfully identified an intervention to curtail dishonesty through introducing a code of conduct in contexts where previously there was none (3, 4), many important transactions already require signatures to confirm compliance to an expected standard of honesty. Nevertheless, as significant economic losses demonstrate (1, 2), the current practice appears insufficient in countering self-interested motivations to falsify numbers. We propose that a simple change of the signature location could lead to significant improvements in compliance.

Even subtle cues that direct attention toward oneself can lead to surprisingly powerful effects on subsequent moral behavior (5–7). Signing is one way to activate attention to the self (8). However, typically, a signature is requested at the end. Building on Duval and Wicklund's theory of objective self-awareness (9), we propose and test that signing one's name before reporting information (rather than at the end) makes morality accessible right before it is most needed, which will consequently promote honest reporting. We propose that with the current practice of signing after reporting information, the "damage" has already been done: immediately after lying, individuals quickly engage in various mental justifications, reinterpretations, and other "tricks" such as suppressing thoughts about their moral standards that allow them to maintain a positive self-image despite having lied (3, 10, 11). That is, once an individual has lied, it is too late to

the extent that written reports feel more distant and make it easier to disengage internal moral control than verbal reports, written reports are likely to be more prone to dishonest conduct (3, 10, 11). However, for both types of reports (verbal or written) we hypothesize a pledge to honesty to be more effective before rather than after self-reporting. Thus, in this work, we test an easy-to-implement method of curtailing fraud in *written* reports: signing a statement of honesty at the beginning rather than at the end of a self-report that people know from the outset will require a signature.

Results and Discussion

Experiment 1 tested this intervention in the laboratory, using two different measures of cheating: self-reported earnings (income) on a math puzzles task wherein participants could cheat for financial gain (3), and travel expenses to the laboratory (deductions) claimed on a tax return form on research earnings. On the one-page form where participants reported their income and deductions, we varied whether participant signature was required at the top of the form or at the end. We also included a control condition wherein no signature was required on the form.

We measured the extent to which participants overstated their income from the math puzzles task and the amount of deductions they claimed. All materials were coded with unique identifiers that were imperceptible to participants, yet allowed us to track each participant's true performance on the math puzzles against the performance underlying their income reported on the tax forms. The percentage of participants who cheated by overclaiming income for math puzzles they purportedly solved differed significantly across conditions: fewer cheated in the signature-at-the-top condition (37%) than in the signature-at-the-bottom and no-signature conditions (79 and 64%, respectively), $\chi^2(2, n = 101) = 12.58$, $P = 0.002$, with no differences between the latter two conditions ($P = 0.17$). The results also hold when analyzing the average magnitude of cheating by condition; Fig. 1 depicts the reported and actual performance, as measured by the number of math puzzles solved, for each condition, $F(2, 98) = 9.21$, $P < 0.001$. Finally, claims of travel expenses followed that same pattern and differed by condition, $F(2, 98) = 5.63$, $P < 0.01$, $\eta 2 = 0.10$. Participants claimed fewer expenses in the signature-at-the-top condition (M = $5.27, SD = 4.43) compared with signature-at-the-bottom (M = $9.62, SD = 6.20; $P < 0.01$) and the no-signature condition (M = $8.45, SD = 5.92; $P < 0.05$), with no differences between the latter two conditions ($P = 0.39$). Thus, signing before reporting

Author contributions: L.L.S., N.M., F.G., D.A., and M.H.B. designed research; L.L.S., F.G., and D.A. performed research; N.M., F.G., and D.A. analyzed data; and L.L.S., N.M., F.G., D.A., and M.H.B. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Did you notice anything funny about this paper?

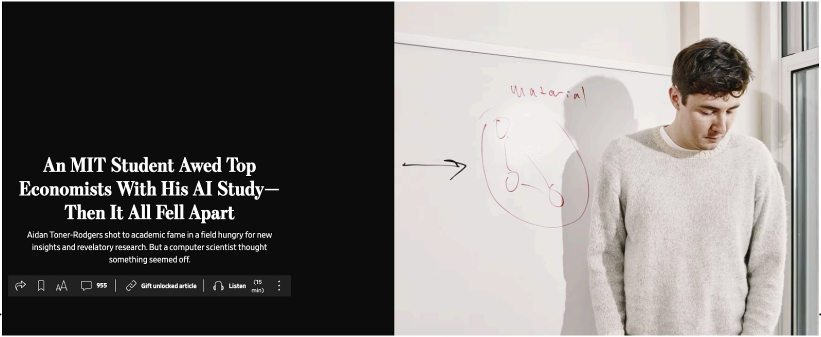
It has been retracted.

Let's see why and how the process unfolded

Yay science!

Yay science?

Another example from Economics



Aidan Toner-Rodgers and AI Productivity

Fox Souder and Johnson (working paper)

- What was this paper about?
- What were the findings?
- What was the method?
- What makes sense? What was confusing?

Penrosian capacity as a constraint on entrepreneurial growth:
An exploratory study employing the dot-com bubble

BRIAN C. FOX (corresponding author)
Management Department
Bentley University
Adamian Academic Center, 175 Forest Street
Waltham, MA 02452
Telephone: (781) 891-2656
e-mail: bfox@bentley.edu

SCOTT G. JOHNSON
Management Department
Ivy College of Business
Iowa State University
2200 Gerdin Business Building
Ames, IA 50011
Telephone: (515) 294-2460
e-mail: sgj68@iastate.edu

DAVID SOUDER
Management Department
2100 Hillside Road Unit 1041
University of Connecticut School of Business
Storrs, CT 06269-1041
Telephone: (860) 486-5747
e-mail: david.souder@uconn.edu

Note: Author list is alphabetical; all authors contributed equally to the manuscript.

Break

Replication Presentation

- Replication: Penrosian capacity as a constraint on entrepreneurial growth: An exploratory study employing the dot-com bubble (working paper)

Skills corner - Class walkthrough in R

Preparation for next class

Next class

Design II: Longitudinal Data

1. Ployhart, R.E. and R.J. Vandenberg. 2010. Longitudinal Research: The Theory, Design and Analysis of Change. Journal of Management, 36(1): 94-120.
2. Mitchell, T. R. & James, L. R. 2001. Building better theory: Time and the specification of when things happen. Academy of Management Review, 26: 530-548.

Next class

Design II: Longitudinal Data

Applications:

3. Certo, S. T., Withers, M. C., & Semadeni, M. 2017. A tale of two effects: Using longitudinal data to compare within- and between-firm effects. Strategic Management Journal, 38(7), 1536-1556.
4. Replication: Firm Repertoires and Performance: The Influence of Complementarity and Competition (working paper)

References

Angrist, J. D., and J. S Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Grant, Adam M., and Toby D. Wall. 2009. "The Neglected Science and Art of Quasi-Experimentation." *Organizational Research Methods* 12 (4): 653–86.

Shadish, WR, and TD Cook. 2009. "The Renaissance of Field Experimentation in Evaluating Interventions." *Annu Rev Psychol* 60: 607–29.

Shu, LL, N Mazar, F Gino, D Ariely, and MH Bazerman. 2012. "Signing at the Beginning Makes Ethics Salient and Decreases Dishonest Self-Reports in Comparison to Signing at the End." *Proc Natl Acad Sci U S A* 109 (38): 15197–200.