Natural Experiments in Management Research: Emerging Practices and Evaluation Guidelines

Simone Santoni[•◦] and Jost Sieweke[⋆]

[•]Bayes Business School (formerly Cass)
[◦]Soundcloud
[⋆]Vrije Universiteit Amsterdam

May 28, 2021

(Structured draft — do not circulate)

.

Keywords: natural experiments; causal inference; research design validity.

INTRODUCTION

The quest for empirical identification in strategic management research has created substantial attention around 'natural experiments,' a form of causal inquiry that has been traditionally popular in economics (Meyer 1995; Rosenzweig and Wolpin 2000) and political science (Dunning 2008). The case for natural experiments is the presence of 'naturally' occurring events—such as new regulations and laws, natural disasters, or economic and political crises—that heterogeneously influence the units of a population (Dunning 2012; Robinson, McNulty, and Krasno 2009). Insofar as these events generate random or as-if random variations in the environment, natural experiments mimic the experimental ideal in which units are split into a treatment and a control group, or, alternatively, receive different levels of the treatment. Ultimately, this opens up the possibility to infer causal effects when the substantive relationship at hand is difficult to investigate in a laboratory setting and/or would require operating costly, impractical, or unethical field experiments.

Despite naturally occurring events can turn into opportunities to conduct causal research, there are limited guidelines that help strategic management scholars to prepare and review papers that implement the natural experiment research design. In order to fill this gap, we highlight the strengths and weaknesses of natural experiments as operated in the strategic management community and propose actionable suggestions to assess and communicate the validity of natural experiments.

To do so, we critically review the population of 50 natural experiments published in the Strategic Management Journal. Particularly, our analysis is motivated by the following research questions: *R1 — How do strategy scholars claim the random or as-if random nature of environmental variation at the core of a natural experiment? R2 — How do they claim the empirical and substantive relevance of a natural experiment? R3 — How do they claim the credibility of the statistical model encapsulated in a natural experiment design?*

This work is organized as follows. The next two sections briefly introduce the key features of the 'standard natural experiment' (i.e., the specific form of natural experiment considered in this review), along with the evaluative framework we use to analyze the individual natural experiments. The following section describes the selection of the reviewed studies. Then, we present the key insights that stem from our analysis and conclude with a suggested check-list that facilitates strategy scholars to exploit the opportunities of causal inference offered by naturally occurring events. The online appendix contains companion software code to operate natural experiments in Python, R, and Stata.

NATURAL EXPERIMENTS AND CAUSAL EMPIRICAL RESEARCH

Foundations and Examples from Strategic Management

The 'standard' natural experiment[1] resembles the design of a ran-

[1] In his comprehensive, cross-disciplinary analysis of the literature, Dunning 2012 identifies three forms of natural experiments: Standard natural experiments; instrumental variables (Angrist, 1990); regression discontinuity designs (Thistlethwaite & Campbell, 1960). In the interest of clarity and integrity, our review concentrates on standard natural experiments, whose

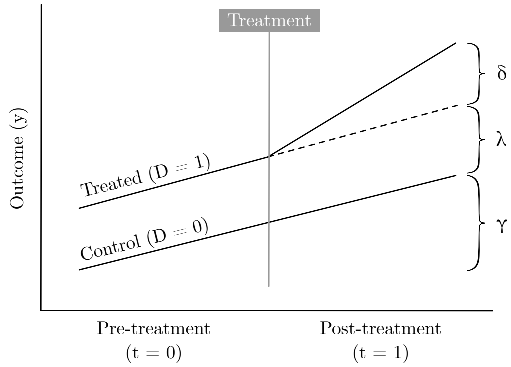

Figure 1: Visual representation of the standard natural experiment. Notes.— The underlying population regression function is $y = \gamma t + \lambda D + \delta t D$, where γ, λ, and δ represent the systematic difference in the outcome across the treated and control cases, the trend effect and the difference in the outcome that is due to the treatment. For sake of clarity, we represent the case in which $\delta > 0$.

domized experiment. In fact, the naturally occurring event (such as an earthquake, see for example Belloc, Drago, and Galbiati 2016) is supposed to determine the treatment status of the statistical units (treated Vs control), each of which has both a pre- and a post-treatment observation. As shown in Figure 1, it is possible to estimate the causal effect of the treatment by contrasting the pre-post change in the outcome variable y across the control group (γ) and the treatment group ($\gamma + \delta$).
!! Comment: One or two examples of natural experiments published in SMJ to be succinctly summarized here!!

Assessing the Validity of Natural Experiment Designs

How do we evaluate a standard natural experiment research design? According to Dunning (2012, page 27), the validity of a natural experiment should be assessed against three criteria (see Figure 2). First, the authors should prove the random nature of the treatment, or, at least, defend the plausibility of as-if random. In the case of the *randomized standard natural experiment*, it is important that the assignment process is truly random. Although this may seem obvious, this condition is sometimes violated, even in the context of lotteries (e.g. Starr 1997). In the case of an *as-if randomization*, it is vital the assignment process, although not truly random, is independent of factors that are related to the outcome and it not affected by unit's self-selection into treatment or control conditions. As Dunning points out, the researcher has to make a very compelling case for this assertion (or to drop the claim to a natural experiment). In depth knowledge of the context (e.g., industry regulatory frameworks), qualitative evidence about the naturally occurring event (e.g., a new law), and quantitative evidence at the event- and unit-level are oftentimes essential ingredients to defend the plausibility of as-if random assignment, and, ultimately, to sustain the natural experiment.

Second, the naturally occurring event should reveal the wider *"theoretical, substantive, and/or policy issues"* (Dunning 2012, page 29) that motivate the study. For example, the sudden, premature death of a star scientist (Azoulay, Zivin, and Wang 2010) create the premises for a natural experiment that quantifies the spill-over effect of collaborating with academics who are very prominent in the field.

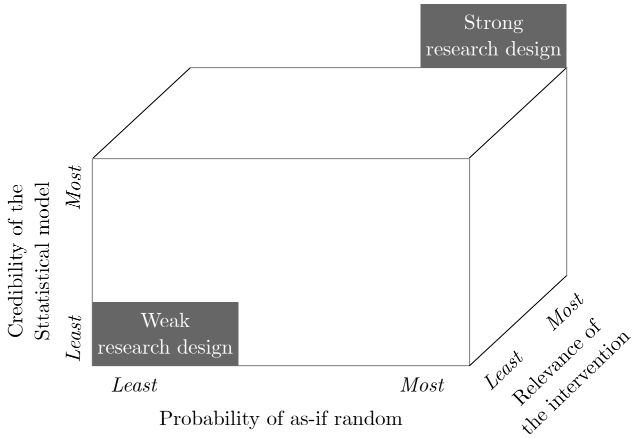

Figure 2: Visual representation of Dunning's validity framework (Source: Dunning 2012, page 31)

Finally, the statistical model should fit with the characteristics of the naturally occurring event. In the case of a randomized standard natural experiment, simplicity and transparency should take precedence in the data analysis stage. Particularly, the Neyman's potential outcomes framework (Splawa-Neyman, Dabrowska, and Speed 1990), namely, a treated Vs. control mean comparison test, should be used *prima facie*. At the same time, some statistical adjustments may be required even in presence of a random (or as-if random) treatment. For example, the Stable-Unit-Treatment-Value-Assumption (SUTVA) may be violated insofar as the treatment status of a unit 'i' interferes with the potential outcome of unit 'j'. This is a concern Belloc and colleagues (2016) seem to have in mind when they estimate the impact of earthquakes on the probability of institutional change at the city-level in the Middle-Ages northern and central Italy. In fact, both the distribution and timing of earthquakes are random. However, the probability a control city will move from autocratic regimes to self-government is also a function of the information key actors are exposed to, such as the transition choices made by a treated, neighbor city. In this case, statistical artefacts would be needed to take into account the correlation of residuals induced by the geographical proximity of any pair of units.

ARTICLES SELECTION

Consistently with previous review of the literature in the field of strategic management (e.g. Haans, Pieters, and He 2016), we decided to survey only studies published in the Strategic Management Journal. Using the search engine available in the journal's web-page, we searched for any article reporting the quote "natural experiment*" in the title, abstract, keywords, *or* full text of any document published since the very first volume of the journal or released as pre-print as of December 31, 2019.

We retrieved 73 studies; 50 were eventually included in the review. We excluded one methodological article (Semadeni, Withers, and Trevis Certo 2014), two articles that recall the empirical evidence

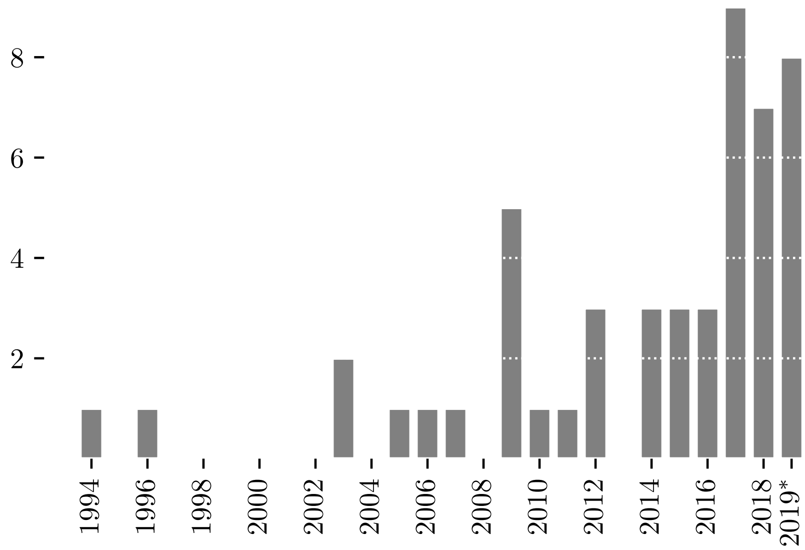

Figure 3: Natural experiment studies published in the Strategic Management Journal over time. The 2019 bucket, marked with an asterisk, contains both published and accepted, online articles.

produced in previous natural experiments (Gallus and Frey 2016; Chakrabarti 2015), seven articles that indicate natural experiments as a possible way to overcome the limitations or expand on the study at hand (e.g. Karim and Williams 2012), and 13 observational studies [2] that use the natural experiment label in a figurative sense to emphasize some appealing properties of the empirical setting (Grigoriou and Rothaermel 2017). Figure 3 illustrates the distribution of the retrieved studies over time.

[2] Consistently with prior works (Dunning 2012; Sieweke and Santoni 2019), we use the expression 'observation study' to denote a study oeprating a correlational research design.

ASSESSING THE AS-IF RANDOM PLAUSIBILITY OF THE TREATMENT

This section focuses on the diagnostics that can be used to assess and argument the (as-if) random nature of the environmental variation at the center of the natural experiment. Specifically, this section surveys and articulates the following:

- diffusion/role of qualitative diagnostics to appreciate:
  - units' information about the treatment
  - units' incentives to self-select into the treatment (control) group
  - unit's capacity to self-select into the treatment (control) group
- diffusion/role of quantitative diagnostics (e.g., balance test) to compare and contrast treated and control units along relevant dimensions

ASSESSING THE RELEVANCE OF THE TREATMENT

This section focuses on the empirical and substantive relevance of the environmental variation at the center of a natural experiment.

Particularly, this section reviews and discusses the following:

- diffusion/role of qualitative diagnostics to show:
  - the empirical, substantive, and policy relevance of the natural experiment
  - the external validity (non idiosyncrasy) of the natural experiment
  - exclusion of 'bundling of treatments' (i.e., environmental variations affecting the outcome through multiple causal pathways)
- diffusion/role of placebo tests supporting the magnitude of the average treatment estimation on the treated (ATT)
- diffusion/role of local average treatment estimation (LATE) considerations

ASSESSING THE CREDIBILITY OF THE STATISTICAL MODEL

This section focuses on the credibility of the statistical model encapsulated in the natural experiment design. Specifically, this section surveys and articulates the following:

- diffusion/rationale of model based adjustments (instead of simple mean-comparison tests):
  - adjustment via control covariates
  - adjustment via matching
- diffusion of SUTVA considerations and associated model adjustments (see derivation of standard errors)

SUMMARY

This section wraps-up around the results of the literature review and provides actionable guidelines in order to better leverage the natural experiment design.
!! Comment: We have already coded the 50 studies. Although further analyses are needed to reveal clear patterns, some interesting elements seem to emerge. Specifically, future studies using a natural experiment design could:

- better integrate qualitative evidence and institutional knowledge in order to establish the as-if random nature of the treatment
- provide a more systematic discussion of the conditions under which a treatment can plausibly be considered as-if random (see the point on units' information, incentives, and capacity to self-select into the treatment group)
- pay equal attention to the empirical and substantive relevance of the treatment (that is, the possibility to reveal and/or detail important theoretical mechanisms by exploiting naturally-occurring events)

- provide a thorough assessment of the strengths and weaknesses of relying on a certain naturally-occurring event — i.e., explaining what the pros and cons are in terms of empirical identification (see LATE aspects) and theorizing opportunities
- use model-based adjustments (such as matching and control covariates) when there is no ground to establish the (as-if) random nature of the treatment. Indeed, the comparative advantage of natural experiments over alternative designs (e.g., quasi-experiments) also comes from the possibility to conduct causal inference by means of simple, transparent statical models. In other words, there should be good reasons to move from a design-based causal inference strategy to a model-based one (e.g., piggybacking on models that jointly use matching, DiD, and a long list of control covariates)
- consider the interactions among as-if random, relevance, and credibility elements. For example, the credibility of a model should be assessed against the nature of the treatement (random, as-if random, not random) and the process through which it is adminstered (see SUTVA)

REFERENCES

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. "Superstar Extinction". In: *Quarterly Journal of Economics* May (2010), pp. 549–589. URL: http://www.nber.org/papers/w14577.

Belloc, Marianna, Francesco Drago, and Roberto Galbiati. "Earthquakes, religion, and transition to selfgovernment in italian cities". In: *Quarterly Journal of Economics* 131.4 (2016), pp. 1875–1926. ISSN: 15314650. DOI: 10.1093/qje/qjw020.

Chakrabarti, Abhirup. "Organizational adaptation in an economic shock: The role of growth reconfiguration". In: *Strategic Management Journal* 36.11 (2015), pp. 1717–1738. ISSN: 10970266. DOI: 10.1002/smj.2309.

Dunning, Thad. "Improving causal inference: Strengths and limitations of natural experiments". In: *Political Research Quarterly* 61.2 (2008), pp. 282–293. ISSN: 10659129. DOI: 10.1177/1065912907306470.

– *Natural Experiments in the Social Sciences*. Cambridge, UK: Cambridge University Press, 2012. ISBN: 9781139084444. DOI: 10.1017/CBO9781139084444. arXiv: arXiv:1011.1669v3. URL: http://ebooks.cambridge.org/ref/id/CBO9781139084444.

Gallus, Jana and Bruno S. Frey. "Awards: A strategic management perspective". In: *Strategic Management Journal* 37.8 (2016), pp. 1699–1714. ISSN: 10970266. DOI: 10.1002/smj.2415.

Grigoriou, Konstantinos and Frank T. Rothaermel. "Organizing for knowledge generation: internal knowledge networks and the contingent effect of external knowledge sourcing". In: *Strategic Management Journal* 38.2 (2017), pp. 395–414. ISSN: 10970266. DOI: 10.1002/smj.2489.

Haans, Richard F.J., Constant Pieters, and Zi-Lin He. "Thinking about U: Theorizing and Testing U- and inverted U-shaped Relationships in Strategy Research". In: *Strategic Management Journal* 37 (2016), pp. 1177–1195. ISSN: 01432095. DOI: 10.1002/smj.2399. URL: http://doi.wiley.com/10.1002/smj.2399.

Karim, Samina and Charles Williams. "Structural knowledge: How executive experience with structural composition affects intrafirm mobility and unit reconfiguration". In: *Strategic Management Journal* 33.6 (2012), pp. 681–709. ISSN: 01432095. DOI: 10.1002/smj.1967.

Meyer, Bruce D. "Natural and quasi-experiments in economics". In: *Journal of Business and Economic Statistics* (1995). ISSN: 15372707. DOI: 10.1080/07350015.1995.10524589.

Robinson, Gregory, John McNulty, and Jonathan S. Krasno. "Observing the Counterfactual? The Search for Political Experiments in Nature". In: *Political Analysis* 17.4 (2009), pp. 341–357. DOI: 10.1093/pan/mpp011.

Rosenzweig, Mark R and Kenneth I Wolpin. "Natural "Natural Experiments" in Economics". In: *Journal of Economic Literature* (2000). ISSN: 0022-0515. DOI: 10.1257/jel.38.4.827.

Semadeni, Matthew, Michael C. Withers, and S. Trevis Certo. "The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations". In: *Strategic Management Journal* 35.7 (2014), pp. 1070–1079. ISSN: 10970266. DOI: 10.1002/smj.2136.

Sieweke, Jost and Simone Santoni. “Natural experiments in leadership research: An introduction, review, and guidelines”. In: *The Leadership Quarterly* (2019), p. 101338.
Splawa-Neyman, Jerzy, Dorota M Dabrowska, and TP Speed. “On the application of probability theory to agricultural experiments. Essay on principles. Section 9.” In: *Statistical Science* (1990), pp. 465–472.
Starr, Norton. “Nonrandom Risk: The 1970 Draft Lottery”. In: *Journal of Statistics Education* 5.2 (1997), p. 1997. ISSN: 1069-1898. DOI: 10.1080/10691898.1997.11910534.

APPENDIX A — SAMPLING

REFERENCES CITED IN THE APPENDIX