Ownership Diversity and Innovation:
Using Split Sample Design to Generate More Credible Findings

Timothy R. Wojan

October 25, 2022 | Federal Committee on Statistical Methodology Conference

National Center for Science and Engineering Statistics

Social, Behavioral and Economic Sciences

National Science Foundation

Disclaimer: This presentation provides results of exploratory research sponsored in part by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF). This information is being shared to inform interested parties of ongoing activities and to encourage further discussion. Any views expressed are those of the authors and not necessarily those of NCSES or NSF.

Essentially, all models are wrong, but some are useful.

--George Box

Outline

- *Pr*(false discovery) increases considerably if the same data are used for specification and hypothesis testing
- Study of diversity particularly vulnerable to false discovery
- Splitting data into exploratory and confirmatory samples can restore validity of hypothesis tests
- Protocol for testing how ownership diversity is associated with innovation in the Annual Business Survey (ABS)
  - Diversity measure selection: axiomatic or inductive?
  - Sample split, sample isolation, pass through, Registered Report
  - Hypothesis testing and submitting final manuscript

Replicability Crisis: Publication Bias, Rush to Regression, and Canned Test Statistics

- Null hypothesis statistical testing makes statistical inference dependent on positive results
- Searches for significance facilitated by low cost of estimation
- *p-value* of single coefficient estimate invalid if preceded by specification tests
- No transparency in specification, and hypothesis testing has resulted in the proliferation of robustness checks
- Faith in robustness checks: what you see is all there is

How specification testing degrades hypothesis test statistics

A significance level of $\alpha = 0.05$ from statistical software assumes an unselected model.

If you have a set of *k* null hypotheses and all are true, then the probability of one erroneous rejection is

$$Pr(\text{at least 1 rejection}) = 1 - Pr(\text{all accepted}) = 1 - (1 - \alpha)^k$$

Number of Hypotheses	Nominal p-value	True p-value
1	0.05	0.05
3	0.05	0.14
6	0.05	0.26
12	0.05	0.46

Statistical analysis of diversity particularly vulnerable to false discovery

- “Essentially, all models are wrong, some models may be useful for reinforcing your priors”
- There are numerous ways to measure diversity as something other than homophily
- In ABS, at least 7 principal owner attributes can capture ownership diversity
- Combinatorial measures using multiple attributes further increase hypotheses that could be tested

Split sample design: Restoring transparency to specification and hypothesis testing

1. 35% of dataset to discover potentially useful models
2. Document the potentially useful models in a public Registered Report
3. 65% of dataset used for hypothesis testing, generating valid test statistics
4. Apply false discovery (FDR) and family-wise error rate (FWER) correction for assessing significance across multiple comparisons
5. Publish full set of hypothesis tests

Selecting a diversity index axiomatically

HOMOPHILY AXIOM: Given other things, all owners belonging to the same group must result in the lowest diversity measure value.

FRACTIONALIZATION AXIOM: Given other things, an increase in the number of groups must increase the diversity measure value.

TEAM SIZE AXIOM: Given other things, larger ownership teams not demonstrating homophily must increase the diversity measure value relative to smaller ownership teams.

CONCENTRATION OF OWNERSHIP AXIOM: Given other things, ownership concentrated in one member of the team must reduce the diversity measure value relative to ownership that is more equally distributed among team members.

Ownership fractionalization index

Derived from the ethno-linguistic fractionalization index (ELF):

$$ELF = 1 - \sum_{i=1}^{n} p_i^2$$

where *p* is the population share of *n* groups

Invariant to population size so violates TEAM SIZE AXIOM

A minor modification of the ELF satisfies all four axioms

$$Ownership\ Fractionalization\ Index = 1 - \sum_{i=1}^{o} p_i^n$$

where *p* represents the ownership share of the i^{th} owner, and *o* is the number of owners

Split sample design protocol

- Anderson and Magruder (2017) NBER Working Paper
- Monte Carlo simulations suggest exploratory share of 0.35
- FSRDC research makes isolation of confirmatory share credible
- Deciding on threshold for passing estimates to confirmation
  - t-statistic of 2 given large size of exploratory sample, and
  - minimum effect size (Cohen's d ≥ 0.2 or odds ratio ≥ 1.44)
- Estimate *New-to-market Innov = f(OF$_i$, Firm Size, NAICS)*
- Publish results in Center for Open Science Registered Report

Split sample design protocol (continued)

- Access to confirmatory sample granted only after Registered Report published
- Hypotheses from Registered Report tested, adjusting test statistics with FDR and FWER correction
  - FDR: $p_i < 0.05*(i/m)$ where *p-values* ranked from smallest to largest (p_m)
  - FWER: $p_i < 0.05/m$ ➢ w/12 tests only $p < 0.00416$ significant
- Publish all statistics from confirmatory tests and include link to Registered Report

Limitations of split sample design

- FSRDC proposals for single use of confidential data ideally suited to split sample design
- Multiple analyses using the same data vulnerable to overfitting and false discovery:
  - Dwork, et al. 2015. The reusable holdout: Preserving validity in adaptive data analysis, *Science.*
  - Access the holdout set only via a differentially private mechanism
  - Ideas are tested against aggregate information, whereas individual dataset components remain confidential

Contentious questions raise cost of false discovery

- What kinds of ownership diversity enable or hinder innovation?
- The focus on Type I errors may exacerbate Type II errors:
  - Large ABS sample size lessens concerns wrt statistical power
  - No inferences wrt negative results in confirmatory analysis
- Annual collection of ABS refreshes data required for sequential investigation of phenomena
- FSRDC can provide strong evidence that claims of restricted access to confirmatory data are valid

Thank you!

Contact: twojan@nsf.gov