Math Doesn't Care About Your Theory: Defining, Describing, and Specifying Instrumental Variables for Content Analysis

John R. Busenbark
Mendoza College of Business
University of Notre Dame
jbusenba@nd.edu

Academy of Management – Digital Edition

Content Analysis Overview

- **Language = Attitudinal, behavioral, and ideological constructs**
  - CEO temporal focus and new product introductions (e.g., Nadkarni & Chen, 2014)
  - CEO regulatory focus and acquisitions (e.g., Gamache et al., 2015)
  - Media coverage severity and executive dismissal (e.g., Busenbark et al., 2019)
  - Board chair orientations and firm performance (e.g., Krause, 2017)
  - Managerial language and capital market reactions (e.g., Konig et al., 2017)

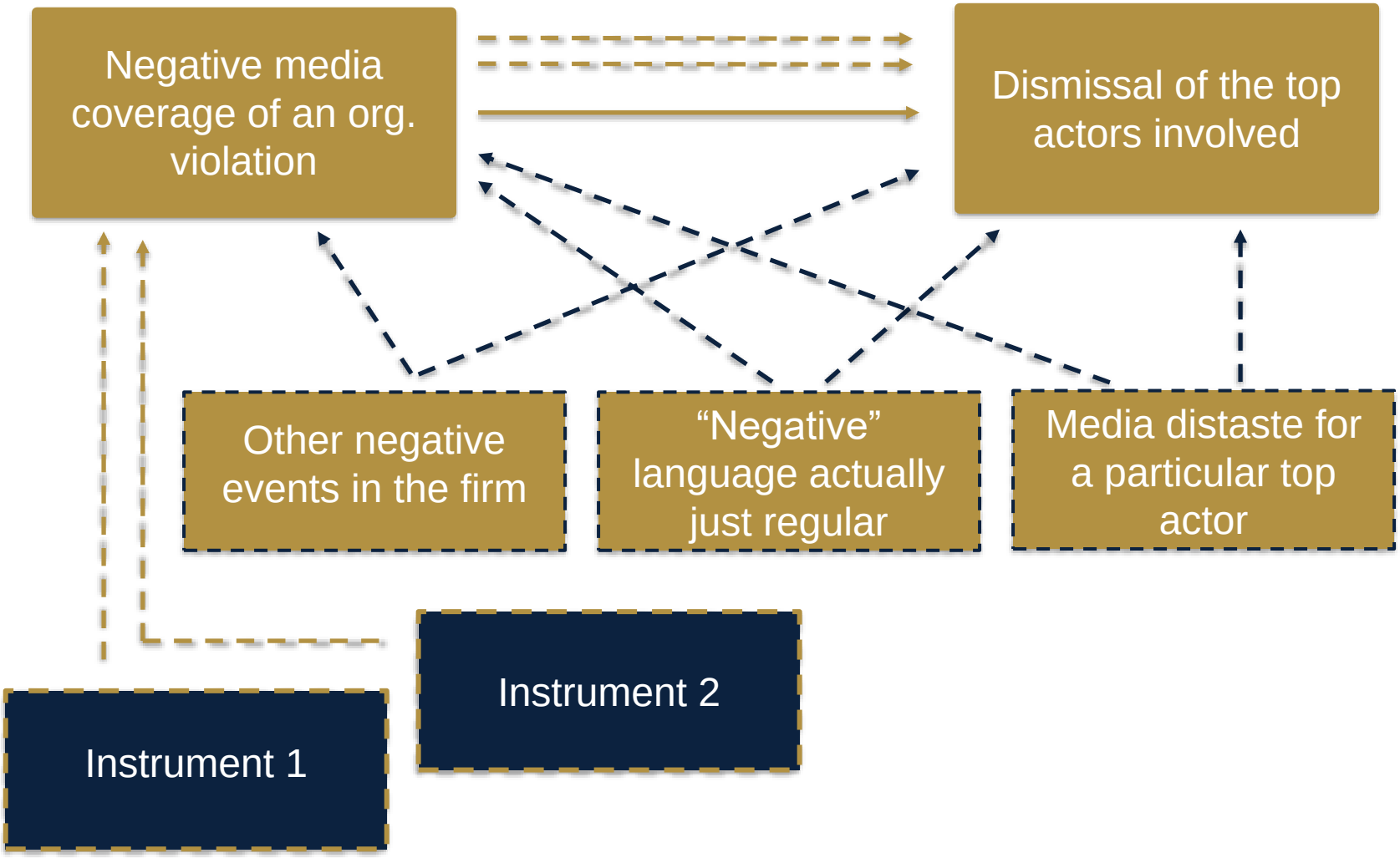
- **Litany of resources and techniques**
  - Computer automated text analysis (CATA)
    - LIWC, Diction, R code, etc.
  - Manual text analysis
    - Dictionaries, procedures, descriptives
  - Resources to help
    - https://www.terry.uga.edu/_contentanalysis/

Rife with Empirical Issues

- **What doesn't get measured does exist** (McKenny, et al., 2016)
  - Measurement error
  - Omitted variables
  - Reverse causality
  - Autocorrelation

- **Sounds a lot like "endogeneity/unexplained heterogeneity"**
  - Strong validation procedures (e.g., McKenny et al., 2016; Konig et al., 2017; Pan et al., 2018)
    - Using expert raters
    - Crafting unique dictionaries
    - Convergent and discriminant validity
    - Basically everything OB scholars do for scales
  - Empirical estimator specification (e.g., Kennedy et al., 2008; Semadeni et al., 2014)
    - Two-stage instrumental variable techniques
    - GMM-related techniques

Example of Endogeneity in CATA



Instruments are a Headache

- **Can do more harm than good**
  - Irrelevance creates measurement error (Stock et al., 2002; Wooldridge, 2010)
  - Endogenous creates more bias (Kennedy, 2008; Semadeni et al., 2014)

- **They are remarkably elusive, but solutions exist!**
  - Natural instruments (Kenney, 2008: 142)
    - Unexpected instruments that happen to exhibit desirable properties
  - Logical connections
    - Instruments informed by general logic, but no theory
  - Data transformations (Kennedy, 2008: 159-160)
    - Creating a dummy variable the size of the observation in two or three groups
    - The Durbin method of rank-ordering variables as instruments

Instruments from CATA

- **TONS of good natural/logical instruments**
  - Total number of words in a document or corpus
  - Total number of documents/articles in a corpus
  - Readability indexes (e.g., Lougran & McDonald, 2014)
  - Organizational activities related to the content of articles
  - Manual content analyses as procedural checks
  - Characteristics of the subjects or authors of text
  - Number of specific characteristics
  - Average length of words in a document
  - Count of long words (words over a certain character threshold)
  - Coverage of related organizations (e.g., industry, social grouping)

The Problem

- **Reviewers, editors, and scholars at large fetishize theory**
  - Ameliorate problems with bad instruments (Kennedy, 2008)
  - Prevent authors from p-hacking instruments (Bettis et al, 2014)

"I understand that your instruments are strongly related to the independent variable, but I do not believe there is a clear theoretical logic for it."

"What is the theoretical reason for the connection between your instruments and [independent variable]?"

"I would like you to consider different instruments with better theoretical rationale."

"Could you please replace [instrument 1] with another that has compelling theory?"

"I am not convinced by the theory for [instrument 2]."

Math Doesn't Care About Theory!

- **Two properties determine strong instruments**
  - Relevance – is the instrument strongly related to the IV?
    - Partial f-stat; the f-stat associated with only the instruments
    - The threshold depends on number of instruments, but anything over 11(ish) is good
  - Exogeneity – is the instrument NOT related to the structural error term?
    - Sargan, Bassman, Hansen tests for exogeneity (overidentification)
    - Involve determining whether the instruments are correlated with the structural error term
    - Higher p-values are better since that implies no relationship

- **Instrument "theory" merely helps support for these two**
  - The first two are requisite features
  - This is a supplement: simply another (non-mathematical) way to gauge strength

Evidence for the Claim

- **Synthetic instruments** (Le Gallo & Paez, 2013)
  - Creating variables that demonstrate desirable properties (Doran & Fingleton, 2018)
  - Uses eigenvector weights that fluctuate as latent constructs (Griffith & Chun, 2016)
  - Shown to outperform weak instruments (Le Gallo & Paez, 2013)

- **Simulations with instruments** (Certo et al., 2016; Semadeni et al., 2014)
  - Instruments generated from randomly drawn data
  - Strong/weak instruments exhibit very different outcomes (Semadeni et al., 2014)
  - Inclusion or absence really matters (Certo et al., 2016)

- **Programming code** (StataCorp, 2017)
  - Stata 2sls: ivregress 2sls y c (x=iv1 iv2)
    - Conspicuously, Stata does not request any “theory” specification
  - Stata extended: eregress y c, endog(x= iv1 iv2)
    - Again, nowhere to specify “theory”

Content Analysis Leads the Way

- **Opportunity for "natural" or "logical" instruments**
  - Perhaps more than with archival data
  - Can use general logic rather than theoretical
  - Often times axiomatic

- **Normalize lack of "theory"**
  - Explicitly highlight natural instruments
  - Cite relevant work from this presentation
  - Help encourage reviewers to think more exhaustively

ANY QUESTIONS?

References

Bettis, R., Gambardella, A., Helfat, C., & Mitchell, W. 2014. Quantitative empirical analysis in strategic management. ***Strategic Management Journal***, 35(7): 949-953.

Busenbark, J. R., Marshall, N., Miller, B. P., & Pfarrer, M. D. 2019. How the severity gap influences the effect of top actor performance on outcomes following a violation. ***Strategic Management Journal***, 40(12): 2078-2104.

Certo, S. T., Busenbark, J. R., Woo, H.-S., & Semadeni, M. 2016. Sample selection bias and Heckman models in strategic management research. ***Strategic Management Journal***, 37(13): 2639-2657.

Doran, J. & Fingleton, B. 2018. US metropolitan area resilience: insights from dynamic spatial panel estimation. ***Environment and Planning A: Economy and Space***, 50(1): 111-132.

Gamache, D., McNamara, G., Mannor, M., & Johnson, R. 2015. Motivated to acquire? The impact of CEO regulatory focus on firm acquisitions. ***Academy of Management Journal***, 58(4): 1261-1282.

Griffith, D. A. & Chun, Y. 2016. Evaluating eigenvector spatial filter corrections for omitted georeferenced variables. ***Econometrics***, 4(2): 29.

Kennedy, P. 2008. ***A Guide to Econometrics*** (2nd Edition ed.). Oxford, UK: Blackwell.

König, A., Mammen, J., Luger, J., Fehn, A., & Enders, A. 2017. Silver bullet or ricochet? CEO's use of metaphorical communication and infomediaries' evaluations. ***Academy of Management Journal***(doi: amj.2016.0626).

Krause, R. 2017. Being the CEO's boss: An examination of board chair orientations. ***Strategic Management Journal***, 38(3): 697-713.

Le Gallo, J. & Páez, A. 2013. Using synthetic variables in instrumental variable estimation of spatial series models. ***Environment and Planning A***, 45(9): 2227-2242.

Loughran, T. & McDonald, B. 2014. Measuring readability in financial disclosures. ***The Journal of Finance***, 69(4): 1643-1671.

McKenny, A. F., Aguinis, H., Short, J. C., & Anglin, A. H. 2016. What doesn't get measured does exist: Improving the accuracy of computer-aided text analysis. ***Journal of Management***: dor: 0149206316657594.

Nadkarni, S. & Chen, J. 2014. Bridging yesterday, today, and tomorrow: CEO temporal focus, environmental dynamism, and rate of new product introduction. ***Academy of Management Journal***, 57(6): 1810-1833.

Pan, L., McNamara, G., Lee, J. J., Haleblian, J., & Devers, C. E. 2018. Give it to us straight (most of the time): Top managers' use of concrete language and its effect on investor reactions. ***Strategic Management Journal***, 39(8): 2204-2225.

Semadeni, M., Withers, M. C., & Certo, S. T. 2014. The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. ***Strategic Management Journal***, 35(7): 1070-1079.

StataCorp. 2017. ***Stata 15 Base Reference Manual***. College Station, TX: Stata Press.

Stock, J. H., Wright, J. H., & Yogo, M. 2002. A survey of weak instruments and weak identification in generalized method of moments. ***Journal of Business & Economic Statistics***, 20(4): 518-529.

Wooldridge, J. M. 2010. ***Econometric analysis of cross section and panel data***. Cambridge, MA: MIT press.