

From Association to Explanation: Strengthening Research Design and Causal Reasoning in Accounting Scholarship

Fabio Yoshio Suguri Motoki¹  Valdemar Pinho Neto²  Victor Rangel³ 

¹ University of Texas Rio Grande Valley, Edinburg, Texas, United States of America

² Escola Brasileira de Economia e Finanças - FGV EPGE, Rio de Janeiro, Rio de Janeiro, Brazil

³ Insper, São Paulo, São Paulo, Brazil



¹fabio.motoki@utrgv.edu

²valdemar.pinho@fgv.br

³victorrs@al.insper.edu.br

Abstract

A recent editorial in this journal (Trevisan, 2024) highlighted the growing importance of credible causal inference in accounting research and the pitfalls of interpreting correlational patterns as causal relationships. Building on that contribution, this editorial advances the discussion by focusing upstream: on how researchers formulate questions, define estimands, articulate mechanisms, justify measurement choices, and construct coherent identification strategies. Rather than revisiting quasi-experimental estimators as a toolbox, we organize the research process as a design workflow that can support descriptive, predictive, associative, and causal inference. We propose a practical checklist for authors and reviewers, discuss common failure modes in empirical accounting designs (controls, influential observations, and misaligned claims), and connect methodological choices to transparency and reproducibility norms in the field.

Keywords: causal inference; research design; econometrics; accounting; mechanisms; measurement; validity.

How to cite:

Motoki, F. Y. S., Pinho Neto, V., & Rangel, V. (2026). From Association to Explanation: Strengthening Research Design and Causal Reasoning in Accounting Scholarship. *Advances in Scientific and Applied Accounting*, 19(1), 001–009/010. <https://doi.org/10.14392/asaa.2026190101>

1 Motivation and Diagnosis: From Association to Explanation

Recent years have seen a notable shift in empirical accounting research toward a more explicit concern with causal inference. The editorial by Trevisan (2024) represented an important milestone in this transition, offering a concise overview of quasi-experimental approaches, such as Difference-in-Differences, Instrumental Variables, and Regression Discontinuity, and highlighting common pitfalls in interpreting regression-based associations as causal claims. This contribution is timely, as Brazilian accounting research has reached a level of technical maturity in which many scholars are ready to engage with identification strategies that go beyond standard panel regressions.

If the previous editorial focused on the methods that support causal identification, this editorial focuses on the reasoning that precedes and guides methodological choices. Before selecting an estimator, researchers should articulate what kind of inference they aim to produce, what theoretical mechanism links their variables, whether their measures faithfully represent the constructs of interest, and what institutional context governs the phenomenon under study (Hernán & Robins, 2020; Imbens & Rubin, 2015; Shmueli, 2010). These foundations matter not only for causal studies but also for descriptive and associational work, since they reduce interpretive ambiguity and make empirical findings more cumulative.

Empirical accounting research in Brazil has benefited from increased data availability, expanded graduate programs, and growing integration with international scholarship. Studies regularly employ panel data methods, sophisticated estimators, and large datasets drawn from regulatory filings, corporate disclosures, and market data. These are meaningful developments. At the same time, persistent tensions shape the methodological landscape. Many empirical studies still rely heavily on conventional regression templates, often treating methodological decisions as steps in a standardized workflow rather than as expressions of an underlying research design. Variables are winsorized by default, proxies for complex constructs are selected based on precedent rather than validity, and control variables accumulate in regressions without a clear theoretical rationale (Kennedy, 2002; Kondylis & McKenzie, 2024; Leamer, 1983).

A key challenge is the gap between the questions scholars claim to answer and the designs they employ. It remains common to find research questions formulated in causal language (“Does X affect Y?”) while the analysis provides only partial correlations, even when endogeneity problems such as omitted variables, selection, and reverse

causality are acknowledged (Larcker & Rusticus, 2010). Consider, for example, the frequent pattern in Brazilian accounting research of regressing earnings management proxies on governance indicators using panel data from listed firms on B3, without articulating why the observed variation in governance identifies a causal effect rather than reflecting simultaneous firm-level choices. Trevisan (2024) highlighted this concern within the specific context of governance research. Our point is broader, that the misalignment between research questions, design, and interpretation is systemic. Noteworthy exceptions exist: studies exploiting the Sarbanes-Oxley Act as a natural experiment for Brazilian cross-listed firms (Funchal & Monte-Mor, 2016) and mandatory IFRS adoption as a regulatory shock using difference-in-differences designs (da Silva & Nardi, 2017; Sampaio et al., 2020) illustrate what design-based accounting research with Brazilian data can look like. The challenge is to make such practices the norm rather than the exception.

This gap is not primarily technical. Rather, it often reflects incentive structures that reward publication volume, short timelines, and adherence to familiar empirical patterns, as well as the constraints imposed by the data that are readily available. A methodological culture that equates sophistication with complexity (the number of controls, the specific estimator used, or the length of robustness sections) can crowd out the deeper requirements of credible scientific inference. For this reason, we propose a design-based framing of empirical work. The core task is not to “choose a method,” but to specify what variation identifies a claim and what assumptions connect the design to an interpretable counterfactual (Angrist & Pischke, 2009; Hernán & Robins, 2020; Imbens & Rubin, 2015).

2 From Question to Design: Conceptual Foundations

This section presents the conceptual foundations of research design. Its purpose is to motivate why each element matters for credible inference, leaving operational details to the workflow in Section 3.

Inference goals. A central step in strengthening empirical design is distinguishing among descriptive, predictive, associative, and causal research (Hernán & Robins, 2020; Shmueli, 2010). Each type of inference relies on different assumptions and serves different scientific goals. Descriptive work may reveal patterns worthy of follow-up; predictive work may support forecasting or risk assessment; associative work may uncover correlations that help

construct indices or map empirical regularities. Causal work seeks to answer “what if” questions by constructing credible counterfactuals—a goal we formalize in Section 3. Many influential contributions in accounting are not causal. The key requirement here is coherence between the claim and the design used to support it.

Mechanisms. Causal claims are meaningful only when anchored in plausible mechanisms. Mechanisms explain why an effect should occur, under what circumstances, and through which channels (Hedström & Ylikoski, 2010). In accounting research, mechanisms frequently involve psychological, informational, governance-related, and institutional channels. For example, the effect of disclosure regulation on firm value is often hypothesized to operate through reductions in information asymmetry, improvements in monitoring, or changes in contracting efficiency. Without articulating these mechanisms, causal claims become difficult to interpret and associational results become ambiguous. Mechanistic reasoning is also important for discriminating between competing explanations. Suppose we observe a negative association between conservative accounting and cost of equity. Is this because conservatism improves monitoring? Because conservative firms differ systematically in risk? Because of omitted variables? Mechanisms clarify these alternatives and guide design choices that are capable of separating them empirically. Graphical tools such as Libby boxes and directed acyclic graphs (DAGs) can be useful for making assumptions explicit.¹

Measurement validity. Accounting research frequently relies on proxies for latent constructs such as audit quality, reporting incentives, earnings management, informational opacity, or corporate culture. The use of proxies is often unavoidable, but their validity must be evaluated rather than assumed (Gow et al., 2016). Measurement error can attenuate coefficients, induce bias, or create misleading patterns that replicate across studies simply because the same flawed proxy is widely adopted. Measurement validity is not ancillary to inference. It is a necessary condition for meaningful interpretation, whether the goal is descriptive, associative, predictive, or causal.

Institutional context and the limits of methodological escalation. Credible empirical claims depend on understanding the institutional context in which accounting outcomes are embedded, including regulation, tax rules, governance structures, and market microstructures (Leuz,

2022). Exploiting institutional specifics is one of the most powerful ways to construct credible identification strategies; contextual understanding is a design input that cannot be replaced by econometric technique. A concerning pattern in empirical work is methodological escalation. When confronted with identification challenges, authors sometimes respond by layering more controls (Cinelli et al., 2024; Whited et al., 2022), instruments (Larcker & Rusticus, 2010), or generic robustness rituals rather than reexamining the core design. Robustness tests are informative only when the identifying variation is credible; otherwise they create a false sense of security (Cinelli & Hazlett, 2020; Roth, 2022).

3 Operational Standards: A Design Workflow for Credible Empirical Claims

This editorial proposes a practical standard for credible empirical claims in observational accounting research. Identification should be treated as a documented workflow that links a research question to an interpretable claim. A credible identification strategy should state (i) the estimand and the counterfactual question; (ii) the assignment mechanism that generates identifying variation; (iii) the measurement strategy that links constructs to proxies; (iv) diagnostics and sensitivity analyses targeted to the design’s main failure modes; and (v) transparency and reproducibility of data construction and code (Gow et al., 2016; Hail et al., 2020; Leuz, 2022).

Step 1: Define the estimand and the counterfactual.

A design begins by stating the causal quantity of interest—the estimand—and the target population. The potential outcomes framework provides a minimal but precise language for doing so (Hernán & Robins, 2020; Imbens & Rubin, 2015; Rubin, 1974).

Let $D_i \in \{0, 1\}$ denote a binary treatment indicator for unit i , and let $Y_i(1)$ and $Y_i(0)$ denote the potential outcomes under treatment and control, respectively. The individual causal effect is $Y_i(1) - Y_i(0)$, which is never jointly observed—the fundamental problem of causal inference (Holland, 1986). Empirical designs therefore target population-level summands of this quantity under stated assumptions. The most common estimands are:

¹Libby boxes (Bloomfield, 2015; Libby, 1981) map theoretical constructs to their empirical proxies and are particularly useful for evaluating measurement validity: whether the proxy faithfully represents the latent concept. DAGs (Cunningham, 2021, Chapter 3; Imbens, 2020; Pearl, 1998), by contrast, encode assumptions about causal relationships among variables and help researchers identify which conditioning sets avoid bias. The two frameworks are complementary. Libby boxes discipline the link between constructs and measures, while DAGs discipline the link between variables and identification.

$$ATE = E [Y_i(1) - Y_i(0)] , \quad (1)$$

$$ATT = E [Y_i(1) - Y_i(0) \mid D_i = 1] , \quad (2)$$

$$LATE = E [Y_i(1) - Y_i(0) \mid \text{complier}] . \quad (3)$$

Each quasi-experimental design identifies a specific estimand under its own assumptions. Difference-in-differences (DiD) identifies the ATT² under a parallel-trends assumption: absent treatment, treated and control units would have evolved similarly (Roth et al., 2023). Instrumental variables (IV) identify the LATE³—the effect on compliers, i.e., units whose treatment status is shifted by the instrument—under relevance, independence, and monotonicity (Angrist et al., 1996). Regression discontinuity (RDD) also identifies a LATE at the assignment threshold under continuity of potential outcomes (Lee & Lemieux, 2010).

Making the estimand explicit has three practical benefits. First, it prevents over-interpretation: a LATE estimated from an IV design is not, in general, the average effect for the full population, and an ATT from a staggered DiD may be a weighted average of heterogeneous cohort-specific effects whose aggregation requires care (Callaway & Sant’Anna, 2021; de Chaisemartin & D’Haultfœuille, 2020; Sun & Abraham, 2021). Second, it forces researchers to define the target population of inference—who is affected, and under what conditions. Third, it clarifies how measurement choices interact with identification: if the proxy \hat{Y} is a noisy measure of the true outcome Y^* , the gap between the estimated and target estimands depends on the structure of measurement error, not only on identification assumptions (Gow et al., 2016).

Step 2: Articulate the assignment mechanism. In design-based research, identification comes from understanding why some units are exposed to the treatment and others are not, and what institutional process governs this exposure (Angrist & Pischke, 2009; Leuz, 2022). In accounting settings, institutional detail about regulation, enforcement, listing rules, disclosure mandates, or auditing regimes can generate quasi-experimental variation that is more credible than residual variation in a saturated regression.⁴

²Average Treatment Effect for the Treated group (ATT), is “the population mean treatment effect for the group of units that had been assigned the treatment” (Cunningham, 2021, Sec. 4.1.2).

³The Average Treatment Effect (ATE) can be expanded to $ATE = E [Y_i(1)] - E [Y_i(0)]$. Notice that it “requires both potential outcomes for each i unit. Since we only know one of these by the switching equation, the average treatment effect, or the ATE, is inherently unknowable” (Cunningham, 2021, Sec. 4.1.2). However, under a few assumptions, we can estimate a Local ATE, the LATE for compliers, i.e., the ones who take the treatment only when encouraged by the instrument.

Step 3: Choose the design, then the estimator. Once the assignment mechanism is plausible, the estimator follows from the design’s identifying assumptions. DiD requires parallel trends (in the relevant counterfactual sense), IV requires relevance and exclusion, and RDD requires continuity in potential outcomes at the threshold (Angrist & Pischke, 2009; Imbens & Rubin, 2015). At this stage, researchers should also revisit the mapping from theoretical constructs to empirical proxies discussed in Section 2, assessing whether measurement error in the chosen proxies could compromise the estimand of interest.

Step 4: Run diagnostics that match the design failure modes. Diagnostics should target the specific ways the design can fail. Adding controls can reduce confounding, but it can introduce bias if variables are post-treatment or colliders (Cinelli et al., 2024; Whited et al., 2022). Theory-based selection and explicit causal diagrams reduce the risk of mechanical over-control (Imbens, 2020; Mändli & Rönkkö, 2025). Extreme observations are common in accounting data, and default winsorization can be ineffective for inference problems created by influential points (Leone et al., 2019).

Step 5: Use falsification and sensitivity tools. The objective is to communicate how conclusions depend on assumptions and potential unobservables, rather than treating robustness as a binary pass/fail criterion. When placebo tests⁵ are weak or ambiguous, negative controls⁶ and sensitivity analysis offer complementary evidence about the plausibility of identifying assumptions (Cinelli & Hazlett, 2020; Lipsitch et al., 2010).

Step 6: Make the design reproducible. Reproducible workflows affect credibility and learning in the field. Survey evidence in accounting highlights that researchers perceive irreproducibility as common and under-addressed (Hail et al., 2020). Transparent documentation of data construction decisions is particularly important in accounting settings where researcher degrees of freedom can be large. Recent

⁴A saturated regression includes a dummy (and thus a separate mean) for every mutually exclusive subgroup (“cell”) defined by the discrete co- variates, i.e., it fully interacts the relevant indicators. As a result, the fitted values are exactly the cell means; coefficients are just differences in those means, and identification comes from comparing observations within the same cell structure rather than from leftover functional-form assumptions (Angrist & Pischke, 2009, Sec. 3.1.4).

many-analysts evidence shows that independent teams replicating the same published results can yield materially different samples and estimates from routine choices in data preparation alone, sometimes altering statistical significance or the sign of the effect (Huntington- Klein et al., 2021).

Interpreting results beyond statistical significance. Even when a design is well-specified, interpretation can fail if results are reported in a way that overemphasizes statistical significance and underemphasizes substantive meaning (Wasserstein et al., 2019). Large samples can make trivial effects statistically detectable, while practically meaningful effects may remain statistically insignificant due to measurement noise or sample constraints. For this reason, accounting research should discuss effect magnitudes in economically interpretable terms (Gelman & Stern, 2006; Hubbard et al., 2019; Wasserstein et al., 2019). Examples of practical interpretation include benchmarking estimates against typical firm-level variation, discussing plausible organizational or market channels required for effects to operate, and clarifying what the estimates imply for investors, regulators, or auditors.

Heterogeneity analysis can illuminate mechanisms, boundary conditions, and policy relevance. When grounded in theory, heterogeneity provides a clearer map of when and for whom effects occur. At the same time, heterogeneity exploration should be disciplined rather than driven by unstructured specification searching. Tools such as multiverse analysis⁷ and specification curve analysis⁸ help clarify how

conclusions depend on modeling choices (Simonsohn et al., 2020; Steegen et al., 2016).

In observational settings common in archival accounting, effect-size interpretation can be strengthened by reporting sensitivity-to-unobservables metrics alongside point estimates and standard errors. For example, Oster (2019) formalizes the widely used “coefficient stability” heuristic by translating coefficient movements and changes in R^2 across baseline and fully controlled specifications into an explicit calibration of how strong selection on unobservables (relative to observables) would need to be to materially change conclusions. This frames robustness as a quantitative statement about remaining confounding, not as a binary checklist tied to statistical significance.

External validity and generalizability. A well-identified study answers a credible causal question, but internal validity alone does not guarantee that results extend beyond the specific setting studied. External validity—the extent to which findings generalize to different populations, institutional environments, and time periods—deserves explicit consideration (List, 2020). This concern is especially relevant in accounting research, where many quasi-experimental designs exploit country-specific regulatory reforms, enforcement changes, or disclosure mandates. A field experiment on tax compliance nudges conducted with Brazilian firms, for example, identifies a local treatment effect shaped by a specific enforcement regime and taxpayer population; whether those results

⁵A placebo test re-estimates the main specification in a setting where the true effect should be zero, e.g., assigning a pseudo treatment date in the pre-period, pseudo-treated units, or an outcome that should not respond mechanically to the intervention. Nonzero “effects” in these placebo settings raise concerns about violations of identifying assumptions (Abadie et al., 2010; Heckman & Hotz, 1989).

⁶Negative controls are variables used to probe residual bias. A negative control outcome shares confounders with the primary outcome but cannot plausibly be affected by the treatment. A negative control exposure is related to the treatment/confounders but cannot affect the outcome except through bias. Evidence of “effects” on negative controls signals remaining confounding, measurement, or selection problems; with additional structure, negative controls can sometimes be used for bias calibration/correction (Lipsitch et al., 2010; Sofer et al., 2016; Tchetgen Tchetgen, 2014).

⁷Multiverse analysis, sometimes “many worlds,” systematically re-estimates the focal effect across a pre-defined set of defensible analytic choices, e.g., alternative variable constructions, sample screens, functional forms, fixed-effects structures, and inference choices—and reports the resulting distribution of estimates, not just a few hand-picked robustness checks. The goal is to make “researcher degrees of freedom” transparent and to show how conclusions depend on reasonable modeling decisions. (Steegen et al., 2016).

⁸Specification curve analysis is a structured multiverse workflow that (i) enumerates a set of “reasonable” specifications, (ii) plots the estimated effect from each specification, often sorted by magnitude, to form a specification curve, and (iii) documents which modeling choices generate which estimates (typically via an accompanying “decision” panel). It can also add global inference (e.g., permutation-based tests) to assess whether the overall pattern of results is inconsistent with a null of no effect across specifications (Simonsohn et al., 2020).

extend to other jurisdictions depends on the similarity of institutional mechanisms, compliance norms, and enforcement capacity (Motoki et al., forthcoming). The question arises equally in observational studies. Research on executive compensation disclosure (Costa et al., 2016) or IFRS adoption by small and medium enterprises (Gonçalves et al., 2022) in Brazil is embedded in a particular regulatory and institutional environment, and its implications for other settings depend on how portable the underlying mechanisms are.

Researchers can strengthen external validity arguments by articulating the theoretical channels through which the treatment operates, assessing whether those channels are plausibly active in other contexts, and comparing the characteristics of the study population with broader populations of interest. When the causal mechanism is well understood, institutional differences become interpretable rather than disqualifying. Even when formal generalization is difficult, well-identified local estimates are more informative than poorly identified global correlations. The goal is not to avoid local designs but to be transparent about the boundaries of inference (List, 2020).

Best practices for authors, reviewers, and editors. The practices outlined in this editorial apply primarily to authors, but their adoption depends on the broader ecosystem in which empirical work is evaluated. Beyond following the suggested workflow, which key points are summarized in Table 1, authors can enhance their contribution by: (i) Clearly stating whether their goal is descriptive, predictive, associative, or causal; (ii) Reporting effect sizes in interpretable metrics and distinguishing between statistical significance and substantive relevance; (iii) Documenting data construction and transformation decisions, including sensitivity to influential observations (Leone et al., 2019); and (iv) Reporting sensitivity analyses when identifying assumptions could plausibly fail (Cinelli & Hazlett, 2020).

These practices support clarity and interpretability. Evans et al. (2015) provide an expanded checklist to assess the quality of research in accounting.

Table 1: Design checklist for identification strategies in empirical accounting research

Design element	What the paper should state explicitly
Estimand and target	Causal question as counterfactual; who is affected; local vs global interpretation (Hernán & Robins, 2020; Imbens & Rubin, 2015).
Assignment mechanism	Institutional process generating variation; why treated vs control differs only in exposure (Leuz, 2022).
Mechanism and alternatives	Main channel + competing explanations; implications for heterogeneity and tests (Hedström & Ylikoski, 2010).
Measurement validity	Construct → proxy mapping; what proxy captures and misses; implications for bias (Bloomfield, 2015; Gow et al., 2016).
Controls and conditioning set	Justification via theory/DAG; avoid post-treatment and colliders (Cinelli et al., 2024; Imbens, 2020; Whited et al., 2022).
Diagnostics	Design-specific checks (e.g., clustering for DiD; continuity checks for RDD) (Bertrand et al., 2004; Roth, 2022; Roth et al., 2023).
Sensitivity / falsification	Negative controls; omitted-variable sensitivity; robustness interpreted as assumption dependence (Cinelli & Hazlett, 2020; Lipsitch et al., 2010).
External validity	Scope of inference; institutional channels supporting generalization; comparison with target populations (List, 2020).
Reproducibility	Data construction logs; code availability; replicable pipeline (Hail et al., 2020).

Reviewers and editors shape methodological standards, and a thorough treatment of their roles would merit a dedicated editorial. Here, we note that the design workflow above provides a natural structure for evaluating empirical manuscripts: reviewers can assess the coherence between research question, design, and claim at each step, rather

than relying on generic robustness demands. Berk et al. (2017) offers a thoughtful reflection on constructive reviewing in economics, and Edmans (2025) provides an editor-facing synthesis of common reasons manuscripts fail, emphasizing contribution, execution, and exposition. Both are directly applicable to accounting. Journals can further support methodological culture by publishing high-quality descriptive studies, replications, null results, and theory-driven explorations, and by encouraging transparency through data and code availability.

4 Concluding Remarks

The evolution of methodological culture in Brazilian accounting research requires investments that strengthen design reasoning, measurement discipline, and transparency as shared professional standards. One important step is to broaden training so that methodological formation goes beyond technical instruction in advanced econometrics and more systematically incorporates research design, causal reasoning, measurement, and the philosophy of science. At the same time, cross-institutional and interdisciplinary collaboration can play a central role in expanding methodological capacity, facilitating access to data, regulatory insights, and institutional knowledge, while also enriching accounting research through dialogue with economists, statisticians, psychologists, and legal scholars. Institutional development is also important. The creation of shared data repositories, the promotion of preregistration when appropriate, and greater support for replication studies can help consolidate stronger methodological expectations within the field. Together, these efforts would contribute to aligning Brazilian accounting research with broader international norms of transparency and reproducibility (Hail et al., 2020).

This editorial has argued for a single, consequential shift: from treating empirical methods as a fixed menu of techniques to treating identification as a reasoning workflow. The practical standard we propose is that empirical work should be evaluated by the coherence between research question, design, and claim, and by the extent to which assumptions are stated and defended with targeted evidence. This standard applies to causal, descriptive, and associative work alike. If one message endures, it should be this: the credibility of an empirical finding is determined not by the sophistication of the estimator employed, but by the clarity with which the researcher explains what variation identifies the result, what could go wrong, and why the reader should find the comparison convincing.

References

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490), 493–505.
- Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434), 444–455.
- Angrist, J. D., & Pischke, J. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Berk, J. B., Harvey, C. R., & Hirshleifer, D. (2017). How to Write an Effective Referee Report and Improve the Scientific Review Process. *Journal of Economic Perspectives*, 31(1), 231–244.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, 119(1), 249–275.
- Bloomfield, R. J. (2015). Rethinking Managerial Reporting. *Journal of Management Accounting Research*, 27 (1), 139–150.
- Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1), 39–67.
- Cinelli, C., Steegen, S., & Tuerlinckx, F. (2024). A Crash Course in Good and Bad Controls. *Sociological Methods & Research*, 53(3), 1071–1104.
- Costa, C. M., Galdi, F. C., Motoki, F. Y. S., & Sanchez, J. M. (2016). Non-Compliance in Executive Compensation Disclosure: The Brazilian Experience. *Journal of Business Finance & Accounting*, 43(3–4), 329–369.
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

- da Silva, R. L. M., & Nardi, P. C. C. (2017). Full Adoption of IFRSs in Brazil: Earnings Quality and the Cost of Equity Capital. *Research in International Business and Finance*, 42, 1057–1073.
- de Chaisemartin, C., & D’Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9), 2964–2996.
- Edmans, A. (2025). Learnings From 1000 Rejections. *Financial Management*, 54(2), 419–444.
- Evans, J. H. I., Feng, M., Hoffman, V. B., Moser, D. V., & van der Stede, W. A. (2015). Points to Consider When Self-Assessing Your Empirical Accounting Research. *Contemporary Accounting Research*, 32(3), 1162–1192.
- Funchal, B., & Monte-Mor, D. S. (2016). Corporate Governance and Credit Access in Brazil: The Sarbanes-Oxley Act as a Natural Experiment. *Corporate Governance: An International Review*, 24(5), 528–547.
- Gelman, A., & Stern, H. (2006). The Difference Between “Significant” and “Not Significant” is Not Itself Statistically Significant. *The American Statistician*, 60(4), 328–331.
- Gonçalves, T., Costa, C., Motoki, F. Y. S., & De Moura, A. A. F. (2022). What Influences the Implementation of IFRS for SMEs? The Brazilian Case. *Accounting & Finance*, 62(2), 2493–2526.
- Gow, I. D., Larcker, D. F., & Reiss, P. C. (2016). Causal Inference in Accounting Research. *Journal of Accounting Research*, 54(2), 477–523.
- Hail, L., Lang, M., & Leuz, C. (2020). Reproducibility in Accounting Research: Views of the Research Community. *Journal of Accounting Research*, 58(2), 519–543.
- Heckman, J. J., & Hotz, V. J. (1989). Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training. *Journal of the American Statistical Association*, 84(408), 862–874.
- Hedström, P., & Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Annual Review of Sociology*, 36(1), 49–67.
- Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Chapman & Hall/CRC.
- Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396), 945–960.
- Hubbard, R., Haig, B. D., & Parsa, R. A. (2019). The Limited Role of Formal Statistical Inference in Scientific Inference. *The American Statistician*, 73(sup1), 91–98.
- Huntington-Klein, N., Arenas, A., Beam, E. A., Bertoni, M., Bloem, J. R., Burli, P., Chen, N., Grieco, P., Ekpe, G., Pugatch, T., Saavedra, M., & Stopnitzky, Y. (2021). The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry*, 59(3), 944–960.
- Imbens, G. W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, 58(4), 1129–1179.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
- Kennedy, P. E. (2002). Sinning in the Basement: What Are the Rules? The Ten Commandments of Applied Econometrics. *Journal of Economic Surveys*, 16(4), 569–589.
- Kondylis, F., & McKenzie, D. (2024). A Curated List of Our Postings on Technical Topics—Your One-Stop Shop for Methodology [Accessed December 10, 2025].
- Larcker, D. F., & Rusticus, T. O. (2010). On the Use of Instrumental Variables in Accounting Research. *Journal of Accounting and Economics*, 49(3), 186–205.
- Leamer, E. E. (1983). Let’s Take the Con Out of Econometrics. *American Economic Review*, 73(1), 31–43.
- Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), 281–355.
- Leone, A. J., Minutti-Meza, M., & Wasley, C. E. (2019). Influential Observations and Inference in Accounting Research. *The Accounting Review*, 94(6), 337–364.
- Leuz, C. (2022). Towards a Design-Based Approach to Accounting Research. *Journal of Accounting and Economics*, 74(2-3), 101550.

- Libby, R. (1981). *Accounting and Human Information Processing: Theory and Applications*. Prentice Hall.
- Lipsitch, M., Tchetgen Tchetgen, E., & Cohen, T. (2010). Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies. *Epidemiology*, 21(3), 383–388.
- List, J. A. (2020). *Non est Disputandum de Generalizability? A Glimpse into The External Validity Trial* (Working Paper N. 27535). National Bureau of Economic Research.
- Mändli, F., & Rönkkö, M. (2025). To Omit or to Include? Integrating the Frugal and Prolific Perspectives on Control Variable Use. *Organizational Research Methods*, 28(1), 114–137.
- Motoki, F. Y. S., Murgia, L. M., & Funchal, B. (forthcoming). The Power and Peril of Nudging Non-Compliant Firms: A Field Experiment on Tax Compliance [Conditionally accepted]. *Journal of Accounting and Public Policy*.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37 (2), 187–204.
- Pearl, J. (1998). Graphs, Causality, and Structural Equation Models. *Sociological Methods & Research*, 27 (2), 226–284.
- Roth, J. (2022). Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review: Insights*, 4(3), 305–322.
- Roth, J., Sant’Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*, 235(2), 2218–2244.
- Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5), 688–701.
- Sampaio, J. O., Gallucci Netto, H., Brunassi Silva, V. A., & Schiozer, R. F. (2020). Mandatory IFRS Adoption, Corporate Governance and Firm Value [Working paper].
- Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289–310.
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification Curve Analysis. *Nature Human Behaviour*, 4, 1208–1214.
- Sofer, T., Richardson, D. B., Colicino, E., Schwartz, J., & Tchetgen Tchetgen, E. J. (2016). On Negative Outcome Control of Unobserved Confounding as a Generalization of Difference-in-Differences. *Statistical Science*, 31(3), 348–361.
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science*, 11(5), 702–712.
- Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.
- Tchetgen Tchetgen, E. J. (2014). The Control Outcome Calibration Approach for Causal Inference with Unobserved Confounding. *American Journal of Epidemiology*, 179(5), 633–640.
- Trevisan, G. (2024). Quebrando o Paradigma: Da Correlação à Inferência Causal. *Advances in Scientific and Applied Accounting*, 17 (3).
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond “ $p < 0.05$ ”. *The American Statistician*, 73(sup1), 1–19.
- Whited, T. M., Moon, T. S., Shipman, J. E., & Swanquist, Q. T. (2022). Out of Control: The (Over)Use of Controls in Accounting Research. *The Accounting Review*, 97 (3), 393–421.