

# Da associação à explicação: fortalecendo o desenho de pesquisa e o raciocínio causal na pesquisa em contabilidade

Fabio Yoshio Suguri Motoki<sup>1</sup>  Valdemar Pinho Neto<sup>2</sup>  Victor Rangel<sup>3</sup> 

<sup>1</sup> University of Texas Rio Grande Valley, Edinburg, Texas, Estados Unidos

<sup>2</sup> Escola Brasileira de Economia e Finanças - FGV EPGE, Rio de Janeiro, Rio de Janeiro, Brasil

<sup>3</sup> Insper, São Paulo, São Paulo, Brasil



<sup>1</sup>fabio.motoki@utrgv.edu

<sup>2</sup>valdemar.pinho@fgv.br

<sup>3</sup>victorrangel@al.insper.edu.br

## Resumo

Um editorial recente desta revista (Trevisan, 2024) destacou a crescente importância da inferência causal na pesquisa em contabilidade e os riscos de interpretar padrões correlacionais como relações causais. Com base nessa contribuição, este editorial avança a discussão ao deslocar o foco para etapas anteriores do processo: como os pesquisadores formulam perguntas, definem estimandos, articulam mecanismos, justificam escolhas de mensuração e constroem estratégias de identificação coerentes. Em vez de retomar estimadores quase-experimentais como uma caixa de ferramentas, organizamos o processo de pesquisa como um fluxo de desenho que pode sustentar inferência descritiva, preditiva, associativa e causal. Propomos um checklist prático para autores e pareceristas, discutimos modos recorrentes de falha em desenhos empíricos na contabilidade (controles, observações influentes e desalinhamento entre desenho e alegações) e conectamos escolhas metodológicas a normas de transparência e reprodutibilidade na área.

**Palavras-chave:** inferência causal; desenho de pesquisa, econometria; contabilidade; mecanismos; mensuração; validade.

## Como Citar:

Motoki, F. Y. S., Pinho Neto, V., & Rangel, V. (2026). Da associação à explicação: fortalecendo o desenho de pesquisa e o raciocínio causal na pesquisa em contabilidade. *Advances in Scientific and Applied Accounting*, 19(1), 001–009/010. <https://doi.org/10.14392/asaa.2026190101>

## 1 Motivação e diagnóstico: da associação à explicação

Nos últimos anos, a pesquisa empírica em contabilidade passou por uma mudança importante em direção a uma preocupação mais explícita com inferência causal. O editorial de Trevisan (2024) representou um marco relevante nessa transição ao oferecer uma visão sintética de abordagens quase-experimentais, como Diferença-em-Diferenças, Variáveis Instrumentais e Descontinuidade de Regressão, além de destacar armadilhas recorrentes na interpretação de associações baseadas em regressão como alegações causais. Essa contribuição é oportuna, pois a pesquisa contábil brasileira atingiu um nível de maturidade técnica em que muitos estudiosos já estão prontos para trabalhar com estratégias de identificação que vão além das regressões em painel convencionais.

Se o editorial anterior se concentrou nos métodos que sustentam a identificação causal, este editorial se concentra no raciocínio que antecede e orienta as escolhas metodológicas. Antes de selecionar um estimador, os pesquisadores devem explicitar que tipo de inferência pretendem produzir, qual mecanismo teórico conecta suas variáveis, se suas medidas representam fielmente os construtos de interesse e qual contexto institucional rege o fenômeno em estudo (Hernán & Robins, 2020; Imbens & Rubin, 2015; Shmueli, 2010). Esses fundamentos importam não apenas para estudos causais, mas também para trabalhos descritivos e associativos, pois reduzem ambiguidades interpretativas e tornam os achados empíricos mais cumulativos.

A pesquisa empírica em contabilidade no Brasil tem se beneficiado do aumento da disponibilidade de dados, da expansão dos programas de pós-graduação e da crescente integração com a literatura internacional. Os estudos empregam regularmente métodos de dados em painel, estimadores sofisticados e grandes bases extraídas de registros regulatórios, divulgações corporativas e dados de mercado. Esses são avanços relevantes. Ao mesmo tempo, tensões persistentes moldam o panorama metodológico. Muitos estudos empíricos ainda dependem fortemente de modelos convencionais de regressão, tratando decisões metodológicas como etapas de um fluxo padronizado, e não como expressões de um desenho de pesquisa subjacente. Variáveis são winsorizadas por padrão, proxies para construtos complexos são escolhidas com base em precedentes e não em validade, e variáveis de controle se acumulam nas regressões sem uma justificativa teórica clara (Kennedy, 2002; Kondylis & McKenzie, 2024; Leamer, 1983).

Um desafio central é a distância entre as perguntas que os pesquisadores afirmam responder e os desenhos que

de fato utilizam. Ainda é comum encontrar questões de pesquisa formuladas em linguagem causal (“X afeta Y?”) enquanto a análise fornece apenas correlações parciais, mesmo quando problemas de endogeneidade, como variáveis omitidas, seleção e causalidade reversa, são reconhecidos (Larcker & Rusticus, 2010). Considere, por exemplo, o padrão frequente na pesquisa contábil brasileira de regredir proxies de gerenciamento de resultados sobre indicadores de governança com dados em painel de empresas listadas na B3, sem explicitar por que a variação observada em governança identifica um efeito causal, em vez de refletir escolhas simultâneas no nível da firma. Trevisan (2024) destacou essa preocupação no contexto específico da pesquisa em governança. Nosso ponto é mais amplo: o desalinhamento entre pergunta de pesquisa, desenho e interpretação é sistêmico. Há exceções dignas de nota: estudos que exploram a Lei Sarbanes-Oxley como experimento natural para firmas brasileiras com cross-listing (Funchal & Monte-Mor, 2016) e a adoção obrigatória das IFRS como choque regulatório em desenhos de Diferença-em-Diferenças (da Silva & Nardi, 2017; Sampaio et al., 2020) ilustram como pode ser uma pesquisa contábil baseada em desenho com dados brasileiros. O desafio é fazer com que tais práticas se tornem a regra, e não a exceção.

Essa lacuna não é primordialmente técnica. Em vez disso, ela frequentemente reflete estruturas de incentivos que recompensam volume de publicação, prazos curtos e adesão a padrões empíricos familiares, além das restrições impostas pelos dados prontamente disponíveis. Uma cultura metodológica que equipara sofisticação à complexidade (o número de controles, o estimador específico utilizado ou a extensão das seções de robustez) pode deslocar requisitos mais profundos da inferência científica crível. Por essa razão, propomos um enquadramento baseado em desenho para o trabalho empírico. A tarefa central não é “escolher um método”, mas especificar qual variação identifica uma alegação e quais pressupostos conectam o desenho a um contrafactual interpretável (Angrist & Pischke, 2009; Hernán & Robins, 2020; Imbens & Rubin, 2015).

## 2 Da pergunta ao desenho: fundamentos conceituais

Esta seção apresenta os fundamentos conceituais do desenho de pesquisa. Seu objetivo é motivar por que cada elemento importa para uma inferência crível, deixando os detalhes operacionais para o fluxo de trabalho da Seção 3.

**Objetivos de inferência.** Um passo central para fortalecer o desenho empírico é distinguir entre pesquisa descritiva, preditiva, associativa e causal (Hernán & Robins, 2020; Shmueli, 2010). Cada tipo de inferência se apoia em pressupostos distintos e serve a objetivos científicos distintos. Trabalhos descritivos podem revelar padrões dignos de investigação posterior; trabalhos preditivos podem apoiar previsão ou avaliação de risco; trabalhos associativos podem descobrir correlações úteis para construir índices ou mapear regularidades empíricas. Trabalhos causais procuram responder a perguntas contrafactuais por meio da construção de contrafactuais críveis — objetivo que formalizamos na Seção 3. Muitas contribuições influentes em contabilidade não são causais. O requisito central, aqui, é a coerência entre a alegação e o desenho usado para sustentá-la.

**Mecanismos.** Alegações causais só são críveis quando ancoradas em mecanismos plausíveis. Mecanismos explicam por que um efeito deveria ocorrer, em que circunstâncias e por quais canais (Hedström & Ylikoski, 2010). Na pesquisa contábil, os mecanismos frequentemente envolvem canais psicológicos, informacionais, ligados à governança e institucionais. Por exemplo, o efeito da regulação de disclosure sobre o valor da firma costuma ser hipotetizado como operando por meio de reduções na assimetria informacional, melhorias no monitoramento ou mudanças na eficiência contratual. Sem explicitar esses mecanismos, alegações causais tornam-se difíceis de interpretar e resultados associativos tornam-se ambíguos. O raciocínio mecanístico também é importante para discriminar entre explicações concorrentes. Suponha que observemos uma associação negativa entre conservadorismo contábil e custo de capital próprio. Isso ocorre porque o conservadorismo melhora o monitoramento? Porque firmas conservadoras diferem sistematicamente em risco? Porque há variáveis omitidas? Os mecanismos esclarecem essas alternativas e orientam escolhas de desenho capazes de separá-las empiricamente. Ferramentas gráficas como Libby boxes e grafos acíclicos direcionados (DAGs) podem ser úteis para tornar os pressupostos explícitos.<sup>1</sup>

**Validade de mensuração.** A pesquisa contábil frequentemente depende de proxies para construtos latentes, como qualidade da auditoria, incentivos de reporte, gerenciamento de resultados, opacidade informacional ou cultura corporativa. O uso de proxies

muitas vezes é inevitável, mas sua validade precisa ser avaliada, e não presumida (Gow et al., 2016). Erro de mensuração pode atenuar coeficientes, induzir viés ou criar padrões enganosos que se replicam entre estudos simplesmente porque a mesma proxy falha é amplamente adotada. A validade de mensuração não é acessória à inferência. Ela é condição necessária para interpretação substantiva, seja o objetivo descritivo, associativo, preditivo ou causal.

**Contexto institucional e os limites da escalada metodológica.** A credibilidade de alegações empíricas depende da compreensão do contexto institucional no qual os resultados contábeis estão inseridos, incluindo regulação, regras tributárias, estruturas de governança e microestruturas de mercado (Leuz, 2022). Explorar especificidades institucionais é uma das maneiras mais poderosas de construir estratégias de identificação críveis; entendimento contextual é um insumo de desenho que não pode ser substituído por técnica econométrica. Um padrão preocupante no trabalho empírico é a escalada metodológica. Diante de desafios de identificação, autores às vezes respondem empilhando mais controles (Cinelli et al., 2024; Whited et al., 2022), instrumentos (Larcker & Rusticus, 2010) ou rituais genéricos de robustez, em vez de reexaminar o desenho central. Testes de robustez só são informativos quando a variação identificadora é crível; caso contrário, produzem uma falsa sensação de segurança (Cinelli & Hazlett, 2020; Roth, 2022).

### 3 Padrões operacionais: um fluxo de desenho para alegações empíricas críveis

Este editorial propõe um padrão prático para alegações empíricas críveis em pesquisa observacional em contabilidade. A identificação deve ser tratada como um fluxo de trabalho documentado que conecta uma pergunta de pesquisa a uma alegação interpretável. Uma estratégia de identificação crível deve explicitar: (i) o estimando e a pergunta contrafactual; (ii) o mecanismo de atribuição que gera a variação identificadora; (iii) a estratégia de mensuração que conecta construtos a proxies; (iv) diagnósticos e análises de sensibilidade direcionados aos principais modos de falha do desenho; e (v) transparência e reprodutibilidade da construção dos dados e do código (Gow et al., 2016; Hail et al., 2020; Leuz, 2022).

**Passo 1:** Definir o estimando e o contrafactual. Um de

<sup>1</sup>Libby boxes (Bloomfield, 2015; Libby, 1981) mapeiam construtos teóricos para suas proxies empíricas e são particularmente úteis para avaliar validade de mensuração: se a proxy representa fielmente o conceito latente. DAGs (Cunningham, 2021, Chapter 3; Imbens, 2020; Pearl, 1998), por sua vez, codificam pressupostos sobre relações causais entre variáveis e ajudam os pesquisadores a identificar quais conjuntos de condicionamento evitam viés. Os dois arcabouços são complementares. Libby boxes disciplinam o vínculo entre construtos e medidas, enquanto DAGs disciplinam o vínculo entre variáveis e identificação.

senho começa com a explicitação da quantidade causal de interesse — o estimando — e da população-alvo. O arcabouço de resultados potenciais oferece uma linguagem mínima, porém precisa, para fazer isso (Hernán & Robins, 2020; Imbens & Rubin, 2015; Rubin, 1974). Seja  $D_i \in \{0, 1\}$  um indicador binário de tratamento para a unidade  $i$ , e sejam  $Y_i(1)$  e  $Y_i(0)$  os resultados potenciais sob tratamento e controle, respectivamente. O efeito causal individual é  $Y_i(1) - Y_i(0)$ , que nunca é observado conjuntamente — o problema fundamental da inferência causal (Holland, 1986). Desenhos empíricos, portanto, visam agregados populacionais dessa quantidade sob pressupostos explicitados. Os estimandos mais comuns são:

$$\begin{aligned} \text{ATE} &= E [Y_i(1) - Y_i(0)], & (1) \\ \text{ATT} &= E [Y_i(1) - Y_i(0) \mid D_i = 1], & (2) \\ \text{LATE} &= E [Y_i(1) - Y_i(0) \mid \text{complier}]. & (3) \end{aligned}$$

Cada desenho quase-experimental identifica um estimando específico sob seus próprios pressupostos. Diferença-em-Diferenças (DiD) identifica o ATT<sup>2</sup> sob um pressuposto de tendências paralelas: na ausência de tratamento, unidades tratadas e de controle teriam evoluído de forma semelhante (Roth et al., 2023). Variáveis Instrumentais (IV) identificam o LATE<sup>3</sup> — o efeito sobre compliers, isto é, unidades cujo status de tratamento é alterado pelo instrumento — sob relevância, independência e monotonicidade (Angrist et al., 1996). A Descontinuidade de Regressão (RDD) também identifica um LATE no limiar de atribuição sob continuidade dos resultados potenciais (Lee & Lemieux, 2010).

Tornar o estimando explícito traz três benefícios práticos. Primeiro, evita sobre-interpretação: um LATE estimado a partir de um desenho de IV não é, em geral, o efeito médio para a população inteira, e um ATT oriundo de um DiD com adoção escalonada pode ser uma média ponderada de efeitos heterogêneos específicos por coorte cuja agregação requer cuidado (Callaway & Sant’Anna, 2021; de Chaisemartin & D’Haultfoeulle, 2020; Sun & Abraham, 2021). Segundo, força o pesquisador a definir a população-alvo da inferência — quem é afetado, e sob quais condições. Terceiro, esclarece como

<sup>2</sup>O efeito médio do tratamento sobre os tratados (ATT) é o “efeito médio do tratamento na população de unidades que recebeu o tratamento” (Cunningham, 2021, Sec. 4.1.2).

<sup>3</sup>O efeito médio do tratamento (ATE) pode ser expandido como  $\text{ATE} = E [Y_i(1)] - E [Y_i(0)]$ . Note que ele requer ambos os resultados potenciais para cada unidade  $i$ . Como conhecemos apenas um deles pela equação de switching, o efeito médio do tratamento, ou ATE, é inerentemente incognoscível” (Cunningham, 2021, Sec. 4.1.2). Contudo, sob alguns pressupostos, podemos estimar um ATE local, o LATE para compliers, isto é, as unidades que recebem o tratamento apenas quando incentivadas pelo instrumento.

escolhas de mensuração interagem com identificação: se a proxy  $\hat{Y}^*$  é uma medida ruidosa do verdadeiro desfecho  $Y^*$ , a diferença entre o estimando estimado e o estimando-alvo depende da estrutura do erro de mensuração, e não apenas dos pressupostos de identificação (Gow et al., 2016).

**Passo 2: Articular o mecanismo de atribuição.** Na pesquisa baseada em desenho, a identificação decorre da compreensão de por que algumas unidades são expostas ao tratamento e outras não, e de qual processo institucional governa essa exposição (Angrist & Pischke, 2009; Leuz, 2022). Em contextos contábeis, detalhes institucionais sobre regulação, enforcement, regras de listagem, exigências de disclosure ou regimes de auditoria podem gerar variação quase-experimental mais crível do que a variação residual em uma regressão saturada.<sup>4</sup>

**Passo 3: Escolher o desenho, depois o estimador.** Uma vez que o mecanismo de atribuição seja plausível, o estimador decorre dos pressupostos identificadores do desenho. DiD requer tendências paralelas (no sentido contrafactual relevante), IV requer relevância e exclusão, e RDD requer continuidade dos resultados potenciais no limiar (Angrist & Pischke, 2009; Imbens & Rubin, 2015). Nesse estágio, os pesquisadores também devem revisar o mapeamento entre construtos teóricos e proxies empíricas discutido na Seção 2, avaliando se erro de mensuração nas proxies escolhidas pode comprometer o estimando de interesse.

**Passo 4: Rodar diagnósticos compatíveis com os modos de falha do desenho.** Os diagnósticos devem mirar as formas específicas pelas quais o desenho pode falhar. Adicionar controles pode reduzir confundimento, mas também pode introduzir viés se as variáveis forem pós-tratamento ou colliders (Cinelli et al., 2024; Whited et al., 2022). Seleção orientada por teoria e diagramas causais explícitos reduzem o risco de sobrecontrole mecânico (Imbens, 2020; Mändli & Rönkkö, 2025). Observações extremas são comuns em dados contábeis, e a winsorização aplicada por padrão pode ser ineficaz para problemas de inferência gerados por pontos influentes (Leone et al., 2019).

<sup>4</sup>Uma regressão saturada inclui uma dummy (e, portanto, uma média separada) para cada subgrupo mutuamente exclusivo (“célula”) definido pelas covariáveis discretas, isto é, interage plenamente os indicadores relevantes. Como resultado, os valores ajustados coincidem exatamente com as médias das células; os coeficientes são apenas diferenças entre essas médias, e a identificação decorre da comparação entre observações dentro da mesma estrutura de células, e não de pressupostos remanescentes de forma funcional (Angrist & Pischke, 2009, Sec. 3.1.4).

### **Passo 5: Usar ferramentas de falsificação e sensibilidade.**

O objetivo é comunicar como as conclusões dependem de pressupostos e de potenciais não observáveis, em vez de tratar robustez como um critério binário de aprovação ou reprovação. Quando testes placebo<sup>5</sup> são fracos ou ambíguos, controles negativos<sup>6</sup> e análise de sensibilidade oferecem evidência complementar sobre a plausibilidade dos pressupostos identificadores (Cinelli & Hazlett, 2020; Lipsitch et al., 2010).

**Passo 6: Tornar o desenho reprodutível.** Fluxos de trabalho reprodutíveis afetam a credibilidade e o aprendizado na área. Evidência de survey em contabilidade mostra que pesquisadores percebem a irreprodutibilidade como um problema comum e subtratado (Hail et al., 2020). A documentação transparente das decisões de construção de dados é particularmente importante em contextos contábeis, nos quais os graus de liberdade do pesquisador podem ser amplos. Evidências recentes do tipo many-analysts mostram que equipes independentes, ao replicarem os mesmos resultados publicados, podem chegar a amostras e estimativas materialmente distintas apenas em função de escolhas rotineiras de preparação dos dados, às vezes alterando a significância estatística ou o sinal do efeito (Huntington-Klein et al., 2021).

**Interpretando resultados para além da significância estatística.** Mesmo quando um desenho é bem especificado, a interpretação pode falhar se os resultados forem repor-

tados de maneira que enfatize em excesso a significância estatística e subestime o significado substantivo (Wasserstein et al., 2019). Amostras grandes podem tornar efeitos triviais estatisticamente detectáveis, enquanto efeitos substantivamente relevantes podem permanecer estatisticamente insignificantes em razão de ruído de mensuração ou restrições amostrais. Por essa razão, a pesquisa contábil deve discutir magnitudes de efeito em termos economicamente interpretáveis (Gelman & Stern, 2006; Hubbard et al., 2019; Wasserstein et al., 2019). Exemplos de interpretação prática incluem comparar estimativas com a variação típica no nível da firma, discutir os canais organizacionais ou de mercado plausíveis necessários para que os efeitos operem e esclarecer o que as estimativas implicam para investidores, reguladores ou auditores.

A análise de heterogeneidade pode iluminar mecanismos, condições de validade e relevância para políticas. Quando ancorada em teoria, a heterogeneidade oferece um mapa mais claro de quando e para quem os efeitos ocorrem. Ao mesmo tempo, a exploração de heterogeneidade deve ser disciplinada, e não guiada por uma busca desestruturada de especificações. Ferramentas como multiverse analysis<sup>7</sup> e specification curve analysis<sup>8</sup> ajudam a esclarecer como as conclusões dependem das escolhas de modelagem (Simonsohn et al., 2020; Steegen et al., 2016).

<sup>5</sup>Um teste placebo reestima a especificação principal em um contexto no qual o verdadeiro efeito deveria ser zero, por exemplo, atribuindo uma pseudo-data de tratamento no período pré-tratamento, pseudo-unidades tratadas ou um desfecho que não deveria responder mecanicamente à intervenção. Efeitos não nulos nesses contextos placebo levantam preocupações sobre violações dos pressupostos identificadores (Abadie et al., 2010; Heckman & Hotz, 1989).

<sup>6</sup>Controles negativos são variáveis usadas para investigar viés residual. Um desfecho controle negativo compartilha confundidores com o desfecho principal, mas não pode plausivelmente ser afetado pelo tratamento. Uma exposição controle negativa está relacionada ao tratamento/confundidores, mas não pode afetar o desfecho exceto por meio de viés. Evidência de “efeitos” sobre controles negativos sinaliza problemas remanescentes de confundimento, mensuração ou seleção; com estrutura adicional, controles negativos às vezes podem ser usados para calibração/correção de viés (Lipsitch et al., 2010; Sofer et al., 2016; Tchetgen Tchetgen, 2014)

<sup>7</sup>Multiverse analysis, às vezes chamada de “many worlds”, reestima

sistematicamente o efeito de interesse ao longo de um conjunto predefinido de escolhas analíticas defensáveis, por exemplo, construções alternativas de variáveis, filtros amostrais, formas funcionais, estruturas de efeitos fixos e escolhas de inferência — e reporta a distribuição resultante das estimativas, e não apenas algumas verificações de robustez escolhidas manualmente. O objetivo é tornar transparentes os graus de liberdade do pesquisador” e mostrar como as conclusões dependem de decisões de modelagem razoáveis. (Steegen et al., 2016).

<sup>8</sup>Specification curve analysis é um fluxo estruturado de análise multiverso que (i) enumera um conjunto de especificações razoáveis, (ii) plota o efeito estimado de cada especificação, muitas vezes ordenado por magnitude, formando uma specification curve, e (iii) documenta quais escolhas de modelagem geram quais estimativas (tipicamente por meio de um painel complementar de decisões). Também pode acrescentar inferência global (por exemplo, testes baseados em permutação) para avaliar se o padrão geral de resultados é incompatível com uma hipótese nula de ausência de efeito ao longo das especificações (Simonsohn et al., 2020).

Em contextos observacionais comuns na pesquisa arquivística em contabilidade, a interpretação do tamanho do efeito pode ser fortalecida pela apresentação de métricas de sensibilidade a não observáveis ao lado das estimativas pontuais e dos erros-padrão. Por exemplo, Oster (2019) formaliza a heurística amplamente usada de “estabilidade do coeficiente” ao traduzir deslocamentos de coeficientes e mudanças no R<sup>2</sup> entre especificações básicas e totalmente controladas em uma calibração explícita do quão forte a seleção em não observáveis (em relação aos observáveis) precisaria ser para alterar materialmente as conclusões. Isso enquadra robustez como uma afirmação quantitativa sobre confundimento remanescente, e não como um checklist binário vinculado à significância estatística.

**Validade externa e generalização.** Um estudo bem identificado responde a uma pergunta causal crível, mas validade interna, por si só, não garante que os resultados se estendam para além do contexto específico analisado. Validade externa — isto é, o grau em que os achados se generalizam para diferentes populações, ambientes institucionais e períodos de tempo — merece consideração explícita (List, 2020). Essa preocupação é especialmente relevante na pesquisa contábil, na qual muitos desenhos quase-experimentais exploram reformas regulatórias específicas de um país, mudanças em enforcement ou exigências de disclosure. Um experimento de campo sobre nudges de conformidade tributária conduzido com firmas brasileiras, por exemplo, identifica um efeito local de tratamento moldado por um regime específico de enforcement e por uma população específica de contribuintes; se esses resultados se estendem a outras jurisdições depende da semelhança entre mecanismos institucionais, normas de conformidade e capacidade de enforcement (Motoki et al., aceito para publicação). A mesma questão surge em estudos observacionais. Pesquisas sobre disclosure de remuneração de executivos (Costa et al., 2016) ou sobre adoção das IFRS por pequenas e médias empresas (Gonçalves et al., 2022) no Brasil estão inseridas em um ambiente regulatório e institucional particular, e suas implicações para outros contextos dependem de quão transportáveis são os mecanismos subjacentes.

Pesquisadores podem fortalecer argumentos de validade externa ao explicitar os canais teóricos pelos quais o tratamento opera, avaliar se esses canais são plausivelmente ativos em outros contextos e comparar as características da população estudada com populações mais amplas de interesse. Quando o mecanismo causal é bem compreendido, diferenças institucionais tornam-se interpretáveis, e não desqualificadoras. Mesmo quando a generalização formal é difícil, estimativas locais bem identificadas são mais informativas do que correlações globais mal identificadas. O objetivo não é evitar desenhos locais, mas ser transparente quanto aos limites da inferência (List, 2020).

**Boas práticas para autores, pareceristas e editores.** As práticas delineadas neste editorial se aplicam primariamente aos autores, mas sua adoção depende do ecossistema mais amplo em que o trabalho empírico é avaliado. Além de seguir o fluxo sugerido, cujos pontos centrais são resumidos na Tabela 1, autores podem ampliar sua contribuição ao: (i) declarar com clareza se seu objetivo é descritivo, preditivo, associativo ou causal; (ii) reportar tamanhos de efeito em métricas interpretáveis e distinguir significância estatística de relevância substantiva; (iii) documentar decisões de construção e transformação dos dados, incluindo sensibilidade a observações influentes (Leone et al., 2019); e (iv) reportar análises de sensibilidade quando os pressupostos identificadores puderem plausivelmente falhar (Cinelli & Hazlett, 2020). Essas práticas promovem clareza e interpretabilidade. Evans et al. (2015) oferecem um checklist ampliado para avaliar a qualidade da pesquisa em contabilidade.

Tabela 1: Checklist de desenho para estratégias de identificação na pesquisa empírica em contabilidade

Elemento do desenho	O que o artigo deve explicar
Estimando e alvo	Pergunta causal como contrafactual; quem é afetado; interpretação local vs. global (Hernán & Robins, 2020; Imbens & Rubin, 2015).
Mecanismo de atribuição	Processo institucional que gera variação; por que tratados e controles diferem apenas em exposição (Leuz, 2022).
Mecanismo e alternativas	Canal principal + explicações concorrentes; implicações para heterogeneidade e testes (Hedström & Ylikoski, 2010).
Validade de mensuração	Mapeamento construído → proxy; o que a proxy capta e o que não capta; implicações para vies (Bloomfield, 2015; Gow et al., 2016).
Controles e conjunto de condicionamento	Justificativa via teoria/DAG; evitar variáveis pós-tratamento e colliders (Cinelli et al., 2024; Imbens, 2020; Whited et al., 2022).
Diagnósticos	Verificações específicas ao desenho (por exemplo, clusterização em DiD; testes de continuidade em RDD) (Bertrand et al., 2004; Roth, 2022; Roth et al., 2023).
Sensibilidade / falsificação	Controles negativos; sensibilidade a variáveis omitidas; robustez interpretada como dependência de pressupostos (Cinelli & Hazlett, 2020; Lipsitch et al., 2010).
Validade externa	Escopo da inferência; canais institucionais que sustentam a generalização; comparação com populações-alvo (List, 2020).
Reprodutibilidade	Registros da construção dos dados; disponibilidade de código; pipeline replicável (Hail et al., 2020).

Pareceristas e editores moldam os padrões metodológicos, e um tratamento aprofundado de seus papéis mereceria um editorial próprio. Aqui, observamos que o fluxo de desenho acima fornece uma estrutura natural para avaliar manuscritos empíricos: pareceristas podem examinar a coerência entre pergunta de pesquisa, desenho e alegação em cada etapa, em vez de depender de exigências genéricas de robustez. Berk et al. (2017) oferece uma reflexão cuidadosa sobre pareceres construtivos em economia, e Edmans (2025) fornece uma síntese voltada a editores sobre razões comuns pelas quais manuscritos fracassam, com ênfase em contribuição, execução e exposição. Ambos são diretamente aplicáveis à contabilidade. Periódicos também podem apoiar a cultura metodológica ao publicar estudos descritivos de alta qualidade, replicações, resultados nulos e explorações orientadas por teoria, além de incentivar transparência por meio da disponibilização de dados e código.

## 4 Considerações finais

A evolução da cultura metodológica na pesquisa contábil brasileira requer investimentos que fortaleçam o raciocínio de desenho, a disciplina de mensuração e a transparência como padrões profissionais compartilhados. Um passo importante é ampliar a formação para que o treinamento metodológico vá além da instrução técnica em econometria avançada e incorpore, de modo mais sistemático, desenho de pesquisa, raciocínio causal, mensuração e filosofia da ciência. Ao mesmo tempo, a colaboração interinstitucional e interdisciplinar pode desempenhar papel central na expansão da capacidade metodológica, facilitando o acesso a dados, insights regulatórios e conhecimento institucional, ao mesmo tempo em que enriquece a pesquisa contábil por meio do diálogo com economistas, estatísticos, psicólogos e juristas. O desenvolvimento institucional também é importante. A criação de repositórios compartilhados de dados, a promoção de pré-registro quando apropriado e maior apoio a estudos de replicação podem ajudar a consolidar expectativas metodológicas mais fortes na área. Em conjunto, esses esforços contribuiriam para alinhar a pesquisa contábil brasileira com normas internacionais mais amplas de transparência e reprodutibilidade (Hail et al., 2020).

Este editorial defende uma mudança específica, mas de grande relevância: deixar de tratar métodos empíricos como um menu fixo de técnicas e passar a tratar identificação como um fluxo de raciocínio. O padrão prático que propomos é que o trabalho empírico seja avaliado pela coerência entre pergunta de pesquisa, desenho e alegação, e pelo grau em que os pressupostos são explicitados e defendidos com evidência direcionada. Esse padrão se aplica igualmente a trabalhos causais, descritivos e associativos. Se uma mensagem permanecer, que seja esta: a credibilidade de um achado empírico

não é determinada pela sofisticação do estimador empregado, mas pela clareza com que o pesquisador explica qual variação identifica o resultado, o que pode dar errado e por que o leitor deve considerar convincente a comparação.

## Referências

- 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. (aceito para publicação). 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.