

# A Endogeneidade e a Importância dos Quase-Experimentos para a Inferência Causal nas Pesquisas em Contabilidade

**Vinicius Gomes Martins**

<https://orcid.org/0000-0001-7401-9570>

**Giuseppe Trevisan**

<https://orcid.org/0000-0001-5165-0597>

Editado em Português e Inglês. Versão original em Português.

Rodada 1: Recebido em 4/7/2023. Pedido de revisão em 7/07/2023. Rodada 2: Resubmetido em 26/07/2023. Pedido de revisão em 26/7/2023. Rodada 3: Resubmetido em 4/08/2023. Aceito em 4/8/2023 por Gerlando Augusto Sampaio Franco de Lima, Doutor (Editor). Publicado em 30/9/2023. Organização responsável pelo periódico: Abracicon.

## 1. Motivação

A pesquisa na área contábil passou por uma significativa mudança de paradigma, notadamente a partir da década de 1960, impulsionada pelos estudos de Ball e Brown (1968) e Beaver (1968). Esses trabalhos pioneiros introduziram a perspectiva positivista na Contabilidade, que passou a enfatizar a abordagem empírica e quantitativa. Com o desenvolvimento da teoria econômica e das finanças, bem como com o avanço da tecnologia da informação e a crescente disponibilidade de dados financeiros, a adoção de uma abordagem baseada em evidências quantitativas ficou ainda mais forte na pesquisa contábil.

Esse choque estrutural produziu impactos profundos na forma como os pesquisadores conduzem seus estudos e, também, na exigência dos principais *journals* de Contabilidade – que passaram a valorizar, cada vez mais, trabalhos empíricos e com enfoque quantitativo. Em muitos periódicos, é praticamente impossível obter uma publicação sem que a pesquisa apresente um desenho quantitativo. Consequentemente, tem havido um expressivo aumento na produção de trabalhos que adotam modelagens econométricas, a fim de identificar relações entre variáveis, prever comportamentos e testar teorias que buscam explicar fenômenos que envolvem a Contabilidade.

Como resultado desse processo, os principais problemas de pesquisa no *mainstream* contábil envolvem o uso de dados observacionais e buscam por evidências de efeitos causais (Gow et al., 2016). Apenas para ilustrar, é comum se deparar com estudos que buscam analisar o “impacto”, o “efeito” ou a “influência” de uma variável *X* sobre uma variável *Y*. Todos esses trabalhos demandam, de forma implícita, a inferência causal. No entanto, os desenhos utilizados em muitos desses estudos têm gerado um crescente número de questionamentos sobre a eficácia da estratégia empírica empregada, e suscitam dúvidas sobre se os resultados obtidos são realmente capazes de revelar efeitos causais ou se estão restritos apenas a apresentar correlações descritivas limitadas.

O principal problema para a interpretação de relação causal associado a essas pesquisas reside na Endogeneidade, algo muito comum em estudos que utilizam dados contábeis. Contornar o problema da Endogeneidade é um desafio recorrente, uma vez que as variáveis contábeis frequentemente estão sujeitas a influências mútuas e a fatores difíceis de se controlar pelo pesquisador. Isso pode levar a correlações espúrias ou vieses nos parâmetros estimados, de forma a comprometer a capacidade de identificação de causalidade e refletir, portanto, na confiabilidade dos resultados.

Os desafios associados aos estudos que buscam capturar relações causais estão cada vez mais evidentes. Felizmente, pesquisadores têm trabalhado para abordar essas questões e aprimorar as estratégias empíricas utilizadas na identificação de relações de causa e efeito. Convém destacar as contribuições dos ganhadores do Prêmio Nobel em Economia de 2021, David Card, Joshua D. Angrist e Guido W. Imbens, reconhecidos pelo uso de técnicas quase-experimentais para inferência causal, especialmente em áreas como economia do trabalho e da educação. As abordagens inovadoras desses economistas se estenderam para além da Economia e influenciaram outras áreas de pesquisa, incluindo a Contabilidade.

A partir disso, neste editorial, propomos apresentar algumas reflexões sobre a pesquisa causal na Contabilidade, com destaque para o problema da Endogeneidade. Procuramos discutir, de forma introdutória, a importância dos quase-experimentos, especificamente de técnicas que podem ser soluções viáveis para muitos projetos de pesquisa na área. Dentre essas técnicas, destacamos o desenho de Diferença-em-Diferenças, o uso de Variáveis Instrumentais e o desenho de Regressão Descontínua. Com isso, buscamos estimular a adoção de métodos mais rigorosos de identificação, no intuito de elevar a preocupação com a qualidade das pesquisas na nossa área e abrir, portanto, caminhos para um ambiente de pesquisa contábil mais sólido e inovador.

## 2. O Problema da Endogeneidade

Afinal, o que é Endogeneidade e por que ela deve receber maior atenção por parte dos pesquisadores no desenho de pesquisas que visam interpretação de relação causal? De forma objetiva, a Endogeneidade ocorre quando a variável explicativa de um modelo de regressão possui correlação com o termo de erro, sendo essa correlação decorrente de omissão de variáveis relevantes, causalidade reversa e/ou erro de medida no regressor. Representa, portanto, a violação da hipótese de Exogeneidade, essencial para a derivação de estimadores não viesados pelo método dos Mínimos Quadrados Ordinários (MQO). É muito importante que o pesquisador esteja alerta a essa questão para a interpretação da estimativa do parâmetro de interesse, uma vez que a existência de viés no estimador compromete a inferência causal entre o regressor e a variável dependente do modelo econométrico.

Para melhor ilustrar o problema causado pela Endogeneidade, trazemos um exemplo didático. Por questões de brevidade, focaremos a discussão nos dois primeiros tipos de Endogeneidade mencionados. Suponha que se esteja interessado em medir o efeito de pertencer à carteira do Índice de Sustentabilidade Empresarial (ISE) da B3 sobre a *performance* de companhias de capital aberto. Denotemos a medida de *performance* da empresa  $i$  como  $Y_{ip}$ , a variável binária  $D_i$  como um indicador de aderência (= 1) ou não (= 0) da empresa ao ISE e, por fim, um vetor que contém “ $k$ ” características observáveis  $X_i$  ( $X_{1p}$ ,  $X_{2p}$ , ...,  $X_{kp}$ ). Inicialmente, partimos de um contexto em que os dados são observados por meio de um corte transversal. O modelo empírico pode ser expresso como:

$$Y_i = \beta_0 + \beta_1 D_i + X_i' \gamma + \varepsilon_i, \quad (1)$$

onde  $\varepsilon_i$  representa o termo de erro da regressão, e o parâmetro de interesse é denotado por  $\beta_1$ , que deveria medir o “efeito” (“impacto” ou “influência”) do ISE sobre a *performance* empresarial. Contudo, não há razão alguma, *a priori*, para se acreditar que, de fato, “ $\beta_1$ ” representa o efeito médio de interesse. Isso porque, sem compreender os fatores que determinam o pertencimento da empresa ao ISE, nada garante que esse parâmetro será estimado sem viés. Nesse sentido, é possível destacar alguns pontos em relação à Equação 1.

Primeiro, as  $k$  variáveis contidas no vetor  $X_i$  em nada contribuirão para minimizar o viés na estimação de  $\beta_1$  se não forem fatores **predeterminados em relação à adesão ao ISE**. Isso ocorre porque as características desse vetor podem refletir um comportamento das empresas em resposta à adesão ao ISE<sup>1</sup> e, inclusive, exercer o papel de variáveis dependentes no modelo. Angrist e Pischke (2009) ressaltam esse problema e denominam essas variáveis como “*bad controls*”. Logo, as variáveis em  $X_i$  podem não ser boas candidatas a variáveis de controle. Segundo, pelo simples fato de o modelo econométrico representar uma simplificação da realidade e refletir o desconhecimento do pesquisador acerca de vários fenômenos em seu termo estocástico  $\varepsilon_i$ , é muito plausível a existência de outras características **não observáveis** que sejam relevantes e correlacionadas com a adesão ao ISE (ou fatores de confusão). Consequentemente, a omissão de variáveis que determinam tanto a *performance* da empresa quanto a probabilidade de aderir ao ISE implicará o viés de  $\beta_1$ . Isso implica que a variação na *performance* das empresas pode, parcial ou totalmente, ser causada por um fator não observado pelo pesquisador e, de forma falaciosa, ser atribuída à adesão da empresa ao ISE. Os dois pontos discutidos refletem bem o problema de Endogeneidade. Vale ainda ressaltar que a magnitude do viés pode ser tão significativa, de modo que venha a afetar a direção da estimativa, comprometendo, assim, tanto a interpretação de “impacto” quanto o “sinal” da correlação.

1 Por exemplo, considerando os verdadeiros critérios de elegibilidade à carteira ISE, é relativamente simples de argumentar que aderir ao ISE é endógeno ao tamanho da empresa, uma vez que empresas maiores tendem tanto a se autosselecionar para participar da carteira quanto a apresentar melhores níveis de *performance*. Ademais, o tamanho das empresas participantes pode ser contemporaneamente impactado como um reflexo da adesão ao ISE. Logo, incluir o logaritmo natural do ativo total (tradicional *proxy* para tamanho da empresa) no modelo empírico como uma variável de controle pode ser bastante problemático. Ahern e Dittmar (2012) trazem uma discussão similar, ao estimarem o efeito de mudanças da composição do Conselho de Administração sobre o *valuation* de corporações. Larcker et al. (2007) promovem outro interessante exemplo, ao discutirem a endogeneidade no uso da variável *leverage* em sua especificação econométrica, em contexto de análise de associação entre governança corporativa e *performance* organizacional.

Outra fonte de viés para a estimação de  $\beta_1$  é a causalidade reversa imbuída na relação de interesse. Em nosso contexto, ambos os argumentos de que “a adesão ao ISE pode afetar a *performance* da empresa” quanto “o nível de *performance* pode determinar uma empresa aderir ao ISE” fazem sentido. Neste ponto, há de se concordar que a questão cronológica é crucial para se entender a relação de causalidade, uma vez que a causa precede o efeito. No entanto, destacamos que explorar informação longitudinal dos dados não é suficiente para contornar o problema de causalidade reversa — que, sob uma perspectiva alternativa, pode ser encarado como um problema de variável omitida — e obter a desejada interpretação causal.

Para motivar essa questão, vamos incluir dimensão temporal nas variáveis anteriormente declaradas do nosso exemplo, supondo periodicidade anual ( $t = \{0, 1, 2, \dots, T\}$ ). Ao se considerarem dados longitudinais, a especificação econométrica de partida mais conveniente para estimação da relação de interesse seria um modelo de efeitos fixos<sup>2</sup>:

$$Y_{it} = \beta_0 + \beta_1 D_{it} + X'_{it} \gamma + \alpha_i + \alpha_t + u_{it}, \quad (2)$$

onde  $u_{it}$  representa o erro da regressão, e os termos  $\alpha_i$  e  $\alpha_t$  são, respectivamente, efeitos fixos de empresa e ano. A inclusão do primeiro efeito fixo é essencial para controlar por fatores idiossincráticos das empresas que são invariantes no tempo e não podem ser observados pelo pesquisador, por exemplo, o ano de constituição e o setor de atividade. O efeito fixo de ano absorve os choques temporais distribuídos de forma homogênea entre as empresas (como inflação, sazonalidades, taxa de juros, choques macroeconômicos, entre vários outros). Recorrer à estrutura longitudinal dos dados não resolve o problema de causalidade reversa, porque é plausível existirem tendências temporais na variável dependente que precedem a participação da empresa no ISE. Ou seja, de maneira dinâmica, o nível de *performance* pode ser responsável pela escolha da empresa em participar da carteira ISE. Por outro lado, faz-se necessário notar que, de fato, o contrário também é uma situação suscetível empiricamente. A questão central é que, nesse contexto, o pesquisador não consegue observar nos dados nem a primeira situação nem a segunda, permitindo que exista uma relação de retroalimentação entre as variáveis.

Ademais, mesmo ao se empregar uma modelagem econométrica mais robusta, como na Equação 2, o problema de identificação do efeito causal permanece, devido aos mesmos problemas anteriormente discutidos. Cabe destacar que o fato de a adesão ao ISE configurar uma decisão da empresa, outros fatores não observáveis variantes no tempo podem estar associados à escolha da empresa em participar da carteira e afetam diretamente sua *performance*. Por exemplo, a empresa pode decidir participar da carteira ao ser motivada pelo comportamento de adesão de seus pares de mesmo setor, o que pode refletir em sua *performance*. Todos os pontos discutidos nesta seção a respeito da dificuldade em se estimar o efeito causal traduzem o fenômeno conhecido como **seleção em não observáveis**, ou seja, viés de seleção ocasionado pela existência de fatores que não podem ser controlados pelo pesquisador. Em suma, enquanto não se resolver o comportamento de autoseleção das empresas à carteira ISE, é impossível inferir causalidade entre *performance* e ISE.

2 É oportuno mencionar que, uma vez que a decisão de participar da carteira ISE parte das empresas, não há motivos para se acreditar que um modelo de efeitos aleatórios (ou mesmo um *pooled* MQO) é empregável neste contexto. Isso acontece porque a heterogeneidade não observada do indivíduo, “”, não é independente da variável , o que reflete, então, a endogeneidade dessa relação. Além disso, o teste de Hausman será pouco informativo quanto à decisão de usar um modelo em detrimento do outro, devido à sua sensibilidade quanto à inclusão de regressores. De fato, Wooldridge (2015) observa que “if we think the unobserved effect  $\alpha_i$  is correlated with any explanatory variables, we should use first differencing or fixed effects”.

A abordagem até aqui é bastante provocativa e um tanto quanto desanimadora em relação às consequências que a Endogeneidade traz para a interpretação causal. Então, torna-se pertinente o questionamento: há maneiras de se contornar esse problema? Felizmente, sim. Contudo, salienta-se que as soluções não são tão triviais e nem sempre factíveis de se implementar. Nesse sentido, a percepção das oportunidades de implementação exige bastante perspicácia do pesquisador sobre o contexto institucional do que se investiga. Antes de apresentarmos as soluções, é conveniente definirmos alguns termos-chave.

Consideremos uma situação **factual** em que um grupo de indivíduos recebe uma intervenção de qualquer natureza (**tratamento**) e uma situação **contrafactual** em que esse mesmo grupo não recebe a intervenção. O propósito é estimar o efeito médio de tratamento sobre um determinado **indicador de impacto** (variável dependente), mensurado após a intervenção. O impacto médio de interesse poderia ser, então, obtido pela diferença das médias do indicador de impacto entre os mundos factual e contrafactual. Porém, note que, enquanto a primeira situação é tangível, o contrafactual não o é (talvez, em um universo paralelo, seja). Essa impossibilidade de observar o mundo contrafactual cria o que se conhece como Problema Fundamental da Inferência Causal<sup>3</sup>. O objetivo do pesquisador é, então, encontrar um grupo de indivíduos no mundo real que mimetize a situação contrafactual do grupo de tratados (denominado **grupo de controle**). Em outras palavras, é necessário identificar indivíduos que sejam comparáveis àqueles que foram selecionados para receber o tratamento. Uma mensagem importante é que nem todas as unidades observacionais que compõem o grupo de não tratados configuram um bom grupo de controle na análise de impacto.

O “padrão-ouro” para a inferência causal é a utilização de **experimentos**, que consistem em intervenções realizadas por pesquisadores ou *polycymakers*, nas quais se manipula a condição de tratamento de maneira aleatória.<sup>4</sup> A randomização dessa condição torna a atribuição do tratamento não correlacionada com as características observáveis e não observáveis dos indivíduos, já que todos têm igual probabilidade de serem selecionados, independentemente de seus traços particulares. Então, a aleatorização cria um grupo de tratamento e um grupo de controle que permite extrair uma relação causal entre a intervenção e um indicador de impacto. O uso de experimentos na área de Ciências Sociais Aplicadas não é tão frequente, devido a questões que envolvem desde conflitos éticos até restrições de ordem financeira e técnica. No nosso exemplo hipotético sobre o ISE, o sorteio de empresas listadas na Bolsa de Valores do Brasil para participação na carteira ISE configuraria um experimento nos moldes aqui discutidos. Porém, na prática, é difícil de se imaginar que isso possa ocorrer. A boa notícia é que o experimento não é o único meio que permite ao pesquisador examinar relações de causalidade.

Na ausência de um desenho experimental, o caminho para a inferência causal é explorar um **quase-experimento** — ou um **experimento natural** —, que constitui um desenho em que a atribuição do tratamento é “tão boa quanto se fosse aleatória”. Para identificar o efeito causal de interesse, é necessário encontrar uma “variação exógena” sobre a variável de tratamento, isto é, uma fonte de variação que contorne questões de autoseleção dos indivíduos ao tratamento. A identificação desse choque exógeno por parte do pesquisador só é possível ao se ter profundo conhecimento acerca do contexto institucional atrelado ao fenômeno estudado. Geralmente, a resposta reside nos critérios de elegibilidade ligados à intervenção. Outras vezes, a fonte de exogeneidade não é tão evidente, situando-se “nas entrelinhas” do contexto. Após identificar um experimento natural, será necessário implementar uma estratégia empírica para realizar inferência causal, ponto que discutimos a seguir.

3 Angrist e Pischke (2009) apresentam de maneira simples e intuitiva o Modelo Causal de Rubin, que aborda esse problema ao derivar o efeito médio de tratamento de uma intervenção. O modelo utiliza a linguagem de *outcomes* potenciais, onde é possível decompor a diferença média dos *outcomes* observados entre os grupos de tratados não tratados em duas partes: o efeito de tratamento e o viés de seleção.

4 Para evitar confusões de terminologia, destaca-se que “atribuição aleatória do tratamento” é algo diferente de “amostragem aleatória”. Enquanto o primeiro determina de forma imparcial os indivíduos a receberem uma dada intervenção, o segundo consiste em selecionar de maneira casual uma porção de elementos que pertencem a um público-alvo. Portanto, obter uma amostra aleatória de uma população não irá resolver problemas de Endogeneidade.

### 3. Estratégias de Identificação

A estratégia de identificação consiste na abordagem metodológica que visa contornar o problema de Endogeneidade e recuperar a interpretação causal da relação de interesse. Os métodos quase-experimentais têm recebido significativas contribuições teóricas ao longo das últimas décadas e são largamente valorizados nas análises empíricas devido ao grau de confiabilidade dos resultados em decorrência do rigor metodológico. Nesta seção, apresentamos os principais métodos usados para identificação de efeitos causais, a fim de discutir resumidamente as principais hipóteses para a identificação e os contextos de aplicação. Nosso objetivo não é desenvolver, com rigor estatístico, cada método quase-experimental, mas, sim, abordar, de forma introdutória e intuitiva, as técnicas.

#### 3.1 Diferença-em-Diferenças

Seguramente, a estratégia de identificação mais utilizada no *mainstream* é a Diferença-em-Diferenças (DD). A aplicação do método está condicionada à disponibilidade de dados longitudinais, para ambos os grupos de tratamento e controle, em pelo menos dois momentos no tempo: um antes de ocorrer o tratamento, e outro, depois. Especificamente, é preciso observar a situação de pré-intervenção para ambos os grupos. O efeito médio de interesse é estimado pela subtração entre as diferenças das médias da variável dependente dos grupos de tratamento e controle, antes e após a intervenção. Essa dupla diferença é responsável por atribuir o nome ao método. Para que o estimador de DD identifique o efeito causal, uma condição precisa ser satisfeita: a de **tendências paralelas**<sup>5</sup>. Essa hipótese de identificação diz que, na ausência da intervenção, as trajetórias do indicador de impacto dos grupos de tratamento e controle teriam dinâmica similar. É importante notar que a condição **não** requer que, antes da intervenção, “as médias do indicador de impacto sejam iguais entre os grupos” ou que “os fatores predeterminados apresentem médias iguais entre os grupos”.

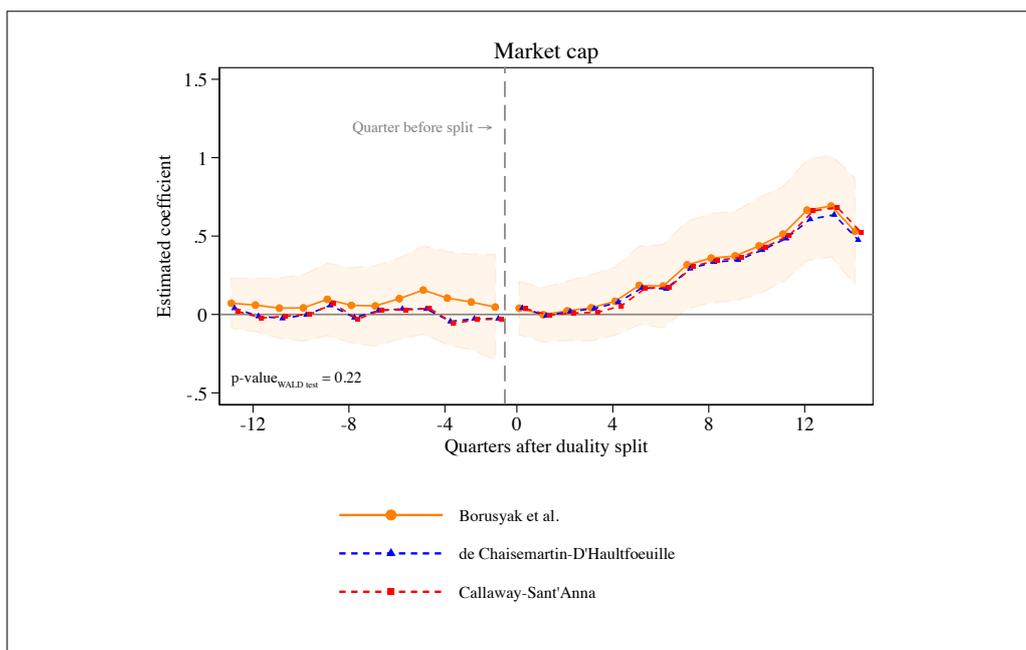
Como não é possível observar a situação contrafactual do grupo de tratamento (ou seja, os indivíduos tratados não recebendo o tratamento), torna-se fácil perceber que essa hipótese de identificação não é diretamente testável. Então, realizar a sequência de diferenças explicadas anteriormente não garante a identificação da relação causal. Como ilustração, aplica-se novamente o exemplo do ISE, considerando que se observa as empresas participantes e não participantes em dois períodos no tempo, antes e após a criação da carteira. A *performance* organizacional poderia se comportar de maneira diferente entre as empresas tratadas antes mesmo de aderirem ao ISE, o que indica a presença de diferenças preexistentes nas trajetórias de *performance* em comparação com as empresas fora da carteira e, portanto, impossibilita a inferência causal.

Nesse sentido, é muito importante que o pesquisador realize testes empíricos que corroborem, de maneira indireta, a sustentação da hipótese, como a verificação de comportamento antecipatório. Por isso, a estratégia de DD é comumente explorada no contexto em que se observam múltiplos momentos no tempo. Isso porque, havendo disponibilidade de dados em vários momentos antes e depois do tratamento, é possível checar se há evidências de pré-tendências na variável dependente. O desenho de DD que explora essa dimensão temporal ampliada num modelo semiparamétrico em relação à variável de tratamento é especificamente conhecido como *event-study*, ou estudo de eventos<sup>6</sup>. Além da utilização de tabelas, o uso de gráficos na descrição das trajetórias do indicador e na apresentação dos resultados é muito bem-vindo e bastante utilizado nos artigos relacionados.

5 Há ainda a hipótese de ausência de efeitos antecipatórios, comumente confundida com a de tendências paralelas devido à estreita ligação entre elas. Por questões didáticas, não nos referimos explicitamente a essa hipótese. Roth et al. (2023) discutem detalhadamente ambas as hipóteses, bem como os recentes avanços da literatura de DD e as perspectivas futuras do método.

6 Destaca-se que este desenho é diferente daquele classicamente aplicado na literatura de finanças.

Outro elemento-chave é a possibilidade de diferenças no momento da adesão ao tratamento entre os indivíduos tratados. Consequentemente, pode-se pensar em basicamente dois casos distintos. O primeiro é o caso canônico, em que todas as unidades tratadas recebem a intervenção ao mesmo tempo. O complementar é caracterizado por uma adoção dos indivíduos ao tratamento que ocorre de forma progressiva no tempo, conhecido como *staggered* DD. O segundo caso merece uma maior atenção, uma vez que é bastante frequente nos desenhos que utilizam DD (Baker et al., 2022) e pode estar sujeito à presença de viés na estimação do efeito de tratamento. Isso ocorre porque a heterogeneidade na entrada ao tratamento — e a possível heterogeneidade do efeito de tratamento — pode distorcer a ponderação utilizada no estimador, introduzindo pesos negativos no cálculo da média do efeito de tratamento. Como consequência, o sinal da correlação pode ser erroneamente invertido. Para contornar essa dificuldade, há uma série de novos estimadores de DD que corrigem os pesos e geram estimativas confiáveis do efeito causal. A Figura 1 ilustra um dos resultados de Guimarães e Trevisan (2022), em que os autores utilizam uma regulação promulgada pela Bolsa de Valores do Brasil para estimar o efeito da separação obrigatória dos cargos de CEO e *chairman* sobre uma medida de valor dos acionistas implementando três diferentes estimadores num *staggered event-study design*. A Figura 1 destaca a ausência de efeitos antecipatórios — estimativas estatisticamente insignificantes e muito próximas de zero, antes da separação — e indica um efeito positivo e progressivo sobre o indicador de impacto.



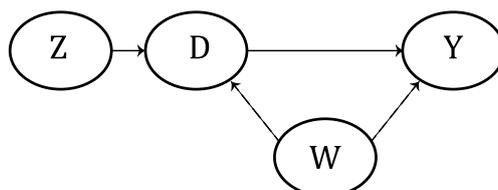
**Figura 1.** Exemplo de *staggered event-study design*.

Fonte: Guimarães e Trevisan (2022).

Convém destacar que, para a identificação do efeito causal no desenho de DD, deve-se recorrer a motivações exógenas aos indivíduos em relação à sua adoção não aleatória ao tratamento. Caso a entrada no tratamento envolva essencialmente a autoseleção das unidades observacionais, será difícil encontrar um bom grupo de controle e extrair causalidade da relação de interesse. Por isso, é interessante explorar regras, regulamentações e diretrizes de naturezas diversas, que configuram um componente alheio à escolha dos indivíduos e que gerem variação temporal e entre grupos. Para melhor compreensão da aplicação do método de DD na pesquisa em Contabilidade, sugerimos a leitura dos artigos de Huang et al. (2020), Chircop et al. (2023) e Lin et al. (2019), que trazem interessantes aplicações nos temas *managerial litigation risk*, *tax avoidance* e impactos da adoção do IFRS.

### 3.2 Variáveis Instrumentais

De todos os métodos quase-experimentais, a Variável Instrumental (IV, na sigla em inglês) pode ser considerada a mais complexa de se identificar situações para sua aplicação. Com o intuito de apresentar o método de maneira intuitiva, vamos recuperar as variáveis  $Y$  e  $D$  já definidas na Seção 2. Consideramos também a existência de outra variável,  $W$ , que não é observada pelo pesquisador. Trazemos, como motivação, o uso de um gráfico direcionado acíclico (DAG, na sigla em inglês) utilizado em todo livro-texto de apresentação de um IV canônico, como a seguir:



**Figura 2.** Gráfico direcionado acíclico.

Fonte: elaborado pelos autores.

A relação causal de interesse é expressa por  $D \rightarrow Y$ . Contudo, a existência do fator  $W$  introduz uma história de seleção em não observáveis, uma vez que ele confunde a relação de interesse, ao determinar tanto a variável dependente quanto a variável endógena de tratamento ( $D \leftarrow W \rightarrow Y$ ). Agora, perceba a existência do fator  $Z$  e como ele atua na cadeia de efeito causal expressa no DAG: uma variação em  $Z$  provoca uma variação em  $D$ , o que causa uma variação em  $Y$ . A percepção mais importante aqui é que, mesmo que  $Y$  varie quando  $Z$  varie, a variação de  $Y$  **só acontece devido à variação em  $D$** . Em outras palavras, a variável  $Z$  afeta  $Y$  **exclusivamente por intermédio de  $D$** . Esse caminho denomina-se **restrição de exclusão**. O fator  $Z$  é conhecido como **variável instrumental** — ou apenas **instrumento** —, termo que dá nome à estratégia de identificação. Então, um bom instrumento é aquele que atende à hipótese de restrição de exclusão (ou seja,  $Z \rightarrow D \rightarrow Y$ ) e é independente de outros fatores de confusão. Nesta cadeia causal, você pode pensar o fator  $D$  assumindo tanto a função de variável mediadora, no caminho  $Z \rightarrow D \rightarrow Y$ , quanto a de um colisor, no sentido  $Z \rightarrow D \leftarrow W$ .

Mas por que, na prática, é tão difícil encontrar bons instrumentos? Porque eles demandam uma consistente fundamentação teórica e lógica para argumentar a validade da restrição de exclusão. Essa, por si só, já é uma boa razão para os pesquisadores evitarem identificar relação causal com o uso de IV, já que a dificuldade aumenta de forma proporcional ao número de instrumentos. Assim como no caso das tendências paralelas no DD, a hipótese de restrição de exclusão não é diretamente testável. Portanto, a defesa da validade de um instrumento requer a aplicação de vários testes de robustez dos resultados e falsificação da estratégia empírica. Por exemplo, é essencial demonstrar que o instrumento apresenta forte correlação — em termos de significância estatística — tanto com a variável de tratamento (ou **primeiro estágio**) quanto com a variável dependente, além de apresentar evidências de ausência de correlação com outras características relevantes. Porém, deve-se ter em mente que o primeiro estágio e a forma reduzida não são suficientes para configurar  $Z$  como um bom instrumento. O pesquisador precisará, também, de uma excelente história para ele.

Como bem destacado por Cunningham (2021), a defesa de um bom instrumento geralmente provoca estranheza ao leitor em relação à argumentação de sua relação com a variável dependente. Tomemos como exemplo o trabalho de Bennedsen et al. (2007), que estimam o efeito da nomeação de CEO familiar ou externo sobre a *performance* da corporação. Claramente, a decisão de manter (ou não) um familiar na gestão da empresa é endógena — ocasionada por problemas anteriores de *performance*, decisões de fusão e aquisição, conexões políticas, entre diversos outros fatores difíceis de se observar. Os autores propõem utilizar, como variação exógena à sucessão familiar ao cargo, o “gênero do primeiro filho do CEO sujeito à sucessão do cargo”. Você deve estar se perguntando: como o gênero do primeiro filho pode “influenciar” a *performance* de uma empresa? A elegância da estratégia está intrinsecamente ligada ao contexto dessa excêntrica e peculiar forma reduzida. Os autores exploram dados únicos de árvores genealógicas de membros de corporações e argumentam/evidenciam que o gênero do primeiro filho é um bom instrumento, porque afeta a probabilidade de incorrer em nomeação de membro familiar **exclusivamente** por questões de primogenitura (primogênitos homens têm maior chance de “herdar” o cargo). Boa parte do artigo é dedicada à defesa da validade dessa restrição de exclusão. Por outro lado, é muito improvável que esse traço familiar esteja correlacionado com outros determinantes da *performance* da empresa, tal como demonstrado pelos autores mediante a implementação de vários testes de robustez. Esse atributo de quase-aleatoriedade do instrumento é o que permite identificar a relação causal de interesse.

As aplicações de IV são viáveis tanto para dados de corte transversal quanto para dados em painel. Além disso, é também muito comum que as pesquisas utilizem tanto tratamentos “contínuos” quanto “binários”. Larcker e Rusticus (2010) realizam uma revisão de literatura acerca do uso do método em pesquisas na área contábil, bem como discutem a não trivialidade de sua implementação e promovem um guia prático para o uso. Alertamos que, qualquer que seja a situação em que o pesquisador acredite na possibilidade de utilização da estratégia de IV, deve-se ter meios (empíricos e argumentativos) de convencer o leitor que o instrumento atende à restrição de exclusão, uma vez que o emprego de maus instrumentos pode ocasionar viés no estimador de forma ainda mais drástica do que utilizar um simples MQO. Alguns interessantes exemplos de artigos que utilizam IV na área de Contabilidade são Fang et al. (2015) e Tseng (2022), que tratam do papel dos investidores estrangeiros sobre práticas de reporte de informações financeiras e dos efeitos de transbordamento de tecnologias em inovação.

### 3.3. Regressão Descontínua

As aplicações com o desenho de Regressão Descontínua (RD) vêm crescendo na área de Ciências Sociais Aplicadas devido à praticidade de implementação e ao grande poder de convencimento quanto à validade interna dos resultados. O método de RD explora uma mudança repentina — de origem exógena — na probabilidade de receber um tratamento, que se dá a partir de um valor específico (ponto de corte, ou *cutoff*) de uma variável quantitativa (*running variable*). Cria-se, numa vizinhança ao redor e próxima desse *cutoff*, um quase-experimento, de forma que as unidades observacionais que se encontram dentro dessa vizinhança podem ser utilizadas como grupos de tratamento e controle, devido à sua similaridade em termos de características predeterminadas.

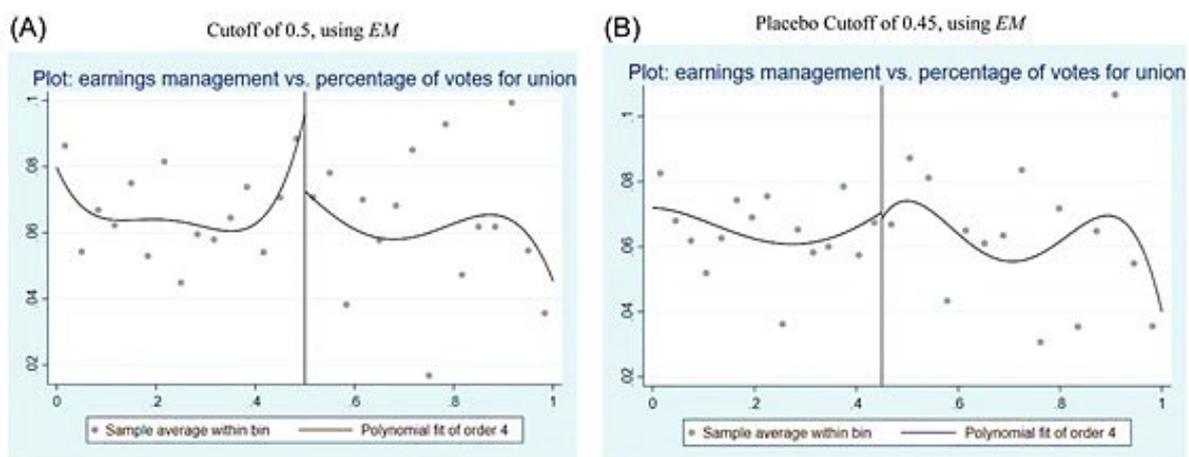
Para ilustrar, vamos supor que a autoridade fiscal implemente uma política de fiscalização tributária que institui auditorias em empresas que atingem um valor específico de receita líquida anual, referente ao exercício social anterior ao ano da política. Nesse caso, a receita líquida anual configura a variável quantitativa (*running variable*), e o ponto de corte (*cutoff*) seria o limite de receita que uma empresa deve alcançar para se qualificar para o processo de fiscalização. Utilizando um desenho de RD, seria possível comparar as empresas próximas a esse limite. As empresas que estivessem ligeiramente acima do limite seriam consideradas o grupo de tratamento, pois seriam elegíveis para a política de fiscalização tributária, enquanto as empresas ligeiramente abaixo do limite seriam consideradas o grupo de controle, pois não seriam selecionadas para a auditoria.

Portanto, em decorrência de uma **descontinuidade** na chance de se receber um tratamento, pode-se estimar o efeito médio de tratamento “local”, que se caracteriza como uma **descontinuidade** na média do indicador de impacto exatamente no *cutoff*. A hipótese de identificação é que a média da variável dependente para o grupo contrafactual possui uma transição suave no *cutoff*. Em outras palavras, na ausência do tratamento, não deveria haver descontinuidade na média do indicador de impacto no *cutoff*.

Não surpreendentemente, essa hipótese não é diretamente testável, porque representa uma situação impossível de se observar. Contudo, existem testes empíricos úteis para corroborar a sustentação da hipótese e que são procedimentos-padrão em todo trabalho que utiliza essa estratégia de identificação. O primeiro deles é o teste de balanceamento, que consiste em checar a semelhança entre os grupos de tratamento e controle a partir de suas características observáveis. Se a intervenção é realmente “tão boa quanto se fosse aleatória”, os grupos devem ser estatisticamente iguais em termos das médias das características mensuradas antes do choque exógeno. Outro teste importante é checar se há manipulação da *running variable*. A intuição é que os indivíduos poderiam se autoselecionar para receber (ou evitar) o tratamento, de forma a manipular os valores da *running variable* para (des)cumprir os critérios de elegibilidade. Para ilustrar esse ponto, considere o exemplo anterior da política de fiscalização tributária. Suponha que, alternativamente, a autoridade fiscal divulgue que as empresas que atingirem um valor mínimo de receita líquida ao fim do exercício social estarão sujeitas à auditoria fiscal. Nesse caso, é provável que as empresas se antecipem e adotem estratégias para evitar ultrapassar o *cutoff* estabelecido pela política, a fim de não serem selecionadas para a auditoria. Caso exista manipulação, a frequência das observações dos grupos de tratamento e controle no *cutoff* apresentará “um salto”, refletindo, assim, um comportamento endógeno dos indivíduos e prejudicando a inferência causal.

O caso em que todos os indivíduos recebem o tratamento ao atenderem o critério de elegibilidade — a probabilidade de receber o tratamento salta de 0% para 100% no *cutoff* — é chamado de desenho *sharp*. Esse é o caso do exemplo apresentado acima, da política de fiscalização tributária. No entanto, é possível que empiricamente nem todos os indivíduos que atendem ao critério de elegibilidade recebam, de fato, a intervenção. Ou seja, pode haver situações em que exista endogeneidade em relação à adoção ao tratamento. Esses casos configuram o desenho *fuzzy*. Para exemplificar esse desenho de RD, vamos continuar utilizando o caso hipotético anterior. Suponha que, devido a restrições de recursos, a autoridade fiscal não consiga auditar todas as empresas elegíveis (ou seja, aquelas que ultrapassaram o valor mínimo de receita líquida). Este caso configura um desenho *fuzzy*, uma vez que, entre as empresas elegíveis, a probabilidade de serem tratadas é inferior a 100%. Nesses casos, a identificação do efeito de tratamento sobre os tratados utiliza a “oportunidade” de receber o tratamento como um IV para determinar a probabilidade de efetivamente receber o tratamento. Vale destacar que a forma reduzida — relação entre o indicador de impacto e a oportunidade de receber o tratamento — representa o que se chama de “efeito de intenção do tratamento”, informação muito útil para quem é responsável por desenhar políticas públicas ou regulações. A estimação do parâmetro de interesse pode ser feita tanto por técnicas não paramétricas, utilizando “bandas ótimas”, quanto paramétricas (modelos de MQO). Cunningham (2021) discute com propriedade os métodos de estimação e testes de validação usados contemporaneamente.

Para ilustrar o uso da técnica de RD na pesquisa em Contabilidade, tomemos o estudo realizado por Lin et al. (2022), que estimam o impacto da sindicalização (*unionization*) no gerenciamento de resultados no contexto norte-americano. A motivação do trabalho está relacionada à possibilidade de que funcionários de empresas sindicalizadas demandem salários mais altos quando as empresas têm bons resultados financeiros. Portanto, o estudo investiga se a sindicalização afeta a prática de gerenciamento de resultados, com o intuito de reduzir os lucros e, assim, mitigar o impacto de maiores despesas com remuneração. Para alcançar esse objetivo, os autores exploram um quase-experimento originado nas eleições sindicais, em que a sindicalização é instituída se a empresa obtiver pelo menos 50% mais “1” dos votos. Logo, para as empresas em que as votações foram acirradas, o evento de sindicalizar ou não é quase-aleatório. Os autores comparam empresas que tiveram êxito na sindicalização por uma margem próxima do limite de 50% com aquelas que não obtiveram sucesso também por uma margem próxima desse limite (desenho *sharp*). O Painel A da Figura 3 apresenta graficamente um dos resultados do artigo. Pode-se perceber que a curva que representa o comportamento médio do gerenciamento de resultados apresenta uma descontinuidade exatamente no *cutoff*, indicando que, após a sindicalização, há uma redução dos *accruals* discricionários. Esse resultado sinaliza que as empresas utilizam da discricionariedade gerencial para reportar lucros menores, de forma a tentar se blindar de demandas dos trabalhadores por maiores salários. O Painel B realiza um teste placebo, simulando que o verdadeiro *cutoff* seria no ponto “0,45”, em vez de “0,50”. O gráfico sugere que não há descontinuidade nesse ponto, validando, assim, a estratégia de identificação do efeito causal da sindicalização.



**Figura 3.** Exemplo de staggered event-study design. .

Fonte: adaptado de Lin et al. (2022).

Em suma, para que o método de RD seja aplicável, é necessário haver uma variação exógena e brusca na chance de se receber um tratamento a partir de um valor específico de um índice quantitativo. Uma grande vantagem do método é o fato de não depender exclusivamente de dados longitudinais para ser implementado. Recomendamos a leitura dos trabalhos de Joshi (2020) e Fan et al. (2021), que utilizam a estratégia de RD para avaliar os efeitos de regulação relacionada ao reporte de informação tributária sobre o *tax avoidance* de empresas e os impactos de propostas relacionadas à governança corporativa sobre o gerenciamento de resultados.

#### 4. Considerações Finais

Um dos aspectos mais admiráveis do pensamento científico é a sua premissa de se permitir quebrar paradigmas e se renovar continuamente. Como pesquisadores, precisamos estar receptivos “ao novo” para continuarmos aprendendo e contribuindo com o avanço do conhecimento sobre os fenômenos nos quais estamos inseridos. Neste editorial, procuramos destacar uma questão de ordem epistemológica no contexto de pesquisas empíricas quantitativas na área contábil que buscam inferir relação de causalidade. Nesse sentido, gostaríamos de incitar algumas reflexões: usamos as ferramentas metodológicas apropriadas para procurar responder às nossas perguntas de pesquisa? Estamos devidamente atualizados e apropriados acerca dos métodos disponíveis? Em nosso ponto de vista, acreditamos que esse conhecimento ainda não está totalmente difundido e consolidado, havendo ainda um caminho a ser trilhado.

O uso dos métodos quantitativos para inferência causal, especialmente no âmbito nacional, precisa ser repensado: é necessária uma quebra estrutural no *modus operandi* das pesquisas em Contabilidade. No *mainstream* (quantitativo) contábil, já se percebe uma mudança de padrão entre os principais *journals* em relação à demanda por trabalhos que implementam métodos mais rigorosos para identificar relações de causalidade, explorando quase-experimentos e desenhos experimentais controlados. Para haver convergência e interação com esse *mainstream*, precisamos direcionar nossa atenção para a apropriação desses métodos. Esperamos, portanto, que este editorial possa esclarecer algumas questões introdutórias em relação às técnicas econométricas quase-experimentais e instigar a comunidade acadêmica quanto à sua “redescoberta” e potencial de utilização.

## Referências

- Ahern, K. R. & Dittmar, A. K. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation. *The Quarterly Journal of Economics*, 127(1), 137–197. <https://doi.org/10.1093/qje/qjr049>.
- Angrist, J. D. & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Baker, A. C., Larcker, D. F., & Wang, C. C. Y. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2), 370–395. <https://doi.org/10.1016/j.jfineco.2022.01.004>.
- Ball, R. & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(2), 159–178. <https://doi.org/10.2307/2490232>.
- Beaver, W. (1968). Market prices, financial ratios, and the prediction of failure. *Journal of Accounting Research*, 6(2), 179–192. <https://doi.org/10.2307/2490233>.
- Bennedsen, M., Nielsen, K. M., Perez-Gonzalez, F., & Wolfenzon, D. (2007). Inside the family firm: The role of families in succession decisions and performance. *The Quarterly Journal of Economics*, 122(2), 647–691. <https://doi.org/10.1162/qjec.122.2.647>.
- Chircop, J., Fabrizi, M., Malaspina, P., & Parbonetti, A. (2023). Anti-mafia police actions, criminal firms, and peer firm tax avoidance. *Journal of Accounting Research*, 61(1), 243–277. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-679X.12455>.
- Cunningham, S. (2021). *Causal inference: The mixtape*. Yale university press.
- Fan, Z., Radhakrishnan, S., & Zhang, Y. (2021). Corporate governance and earnings management: Evidence from shareholder proposals. *Contemporary Accounting Research*, 38(2), 1434–1464. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1911-3846.12640>.
- Fang, V. W., Maffett, M., & Zhang, B. (2015). Foreign institutional ownership and the global convergence of financial reporting practices. *Journal of Accounting Research*, 53(3), 593–631. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-679X.12076>.
- Gow, I. D., Larcker, D. F., & Reiss, P. C. (2016). Causal inference in accounting research. *Journal of Accounting Research*, 54(2), 477–523. <https://doi.org/10.1111/1475-679X.12116>.
- Guimarães, R. & Trevisan, G. (2022). Mandatory CEO Non-Duality, Managerial Agency, and Shareholder Value. Working Paper, SSRN. <https://ssrn.com/abstract=4122646>.
- Huang, Y., Li, N., Yu, Y., & Zhou, X. (2020). The effect of managerial litigation risk on earnings warnings: Evidence from a natural experiment. *Journal of Accounting Research*, 58(5), 1161–1202. <https://doi.org/10.1111/1475-679X.12336>.
- Joshi, P. (2020). Does private country-by-country reporting deter tax avoidance and income shifting? evidence from beps action item 13. *Journal of Accounting Research*, 58(2), 333–381. <https://doi.org/https://doi.org/10.1111/1475-679X.12304>.
- Larcker, D. F., Richardson, S. A., & Tuna, I. (2007). Corporate governance, accounting outcomes, and organizational performance. *The Accounting Review*, 82(4), 963–1008. <https://doi.org/10.2308/accr.2007.82.4.963>.
- 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. <https://www.sciencedirect.com/science/article/pii/S0165410109000718>.

- Lin, G., Brown, A. B., & Wu, W. (2022). The effect of unionization on firms' earnings management and employee compensation: Evidence from a regression discontinuity design. *Journal of Business Finance Accounting*, (pp. 1–28.). <https://doi.org/10.1111/jbfa.12665>.
- Lin, S., Riccardi, W. N., Wang, C., Hopkins, P. E., & Kabureck, G. (2019). Relative effects of ifrs adoption and ifrs convergence on financial statement comparability. *Contemporary Accounting Research*, 36(2), 588–628. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1911-3846.12475>.
- 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. <https://doi.org/10.1016/j.jeconom.2023.03.008>.
- Tseng, K. (2022). Learning from the joneses: Technology spillover, innovation externality, and stock returns. *Journal of Accounting and Economics*, 73(2), 101478. <https://www.sciencedirect.com/science/article/pii/S0165410122000015>.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.