

# AQUILO QUE CONFUNDE AO EXPLICAR\*

## ANOTAÇÕES INTRODUTÓRIAS SOBRE ENDOGENEIDADE

# 6

Marcus Paulo Lisboa Barbosa<sup>†</sup>

Luiza Jardim<sup>‡</sup>

**Resumo:** A busca por inferências válidas é um dos maiores desafios das ciências sociais. O objetivo deste artigo é introduzir o conceito da endogeneidade e os problemas relacionados às explicações causais. Para tanto, apresenta-se a endogeneidade como a expressão da existência de pelo menos um desses três problemas: i) erros em variáveis; ii) variáveis omitidas; e iii) causalidade reversa. Além disso, discutem-se estratégias para contornar estes problemas e alguns exemplos aplicados.

**Palavras-chave:** Endogeneidade. Causalidade reversa. Erros em variáveis. Variáveis omitidas.

### What Is Confusing When Explaining: Introductory Notes On Endogeneity

**Abstract:** The search for valid inferences is one of the biggest challenges of the social sciences. The purpose of this article is to introduce the concept of endogeneity and the problems related to causal explanations. Therefore, endogeneity is presented as the expression of the existence of at least one of these three problems: i) errors in variables; ii) omitted variables, and, iii) reverse causality. In addition, strategies to overcome these problems and examples are discussed.

**Keywords:** Endogeneity. Reverse causality. Errors in variables. Omitted variables.

## 1 O REAL E SEUS RUÍDOS

Democracia gera desenvolvimento ou desenvolvimento leva à democracia? (PRZEWORSKI, 2007). Se locais com maior efetivo policial são simultaneamente os mais violentos, como determinar o impacto da polícia sobre índices de criminalidade? (LEVITT, 1997; 2002). Um nível mais alto de escolaridade leva a um maior nível de participação política ou ambos são afetados por variáveis

---

\* O título deste artigo faz referência aos versos “Eu tô te explicando / Prá te confundir / Eu tô te confundindo / Prá te esclarecer” da canção *Tô*, do álbum *Estudando o Samba* (1976), composta por Tom Zé e Elton Antônio de Medeiros.

<sup>†</sup> Graduado em Direito pela Universidade Federal de Ouro Preto (UFOP). Mestre em Ciência Política pela Universidade Federal de Minas Gerais (UFMG). Quando da escrita do artigo, foi bolsista da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Orcid: <https://orcid.org/0000-0001-5293-2975>

<sup>‡</sup> Administradora Pública pela Fundação Getúlio Vargas (FGV-EAESP), Mestre em Ciência Política pela Universidade Federal de Minas Gerais (UFMG) e doutoranda em Administração Pública e Governo pela FGV-EAESP.

Orcid: <https://orcid.org/0000-0002-5981-8275>

não consideradas? (KING; KEOHANE; VERBA, 1994). A relação entre o tipo de plano de saúde escolhido por um indivíduo e seus gastos com saúde poderia ser analisada sem que se avaliassem condições físicas prévias? (DOWD; TOWN, 2002). Todas essas perguntas preocupam-se com o problema da endogeneidade, e, neste artigo, que visa a introduzir o conceito da endogeneidade e possíveis abordagens, servem como exemplo e ilustração do problema.

Embora entre parte expressiva dos teóricos da Ciência Política (KING; KEOHANE; VERBA, 1994; PRZEWORSKI, 2007; FRANZESE, 2007) a endogeneidade seja tratada como sinônimo de causalidade reversa, abordagem que não incorre em erro, este artigo adotará conceito mais amplo do fenômeno, proposto pelo campo econométrico (ANTONAKIS *et al.*, 2014; BASCLE, 2008; WOOLDRIDGE, 2006). Tal abordagem, além da causalidade reversa, compreende também outras duas espécies endógenas, os erros em variáveis e as variáveis omitidas.

Entre fenômenos com distintos graus de sistematicidade e apreensão, as ciências sociais têm, entre seus objetivos, a busca por inferências causais (REZENDE, 2015). A fim de alcançar inferências metodologicamente válidas, o campo dispõe de ferramentas cada vez mais sofisticadas. Do impacto da obra de King, Keohane e Verba (1994) aos avanços de técnicas econométricas e qualitativas (MAHONEY, 2010), o repertório de caminhos para investigações sociais cresce em amplitude e exigências. Em suma, para saber se X causa Y, existem rotas diversas — nem sempre pavimentadas (REZENDE, 2017).

Embora a amplitude das ferramentas cresça, questões fundamentais permanecem. As ciências sociais se debruçam sobre um campo mais incerto e nuançado que as ciências naturais (ALMOND; GRECO, 1977). O mundo político, se não tem a incerteza de nuvem ou a ordem estrita de relógio<sup>1</sup>, trafega num ambiente de “controle plástico” (POPPER *apud* ALMOND; GRECO, 1977, tradução nossa), no qual “quase tudo causa quase tudo na sociedade, economia e política” e com frequência “variáveis dependentes e independentes de um estudo invertem papéis em outro” (FRANZESE, 2007, p. 62, tradução nossa).

---

<sup>1</sup> Referência à metáfora de Karl Popper, comentada por Almond e Greco (1977), sobre as extremidades de fenômenos nas ciências naturais, das mais ordenadas, relógios, para as mais anárquicas, nuvens.

Diante do complexo cenário para a produção de inferências causais válidas, investigaremos, neste artigo, um dos problemas mais comuns da pesquisa em Ciência Política (PRZEWORSKI, 2007), crescentemente abordado nas revistas acadêmicas (ULLAH; ZAEFARIAN; ULLAH, 2020): a endogeneidade. Nosso objetivo consiste em apresentar, introdutoriamente, o debate sobre endogeneidade, seus problemas e possibilidades de correção em nível pré-operacional e teórico. Para tanto, é preciso definir o conceito — frequentemente tomado apenas por uma de suas espécies, a causalidade reversa — e entender suas implicações, tarefas desenvolvidas na próxima seção. No passo seguinte, abordaremos algumas estratégias para contornar o problema — do bom uso da teoria às variáveis instrumentais. Em direção ao fim, trataremos de pesquisas que souberam lidar com a endogeneidade e refletiremos sobre como agir frente ao surgimento de problemas endógenos.

## 2 ENDOGENEIDADE: CONCEITO E PROBLEMAS

Não é possível saber se um indivíduo A submetido ao tratamento X reagiria de modo diferente de A, ele próprio, sem a submissão ao tratamento imputado. Nesse cenário, “a infeliz conclusão é de que, em geral, efeitos causais individuais não podem ser identificados — isto é, não podem ser expressados como uma função de um dado observado — por causa do dado faltante” (tradução nossa) (ROBINS; HERNÁN, 2010, p. 4). A impossibilidade física de observar a reação de um indivíduo, num mesmo espaço de tempo, com e sem tratamento, refere-se ao *problema fundamental da inferência causal* (HOLLAND, 1986, *apud* KING; KEOHANE; VERBA, 1994). A partir disso, “toda inferência causal é, no limite, incerta: a mera definição de causalidade implica o fato de que nós nunca poderemos ter certeza sobre um efeito causal” (SÁTYRO; REIS, 2014, p. 18). Amparados no pressuposto de incerteza, cientistas políticos buscam desenhos para entender o que implica o quê.

Nessa perspectiva, experimentos — controlados e naturais — e pesquisas com dados observacionais são alternativas sedimentadas. A diferença entre cada rota permite entender parte da endogeneidade. Enquanto experimentos dependem, por definição, de um elemento randômico e, conseqüentemente, exógeno para diferenciar categorias comparativas — grupo de tratamento e controle —, os dados observacionais, frequentemente, encontram-se “à mercê da história” (PRZEWORSKI, 2007, p. 167).

Em maior detalhe, a lógica experimental permite inferir, a partir do cotejo de dois grupos semelhantes, o impacto causado por uma variável exógena. Ao submeter um grupo (tratamento) à exposição de um fator de impacto e resguardar o outro grupo (controle), em tudo similar, do mesmo procedimento, o desenho metodológico permite a comparação de resultados (PRZEWORSKI, 2007; BASCLE, 2008; ANTONAKIS *et al.*, 2010). Além disso, conforme aponta Przeworski (2007, p. 150, tradução nossa), a “atribuição aleatória significa que o valor da variável causal é independente de todas as características de uma unidade”, circunstância que protege os resultados encontrados de vieses endógenos, em termos de causalidade reversa. Após o procedimento, a diferença entre os dois grupos indicará a força causal da variável explicativa.

No entanto, mesmo na lógica experimental, algumas nuances demandam atenção. A capacidade de controle sobre a pesquisa empreendida determina diferentes abordagens. *Experimentos controlados*, conforme Dunning (2012), reúnem três características: a) a resposta do grupo assinalado como tratamento é comparável com a do grupo de controle; b) a atribuição das características *tratamento* e *controle* é feita aleatoriamente e; c) a manipulação do tratamento pode ser controlada pela pesquisadora.

Diante da dificuldade de viabilizar agendas com tamanha complexidade metodológica, sejam por motivos éticos ou práticos, pesquisadores se valem de *experimentos naturais* — desenho que não permite a manipulação do tratamento, mas que também conta com um elemento randômico capaz de diferenciar grupos de *tratamento* e *controle*, logo compartilha os já mencionados itens *a* e *b* com os experimentos controlados, mas não alcança o requisito *c* (DUNNING, 2012)<sup>2</sup>. Nesse sentido, as duas modalidades de experimentos, desde que con-

---

<sup>2</sup> A fim de pontuar a diferença entre experimentos controlados e experimentos naturais, dois estudos podem ser ilustrativos. Em exemplo de experimento controlado, Bertrand e Mullainathan (2003) analisaram o impacto do preconceito racial sobre as oportunidades de trabalho nos Estados Unidos. Para tanto, selecionaram, a partir de registros cartorários e *surveys*, nomes que soavam racializados — ou seja, atribuíveis a pessoas negras (Lakisha e Jamal) e a pessoas brancas (Emily e Greg). No passo seguinte, enviaram currículos semelhantes para empresas que anunciavam vagas em jornais, com uma única diferença: os nomes dos postulantes. Nesse trabalho, que ilustra o controle de pesquisadores sobre o tratamento, Bertrand e Mullainathan concluíram que pessoas negras têm 50% a menos de chance de serem chamadas para entrevistas de empregos quando comparadas às pessoas brancas. Na outra chave, em exemplo de experimento natural, em que os autores não possuíam controle sobre o tratamento, Ferraz e Finan (2008) investigaram o impacto da corrupção nas perspectivas de reeleição de prefeitos investigados. Na pesquisa, os autores se valeram de um programa iniciado em 2003 pela Controladoria-Geral da União, o qual, por sorteio,

duzidas em suas melhores práticas, permitem diluir eventuais influências de aspectos alheios ao tratamento por, aleatoriamente, separar dois grupos iguais, igualmente afetados por quaisquer fatores — exceto o *tratamento*, que fornecerá o resultado de interesse (DUNNING, 2012; ROBINS; HERNÁN, 2019).

De modo distinto, os dados observacionais não permitem tamanho controle sobre a cadeia causal, pois não compartilham nem *b*, a atribuição aleatória dos grupos de tratamento e controle, nem *c*, a possibilidade de manipular o tratamento imputado (DUNNING, 2012). Em face da frequente impossibilidade de realizar experimentos, portanto, resta um espaço fluido, em que a extração de causalidades tem de conviver com riscos inferenciais de toda ordem. Tal circunstância se traduz no problema da identificação, descobrir “se as causas realmente são causas e se os efeitos são efetivamente gerados pelas causas supostas, diante de dados observacionais” (REZENDE, 2017, p. 35). Nesse sentido, a endogeneidade, “a onipresente questão da endogeneidade” (FRANZESE, 2007, p. 61, tradução nossa), aparece como uma das principais armadilhas. Na síntese de Przeworski (2007, p. 168, tradução nossa), “a dificuldade apresentada pela endogeneidade é distinguir entre os efeitos de causas e os efeitos de condições sob as quais estes operam”. Sendo assim, olhar para questões de endogeneidade em estudos causais é um passo fundamental para produzir inferências válidas.

Definida na econometria como *a relação entre o regressor e o termo de erro*, ocorrência que, por meio de vieses e problemas de direção causal, impede a produção de inferências precisas, a endogeneidade pode se apresentar de formas variadas (ULLAH; ZAEFARIAN; ULLAH, 2020). Em diálogo com a violação da exogeneidade, condição própria dos experimentos mencionada anteriormente, Guilhem Bascle (2008, p. 288, tradução nossa) aponta que “há três instâncias em que a condição de exogeneidade é violada (e, portanto, a endogeneidade está presente): erros nas variáveis, variáveis omitidas e causalidade simultânea” — cada uma delas detalhada nos subtópicos desta seção. Percebe-se, portanto, que problemas endógenos não se resumem a questões de

---

escolhia municípios que teriam o uso de verbas federais investigado. A partir desse elemento aleatório, alheio à qualquer manipulação, os autores compararam os índices de reeleição de municípios auditados antes e depois das eleições de 2004. Concluíram que prefeitos que incorreram em ilegalidades tiveram índices de reeleição menores do que aqueles que não cometeram malfeitos.

causalidade reversa, configurando ampla gama de problemas à eventual generalização dos resultados encontrados (ULLAH; ZAEFARIAN; ULLAH, 2020).

## 2.1 Erros em variáveis

Os erros em variáveis, também conhecidos como erros de medição/mensuração, apesar de não receberem tanta atenção quanto as outras espécies de endogeneidade (BASCLE, 2008), são corriqueiros nas ciências sociais, campo em que “as categorias ou medidas usadas são geralmente artefatos criados pelo investigador e não ‘dadas’ pela natureza” (KING; KEOHANE; VERBA, 1994, p. 152). Nem todos os erros de mensuração, porém, acarretam os mesmos problemas. Erros em variáveis dependentes não apresentam endogeneidade, porque são capturados pelo termo de erro de um modelo, ou seja, pela parte que o próprio modelo reconhece não explicar. Desse modo, resultam apenas em perda de eficiência. De modo diferente, medições equivocadas da variável independente implicam endogeneidade, vez que conectam o regressor (aferido erroneamente) ao termo de erro da equação, provocando problemas na estimação dos resultados (GUIDE JR.; KETOKIVI, 2015).

Frutos da dificuldade inerente dos objetos de estudo da área, que muitas vezes não podem ser adequadamente observados e mensurados (ANTONAKIS *et al.*, 2014), os erros de medição “surgem quando o verdadeiro valor do regressor,  $X$ , não é observado” (BASCLE, 2008, p. 289). Por trás desses erros, encontram-se circunstâncias diversas, como escalas impróprias, agregações equivocadas de itens em um índice e imprecisões em *surveys* (ZAEFERIAN *et al.*, 2017). Em síntese, os valores imprecisos das variáveis explicativas afetariam a qualidade dos resultados encontrados, inviabilizando a exata compreensão da causalidade e de sua respectiva intensidade.

Antonakis *et al.* (2014) exemplificam a dificuldade endógena trazida por erros de mensuração de variáveis. Em uma pesquisa que busque mensurar o impacto da inteligência de um líder na motivação de seus seguidores, a variável explicativa *inteligência do líder* não poderá ser adequadamente mensurada, pelo menos se considerada em termos puros, sem uma *proxy*. Consequentemente, um modelo que utilize a variável *inteligência do líder*, por se valer de um construto com erro de medição, produzirá resultados inadequados, porque endógenos (ANTONAKIS *et al.*, 2014). Ilustrando o exemplo de Antonakis

*et al.* (2014), dizemos que a inteligência observada ( $X$ ) é diferente da variável construída em plano teórico ( $X^*$ ) para constar no modelo preditivo, fator que implica endogeneidade. Em outras palavras,  $X$  (a inteligência observada) corresponde à variável inteligência construída ( $X^*$ ) mais o termo de erro, a diferença entre o valor desejado para a variável e o valor efetivamente encontrado<sup>3</sup>.

## 2.2 Variáveis omitidas

Outro problema relacionado à endogeneidade diz respeito às variáveis omitidas. Na tentativa de inferir causalidade, pesquisas mapeiam variáveis independentes (explicativas) que impactam em variáveis dependentes (a serem explicadas). No entanto, não é raro que variáveis não selecionadas para a confecção de um modelo acabem por prejudicar a capacidade preditiva. As variáveis não selecionadas, denominadas omitidas, representam um problema de endogeneidade ao causar um efeito tanto na variável independente quanto na dependente (BASCLE, 2008; DOWD; TOWN, 2002).

Em pesquisas observacionais, a pesquisadora precisa reunir argumentos e evidências de que não há variáveis omitidas interferindo nas relações causais apresentadas. Porém Cinelli e Hazlett (2020) alertam que, embora em alguns casos os testes de equilíbrio, que comparam as características de partida dos grupos analisados, mostrem resultados nulos, eles não podem ser tomados como garantias da inexistência dessas variáveis — dado que são, como o nome ilustra, omitidas. Conforme os autores afirmam, em referência a Pearl (2009), “fundamentalmente, a inferência causal sempre requer suposições que não são verificáveis a partir dos dados” (CINELLI; HAZLETT, 2020, p. 1) — desse modo, há risco considerável de que variáveis omitidas perturbem qualquer

---

<sup>3</sup> A título ilustrativo, considere-se  $Xi^*$  como a variável inteligência construída como proxy para inteligência efetivamente observada  $Xi$ . Uma vez que há incertezas sobre essa variável construída, ela carrega consigo um termo de erro próprio ( $\mu i$ ). Assim,  $Xi = Xi^* + \mu i$ . Em um modelo de regressão tradicional, a notação seria:  $yi = \alpha + \beta Xi^* + \varepsilon i$ . Logo, substituindo os valores:  $yi = \alpha + \beta(Xi - \mu i) + \varepsilon i \Rightarrow yi = \alpha + \beta Xi - \beta \mu i + \varepsilon i$  e reagrupando os termos de erro em uma nova variável,  $vi = \varepsilon i - \beta \mu i$ , tem-se que:  $yi = \alpha + \beta Xi + vi$ . Como a variável original  $Xi$  depende do seu termo de erro ( $\mu i$ ) e o próprio termo de erro da regressão modificada depende desse mesmo erro, tem-se um problema de viés na estimação OLS padrão (para maior detalhamento dessa construção, ver Antonakis et al., 2014).

afirmação feita a partir de dados observacionais, caso não sejam adotados os cuidados devidos.

Dowd e Town (2002), ao tratarem da conexão entre tipos de planos de saúde e despesas de saúde (*managed care* e *fee-for service*), percebem um problema de variável omitida — no caso em questão, a ausência da variável *doenças crônicas* no modelo. A gravidade do problema consiste no fato de que tanto a escolha do plano de saúde quanto eventuais despesas médicas são impactadas pela variável omitida — *doenças crônicas*. Diante disso, os autores apontam que os resultados encontrados pelo modelo que relaciona tipos de planos de saúde e despesas médicas, mas que ignora doenças crônicas, seriam enviesados, refletindo correlações espúrias.

### 2.3 Causalidade simultânea

A endogeneidade também está presente quando há dificuldade em determinar o sentido das relações causais. Na Ciência Política, há diversos fenômenos em que variáveis têm influência umas sobre as outras (FRANZESE, 2007). A causalidade reversa, terceira espécie de endogeneidade apresentada neste artigo, ocorre quando a variável tida como dependente impacta a variável independente — ou quando as duas variáveis apresentam influências recíprocas, dificultando o entendimento da direção causal. Em vista disso, a causalidade reversa pode estar presente “em adição ao efeito causal que estamos tentando detectar, ou pode estar presente ao invés do efeito que estamos tentando detectar” (DOWD; TOWN, 2002, p. 10, tradução nossa).

Supondo que X seja o que, em hipótese, temos como variável independente e Y seja a variável dependente, a causalidade reversa, também chamada de causalidade simultânea, ocorre quando Y causa X ao mesmo tempo em que X causa Y, o que pode ser ilustrado pela relação  $X \leftrightarrow Y$  (FRANZESE, 2007, p. 61). Ainda sobre os planos de seguro de saúde, Dowd e Town (2002) apontam que, em um estudo sobre o acesso a um bom plano de saúde e seu efeito no nível de saúde da população, é possível enxergar uma relação de causalidade simultânea entre as variáveis dependente e independente. Isso acontece porque um bom plano de saúde pode influenciar diretamente o acesso a tratamentos adequados, que aumentam a saúde da população. Ao mesmo tempo, a saúde individual pode ser determinante para que a pessoa tenha recursos próprios

para pagar um seguro particular — ou para que consiga um emprego que garantirá um plano de saúde corporativo.

Os obstáculos endógenos percorridos nestes subtópicos — erros em variáveis, impacto das variáveis omitidas e causalidade reversa —, que impactam a precisão das inferências, demandam soluções. A seguir, apresentamos, de modo panorâmico, algumas possibilidades para encontrá-las — teoria, que apresenta rotas iniciais, experimentos, coleta de dados adicionais, informações extraempíricas, abordagens multimétodos, variáveis instrumentais, análise comparativa, decomposição de variáveis (dependente e independente).

### 3 COLETÂNEA DE FERRAMENTAS PARA CONTORNAR A ENDOGENEIDADE: UM BREVE PANORAMA

Buscar causalidades e entender quais variáveis devem ser levadas em consideração para explicar a ocorrência de um fenômeno são tarefas que dependem de aporte teórico substantivo (SÁTYRO; REIS, 2014; ANTONAKIS *et al.*, 2014). Conhecer o estado da arte bibliográfico pode se tornar um escudo contra direções causais equivocadas e omissões de variáveis importantes para um modelo preditivo. A presença de causalidade reversa, por exemplo, expõe insuficiência teórica (KING; KEOHANE; VERBA, 1994), vez que são raros os fenômenos da Ciência Política que não têm em seu horizonte pesquisas prévias sobre caminhos inferenciais confiáveis.

No incerto horizonte causal dos dados observacionais, em que “longas sequências de decisões, ações, padrões institucionais e outros fatores conectam o tratamento ao resultado” (SEAWRIGHT, 2016, p. 57), as abordagens multimétodos colaboram tanto para avaliar trajetórias causais quanto para identificar variáveis omitidas. Antes de testarem suas hipóteses em estudos quantitativos, pesquisadores podem empregar métodos qualitativos, como o *process tracing*, ferramenta capaz de analisar em profundidade o desenvolvimento histórico de um fenômeno e, assim, indicar possíveis fatores causais presentes em determinada dinâmica (SEAWRIGHT, 2016). Além da identificação causal, o *process tracing* também é importante na análise de eventuais variáveis omitidas. Novamente na trilha de Seawright (2016), a capacidade preditiva de modelos depende da inexistência de uma variável que impacte tanto

na dependente quanto na independente. Nesse sentido, o detalhamento oferecido pelo *process tracing*, ao focar no percurso de uma ocorrência, contribui para revelar possíveis variáveis omitidas e evitar vieses nos resultados.

Uma pesquisa com vieses decorrentes de variáveis omitidas também pode indicar desatenção à teoria, que, em regra, apresenta um ponto de partida sistemático sobre o que importa e sobre o que não importa para detectar variações de resultados (CRESWELL, 1994). King, Keohane e Verba (1994) discutem a importância da teoria para a descoberta de variáveis omitidas. Tomando como exemplo uma pesquisa hipotética sobre níveis educacionais e participação política nos Estados Unidos da América, os autores analisam a possibilidade de variáveis inicialmente não consideradas afetarem tanto a variável dependente (participação) quanto a variável independente (educação) — configuração que enviesaria os resultados. Na discussão, consideram a possibilidade de que raça e envolvimento político dos pais sejam variáveis omitidas. De onde extraem essa possibilidade e de onde poderiam extrair outras? Da teoria, vez que “não há outra maneira para escolher quais variáveis usar em nossa análise” (KING; KEOHANE; VERBA, 1994, p. 177, tradução nossa).

Detectar a influência de variáveis omitidas é tarefa mais complexa nas pesquisas observacionais. Enquanto o elemento randômico dos experimentos permite a manipulação controlada de causas e efeitos, nas pesquisas observacionais surgem inúmeros elementos que desafiam a correta identificação da causalidade (REZENDE, 2015, p. 35). Diante de tal desafio, e considerando a impossibilidade de realizar experimentos, pesquisadores podem também recorrer à coleta de dados adicionais sobre as variáveis omitidas em tentativa de compreender e mitigar a sua interferência na pesquisa. Aumentar o número de observações é fundamental para identificar e solucionar erros de pesquisa — embora o valor desta estratégia não deva ser subestimado — convém reconhecer que ele, em regra, vem associado a recursos, às vezes, indisponíveis (DOWD; TOWN, 2002).

Franzese (2007) aponta que todas as ferramentas de pesquisa utilizadas para identificar causalidade, em última instância, demandam a adição de observações extraempíricas às análises empíricas, sejam tais informações pressupostos teóricos ou variáveis inicialmente não previstas. Acrescenta ainda que tal estratégia apenas funciona sob determinadas condições, nas quais, em suas

palavras, são adicionadas observações extraempíricas adequadas. O autor ressalva, porém, o fato de que “a correção da logicamente necessária informação extra-empírica não pode ser testada de forma empírica diretamente” (FRANZESE, 2007, p. 63), situação que exige multiplicação de cuidados para que as informações acrescidas possam, de fato, ampliar a capacidade inferencial da pesquisa. Informações extraempíricas podem ser relevantes para contornar questões endógenas, como é o caso das variáveis instrumentais.

Recorrentes problemas de endogeneidade podem ser solucionados pelo uso de variáveis instrumentais (Z) (ULLAH; ZAEFARIAN; ULLAH, 2020; ZAEFARIAN *et al.*, 2017). No que trata especificamente da causalidade reserva, uma variável tem de satisfazer dois requisitos para que seja considerada instrumental: não se relacionar com o termo de erro da equação, ou seja, não padecer dos mesmos problemas da variável independente endógena (X); e relacionar-se com a variável independente (WOOLDRIDGE, 2006; ANTONAKIS *et al.*, 2014; BASCLE, 2008; ULLAH; ZAEFARIAN; ULLAH, 2020). Nesse sentido, uma variável instrumental se relaciona apenas com a variável independente (X), sem afetar, em um primeiro momento, a variável dependente (Y). Em outras palavras, a variável instrumental Z impactará Y apenas na medida em que influenciar X (ZAEFARIAN *et al.*, 2017). A escolha das variáveis instrumentais, passo decisivo para a correção do modelo, dependerá de aporte teórico que forneça motivos para sua utilização (ULLAH; ZAEFARIAN; ULLAH, 2020). Com o uso da variável instrumental, um componente exógeno, é possível aferir causalidade em um fenômeno.

Outro caminho para tratar a questão da endogeneidade na Ciência Política é a análise comparativa. De acordo com King, Keohane e Verba (1994), tal método é utilizado para identificar a presença de elementos endógenos e exógenos no fenômeno que se busca investigar. Os autores mostram que a alternativa de solução, empregada por Halpern em um estudo sobre “o papel das ideias na adoção de políticas econômicas stalinistas em outros países socialistas” (KING; KEOHANE; VERBA, 1994, p. 191–192, tradução nossa), baseia-se em enxergar a endogeneidade como potencialmente um problema de vieses de variáveis omitidas e, a partir daí, analisar casos em que não haveria estes mesmos vieses. Nessa comparação com outros subconjuntos, Halpern encontrou evidências indicando que as ideias em análise “não eram inteiramente

endógenas nessa situação — eles desempenharam um papel causal” (KING; KEOHANE; VERBA, 1994, p. 193).

Decompor variáveis também pode ser uma alternativa para eliminar impasses oriundos de causalidade reversa (KING; KEOHANE; VERBA, 1994). Em determinadas circunstâncias, é possível fazê-lo na variável dependente, ao organizá-la em dois segmentos — um de fato dependente da variável explicativa e outro que não escape do problema endógeno, influente sobre a variável explicativa. Com essa estratégia, torna-se viável entender variações pelo menos na parte exclusivamente dependente da variável a ser explicada. Em outras ocasiões, o caminho pode se inverter, exigindo a decomposição da variável independente numa parte exógena, capaz de se livrar dos problemas em questão, e numa parte endógena, ainda vulnerável ao problema da causalidade reversa. Dessa forma, a pesquisadora poderá se valer apenas do conteúdo exógeno da variável independente na busca pela inferência. Ambas as possibilidades de decomposição podem salvar, no mínimo parcialmente, a capacidade preditiva de uma pesquisa.

Os problemas endógenos, entretanto, não se extinguem com a resolução da trajetória causal ou com a identificação de variáveis omitidas. Embora nucleares, tais problemas dividem espaço com os erros em variáveis no amplo guarda-chuva da endogeneidade. Seawright (2016), ao tratar da abordagem multimétodo, sinaliza a importância dos estudos de casos para afinar mensurações de variáveis. Notáveis pelo grau de detalhe com que investigam um fenômeno, estudos de caso podem auxiliar na descoberta de imprecisões relacionadas à variável medida, “permitindo uma avaliação da extensão em que um dado indicador captura a informação relevante ou a distorce pelas características do processo que conduz ao índice final para o caso” (SEAWRIGHT, 2016, p. 49, tradução nossa). Modelos que se baseiam em variáveis coletadas em *surveys* ilustram os benefícios dos estudos de caso em termos de mensuração. Na eventualidade de um conjunto de entrevistados não compreender adequadamente uma pergunta, os resultados encontrados terão pouco a dizer. A realização de estudos de caso em profundidade, anteriores ao *survey*, poderia mapear como a questão é entendida pelos integrantes da amostra — situação que permitiria o ajuste da pergunta e a consequente medição do objeto investigado (SEAWRIGHT, 2016).

Problemas endógenos não devem, em um primeiro momento, inviabilizar uma pesquisa. Traçado o cenário apresentado nos últimos parágrafos, é possível visualizar algumas alternativas. A próxima seção dedica-se a apresentar alguns estudos que lidaram de maneira adequada com questões de endogeneidade, sobretudo de causalidade reversa —seja reconhecendo a necessidade de considerá-la para seguir adiante, seja usando um dos métodos apresentados acima, as variáveis instrumentais<sup>4</sup>, para garantir a produção de inferências causais válidas.

#### 4 SE NÃO CONTORNAR, PELO MENOS CONSIDERE

Instituições causam desenvolvimento econômico ou o desenvolvimento econômico é responsável pela ascensão de instituições? Ambos. De acordo com Przeworski (2004), não é possível superar a endogeneidade ao tratar do tema. Não existe causa primária identificável. Causalidades recíprocas impedem a compreensão de efeitos isolados. Diante da conclusão de que “o único motor da história é a endogeneidade” (2004, p. 5), Przeworski sugere que os estudos sobre o tema devam mudar de foco — não priorizar a busca pela causa primária, mas o correto entendimento dos efeitos recíprocos. Nesse caso, a endogeneidade não é um problema a ser resolvido, mas uma condição a ser considerada — o que permite abordagens igualmente relevantes no estudo do fenômeno, redirecionadas a partir da percepção de que a causalidade original não pode ser transposta.

Embora isso se aplique a alguns casos, é possível avançar em determinadas ocasiões, quando empregadas as ferramentas para enfrentar a endogeneidade. Os textos de Steven Levitt (1997 e 2002) sobre policiamento e taxas de criminalidade apresentam modelos de superação de problema endógeno — a causalidade reversa/simultânea. Trabalhos anteriores sobre o tema não conseguiam apresentar relação significativa entre o aumento do efetivo de policiais e diminuição dos índices criminais. Levitt observa que tal dificuldade tem como causa provável uma questão endógena: a presença de maiores índices de criminalidade em determinado local faz com que o poder público mobilize maior

---

<sup>4</sup> Vale dizer que o uso de variáveis instrumentais não é a única estratégia capaz de minimizar problemas de endogeneidade. Outras estratégias, que ultrapassam o escopo deste artigo, como Matching, Regressões Descontínuas e Diferenças em Diferenças também podem ser úteis ao contorno do problema.

número de agentes, criando uma correlação entre as duas ordens de grandeza — criminalidade e tamanho do efetivo policial.

Para superar a simultaneidade presente no fenômeno, Levitt vale-se de uma variável instrumental, qual seja o *timing* das eleições para prefeitos e governadores — variável que não se relaciona com a criminalidade e, portanto, não incorre em endogeneidade. Inicialmente, os autores buscam estabelecer conexão entre eleições e crescimento do efetivo policial. Conseguem. Os efetivos policiais são ampliados em anos eleitorais. A partir disso, verificam, por meio de modelos econométricos, se, no período subsequente à ampliação da força policial, os índices de criminalidade foram reduzidos. Concluem que boa parte dos crimes analisados diminuíram após o aumento do efetivo, demonstrando a utilidade de variáveis instrumentais para corrigir defeitos relacionados à endogeneidade.

Outro exemplo bem sucedido no trato de questões endógenas pode ser encontrado em Gerber (1998). O estudo se dedica a analisar os impactos dos gastos de candidatos incumbentes e opositoristas nas eleições para o Senado norte-americano. Ao retomar a literatura sobre tema, Gerber aponta um pressuposto teórico: gastos de candidatos opositoristas produziram maior impacto no número de votos do que gastos dos incumbentes, pelas seguintes razões, defendidas por Gary Jakobson<sup>5</sup>: a) os incumbentes já contam com estruturas de campanha bem montadas; b) os incumbentes, pela visibilidade do cargo, são mais conhecidos do que seus oponentes. As duas razões fariam com que os gastos do incumbente não implicassem efeitos relevantes, pois já partiriam de condições vantajosas, ao passo que os oponentes, em tese desconhecidos e com estruturas de campanha menos sedimentadas, teriam nas despesas eleitorais uma alternativa relevante para ampliar seu eleitorado. Mas o modelo teórico, em linha com estudos posteriores, desconsiderou fatores que ensejam endogeneidade. Conforme aponta Gerben: a) a probabilidade de vitória tende a facilitar o recebimento de doações; b) em eleições apertadas, apoiadores podem doar maiores quantidades; c) a certeza de vitória pode levar o incumbente a não buscar grandes arrecadações. As três questões, eminentemente endógenas no sentido de simultaneidade, trazem vieses às conclusões sobre a efetividade dos gastos de campanha dos candidatos. Havia, portanto, questões

---

<sup>5</sup> Gerben sinaliza que o argumento de Jakobson foi criado para explicar eleições para a Câmara dos Deputados, mas, em termos teóricos, poderia se aplicar também às eleições senatoriais.

não consideradas que dificultavam inferências válidas no pressuposto teórico sobre gastos de candidaturas.

Para responder aos problemas endógenos, Gerben se vale de variáveis instrumentais — relacionadas aos gastos de campanha, a variável independente, mas, inicialmente, não relacionadas à variável dependente, votos recebidos pelos candidatos incumbentes. Essas variáveis instrumentais são: (i) riqueza do candidato opositor, capaz de influenciar sua arrecadação de fundos para a campanha; (ii) população votante do estado, importante para mensurar o gasto por cada voto; e (iii) despesas em eleições anteriores (soma dos dois candidatos), tendo em vista que as eleições para Senado, de modo geral, não apresentam os mesmos incumbentes e opositores, logo, comparar os gastos de eleições anteriores não revelaria a sobreposição da capacidade de arrecadação de um candidato em relação aos demais — o que poderia implicar endogeneidade. A partir de modelos econométricos que levam em conta as variáveis instrumentais, Gerben encontra que o impacto dos gastos de candidatos incumbentes é semelhante ao impacto dos gastos de candidatos desafiantes — resultado que os modelos econométricos desatentos às questões de causalidade reversa não encontraram em estudos anteriores, estimando efeitos maiores para gastos dos candidatos desafiantes.

Também por meio do uso de variáveis instrumentais, Kern e Hainmueller (2009) avaliaram o impacto de canais televisivos da Alemanha Ocidental, capitalista, sobre as percepções dos cidadãos da Alemanha Oriental, comunista, em relação ao seu próprio regime. Com base em literatura acerca da influência de mídias ocidentais na erosão do apreço social pelos regimes comunistas, Kern e Hainmueller esperavam encontrar padrão semelhante na Alemanha Oriental entre os espectadores de canais televisivos da Alemanha Ocidental. Para tanto, eles empregaram, como desenho de pesquisa, um experimento natural, o qual, por meio de um fator randômico, alheio ao controle dos pesquisadores, contribuiu para a validade das conclusões. Partindo de um *survey* conduzido com a população jovem do país, os autores perceberam que a mera comparação daqueles que assistiram às mídias ocidentais com aqueles que não assistiram seria insuficiente para garantir inferências, visto que poderia estar sujeita a numerosos vieses. No entanto a utilização do desenho experimental, viabilizado no caso pelo uso de uma variável instrumental — residir ou não em região com

acesso às redes<sup>6</sup> —, permitiu a comparabilidade segura entre os resultados. É dizer, os autores compararam as respostas daqueles que residiam em regiões sem acesso às redes e que, por consequência, não assistiam à mídia oriental com as respostas dos que residiam em regiões com acesso à rede e viam os programas do país vizinho.

De acordo com os resultados encontrados, assistir às mídias da Alemanha Ocidental se traduziu em maior apoio ao regime comunista da Alemanha Oriental. Na hipótese dos autores, construída a partir das respostas presentes no *survey*, o escapismo explicaria o maior apoio ao regime entre aqueles que assistiam às mídias da Alemanha Ocidental, dado que o conteúdo consumido privilegiaria o entretenimento e aliviaria, em alguma medida, as dificuldades vividas dentro do regime Oriental.

## 5 CONSIDERAÇÕES FINAIS

Nas palavras de Franzese (2008), a inferência causal em Ciência Política tem de conviver com um mundo em que *quase tudo causa quase tudo*. Com exceção dos experimentos, que permitem controle mais estrito sobre causas e efeitos, em razão de seu componente aleatório (REZENDE, 2015), há um cenário de fenômenos incertos — mensurações imprecisas, omissões metodologicamente problemáticas e causalidades sem direção precisa, confusas em setas recíprocas.

É do que se preocupa a endogeneidade. Não apenas de causalidade reversa, como frequentemente entendida, mas da relação entre regressor e termo de erro, que também engloba erros em variáveis e variáveis omitidas. Como discutido nas seções anteriores, a endogeneidade trata de problemas que, muitas vezes, são passíveis de solução por meio do bom uso da teoria existente, de análises comparativas, da decomposição de variáveis (dependentes e independentes), de informações extraempíricas e de variáveis instrumentais.

Mas, antes de tudo, o trato da endogeneidade exige duas coisas: tomada de consciência e postura cientificamente responsável, nessa ordem. King, Keohane e Verba (1994) lembram que, frequentemente, pesquisadores sequer têm consciência dos problemas endógenos implicados nas hipóteses que desenvolvem.

---

<sup>6</sup> Tal variável foi aferida por meio do local de trabalho do respondente. Na antiga Alemanha Oriental, fatores topográficos faziam variar o alcance das redes de mídia. A região de Dresden, por exemplo, não tinha acesso à mídia da Alemanha Ocidental, pois ficava distante das torres de transmissão.

O primeiro passo, portanto, é perceber o problema — o que não acontecerá sem um amplo domínio da literatura que cobre o fenômeno analisado. Embora frustrante, descobrir limitações é o preço da ciência. Alberto Caeiro, o heterônimo de Fernando Pessoa (2011), já disse n’*O Guardador de Rebanhos*: “pensar incomoda como andar à chuva quando o vento cresce e parece que chove mais”.

Depois da consciência, a postura. Não é necessário abandonar o projeto quando diante de questão endógena. Mas, sob pena de concluir um trabalho cientificamente frágil, é preciso agir — seja buscando alternativas para contornar os problemas, algumas delas apresentadas de maneira introdutória neste artigo, seja alterando o foco da questão de pesquisa, levando em conta a endogeneidade e buscando novos caminhos — que até podem partir do reconhecimento de sua existência, mas que devem chegar em locais metodologicamente mais seguros. Independentemente da rota escolhida, impõe-se uma necessidade subjacente — explicar cada passo, deixar transparentes as decisões e os desenhos escolhidos para perseguir a questão de pesquisa — com todas as limitações pendentes explicitadas.

#### REFERÊNCIAS

- ALMOND, Gabriel; GRECO, Stephen. Clouds, clocks, and the study of politics. *World Review*, v. 29, n. 4, p. 489–522, 1977.
- ANTONAKIS, John *et al.* Causality and endogeneity: Problems and solutions. In: DAY, David V. (ed.). *The Oxford Handbook of Leadership and Organizations*. New York: Oxford University Press, 2014. p. 93–117.
- BASCLE, Guilhem. Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, v. 6, n. 3, p. 285–327, 2008.
- BERTRAND, Marianne; MULLAINATHAN, Sendhil. Are Emily and Greg more employable than Lakisha and Jamal?: a field experiment on labor market discrimination. *National Bureau of Economic Research*, Cambridge, MA, Working Paper n. 9873, 2003. Disponível em: <http://www.nber.org/papers/w9873>. Acesso em: 5 mar. 2022.

- CINELLI, Carlos; HAZLETT, Chad. Making sense of sensitivity: extending omitted variable bias. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, v. 82, n. 1, p. 39–67, 2020.
- CRESWELL, John. The use of theory. In: CRESWELL, John. *Research Design qualitative and quantitative approaches*. London: Sage Publications, 1994. p. 81–104.
- DOWD, Bryan; TOWN, Robert. Does X Really Cause Y? *Academic Health*: v. 1, p. 21, 2002.
- DUNNING, Thad. *Natural experiments in Social Sciences: a design-based approach*. Cambridge: Cambridge University Press, 2012.
- FERRAZ, Claudio; FINAN, Frederico. Exposing corrupt politicians: the effect of Brazil’s publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, v. 123, n. 2, p. 703–745, 2008.
- FRANZESE, Robert J. Context matters: the challenge of multicausality, context-condicionality and endogeneity for empirical evaluation of positive theory in comparative politics. In: FRANZESE, Robert J. *The Oxford Handbook of Comparative Politics*. Chapter 2. 2007. p. 26–72.
- GERBER, Alan. Estimating the effect of campaign spending on senate election outcomes using instrumental variable. *American Political Science Review*, v. 92, n. 2, p. 401–411, 1998.
- GUIDE JR., V. Daniel R.; KETOKIVI, Mikko. Notes from the editors: redefining some methodological criteria for the journal. *Journal of Operations Management*, v. 37, n. 1, p. v–viii. 2015.
- KERN, Holger Lutz; HAINMUELLER, Jens. Opium for the masses: how foreign media can stabilize authoritarian regimes. *Political Analysis*, v. 17, p. 377–399, 2009.
- KING, Gary; KEOHANE, Robert; VERBA, Sidney. *Designing social inquiry*. Princeton, N.J.: Princeton University Press, 1994.

- LEVITT, Steven. Using electoral cycles in police hiring to estimate the effect of police on crime. *The American Economic Review*, v. 87, n. 3, p. 270–290, 1997.
- LEVITT, Steven. Using electoral cycles in police hiring to estimate the effect of police on crime: reply. *The American Economic Review*, v. 92, n. 4, p. 1244–1250, 2002.
- MAHONEY, James. After KKV: The new methodology of qualitative research. *World Politics*, v. 62, n. 1, p. 120–147, 2010.
- PESSOA, Fernando. *Antologia poética: texto integral*. Organização Walmir Ayala. Coleção Saraiva de Bolso. Rio de Janeiro: Editora Nova Fronteira, 2011.
- PRZEWORSKI, Adam. Is the science of comparative politics possible? *In: The Oxford Handbook of Comparative Politics*, 2007. p. 147–171.
- PRZEWORSKI, Adam. The last instance: are institutions the primary cause of economic development? *European Journal of Sociology*, v. 45, n. 2, p. 165–188, 2004.
- REZENDE, Flávio. O pluralismo inferencial na Ciência Política pós-KKV (2005– 2015): argumento e evidências. *Revista Política Hoje*, v. 26, n. 1, p. 241–277, 2017.
- REZENDE, Flávio. Transformações metodológicas na Ciência Política contemporânea. *Revista Política Hoje*, 2. Ed., v. 24, p. 13–45, 2015.
- SÁTYRO, Natália. G. D.; REIS, Bruno P. W. Reflexões sobre a produção de inferências indutivas válidas em Ciências Sociais. *Revista Teoria & Sociedade*, n. 22.2, p. 19–39, 2014.
- SEAWRIGHT, Jason. *Multi-Method Social Science: combining qualitative and quantitative tools*. Cambridge: Cambridge University Press, 2016.
- TOM ZÉ; ELTON MEDEIROS. Tô. Rio de Janeiro: Continental: 1976. Disponível em: (<https://www.cifraclub.com.br/tom-ze/to/>). Acesso em: 30 mar. 2023.

ULLAH, Subhan; ZAEFARIAN, Ghasem; ULLAH, Farid. How to use instrumental variables in addressing endogeneity?: a step-by-step procedure for non-specialists. *Industrial Marketing Management*, v. 96, p. A1-A6, 2020.

WOOLDRIDGE, Jeffrey. *Uma introdução à econometria: uma abordagem moderna*. 2. ed. São Paulo: Editora Thomson, 2006.

ZAEFARIAN, Ghasem *et al.* Endogeneity bias in marketing research: problem, causes and remedies. *Industrial Marketing Management.*, v. 96, p. 39–46, 2017.