Deferência *vs.* intervenção: análise empírica da relação entre o controle externo do TCU sobre as agências reguladoras e os níveis de governança destas*

Deference vs. intervention: an empirical analysis of the relationship between the TCU's external control on regulatory agencies and their governance

*Bruno Dantas***
*Leonardo Ávila****

* Artigo recebido e aprovado em 13 de abril de 2026. DOI: https://doi.org/10.12660/rda.v285.2026.98378.

** Ministro do TCU. Doutor e mestre pela Pontifícia Universidade Católica de São Paulo (PUC-SP). Foi *senior hauser global fellow from government and practice* na NYU School of Law. *Visiting researcher* na Cardozo School of Law e no *Institute de Recherche Juridique* da Universidade Paris 1 Panthéon-Sorbonne. *Post doctoral scientific guest* do Max Planck Institute for Regulatory Procedural Law. Professor da Fundação Getulio Vargas (FGV-Rio), Universidade do Estado do Rio de Janeiro (Uerj) e Universidade Nove de Julho (Uninove). País: Brasil. E-mail: bruno.dantas@fgv.br.

*** Doutorando em direito pela Universidade de São Paulo (USP). Mestre em direito pela Universidade de Brasília (UnB). Pesquisador na FGV Consenso. País: Brasil. E-mail: leonardoavila@usp.br.

CC BY NC ND

RESUMO

Este artigo investiga, por meio de desenho empírico, se há associação entre o nível de governança de agências reguladoras federais e o perfil de intervenção do controle externo do Tribunal de Contas da União (TCU). A governança é mensurada pelo Índice Integrado de Governança e Gestão Públicas (IGG), com defasagem temporal, e o controle é aproximado por determinações e recomendações, tipificadas pela Resolução-TCU nº 315/2020, agregadas em *score* ponderado e normalizadas por exposição. Com base em 33 observações agência-ano e cinco testes complementares, os resultados sugerem associação compatível com duas hipóteses: maior governança tende a coexistir com menor carga de comandos e com maior participação relativa de recomendações.

PALAVRAS-CHAVE

governança — TCU — IGG — controle externo — recomendações — determinações

ABSTRACT

This article investigates, through an empirical design, whether there is an association between the level of governance of federal regulatory agencies and the intervention profile of the external oversight conducted by the Federal Court of Accounts. Governance is measured by the Integrated Index of Public Governance and Management (IGG), with a time lag, and oversight is approximated by determinations and recommendations, as defined by TCU Resolution No. 315/2020, aggregated into a weighted score and normalized by exposure. Based on 33 agency-year observations and five complementary tests, the results suggest an association consistent with two hypotheses: higher governance tends to coexist with a lower burden of directives and a higher relative share of recommendations.

KEYWORDS

governance — TCU — IGG — external control — recommendations — determinations

1. Introdução

A discussão sobre discricionariedade, vinculação e controle busca, em grande medida, delimitar quando e com que intensidade o controlador deve ou pode interferir no mérito administrativo.[1] No campo regulatório, essa tensão é agravada pela alegação, inclusive reafirmada inúmeras vezes pelo Judiciário,[2] de um "mérito técnico" e pela necessidade de deferência a escolhas baseadas em *expertise*.[3] Deferência essa que, contudo, não elimina o dever de controlar procedimentos, motivação, consistência das evidências e conformidade com *standards* e com a lei.[4]

Se a deferência é condicionada à qualidade institucional do processo decisório, então é plausível tratar a governança como variável explicativa do padrão de intervenção do controle.[5] Esta é exatamente a proposta deste

---

1 Essa discussão não é nova. Por exemplo, no primeiro livro de direito administrativo escrito no Brasil já é possível encontrar explicações sobre os dois poderes da administração, sendo um deles "discricionário, inteligente e livre" e o outro "passivo, regulado e dominado pela lei, regulamento, decreto, ou contrato que o rege". REGO, Vicente Pereira do. *Elementos de direito administrativo brasileiro comparado com o direito administrativo francez*. [*S. l.*]: Typographia Universal, 1857. Disponível em: https://bibliotecadigital.stf.jus.br/xmlui/handle/123456789/1215. Acesso em: 13 fev. 2026. Como exemplo da atualidade do debate, ver: PEREZ, Marcos Augusto. *Testes de legalidade*: métodos para o amplo controle da discricionariedade administrativa. Belo Horizonte: Fórum, 2020.

2 Por todos os exemplos, o seguinte: Ementa: Direito constitucional e administrativo. Agravo Regimental em Mandado de segurança. Desapropriação para reforma agrária. Produtividade da propriedade. Fixação dos índices para aferição do grau de eficiência na exploração. Imposição de multa. 1. A determinação dos índices para a aferição dos graus de utilização da terra e de eficiência na exploração, para fins de desapropriação, insere-se no âmbito de ***discricionariedade técnica*** do órgão federal competente, de modo que não cabe ao Judiciário intervir, salvo nas hipóteses de extrapolação da atribuição legal conferida (art. 6º da Lei nº 8.629/1993), o que não é o caso dos autos. Informações que demonstram a razoabilidade dos índices estabelecidos. Verificação da alegada violação à isonomia que demandaria dilação probatória, incabível em sede de mandado de segurança. Precedentes. 2. Agravo a que se nega provimento por manifesta improcedência, com aplicação de multa de 1% (um por cento) do valor corrigido da causa, ficando a interposição de qualquer outro recurso condicionada ao depósito do respectivo valor. (Grifo nosso). (BRASIL. Supremo Tribunal Federal. MS 24883 AgR. rel. min. Roberto Barroso, Tribunal Pleno, julgado em 24 fev. 2017, *DJe* 19 abr. 2017).

3 Cf. ROMAN, Flavio José. *Discricionariedade técnica na regulação econômica*. São Paulo: Saraiva, 2013.

4 "Não vemos fundamento para afastamento, a priori, do controle judicial pelo simples fato de se tratar de ato produzido por agência especializada na matéria." (DI PIETRO, Maria Sylvia Zanella. *Direito administrativo*. 37. ed. Rio de Janeiro: Forense, 2024. p. 523).

5 A ideia não é inovadora. Vários autores já analisaram a relação entre governança e o perfil de atuação do TCU. O diferencial deste trabalho está nos testes de avaliação e na base de dados, que foi construída especificamente para a pesquisa. Cf. MERLOTTO, Nara Carolina. *A atuação do Tribunal de Contas da União sobre as agências reguladoras*: entre a independência e o controle.

CC BY NC ND

artigo, testar, com indicadores observáveis e replicáveis, se variações no nível de governança (medidas pelo IGG com defasagem) se associam a variações na intensidade e no perfil dos comandos do TCU (determinações e recomendações) dirigidos às agências reguladoras.

Aqui se insere o primeiro recorte metodológico. Por mais que normalmente se debata sobre o controle da administração pelo Poder Judiciário, para as finalidades deste artigo, toma-se como parâmetro o controle externo da administração pública federal exercido no âmbito do Congresso Nacional com auxílio do TCU.

A partir disso, formula-se o problema de pesquisa nos seguintes termos: existe associação sistemática entre o nível de governança das agências reguladoras e a intensidade do controle exercido pelo TCU?

Trabalha-se com duas hipóteses. A hipótese principal (H1) sustenta que, quanto maior a governança das agências reguladoras, menor a intensidade do controle exercido pelo TCU. A hipótese concorrente (H2), de "controle como capacitação", admite que maior governança possa reduzir sobretudo determinações e elevar a participação relativa de recomendações (mais orientativas), sem necessariamente diminuir o volume total de tarefas demandadas pelo controlador.

Essas hipóteses necessitam de duas parametrizações. A primeira delas é como se mede a governança de agências reguladoras. A segunda, como se mede a intensidade do controle. Em contraste com estudos anteriores,[6] que investigam padrões comportamentais do TCU por meio de uma tipologia de deliberações com foco em setores de infraestrutura, este artigo adota testes quantitativos específicos para examinar associações entre governança (IGG_lag) e perfis de intervenção, sem ênfase em classificações qualitativas de intensidade.

Dissertação (mestrado em direito) — Faculdade de Direito, Universidade de São Paulo, São Paulo, 2020. Disponível em: https://www.teses.usp.br/teses/disponiveis/2/2134/tde-25092020-144220/publico/6490332_Dissertacao_Parcial.pdf. Acesso em: 22 fev. 2026; FERREIRA, R. F. *Intensidade do controle e governança*: um (possível) padrão de comportamento do TCU no controle das agências reguladoras. Dissertação (mestrado em direito) — Escola de Direito do Rio de Janeiro, Fundação Getulio Vargas, Rio de Janeiro, 2024. Disponível em: https://repositorio.fgv.br/bitstreams/6dc9e7a4-0299-4a77-8dfc-2ea3c34018d0/download. Acesso em: 22 fev. 2026.

6 Ferreira, *Intensidade do controle e governança*, *op. cit.*

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

2. Governança e controle: variáveis do estudo e escolhas de mensuração

2.1. Governança: IGG e construção do IGG lag

Como padrão de medição do nível de governança, este artigo utilizou o Índice Integrado de Governança e Gestão Públicas (IGG), que é um indicador-síntese utilizado pelo TCU para avaliar a adoção de práticas de governança organizacional e de gestão pública.

O IGG, conforme se pode ler no Acórdão nº 1.913/2024[7] — Plenário, resume práticas avaliadas por questionário e se forma pela agregação de outros indicadores de governança e de gestão, tais como: Índice de Governança Pública Organizacional (iGovPub), Perfil de Governança e Gestão de Pessoas, Perfil de Governança e Gestão de TI e de Segurança da Informação, Perfil de Governança e Gestão de Contratações, e Perfil de Governança e Gestão Orçamentárias.

De modo didático, o IGG combina várias dimensões, subconjuntos de práticas, em um número único. Esse número resulta de um instrumento padronizado, aplicado em larga escala, com a finalidade de diagnóstico e indução de melhorias na administração pública federal.

Neste artigo, utilizam-se apenas os valores consolidados do IGG fornecidos para as 11 agências previstas na Lei nº 13.848/2019, sem a Agência Nacional de Proteção de Dados que só passou a constar no rol em 2025,[8] e três ciclos (2018, 2021 e 2024), sem reconstruir os microcomponentes do questionário.

Assim, partindo dos dados fornecidos pelo TCU, foi possível compilar por ano o nível de governança de cada agência. Com isso, fica consolidado o mecanismo utilizado por este artigo para medição da governança no Anexo I, Tabela 3 — IGG individual das agências por ano de controle (2018, 2021 e 2024).[9]

---

[7] O mesmo acórdão registra a evolução metodológica do instrumento e a incorporação, em 2024, de um índice ampliado (iESGo) que agrega governança, gestão e sustentabilidade, sendo o IGG a parte que sintetiza governança e gestão (com exceção das novas práticas de sustentabilidade). BRASIL. Tribunal de Contas da União. *Acórdão nº 1.913/2024 — Plenário*. Relatório de levantamento para avaliar e promover as práticas de Governança, Sustentabilidade e Gestão nas organizações públicas federais. Relator: ministro Vital do Rêgo. Brasília, DF: TCU, 18 set. 2024. Disponível em: https://pesquisa.apps.tcu.gov.br/. Acesso em: 25 fev. 2026.

[8] BRASIL. Lei nº 13.848, de 25 de junho de 2019. Diário Oficial da União, Brasília, DF, 26 jun. 2019.

[9] A Tabela 3 pode ser acessada a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

CC BY NC ND

2.2. Determinações e recomendações como proxy *de intensidade e perfil*

Para medir o nível de intensidade do controle do TCU nas agências reguladoras, foi adotada uma *proxy* objetiva baseada no conteúdo das deliberações, distinguindo-se entre determinações, comandos de natureza mandamental, com imposição de providências e prazos, e recomendações, orientações colaborativas de aperfeiçoamento, conforme consolidado na Resolução-TCU nº 315/2020.[10]

Nesse sentido, usando a referida Resolução como base, é possível diferenciar as deliberações em três espécies: (i) determinação, (ii) ciência e (iii) recomendação. Neste trabalho, interessam somente determinações e recomendações, pois constituem indicadores diretamente observáveis do grau e do perfil de intervenção do controle externo, além do fato de que as ciências têm natureza meramente declaratória.

Nos termos do art. 2º, I, determinação é deliberação de natureza mandamental que impõe ao destinatário, em prazo fixado, a adoção de providências concretas e imediatas voltadas a prevenir ou corrigir irregularidade, remover seus efeitos ou abster-se de executar atos irregulares. A recomendação, por sua vez, definida no art. 2º, III, tem natureza colaborativa, identifica oportunidades de melhoria e sugere medidas para aperfeiçoar a gestão e a execução de políticas, programas e ações governamentais.

Essa distinção é fulcral para o desenho empírico adotado. Determinações tendem a expressar um nível mais elevado de intrusão no plano decisório ao passo que recomendações operam, em regra, como orientação qualificada, preservando maior espaço para a escolha de meios pelo gestor, no caso, das agências reguladoras.

Por isso, além de mensurar o volume total de intervenções, o artigo decompõe a intensidade do controle em dois componentes, mandamental (determinações) e orientativo (recomendações), o que permite testar se o aumento da governança se associa a um controle menos intenso (H1), bem como se ela se associa a um deslocamento do tipo de intervenção, com redução relativa de comandos mandamentais e maior ênfase em recomendações (H2).

---

[10] BRASIL. Tribunal de Contas da União. Resolução TCU nº 315, de 1º de julho de 2020. Dispõe sobre a elaboração de deliberações que contemplem medidas a serem tomadas pelas unidades jurisdicionadas no âmbito do Tribunal de Contas da União e dá outras providências. *Diário Oficial da União*, Brasília, DF, 6 jul. 2020.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

2.3. Recorte temporal e pareamento com defasagem

Como já se disse, a unidade de observação foi agência-ano (11 agências em três anos: 2019, 2022 e 2025) e a intensidade do controle foi aproximada pela quantidade de comandos endereçados a cada agência nos acórdãos que a têm como "entidade jurisdicionada", normalizada pela "exposição" de cada agência no ano (número de acórdãos anuais em que ela aparece como entidade jurisdicionada), para evitar que maior volume de menções fosse confundido com maior severidade do controle.

Especificamente, construiu-se um *score* ponderado de intervenção, que atribui maior peso às determinações do que às recomendações, refletindo a maior carga interventiva típica dos comandos mandamentais, e que é normalizado pela exposição anual de cada agência. Esse *score* é ainda decomposto em um núcleo mandamental e um núcleo orientativo, além de uma medida de participação relativa das recomendações, permitindo examinar não apenas "quanto" o controle incide, mas também "como" ele incide. As definições operacionais de todas essas variáveis são detalhadas na seção 3.2. Dito de outro modo, buscou-se entender se a intervenção assume perfil mais corretivo/mandamental ou mais orientativo/capacitador (ver: Tabela 4 — Indicadores secundários por agência e ano).[11]

Como o IGG mede a governança de um órgão em um determinado ano, para se compreender corretamente o impacto das alterações no nível de governança no perfil de controle sobre aquele órgão, é necessário, evidentemente, utilizar-se os anos prospectivos da mensuração.

Para este estudo foi adotado o ano imediatamente posterior e um intervalo igual de três anos entre cada marco observado. Adotou-se, portanto, um pareamento com defasagem temporal (lag), o IGG 2018 teve seus reflexos no controle observados em 2019; o IGG 2021 teve seus reflexos no controle observados em 2022; o IGG 2024 teve seus reflexos no controle observados em 2025.

[11] A Tabela 4 pode ser acessada a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

3. Dados: obtenção, verificação e base primária

3.1. Coleta e universo de acórdãos

Para se adquirir o número de determinações e recomendações específicos de cada agência, primeiro se acedeu ao sítio eletrônico do TCU,[12] na seção de "Pesquisa integrada", subseção de "Acórdãos e Jurisprudência", item "Acórdãos", onde foi feita a busca no buscador de "nome do órgão ou entidade" ao mesmo tempo que se delimitou o período da busca sendo de 1º de janeiro de 2019 até 31 de dezembro de 2025.

Com esse período fixo, foram feitas 11 buscas individuais, uma para cada agência reguladora. O resultado dessas buscas foi expresso em 11 tabelas contendo todos os metadados (título; data; relator; sumário; número do processo; tipo de processo; interessado/responsável/recorrente; entidade jurisdicionada; representante do Ministério Público; unidade técnica; representante legal; endereço do arquivo) de todos os acórdãos que tinham por unidade jurisdicionada cada uma das agências (ver: Tabela 2 — Número integral de determinações, recomendações e acórdãos).[13]

De posse dessas 11 tabelas e, consequentemente, de todos os *links* das decisões, foi feito o *download* individual de cada um dos acórdãos mencionados.

Com os 1.143 acórdãos baixados, foi feita uma dupla verificação do teor dos dispositivos de cada acórdão individual para avaliação da quantidade de determinações e recomendações expressos em cada acórdão (ver: Tabela 1 — Número total de acórdãos por agência reguladora).[14]

3.2. Codificação dos comandos e regras operacionais

A primeira verificação consistiu em leitura dirigida do bloco deliberativo (em geral introduzido por fórmulas do tipo "ACORDAM… em:" ou "determinar/recomendar…"), com seleção dos trechos em que o TCU (i) impunha providências com prazo e caráter mandamental (determinações) ou (ii) formulava orientação de aperfeiçoamento (recomendações).

---

[12] BRASIL. Tribunal de Contas da União. *Portal do TCU*. Brasília: TCU, 2026.

[13] A Tabela 2 pode ser acessada a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

[14] A Tabela 1 pode ser acessada a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

Em seguida, realizou-se uma segunda verificação, de controle de qualidade, destinada a reduzir erros típicos desse tipo na tabulação:

(a) contagem indevida de subitens como comandos autônomos quando se tratava, na verdade, de desdobramentos de um único comando;

(b) imputação do comando ao destinatário errado (por exemplo, quando o acórdão trata de múltiplos atores no mesmo dispositivo); e

(c) situações em que o comando é dirigido ao órgão supervisor, ao ministério setorial, a entidades do setor ou a outras unidades jurisdicionadas, sem se endereçar à agência reguladora, hipótese em que o item não foi contabilizado como intervenção "na agência".

Para fins de padronização, foram adotadas quatro regras operacionais: (1) somente foram contabilizados comandos explicitamente qualificados como determinação ou recomendação, ou redigidos em termos inequivocamente equivalentes (por exemplo, "determinar que…"; "recomendar que…"), com destinatário identificável; (2) quando uma determinação era formulada como "determinar que adote as seguintes medidas: (a)… (b)…", os subitens foram tratados como desdobramentos do comando principal, salvo quando o próprio texto distinguisse deveres independentes com autonomia material e prazos diversos; (3) determinações e recomendações repetidas em decisões distintas foram contabilizadas separadamente, porque representam incidências distintas do controle no tempo; e (4) em caso de dúvida razoável sobre o destinatário do comando (como quando se referencia a entidades previstas em determinada tabela que não está disponível para o público), adotou-se um critério conservador e não foram computadas tais recomendações e/ou determinações.

A partir dessa codificação, foi possível construir, para cada combinação agência-ano (2019, 2022, 2025), o total de determinações (D) e recomendações (R) extraídas dos dispositivos, bem como o número anual de acórdãos em que a agência figurou como unidade jurisdicionada (disponível no Anexo I — Tabela 2 — Número integral de determinações, recomendações e acórdãos).[15]

Essa última variável foi utilizada como medida de exposição ao controle no respectivo ano; justamente para evitar inferências indevidas, uma agência pode receber mais comandos simplesmente porque aparece em mais decisões, e não necessariamente porque o TCU foi efetivamente mais interventivo em termos proporcionais.

---

[15] A Tabela 2 pode ser acessada a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

Por essa razão, todos os indicadores de intensidade foram apresentados tanto em números absolutos (D, R e Score) quanto em forma de taxas por acórdão (Taxa, Taxa_D e Taxa_R), de maneira a permitir comparações minimamente simétricas entre agências com níveis muito diferentes de exposição anual (ver: Tabela 4 — Indicadores secundários por agência e ano).

Convém distinguir, antes dos testes, os dados primários dos secundários. De dados primários nós temos:[16]

(i) os valores do IGG para cada agência nos ciclos de 2018, 2021 e 2024 (ver: Tabela 3 — IGG individual das agências por ano de controle (2018, 2021 e 2024)).
(ii) o número de determinações (D) e o número de recomendações (R) dirigidas a cada agência em cada um dos anos analisados (2019, 2022 e 2025) (ver: Tabela 2 — Número integral de determinações, recomendações e acórdãos).
(iii) o número de acórdãos do ano que tem a agência como órgão jurisdicionado, igualmente para 2019, 2022 e 2025 (ver: Tabela 1 — Número total de acórdãos por agência reguladora).

Esses três blocos de informação, governança (IGG), comandos (D e R) e exposição (nº de acórdãos) são os insumos básicos do estudo e foram extraídos diretamente das tabelas de resultados e dos acórdãos baixados a partir da pesquisa no sítio do TCU, sem qualquer transformação substantiva além da organização por agência e ano.

A partir desses dados primários, foram construídos os seguintes dados secundários, que foram utilizados nos testes que serão mais bem descritos a seguir:

(a) *Score* ponderado de intervenção, definido como *Score* = 2×D + 1×R, atribuindo maior peso às determinações por seu caráter mandamental; com os seguintes resultados por agência:
(b) a Taxa de intensidade do controle, definida como Taxa = *Score*/Acórdãos, isto é, o *score* normalizado pela exposição anual, permitindo comparação proporcional entre agências com volumes distintos de decisões; sintetizados pelas seguintes tabelas;
(c) a decomposição do controle em componentes, por meio de Taxa_D = (2×D)/Acórdãos (núcleo mandamental) e Taxa_R = R/Acórdãos (núcleo orientativo);
(d) a Taxa_total_simples = (D+R)/Acórdãos, que serve como medida alternativa não ponderada; e

[16] Todas as tabelas podem ser acessadas a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

(e) a variável de perfil da intervenção, Particip_R = R/(D+R), que expressa a participação relativa de recomendações entre as deliberações (quando D+R>0), permitindo verificar se, para níveis mais altos de governança, há deslocamento do controle de comandos mandamentais para orientações colaborativas.

Todos esses dados estão compilados na Tabela 4 — Indicadores secundários por agência e ano. Com esses dados em mãos, passa-se da etapa de construção e padronização dos dados e das variáveis para a etapa inferencial do trabalho, em que se verifica se os padrões observados são compatíveis com as hipóteses formuladas. A partir daqui, portanto, não se "criam" novos dados; apenas se aplicam critérios de comparação às variáveis já definidas (IGG_lag, D, R, Score, Taxa, Taxa_D, Taxa_R, Particip_R e, como medida alternativa não ponderada, Taxa_total_simples).

4. Estratégia empírica: cinco testes

A estratégia empírica deste trabalho adota cinco testes. Eles não pretendem substituir modelos estatísticos avançados, a finalidade é mais modesta e, ao mesmo tempo, metodologicamente coerente: verificar se, dadas as definições adotadas para governança (IGG_lag) e intensidade do controle (Taxa, Taxa_D, Taxa_R e Particip_R), os dados se organizam de maneira sistemática na direção prevista pelas hipóteses.

Em outras palavras, os testes respondem à pergunta: "os números caminham como esperaríamos se as hipóteses fossem plausíveis?", distinguindo dois níveis de resultado. O primeiro nível diz respeito à hipótese principal, H1: "quanto maior a governança, menor a intensidade do controle". O segundo nível diz respeito à hipótese complementar, H2: "maior governança pode deslocar o perfil do controle, com menos determinações (núcleo mandamental) e relativamente mais recomendações (núcleo orientativo)".

Os cinco testes são complementares porque atacam o mesmo problema por ângulos distintos, reduzindo o risco de que uma conclusão dependa de uma única forma de olhar os dados.

O Teste 1 é uma comparação por grupos, que pergunta se, ao dividir a amostra em "IGG alto" e "IGG baixo" por um critério objetivo (a mediana do IGG_lag), o grupo de governança mais elevada exibe, como padrão típico, menor intensidade de controle por acórdão. Esse teste afirma H1 se as taxas de intensidade (especialmente a Taxa e, com ainda mais relevância, a Taxa_D)

CC BY NC ND

forem menores no grupo IGG alto; e afirma H2 se, além disso, a participação relativa de recomendações (Particip_R) for maior no grupo IGG alto.

O Teste 2 é um teste de extremos, em que se compara diretamente as observações no topo e na base da governança (as mais bem colocadas *vs*. as menos bem colocadas em IGG_lag), verificando se a distância entre os extremos reforça o mesmo sinal. O que se espera é que o extremo de maior IGG apresente menor Taxa e menor Taxa_D, e, potencialmente, maior Particip_R; caso o oposto ocorra de maneira recorrente, isso desafirma H1 e enfraquece H2.

O Teste 3 é um teste por posições (*rankings*), em que se ordenam as agências em cada ano de duas formas: **(i)** pela governança (IGG_lag, do maior para o menor) e **(ii)** pela intensidade do controle (Taxa, do menor para o maior, pois "menor Taxa" significa menor intensidade).[17] Se H1 for plausível, as ordens tendem a caminhar em sentidos opostos: agências no topo do IGG aparecem com mais frequência no "lado baixo" de intensidade; se a ordenação não mostrar qualquer alinhamento, ou mostrar alinhamento contrário, isso contraria a hipótese.[18]

Por sua vez, o Teste 4 é uma verificação de robustez por métrica alternativa, que repete as comparações usando uma medida não ponderada da intensidade, Taxa_total_simples = (D+R)/Acórdãos. A função desse teste é controlar o argumento de que o resultado poderia ser "artefato do peso 2" atribuído às determinações. Se H1 também aparecer, ao menos em termos de sinal (IGG mais alto associado a menor taxa), quando se usa a métrica simples (sem pesos), a conclusão ganha robustez; se o sinal desaparecer completamente ou inverter, isso indica que a evidência depende crucialmente do modo de mensuração e deve ser interpretada com mais cautela.

O Teste 5, por fim, é um teste de variação intra-agência em janelas sucessivas (2019→2022 e 2022→2025), no qual cada agência é comparada contra si mesma no tempo, em vez de ser comparada com outras agências. Calcula-se, para cada janela, a mudança na governança (ΔIGG_lag) e as mudanças correspondentes na intensidade e no perfil do controle (ΔTaxa, ΔTaxa_D e ΔParticip_R).

Esse teste afirma H1 quando aumentos de governança dentro da própria agência (ΔIGG_lag>0) são acompanhados, de maneira recorrente, de redução da intensidade proporcional do controle (ΔTaxa<0), sobretudo do

---

[17] EPSTEIN, Lee; MARTIN, Andrew D. *An introduction to empirical legal research*. Oxford: Oxford University Press, 2014. p. 164-170.

[18] *Ibid.*, p. 42-50.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

núcleo mandamental (ΔTaxa_D<0); e afirma H2 quando, mesmo sem queda uniforme da intensidade total, a elevação de governança se associa a um deslocamento do perfil do controle, indicado por aumento da participação relativa de recomendações (ΔParticip_R>0). A relevância do Teste 5 é reduzir o risco de que o resultado seja apenas produto de diferenças estruturais entre setores e agências, fornecendo uma checagem temporal interna da coerência das hipóteses.

Registre-se que os testes não comprovam causalidade entre a variação da governança e a alteração do perfil de controle do TCU.[19]

Cabe ainda dizer que cada teste é municiado por uma série de tabelas. Para melhorar a fluidez do texto, optou-se por colocar as tabelas em anexos específicos cujos nomes são os mesmos dos testes.[20]

4.1 Teste 1 — Comparação por grupos: IGG alto vs. *IGG baixo*

O Teste 1, com dados disponíveis no "Tabela 3 — IGG individual das agências por ano de controle (2018, 2021 e 2024)", parte do painel completo de 33 observações agência-ano e o divide em dois grupos com base na mediana do IGG_lag do próprio painel. A mediana é o valor central que separa a amostra ao meio, metade das observações fica abaixo e metade, acima.

A vantagem desse corte é ser objetivo, afastando o risco de "escolha por conveniência" de um ponto de corte. Neste painel, a mediana do IGG_lag é 0,693734. Assim, o grupo "IGG alto" é composto pelas observações com IGG_lag igual ou superior a 0,693734 (N=17) e o grupo "IGG baixo", pelas observações abaixo desse valor (N=16).

Em seguida, compara-se a intensidade do controle entre os dois grupos por meio das quatro famílias de indicadores, sempre normalizados por exposição (número de acórdãos no ano), o que permite comparar agências em bases proporcionais. Para evitar que a conclusão dependa de um único resumo numérico, o teste compara tanto medianas quanto médias em cada um desses indicadores.

---

[19] Cf. TREVISAN, Giuseppe. Quebrando o paradigma: da correlação à inferência causal. *Advances in Scientific and Applied Accounting*, São Paulo, v. 17, n. 3, p. 1-5, set./dez. 2024. DOI: https://doi.org/10.14392/asaa.2024170301.

[20] Todas as tabelas podem ser acessadas a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

CC BY NC ND

A mediana funciona como o "caso típico" do grupo e é menos sensível a valores muito extremos; a média mostra o nível médio geral, embora possa ser influenciada por observações atípicas. Quando mediana e média apontam na mesma direção, ganha-se confiança de que o padrão não é meramente acidental.

A leitura da "Tabela 5.3 — Comparação por grupos (painel completo): intensidade do controle por IGG alto/baixo" é consistente em dois planos.[21] No plano da hipótese principal (H1), o grupo de IGG alto apresenta intensidade proporcional menor do que o grupo de IGG baixo: a taxa mediana (Taxa) no grupo IGG alto é 0,227, ao passo que no grupo IGG baixo é 0,545, e a mesma direção aparece na média (0,344 no IGG alto contra 0,878 no IGG baixo).

Ou seja, quando se normaliza por acórdão, o conjunto de observações associado a maior governança recebe menos comandos ponderados por decisão, o que é compatível com a ideia de que governança mais elevada se relaciona a menor "necessidade" de intervenção corretiva pelo controlador.

Esse resultado se torna ainda mais significativo quando se observa o núcleo mandamental do controle, isto é, as determinações. A taxa mediana de determinações (Taxa_D) cai de 0,411 no grupo IGG baixo para 0,131 no grupo IGG alto; a média também cai, de 0,712 para 0,203.

Assim, o Teste 1 não indica somente que houve menos deliberações no agregado, é possível observar, também, especificamente um recuo acentuado do componente mais interventivo do controle (determinações) nos casos associados a maior governança, o que reforça a leitura favorável a H1.

No plano da hipótese complementar (H2), a mesma tabela (Tabela 5.3 — Comparação por grupos (painel completo): intensidade do controle por IGG alto/baixo) sugere deslocamento de perfil, pois a participação relativa de recomendações (Particip_R) é maior no grupo IGG alto, tanto na mediana (0,524 contra 0,214) quanto na média (0,559 contra 0,297). Isso significa que, onde a governança é mais elevada, as recomendações ocupam uma parcela maior do conjunto de deliberações (D+R), compatibilizando-se com a ideia de controle como capacitação: menos ênfase em comandos mandamentais e relativamente mais espaço para orientação qualificada.

Note-se que a Taxa_R, por si só, não precisa necessariamente ser menor no grupo IGG alto para que H2 seja plausível, porque a hipótese fala em perfil relativo; ainda assim, a evidência de Particip_R mais alta no IGG alto cumpre

[21] A Tabela 5.3 pode ser acessada a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

exatamente a função de mostrar "como" o controle se distribui, e não apenas "quanto" ele ocorre.

Por cautela metodológica, é importante registrar expressamente o alcance do teste. A comparação por grupos demonstra uma associação consistente entre governança defasada (IGG_lag) e padrões de intensidade/perfil do controle no painel analisado, mas não prova causalidade. Não se pode excluir, apenas com esse teste, que fatores setoriais, momentos regulatórios, prioridades do TCU ou materialidade de políticas públicas influenciem simultaneamente governança e controle.

Ainda assim, como primeiro teste, a evidência é no sentido de que os dados caminham na direção prevista por H1 e, de modo adicional, exibem um padrão compatível com H2, justificando avançar para os testes de extremos, de *rankings* e de robustez por métrica alternativa, e de variação interna que funcionam como verificações independentes do mesmo raciocínio com lentes igualmente transparentes.

4.2 Teste 2 — Comparação de extremos: Top 6 x Bottom 6

O Teste 2 é um "teste de extremos". Ao invés de comparar toda a amostra por um corte central como no Teste 1, agora se comparam diretamente as observações do topo e da base da governança. A lógica é que, se a hipótese H1, "maior governança → menor intensidade do controle", for verdadeira como tendência empírica, então ela deve aparecer com mais nitidez quando comparamos os casos mais governados com os menos governados. Para manter o critério objetivo e replicável, definiu-se o extremo superior como as seis observações (agência-ano) com maior IGG_lag e o extremo inferior como as seis observações com menor IGG_lag, dentro do painel de 33 observações (ver: ANEXO III — Tabelas essenciais para o Teste 2[22] (Comparação de extremos: Top 6 x Bottom 6)).

Neste teste, H1 é afirmada se o extremo superior de IGG_lag apresentar Taxa e, especialmente, Taxa_D menores do que o extremo inferior. H2 é afirmada se, além disso, o extremo superior apresentar Particip_R maior do que o extremo inferior (mais recomendações relativamente às deliberações), mesmo que o volume total de tarefas não caia de modo uniforme.

[22] O Anexo III pode ser acessado a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

Os resultados aparecem com clareza na tabela principal de comparação entre extremos (Tabela 6.3 — Resumo comparativo (mediana e média)). No extremo superior (Top 6), o IGG_lag varia de 0,823518 a 0,903155; no extremo inferior (Bottom 6), de 0,084165 a 0,531559. Quando se olha a intensidade proporcional do controle, a taxa mediana no Top 6 é 0,321 contra 0,586 no Bottom 6; e a taxa média é 0,503 contra 1,287.

Isso significa que, nos casos do topo de governança, o "peso" dos comandos tende a ser menor do que nos casos da base. O componente mais relevante para H1, por expressar maior intensidade decisória, é a Taxa_D, cuja mediana é 0,271 no Top 6 e 0,375 no Bottom 6; a média é 0,351 no Top 6 e 1,071 no Bottom 6. O contraste nas médias é particularmente importante porque evidencia que, entre os casos de IGG muito baixo, há observações com intervenção mandamental muito elevada que puxam o nível médio para cima (o que é compatível com a ideia de que baixa governança se associa, ao menos em parte dos casos, a maior carga mandamental).

Quanto à hipótese H2 (controle como capacitação), o indicador mais direto aqui é a Particip_R. A mediana da participação relativa de recomendações no Top 6 é 0,417, enquanto no Bottom 6 é 0,126; e a média é 0,453 contra 0,265. Em linguagem substantiva: nos extremos de maior governança, as recomendações tendem a ocupar uma fatia maior das deliberações (quando existem deliberações), o que é coerente com a leitura de um perfil relativamente mais orientativo/capacitador.

É verdade que as taxas de recomendações por acórdão (Taxa_R) não necessariamente sobem no Top 6 (a mediana é 0,074 no Top 6 contra 0,107 no Bottom 6), mas isso não contradiz H2. O que se traz como hipótese não é "mais recomendações em termos absolutos", e sim maior participação relativa das recomendações entre as deliberações (menos mandamentos e, proporcionalmente, mais recomendações qualificadas), o que é exatamente o padrão observado.

Para fins de transparência e controle de leitura, o documento também traz duas tabelas de checagem por totais. No Top 6, há 205 acórdãos somados, com 53 determinações e 57 recomendações (*Score* total 163) (Tabela 6.1 — Top 6 (maiores valores de IGG_lag));[23] no Bottom 6, há 167 acórdãos, com 109 determinações e 47 recomendações (*Score* total 265) (Tabela 6.2 — Bottom 6 (menores valores de IGG_lag)).

[23] A Tabela 6.1 pode ser acessada a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

Essa checagem é útil porque mostra que o Bottom 6 concentra não apenas maior intensidade proporcional, mas também maior carga mandamental em termos absolutos (D_total), ainda que com menos acórdãos no total, reforçando a leitura de que o padrão não é um simples artefato de "aparecer mais".

Em síntese, o Teste 2 reforça H1 na medida em que a comparação entre extremos aponta que, nos casos de IGG_lag alto, a intensidade proporcional do controle é menor, sobretudo no núcleo mandamental (Taxa_D), enquanto nos casos de IGG_lag muito baixo aparecem níveis mais altos de controle proporcional e mandamental. O mesmo teste também fornece evidência favorável a H2, porque a participação relativa de recomendações é significativamente maior no extremo superior de governança.

*4.3 Teste 3 — Posições (*rankings*) por ano: governança vs. intensidade do controle*

O Teste 3 é um teste por posições (*rankings*): em cada ano (2019, 2022 e 2025), as agências são ordenadas em três dimensões e, depois, observa-se se essas ordenações "conversam" entre si de modo compatível com as hipóteses (ver: ANEXO IV[24] — Tabelas essenciais para o Teste 3 — Posições (*rankings*) por ano: governança *vs.* intensidade do controle).

A vantagem do teste é sua transparência porque ele não depende de pesos adicionais, nem de suposições estatísticas; ele apenas pergunta se, quando colocamos as agências "em fila", a fila da governança se aproxima (ou não) da fila da intensidade do controle. A limitação é igualmente clara, *rankings* são uma medida "grosseira" (uma posição pode mudar por diferenças pequenas) e não capturam bem magnitudes; por isso ele funciona como evidência complementar.

A primeira ordenação é a de governança, medida pelo IGG_lag (isto é, o IGG do ciclo imediatamente anterior pareado ao ano do controle: 2018→2019; 2021→2022; 2024→2025). Nessa fila, Rank_IGG = 1 significa "maior governança naquele ano". A segunda ordenação é a de intensidade do controle, medida pela Taxa = *Score*/Acórdãos, em que *Score* = 2×D + R.

[24] O Anexo IV pode ser acessado a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

Como "menor taxa" significa "menor intensidade proporcional", definiu-se Rank_Taxa = 1 para a menor Taxa (isto é, menor intensidade). A terceira ordenação, mais estritamente vinculada à ideia de intrusão decisória, é a do núcleo mandamental: Taxa_D = (2×D)/Acórdãos, também com Rank_Taxa_D = 1 para a menor Taxa_D. Por fim, para dialogar com a hipótese H2 (perfil capacitador), usa-se Particip_R = R/(D+R) (quando há deliberações), e, aqui, o "melhor" é o maior valor (isto é, maior presença relativa de recomendações), de modo que Rank_Particip_R = 1 significa maior participação de recomendações.

Com essas filas construídas, o teste afirma H1 quando, de maneira recorrente, as agências com melhores posições em governança (Rank_IGG mais baixo) aparecem também com melhores posições em intensidade (Rank_Taxa e Rank_Taxa_D mais baixos). Se, ao contrário, as agências mais bem governadas aparecem sistematicamente entre as mais intensamente controladas (Ranks de intensidade altos), o teste desafirma H1.

Já H2 tende a ser corroborada quando agências com melhor Rank_IGG aparecem também com melhores posições em Particip_R (Rank_Particip_R mais baixo), sugerindo deslocamento do perfil de intervenção para um padrão relativamente mais recomendatório.

Os resultados do Teste 3, quando olhados ano a ano, são mistos. A tabela de síntese mostra, como medida simples de alinhamento, a média da distância entre posições (por exemplo, a média do módulo da operação Rank_IGG − Rank_Taxa) (Tabela 7.1 — Síntese de alinhamento entre *rankings* (por ano)).

Em 2019 (ver: Tabela 7.2 — *Rankings* por agência no ano de 2019), essa distância média é menor (aprox. 1,909 para Taxa e 1,727 para Taxa_D), sugerindo um alinhamento moderado entre governança e menor intensidade naquele ano. Em 2022 e 2025 (ver: Tabela 7.3 — *Rankings* por agência no ano de 2022 e Tabela 7.4 — *Rankings* por agência no ano de 2025),[25] entretanto, a distância média cresce (por exemplo, 3,636 e 3,818 para Taxa; 3,545 e 3,909 para Taxa_D), o que indica que, nesses anos, as filas não caminham tão próximas: há mais casos em que a posição de governança não "prediz" bem a posição de intensidade.

Uma forma ainda mais direta de ler esse teste é olhar os Top 3 de governança e perguntar se eles aparecem também entre os Top 3 de menor

[25] As Tabelas 7.1; 7.2 e 7.3 podem ser acessadas a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

intensidade. Essa checagem aparece na Tabela 7.1 — Síntese de alinhamento entre *rankings* (por ano).

O padrão encontrado é que, em cada ano, a interseção entre Top 3 (IGG) e Top 3 (menor Taxa) é 1; e o mesmo ocorre para Top 3 (menor Taxa_D), que também é 1. Isso significa que existe, em cada ano, ao menos um caso em que "alta governança" anda junto com "baixa intensidade proporcional", mas esse alinhamento não é amplo o suficiente para dizer que, por *rankings* anuais, o padrão é geral para todas as agências.

Na prática, há um "caso-âncora" de alinhamento consistente (que aparece repetidamente) e um conjunto maior de agências em que o ordenamento não coincide.

Quanto à H2, o mesmo raciocínio se aplica. Se a governança mais alta implicasse, de modo robusto, um perfil mais recomendatório, seria esperado que Top 3 (IGG) tivesse interseção alta com Top 3 (maior Particip_R). O resultado é novamente heterogêneo: em 2019 e 2022 a interseção é 1, e em 2025 é 0 (Tabela 7.1). Isso não invalida H2 (porque o Teste 3 é mais sensível a mudanças pequenas e a empates/variações discretas), mas indica que, por *rankings* anuais estritos, o deslocamento do perfil não aparece de forma estável no topo da governança em todos os anos.

Em síntese, o Teste 3 não é o teste mais favorável às hipóteses quando formulado como ordenação anual pura, posto que ele sugere que a relação entre governança e intensidade/perfil do controle é mais heterogênea do que uma leitura linear deixaria supor, variando entre anos e entre agências.

Ainda assim, ele é útil metodologicamente porque impede que a narrativa dependa apenas de médias/medianas e mostra, com nitidez, que o fenômeno pode ser condicionado por fatores não observados aqui.

4.4. Teste 4 — Robustez por métrica alternativa não ponderada

O Teste 4 tem uma função de checagem de robustez do resultado. Nos testes anteriores, a intensidade do controle foi medida principalmente por uma métrica ponderada (*Score* = 2×D + R), que atribui peso maior às determinações por refletirem, em regra, um comando mais intrusivo e mandamental.

Como essa escolha de pesos pode ser questionada ("e se o resultado for apenas um efeito do peso 2?"), o Teste 4 refaz as comparações usando uma métrica deliberadamente mais simples e neutra, usando a Taxa_total_simples = (D+R)/Acórdãos.

CC BY NC ND

Em termos substantivos, essa taxa responde à pergunta mais direta possível: "em média, quantas deliberações (determinações + recomendações) aparecem por acórdão que menciona a agência?", sem diferenciar intensidade mandamental *versus* orientativa.

Se a hipótese H1 (maior governança associada a menor intensidade) permanecer com o mesmo sinal sob essa mensuração alternativa, a evidência ganha força, porque deixa de depender crucialmente da ponderação escolhida. Os dados relacionados ao Teste 4 foram reunidos no anexo V — Tabelas essenciais para o Teste 4 — Métrica alternativa não ponderada.[26]

A primeira leitura do Teste 4 replica a lógica do Teste 1, mas substituindo a Taxa ponderada pela Taxa_total_simples. Ao dividir o painel pelo mesmo critério objetivo da mediana do IGG_lag (0,693734), observa-se que o grupo de IGG alto apresenta Taxa_total_simples mediana de 0,213 (aprox.) e média de 0,242, enquanto o grupo de IGG baixo apresenta mediana de 0,359 e média de 0,522. Ou seja: mesmo quando se ignora completamente a diferença de "peso" entre determinação e recomendação, a governança mais alta continua associada a menos deliberações por acórdão, em termos típicos (mediana) e em termos médios.

Isso constitui evidência favorável à H1 por um caminho menos dependente das escolhas normativas embutidas na ponderação.

A segunda leitura aplica a lógica do Teste 2 (extremos), novamente usando a Taxa_total_simples. Comparando-se as seis observações com maior IGG_lag com as seis de menor IGG_lag, a diferença permanece nítida: no Top 6, a mediana da Taxa_total_simples é 0,186 e a média é 0,328; no Bottom 6, a mediana é 0,427 e a média é 0,751. Em termos interpretativos, isso significa que, nos polos de governança, a "distância" entre o topo e a base também aparece quando se mede intensidade de modo não ponderado, reforçando a leitura de que o sinal favorável à H1 não é um fruto do peso 2 atribuído às determinações (Ver: Tabela 8.4 — Observações do extremo superior (Top 6 IGG_lag) — Taxa_total_simples; e Tabela 8.5 — Observações do extremo inferior (Bottom 6 IGG_lag) — Taxa_total_simples).[27]

Esse teste fala principalmente sobre H1 (volume proporcional de deliberações) e não é o melhor instrumento para examinar H2 porque a

[26] O Anexo V pode ser acessado a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

[27] As Tabelas 8.4 e 8.5 podem ser acessadas a partir do seguinte link: https://drive.google.com/drive/folders/1ir5--SX0wdz1yFRsf6aYvi-tRNkJbm7U?usp=sharing.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

Taxa_total_simples soma D e R sem decompor o conteúdo. Ainda assim, ele cumpre um papel importante no conjunto ao proteger a conclusão de H1 contra a objeção metodológica de que o resultado poderia estar excessivamente ancorado na ponderação adotada.

Dito de outra maneira, se H1 continuasse aparecendo apenas com a métrica ponderada e desaparecesse com a métrica simples, seria necessário tratar a evidência com grande cautela; como o sinal se mantém também na Taxa_total_simples, a inferência associativa se torna mais estável.

4.5. Teste 5 — Variação intra-agência em janelas sucessivas

O Teste 5 foi concebido para responder a uma pergunta que não é plenamente capturada quando se comparam agências diferentes entre si: quando a mesma agência melhora (ou piora) sua governança, o padrão de intensidade e o perfil do controle do TCU se alteram na direção prevista pelas hipóteses?

Em vez de contrastar níveis de controle entre entidades distintas, que podem ser afetados por diferenças relativamente estáveis entre setores regulados, materialidade econômica, complexidade e agenda fiscalizatória, o teste observa mudanças internas no tempo, tomando cada agência como seu próprio parâmetro.

Trata-se, então, de um exame de consistência direcional que verifica se as variações observadas caminham, de modo recorrente, na direção esperada por H1 (maior governança associada a menor intensidade do controle) e por H2 (maior governança associada a deslocamento do perfil do controle, com menor peso de determinações e maior participação relativa de recomendações).

Conforme o pareamento defasado descrito na seção 2.3, o Teste 5 é aplicado em duas janelas sucessivas: 2019→2022 e 2022→2025. Essa opção evita misturar movimentos de ciclos distintos e preserva a lógica de comparação entre ciclos.

Na operacionalização, o Teste 5 utiliza as mesmas variáveis já definidas no desenho do estudo, agora expressas como variações (Δ). Para cada agência, calcula-se a variação de governança ΔIGG_lag na janela (ano final − ano base). Em paralelo, calcula-se a variação da intensidade proporcional do controle por acórdão, com base em taxas normalizadas pela exposição anual (número de acórdãos que mencionam a agência no respectivo ano). Os resultados estão expostos como tabelas no Anexo VI — Tabelas essenciais para o Teste 5 — Variação intra-agência em janelas sucessivas.

CC BY NC ND

Em termos interpretativos, H1 é favorecida quando aumentos de governança dentro da agência (ΔIGG_lag > 0) se acompanham de queda da intensidade proporcional (ΔTaxa < 0), sobretudo do componente mandamental (ΔTaxa_D < 0). Já H2 é favorecida quando aumentos de governança se acompanham de deslocamento do perfil do controle, indicado por queda do componente mandamental e aumento da participação relativa de recomendações (ΔParticip_R > 0), ainda que o volume agregado ponderado não caia de modo uniforme.

Como se trata de um teste descritivo, a leitura combina a análise agência a agência com uma síntese por recorrência de sinais, isto é, quantas vezes, diante de aumento de IGG_lag, as métricas caminham na direção prevista.

Na janela 2019→2022 (ver: Tabela 9.1.a — Variações (2019→2022) — Bloco 1; e Tabela 9.1.b — Variações (2019→2022) — Bloco 2), o padrão observado é fortemente compatível com H1. Sete agências registram aumento de governança (ΔIGG_lag > 0): ANA, Anatel, Aneel, ANM, ANP, Antaq e ANTT. Dentro desse conjunto, em seis há redução da intensidade proporcional do controle (ΔTaxa < 0), e também redução do núcleo mandamental (ΔTaxa_D < 0): Anatel, Aneel, ANM, ANP, Antaq e ANTT.

A única exceção relevante é ANA, na qual a governança aumenta, mas a intensidade ponderada por acórdão e o componente mandamental também aumentam (ΔTaxa > 0 e ΔTaxa_D > 0), sinal contrário ao previsto por H1.

O resultado desse intervalo, portanto, sugere que, para a maior parte das agências, a elevação de governança entre 2019 e 2022 se associa à redução da intensidade do controle em termos proporcionais, sobretudo no eixo mais mandamental, o que fornece evidência direcional robusta em favor de H1 nesse período.

Ainda em 2019→2022, a leitura de H2 é mais heterogênea, mas registra sinais que não podem ser desconsiderados. Entre as agências cuja governança aumentou, a participação relativa de recomendações cresce em Anatel, Aneel, Antaq e ANTT (ΔParticip_R > 0), permanece inalterada em ANM (ΔParticip_R = 0) e diminui em ANA (ΔParticip_R < 0).

No caso de ANP, a participação relativa não é computável em 2022 porque, naquele ano, não houve deliberações relevantes (D+R = 0), de modo que não existem determinações e recomendações.

Esses resultados são compatíveis com a hipótese de que, em parte significativa dos casos, aumentos de governança se associam não apenas a menor intensidade, mas também a maior presença relativa de recomendações; ao mesmo tempo, as exceções indicam que o perfil do controle pode variar conforme condições setoriais e prioridades de fiscalização no período.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

Na janela 2022→2025 (ver: Tabela 9.2a — Variações (2022→2025) — Bloco 1, Tabela 9.2b — Variações (2022→2025) — Bloco 2), o comportamento de H1 se torna menos uniforme, mas os dados ainda assim apontam na direção prevista. Nove agências apresentam aumento de governança (ΔIGG_lag > 0): ANA, Anac, Anatel, Ancine, ANM, ANS, Antaq, ANTT e Anvisa.

Entre elas, a intensidade proporcional do controle diminui em cinco (ΔTaxa < 0): ANA, Anac, Ancine, ANTT e Anvisa; e aumenta em quatro (ΔTaxa > 0): Anatel, ANM, ANS e Antaq.

Assim, nesse segundo intervalo, o aumento de governança não vem acompanhado, de modo predominante, de queda da intensidade total ponderada por acórdão, o que recomenda qualificar o suporte empírico a H1 como parcial nessa janela.

O dado relevante, além da existência de casos contrários, é também a própria heterogeneidade temporal, a coerência direcional forte observada em 2019→2022 se enfraquece em 2022→2025, sugerindo a influência de fatores contextuais que variam no tempo. Esse é um elemento que esta pesquisa não se propõe a analisar, mas que não pode ser deixado de lado.

Por outro lado, em 2022→2025 (ver: Tabela 9.2a — Variações (2022→2025) — Bloco 1, Tabela 9.2b — Variações (2022→2025) — Bloco 2), os dados favorecem com mais clareza a hipótese H2, pois o deslocamento de perfil aparece mesmo quando a intensidade total não cai de forma uniforme.

Entre as nove agências com aumento de governança, a participação relativa de recomendações aumenta em seis (ΔParticip_R > 0): ANA, Anac, ANM, ANS, Antaq e ANTT; permanece inalterada na Anvisa (ΔParticip_R = 0); e diminui em outras duas (Anatel e Ancine). Em termos substantivos, isso indica que, em boa parte dos casos, aumentos de governança se associam a um controle relativamente mais orientativo, com maior peso de recomendações no conjunto das deliberações relevantes, ainda que o volume ponderado de intervenções não apresente queda monotônica.

Em síntese, o Teste 5 pode ser feito pela contagem de alinhamentos esperados. Na janela 2019→2022, entre as sete agências em que a governança subiu, seis exibiram queda de intensidade (ΔTaxa < 0) e seis exibiram queda do núcleo mandamental (ΔTaxa_D < 0), o que configura alinhamento forte com H1 nesse intervalo. Na janela 2022→2025, entre as nove agências em que a governança subiu, cinco exibiram queda de intensidade (ΔTaxa < 0), mas seis exibiram aumento da participação relativa de recomendações (ΔParticip_R > 0), reforçando sobretudo H2.

CC BY NC ND

Em conjunto, o teste sugere que a governança se associa à redução clara da intensidade proporcional do controle (especialmente mandamental) e, em outros, a um deslocamento do perfil para uma atuação relativamente mais orientativa.

Em qualquer caso, deve-se registrar a cautela de que o Teste 5 identifica associações temporais internas consistentes com as hipóteses, mas não prova, nem poderia provar, causalidade, pois diversos outros fatores podem alterar simultaneamente governança e padrão de controle no período analisado, fatores esses que não foram contabilizados nesta pesquisa.

5. Limitações e alcances

As conclusões deste artigo devem ser lidas com ambição limitada, afinal, trata-se de um desenho empírico voltado a descrever padrões, não a estimar efeitos causais. A utilidade está em traduzir um debate dogmático recorrente em indicadores observáveis de intervenção do controle externo e por isso mesmo é necessário indicar alguns limites da pesquisa.

O primeiro limite é o tamanho e o formato do painel (11 agências, três anos). Em base pequena, observações atípicas influenciam médias e posições, e a análise retrata apenas três "fotografias" temporais (2019, 2022 e 2025), não uma dinâmica contínua do controle ao longo de muitos anos.

O segundo limite é a mensuração de governança: o IGG é indicador-síntese por ciclos e questionário, útil e padronizado, mas ainda um *proxy* agregado; ele não captura integralmente a complexidade setorial nem a qualidade substantiva de cada decisão regulatória. O uso de defasagem (lag) melhora a coerência temporal, contudo não elimina endogeneidade nem fatores omitidos que podem afetar simultaneamente governança e controle.

O terceiro limite é a mensuração do desfecho visto que determinações e recomendações são *proxies* replicáveis (e tipificadas), que podem não capturar integralmente o alcance da intervenção do TCU nem a severidade substantiva de cada comando, além do mais, a normalização por número de acórdãos corrige diferenças de exposição, mas não mede risco/materialidade real e pode ser sensível à composição do denominador.

O alcance, apesar disso, é relevante posto que o artigo oferece um modo transparente de observar quando o controle tende a recuar no mandamento e avançar na orientação, contribuindo para qualificar a discussão sobre deferência condicionada à qualidade institucional e procedimental das decisões.

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

6. Conclusão

O problema de pesquisa deste trabalho foi formulado de modo a identificar se há associação sistemática entre a governança institucional das agências reguladoras federais, medida pelo IGG com defasagem (IGG_lag), e a intensidade e o perfil do controle exercido pelo TCU, aproximados pela incidência de determinações e recomendações normalizadas pelo número anual de acórdãos em que a agência figura como unidade jurisdicionada. Ao traduzir um debate jurídico clássico em indicadores observáveis, o artigo buscou produzir evidência empírica sobre padrões de intervenção do TCU nas agências reguladoras, sem pretender substituir a dogmática nem operar como prova de causalidade.

No conjunto do painel (33 observações agência-ano), os resultados apontam para uma associação consistente na direção da hipótese H1: níveis mais altos de governança tendem a coexistir com menor intensidade proporcional do controle, sobretudo quando se observa o núcleo mandamental (Taxa_D) e não apenas o volume agregado de deliberações. Essa leitura é favorecida principalmente pelos testes que comparam níveis — comparação por grupos (Teste 1) e comparação de extremos (Teste 2) — e é reforçada pela métrica alternativa não ponderada (Teste 4), o que reduz a possibilidade de que o sinal seja mero efeito do peso atribuído às determinações no *score*.

No que se refere à hipótese H2, os achados são, em geral, mais convergentes quando se focaliza o "como" do controle. De modo recorrente, maior governança aparece associada a menor peso do componente mandamental e a maior participação relativa de recomendações entre as deliberações (Particip_R), sugerindo deslocamento do perfil do controle para uma atuação comparativamente mais orientativa e capacitadora, ainda que não se observe queda uniforme do volume total em todas as leituras e janelas.

O ponto de maior cautela está no teste por *rankings* anuais (Teste 3) e na heterogeneidade temporal captada na variação intra-agência (Teste 5). A evidência não é uniforme quando o fenômeno é observado por ordenações anuais estritas ou ao longo do tempo dentro de cada agência, o que impõe qualificação: o padrão existe, mas é sensível a fatores setoriais e conjunturais não mensurados neste desenho. A leitura adequada, portanto, é de associação empírica compatível com H1 e H2, com variações relevantes entre anos e entre agências, predominando o suporte a H1 nos testes de nível e um suporte mais estável a H2 no perfil do controle.

CC BY NC ND

É indispensável explicitar o alcance inferencial do trabalho: os testes aplicados descrevem associações e consistências direcionais, mas associação não implica causalidade. O painel é pequeno (11 agências em três anos), o IGG é *proxy* de governança agregada por questionário e ciclos, e a intensidade do controle foi medida por deliberações que, embora observáveis e tipificadas, não capturam integralmente a severidade substantiva de cada comando nem todas as formas de intervenção relevantes do TCU. Além disso, diferenças de materialidade, risco setorial, complexidade regulatória, agenda de fiscalização e eventos críticos podem afetar simultaneamente governança e padrões de controle, limitando qualquer leitura causal forte com o desenho disponível.

Ainda assim, o trabalho oferece duas contribuições. A primeira é metodológica: demonstra que o debate sobre deferência, controle do mérito e limites da discricionariedade técnica pode ser complementado por um olhar empírico, com regras explícitas de contagem e normalização, aproximando a discussão de padrões observáveis de intervenção. A segunda é substantiva: os dados sugerem que a governança institucional, medida pelo IGG_lag, é variável relevante para qualificar como o controle externo se manifesta, especialmente pela distinção entre comandos mandamentais e orientações colaborativas, o que pode ajudar a refinar argumentos sobre deferência condicionada à qualidade institucional e procedimental da decisão administrativa.

Em suma, maior governança aparece associada a menor carga mandamental e a um perfil relativamente mais recomendatório, com heterogeneidades que recomendam prudência interpretativa.

Referências

BRASIL. Lei nº 13.848, de 25 de junho de 2019. Dispõe sobre a gestão, a organização, o processo decisório e o controle social das agências reguladoras. *Diário Oficial da União*, Brasília, DF, 26 jun. 2019. Disponível em: www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/l13848.htm. Acesso em: 15 fev. 2026.

BRASIL. Supremo Tribunal Federal. Mandado de Segurança nº 24.883 AgR. Rel. min. Roberto Barroso. Tribunal Pleno, julgado em 24 fev. 2017. *Diário da Justiça Eletrônico*, n. 78, 19 abr. 2017.

BRASIL. Tribunal de Contas da União. *Acórdão nº 1.913/2024 — Plenário*. Relatório de levantamento para avaliar e promover as práticas de Governança, Sustentabilidade e Gestão nas organizações públicas federais. Relator:

ISSN 0034-8007 • e-ISSN 2238-5177 • DOI: 10.12660/rda.v285.2026.98378

ministro Vital do Rêgo. Brasília, DF: TCU, 18 set. 2024. Disponível em: https://pesquisa.apps.tcu.gov.br/.

BRASIL. Tribunal de Contas da União. *Portal do TCU*. Brasília, DF: TCU, 2026. Disponível em: https://portal.tcu.gov.br/. Acesso em: 22 fev. 2026.

BRASIL. Tribunal de Contas da União. Resolução TCU nº 315, de 1º de julho de 2020. Dispõe sobre a elaboração de deliberações que contemplem medidas a serem tomadas pelas unidades jurisdicionadas no âmbito do Tribunal de Contas da União e dá outras providências. *Diário Oficial da União*, Brasília, DF, 6 jul. 2020. Disponível em: https://pesquisa.apps.tcu.gov.br/doc/norma/Resolu%C3%A7%C3%A3o/TCU/315/2020. Acesso em: 15 fev. 2026.

DI PIETRO, Maria Sylvia Zanella. *Direito Administrativo*. 37. ed. Rio de Janeiro: Forense, 2024. Edição do Kindle.

EPSTEIN, Lee; MARTIN, Andrew D. *An introduction to empirical legal research*. Oxford: Oxford University Press, 2014.

FERREIRA, R. F. *Intensidade do controle e governança*: um (possível) padrão de comportamento do TCU no controle das agências reguladoras. Dissertação (mestrado em direito) — Escola de Direito do Rio de Janeiro, Fundação Getulio Vargas, Rio de Janeiro, 2024. Disponível em: https://repositorio.fgv.br/bitstreams/6dc9e7a4-0299-4a77-8dfc-2ea3c34018d0/download. Acesso em: 22 fev. 2026.

MERLOTTO, Nara Carolina. *A atuação do Tribunal de Contas da União sobre as agências reguladoras*: entre a independência e o controle. Dissertação (mestrado em direito) — Faculdade de Direito, Universidade de São Paulo, São Paulo, 2020. Disponível em: www.teses.usp.br/teses/disponiveis/2/2134/tde-25092020-144220/publico/6490332_Dissertacao_Parcial.pdf. Acesso em: 22 fev. 2026.

PEREZ, Marcos Augusto. *Testes de legalidade*: métodos para o amplo controle da discricionariedade administrativa. Belo Horizonte: Fórum, 2020.

REGO, Vicente Pereira do. *Elementos de direito administrativo brasileiro comparado com o direito administrativo francez*. [*S. l.*]: Typographia Universal, 1857. Disponível em: https://bibliotecadigital.stf.jus.br/xmlui/handle/123456789/1215. Acesso em: 13 fev. 2026.

ROMAN, Flavio José. *Discricionariedade técnica na regulação econômica*. São Paulo: Saraiva, 2013.

TREVISAN, Giuseppe. Quebrando o paradigma: da correlação à inferência causal. *Advances in Scientific and Applied Accounting*, São Paulo, v. 17, n. 3, p. 1-5, set./dez. 2024. DOI: https://doi.org/10.14392/asaa.2024170301.

Informações adicionais
Additional information

Editores responsáveis ***Handling editors***	
Editor-chefe / ***Editor-in-chief***	Sérgio Guerra
Editor assistente / ***Assistant editor***	Luis Lopes Martins

Declaração de disponibilidade de dados
Data availability statement
Não foram gerados ou analisados conjuntos de dados no âmbito deste estudo. O trabalho baseia-se exclusivamente em pesquisa bibliográfica e documental. Informações adicionais podem ser solicitadas aos autores, quando aplicável.
No datasets were generated or analyzed in the course of this study. The research is based exclusively on bibliographic and documentary sources. Additional information may be requested from the authors, where applicable.

Licenciamento
Licensing
Este artigo está licenciado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivação 4.0 Internacional (CC BY-NC-ND 4.0).
This article is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).