

1 METODOLOGIA

Este capítulo dedica-se a apresentar a proposta metodológica da tese. Como vimos, estudar os efeitos do lobby apresenta desafios metodológicos importantes. Primeiro, sistematizemos os seis principais complicadores para identificar os efeitos do lobby; em seguida, apresento a proposta de desenho metodológico para buscar essa identificação.

O primeiro desafio, que podemos chamar de "o problema da persistência temporal", relaciona-se com o fato de que o montante dos esforços de lobby ao longo do tempo tende a ser constante (FIGUEIREDO; RICHTER, 2014). Por conta disso, alguns métodos de análise de dados em painel, que exigem variação nas variáveis de interesse, cujo objetivo seja medir os efeitos do esforço de lobby, pode não obter a variabilidade necessária.

A segunda complicação metodológica, ou "o problema do viés de seleção", refere-se ao fato de que a escolha dos alvos dos lobistas não é aleatória. Os lobistas tendem a escolher como alvos prioritários parlamentares aliados (KOLLMAN, 1998; CALDEIRA; HOJNACKI; WRIGHT, 2000; HOJNACKI; KIMBALL, 2001) ou neutros (HOLYOKE, 2003; KELLEHER; YACKEE, 2009; BERTRAND; BOMBARDINI; TREBBI, 2014; GAWANDE; KRISHNA; OLARREAGA, 2012) que tenham posições importantes na tramitação de determinado projeto (MARSHALL, 2010).

O terceiro desafio, que pode ser denominado de "o problema do lobby contra-ativo", diz respeito a possibilidade de que, mesmo que um grupo não obtenha sucesso, não quer dizer que ele não tenha influência. Pode significar que houve um esforço de lobby no sentido contrário que obteve sucesso (AUSTEN-SMITH; WRIGHT, 1996).

O quarto obstáculo, ou "o problema das faces do poder", dispõe que o lobby pode agir em diferentes etapas do processo de decisão, seja na formação da agenda (BACHRACH; BARATZ, 1962), seja na sua implementação.

A quinta dificuldade, ou "o problema das variáveis omitidas" (FIGUEIREDO; RICHTER, 2014), está relacionado a possibilidade de que variáveis importantes podem não ser observadas. Como exemplo, podemos citar a habilidade de negociação de cada lobista. Outra possibilidade relacionada a esse desafio é caso os lobistas queiram pressionar de maneira não transparente.

Por fim, a sexta complicação, que chamemos de "o problema dos canais de influência", refere-se a capacidade de que os lobistas tem de agir em diferentes canais (DÜR, 2008). Eles podem optar desde lobby direto a campanhas de pressão perante a opinião pública, ou *outside lobbying* (KOLLMAN, 1998). Além disso podem mesmo agir na própria seleção dos tomadores de decisão (FORDHAM; MCKEOWN, 2003). A seguir, analisemos a proposta para a identificação do efeito do lobby no comportamento parlamentar.

1.1 ESTRATÉGIA DE IDENTIFICAÇÃO E DE OPERACIONALIZAÇÃO

Muitos trabalhos sobre os efeitos lobby preocupam-se em medir os resultados de políticas (como impacto em tarifas comerciais), de volume de contratos com o poder público, benefícios fiscais, entre outras variáveis dependentes amplas. Ora, os determinantes desses resultados são complexos. Ao focarmos no comportamento parlamentar conseguimos controlar melhor os seus determinantes.

O comportamento parlamentar pode se dar de diferentes formas. Os trabalhos que buscam estudar os efeitos sobre os votos dos parlamentares, caem no problema de que o voto possui, também, determinantes complexos, tais como o partido, relação com os pares, entre outros. Buscando minimizar esse desafio, este trabalho analisa a Atividade Legislativa (AL), isto é, o quão ativo determinado parlamentar é em determinado tema. A atividade, portanto, envolve a proposição de projetos, requisições, emendas e discursos. a AL, então, é majoritariamente determinada pelos interesses dos parlamentares. Não quero dizer que as variáveis como partido e relação entre os pares não influenciam, apenas defendo que a AL está mais fortemente correlacionada com os objetivos de um parlamentar, como

visto no *framework* de análise do comportamento parlamentar (??). Assim buscamos minimizar os impactos do problema das variáveis omitidas. De tal forma que esperamos que o AL seja uma função de características individuais dos parlamentares, dos países pelos quais os eurodeputados são eleitos e dos partidos dos quais fazem parte.

Consideremos então que:

$AL_{1,idt}$ = o ativismo legislativo do parlamentar i , caso tenha recebido esforço de lobby significativo no domínio temático d no tempo t ; e

$AL_{0,idt}$ = o ativismo legislativo do parlamentar i , caso *não* tenha recebido esforço de lobby significativo no domínio temático d no tempo t .

Esses resultados são, contudo, potenciais, isto é, não observamos AL_{1idt} e AL_{0idt} ao mesmo tempo, apenas um ou outro. Partindo do *framework* de análise do comportamento parlamentares, esperamos que:

$$E(AL_{0,idt}|c,p,d,t,X) = \gamma_{ct} + \lambda_{pt} + \theta_{dt} + X'_{it}\delta \quad (1.1)$$

A equação 1.1 denota que, na ausência de pressão significativa de lobby, a AL é determinada pela soma dos efeitos:

- γ_{ct} : captura o efeito específico do país no tempo t . Ele leva em conta fatores que podem influenciar o ativismo legislativo em todos os partidos dentro de um determinado país em um momento específico (por exemplo, sistema eleitoral, características do eleitorado, condições econômicas de longo prazo, etc.);
- λ_{pt} : representa o efeito específico do partido no tempo t . Ele leva em conta fatores que podem influenciar o ativismo legislativo em todos os países para um determinado partido em um momento específico (por exemplo, ideologia do partido.);
- θ_{dt} : captura o efeito do domínio temático no tempo t , tais como a competitividade, saliência, etc.

- $X'_{cpt}\delta$: este termo leva em conta os efeitos de outros fatores relevantes (variáveis de controle) capturados no vetor X específicos de cada parlamentar, como expertise, experiência, gênero, etc.

Esperamos que os efeitos de interação entre país e partido seja zero, uma vez que os grupos políticos se organizam transnacionalmente. Por conta disso, não incluí um termo que capturasse um efeito do partido no ativismo legislativo que variasse em diferentes países.

Considerando que o esforço de lobby varia de acordo com o domínio temático, podemos utilizar o método da Tripla Diferença (DDD), cuja especificação se daria por:

$$\begin{aligned}
 AL_{idt} = & \gamma_{ct} + \lambda_{pt} + \theta_{dt} \\
 & + \beta_1 L_{idt} \\
 & + \beta_2 (L_{idt} * T_{dt}) \\
 & + \beta_3 (L_{idt} * D_d) \\
 & + \beta_4 (L_{idt} * T_{dt} * D_d) \\
 & + X'_{idt}\delta
 \end{aligned} \tag{1.2}$$

Onde:

- AL_{idt} : denota o ativismo legislativo esperado do parlamentar i , no domínio d no tempo t ;
- L_{idt} : o esforço de lobby sobre o parlamentar i no domínio d no tempo t ;
- T_{dt} : variável *dummy* que recebe o valor 1 a partir do momento em que o esforço do lobby é realizado no domínio d ; e
- D_d : variável *dummy* que recebe o valor 1 nos temas em que houve esforço significativo de lobby.

De tal forma que β_1 representa o efeito médio do lobby no ativismo legislativo em todos os domínios e períodos, independente do tratamento (receber ou não

lobby); β_2 captura o efeito do lobby após o tratamento; β_3 mensura a diferença no efeito do lobby entre tratados e grupo de controle antes do tratamento; e β_4 é estimador principal do DDD, pois captura o efeito médio do tratamento sobre os tratados.

Ocorre, porém, dois problemas com essa especificação. Primeiro, o efeito do lobby não é binário, mas contínuo. Estabelecer limiares para o efeito do lobby pode gerar vieses. Segundo, as reuniões ocorrem a todo momento, tornando impossível definir um período de tratamento. Cada parlamentar pode receber lobby em qualquer momento.

$$\begin{aligned} AL_{idt} = & \gamma_{ct} + \lambda_{pt} + \theta_{dt} \\ & + \beta_1 L_{idt} \\ & + X'_{idt} \delta \end{aligned} \tag{1.3}$$

Para operacionalização, irei mensurar o AL por meio de um indicador de ativismo legislativo, o qual levará em consideração a autoria de requisições, discursos e emendas realizadas pelo parlamentar por domínio temático. Para a captação de dados, utilizarei a API disponibilizada pelo próprio Parlamento e, adicionalmente, lançarei mão de *scraper* em Python a fim de pegar os dados mais detalhados dos trâmites que não estão disponíveis na API, porém estão no site da PE.

Os esforços de lobby (L) será mensurado pela quantidade de reuniões que determinado parlamentar realizou com representantes de grupos de pressão. Os eurodeputados são obrigados a publicizá-las, quando forem relatores. Esses dados também são possíveis de serem extraídos por meio de *scraper*. Com isso, teremos o registro de quem se reuniu com quem e a data do encontro

Para encontrar o domínio em que o lobby foi realizado, faremos uma inferência a partir do representante de interesse lobista. Ou seja, partiremos do pressuposto que as organizações tem interesses específicos em temas específicos, que tem relação com a sua natureza, de tal modo que, por exemplo, uma associação comercial tem interesses em temas comerciais e econômicos, já uma organização da sociedade civil de defesa do meio ambiente tem interesses no domínio de meio ambiente. O Registro de Transparência da UE nos permite saber quem são os atores registrados,

bem como suas áreas de interesse. Assim, ao cruzarmos com os dados dos encontros publicados, poderemos saber em qual área determinado parlamentar recebeu pressão de lobby. Demais dados podem ser obtidos pela API, tais como grupo político, gênero, ocupação, país eleito de um eurodeputado.

2 RESULTADOS

Capítulo dedicado a apresentar os resultados da tese.

2.1 ANÁLISE DESCRITIVA DOS LOBISTAS

Avaliação da distribuição dos lobistas registrados junto às instituições europeias.

Tabela 1 – Distribuição de organizações por categoria

Categoria	Total	(%)
Empresas	5.770	46,3
ONGs	3.480	27,9
Outros	3.218	25,8

A Tabela acima apresenta a composição do universo de lobistas por categoria organizacional. Observa-se o predomínio de entidades empresariais e organizações da sociedade civil. A categoria *Outros* abrange uma variedade de organizações, incluindo representações governamentais, universidades, sindicatos, associações profissionais, *think tanks*, consultorias profissionais, instituições acadêmicas, redes de autoridades públicas, organizações religiosas, escritórios de advocacia e entidades estabelecidas por países terceiros, conforme a classificação detalhada no registro de transparência.

O resultado reforça assimetrias moderadas entre categorias e não sugere concentração extrema em um único tipo organizacional. Em termos substantivos, isso indica competição horizontal por acesso e agenda entre perfis empresariais e societais.

A distribuição geográfica evidencia forte concentração em Estados-Membros centrais: Bélgica (18,2%), Alemanha (14,0%) e França (9,3%). Países como Países Baixos, Espanha, Reino Unido e Itália aparecem na sequência (6% cada). Nota-se presença extracomunitária não desprezível. Cerca de 19% são de países não

membros da UE (com destaque para os Estados Unidos, com 4,5% dos lobistas registrados), sinalizando a atratividade regulatória do mercado europeu.

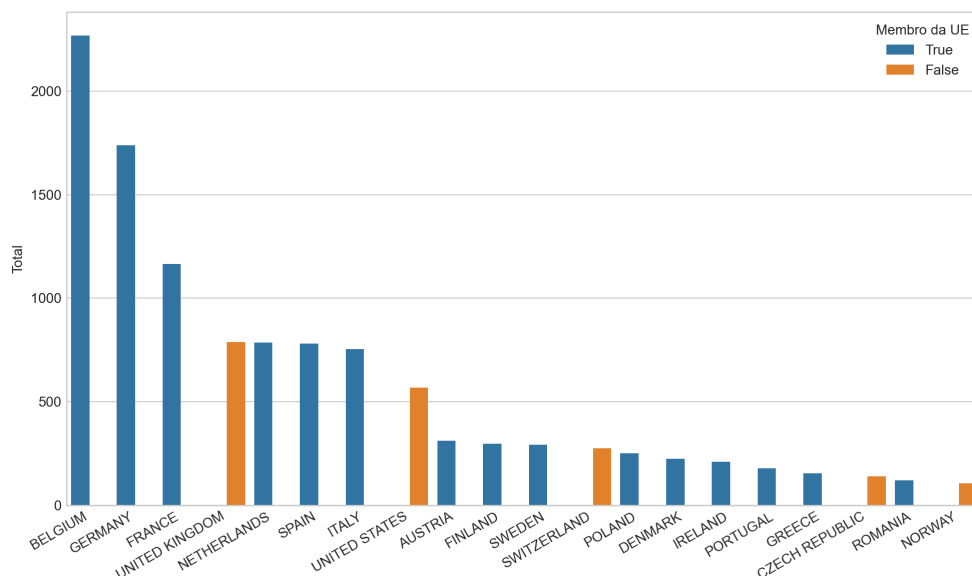


Figura 1 – Top 20 países-sede (n = 12.468 organizações, totalizando 91% do total de organizações)

O recorte dos *top 20* evidencia uma cauda longa: muitos países com baixas frequências, consistentes com a internacionalização seletiva do *lobbying*. O padrão é coerente com hipóteses de *venue shopping* e vantagens de proximidade institucional em Bruxelas/Strasburgo.

Temporalmente, observa-se aceleração do registro de entidades após meados da década de 2010, com 2023 concentrando 17,9% do total. Picos intermediários (2015–2016; 2020–2022) são compatíveis com ciclos legislativos, janelas regulatórias e alterações incrementais nos mecanismos de transparência.

O padrão visual sugere crescimento estrutural recente do ecossistema de representação de interesses, possivelmente associado às agendas de transição digital e verde e à recomposição pós-pandemia.

As incidências por domínio destacam *Infraestrutura e Indústria* (68,3%), *Tecnologia* (67,9%) e *Economia e Comércio* (67,4%), seguidas por *Ambiente e Clima* (64,7%) e *Assuntos Externos e Segurança* (51,3%). Temas como *Saúde* (43,7%)

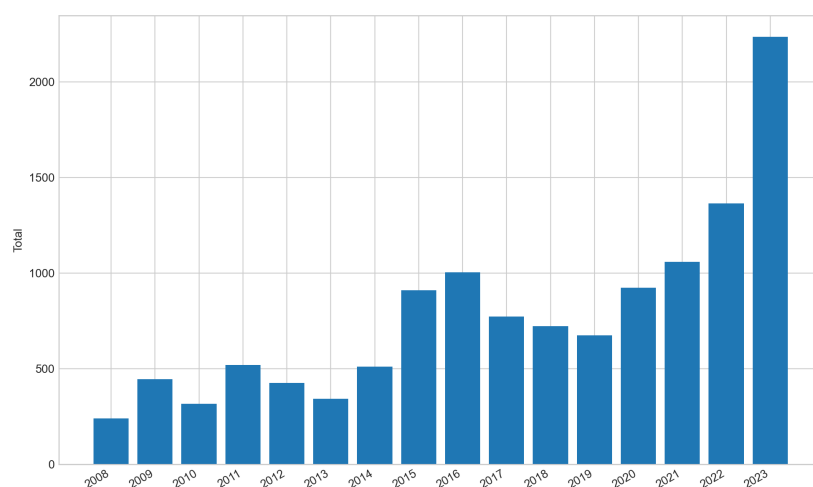


Figura 2 – Ano de registo

e *Educação* (41,2%) são intermediários; *Agricultura* (35,3%) e *Direitos Humanos* (25,3%) têm menor incidência relativa.

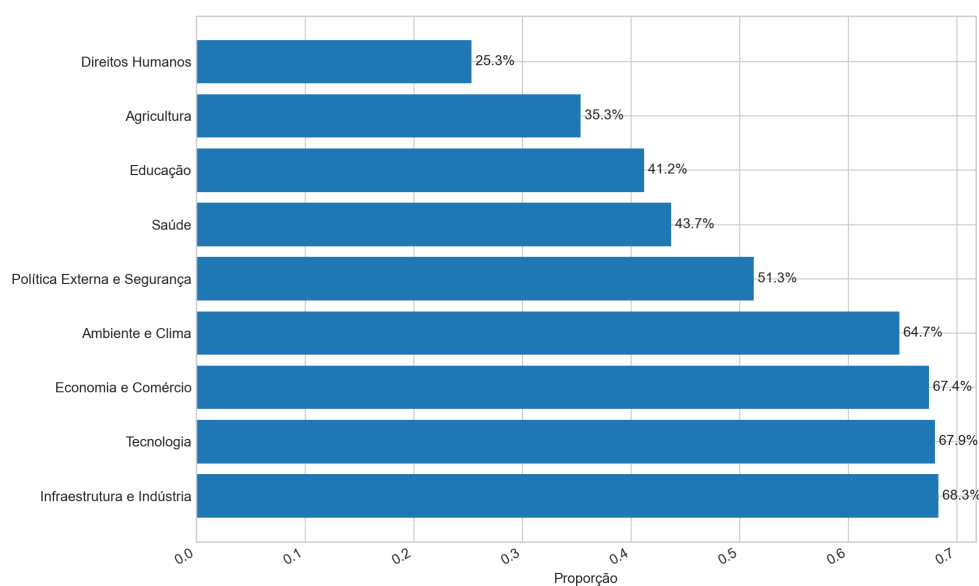


Figura 3 – Cobertura temática (proporção de entidades)

O ordenamento por proporção sugere centralidade de agendas de competitividade industrial, digitalização e cadeias de valor, bem como a transversalidade da pauta ambiental.

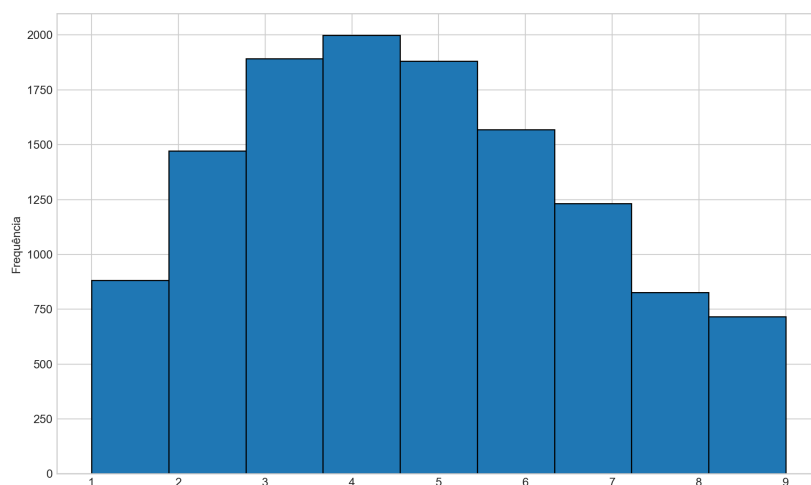


Figura 4 – Número de temas por lobista

A distribuição do número de temas por entidade indica a coexistência de atores multi-temáticos e especializados. Esse traço é relevante para a modelagem, pois sugere que a intensidade de esforço (extensivo vs. intensivo) varia com o perfil organizacional e o ambiente regulatório dos domínios.

As estatísticas de orçamento máximo declarado (em escala logarítmica) indicam mediana em torno de 11,5 e quartis aproximadamente entre 10,1 e 12,9. A presença de valores inválidos/extremos na base administrativa (por exemplo, ocorrências infinitas) distorce a média e o desvio-padrão, recomendando foco em medidas robustas (mediana, intervalos interquartílicos) e rotinas de limpeza nos exercícios inferenciais.

A distribuição apresenta assimetria e cauda à direita compatíveis com heterogeneidade de porte organizacional, sugerindo coexistência de grandes associações/empresas e organizações menores.

Em conjunto, os resultados descritivos apontam para um ecossistema plural, geograficamente ancorado em polos institucionais centrais, com dinamismo temporal recente e agendas orientadas por digitalização, competitividade industrial e sustentabilidade. Esses padrões informam as escolhas de especificação nos capítulos seguintes, notadamente a estratificação por perfis organizacionais, a construção de domínios temáticos e o controle para tendências temporais.

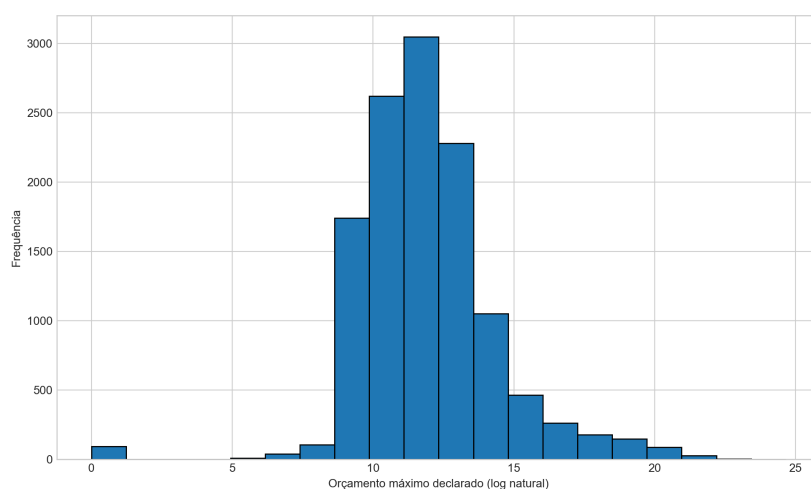


Figura 5 – Distribuição de orçamento máximo declarado ($\ln(budget)$)

Os resultados descritivos delineiam um panorama abrangente do universo de lobistas registrados junto às instituições europeias. Em primeiro lugar, a distribuição por categoria revela a coexistência de diferentes perfis organizacionais (*Business*, *NGOs* e *Other*), com magnitudes comparáveis entre atores empresariais e organizações da sociedade civil. Essa composição é compatível com a literatura sobre pluralismo organizacional e competição por acesso institucional no contexto da União Europeia, sugerindo um campo de ação onde interesses difusos e concentrados buscam simultaneamente agenda e influência.

No plano geográfico, observa-se forte concentração em Estados-Membros centrais e com infraestrutura institucional robusta. Destacam-se Bélgica (18,2%), Alemanha (14,0%) e França (9,3%), seguidas por Países Baixos, Espanha, Reino Unido e Itália (6% cada). Há ainda presença extracomunitária não desprezível (Estados Unidos 4,5%), o que evidencia a atratividade regulatória do mercado europeu e a permeabilidade do *lobbying* transnacional. Esses achados são consistentes com hipóteses de *venue shopping* e vantagens de proximidade institucional (Bruxelas/Strasburgo) para atividades de representação de interesses.

Temporalmente, as frequências por ano indicam aceleração recente dos registros, com 2023 concentrando 17,9% das entradas no período observado. Picos intermediários (2015–2016; 2020–2022) são compatíveis com ciclos legislativos, janelas

regulatórias e mudanças incrementais no regime de transparência, fatores que tendem a alterar a propensão ao registro. A expansão no pós-2020 pode refletir a reconfiguração de estratégias após as restrições pandêmicas, além da ênfase em agendas de transição digital e verde.

Quanto à cobertura temática, a incidência é mais elevada em *Infraestrutura e Indústria* (68,3%), *Tecnologia* (67,9%) e *Economia e Comércio* (67,4%), seguidas por *Ambiente e Clima* (64,7%) e *Política Externa e Segurança* (51,3%). Temas como *Saúde* (43,7%) e *Educação* (41,2%) ocupam posição intermediária, ao passo que *Agricultura* (35,3%) e *Direitos Humanos* (25,3%) apresentam menor incidência relativa. Em conjunto, esse perfil sugere: (i) centralidade de agendas de competitividade industrial, digitalização e cadeias de valor; (ii) transversalidade da pauta ambiental como condicionante regulatória; e (iii) segmentação de atores com missões setoriais mais estreitas ou normativas, potencialmente menos numerosos.

A distribuição do número de temas por lobista sugere coexistência de atores multi-temáticos, capazes de cobrir diversas frentes de política pública, e de atores especializados com foco estreito. Tal heterogeneidade é relevante para a modelagem empírica, pois a intensidade de esforço (extensivo vs. intensivo) pode variar sistematicamente com o tipo de organização e com o ambiente regulatório dos diferentes domínios.

No que se refere ao orçamento máximo declarado (em log natural), as medidas-resumo apontam mediana próxima de 11,5 e quartis aproximados entre 10,1 e 12,9. Identificam-se valores inválidos/extremos na base administrativa (por exemplo, ocorrências infinitas), que distorcem a média e o desvio-padrão; por isso, a interpretação deve privilegiar estatísticas robustas (mediana e intervalos interquartílicos) e, quando pertinente, rotinas de limpeza e estratégias robustas nos exercícios inferenciais. Substantivamente, a dispersão é compatível com a coexistência de grandes associações/empresas e organizações de menor porte, com implicações para capacidades de acesso e agenda-setting.

Em síntese, as evidências descritivas apontam para um ecossistema plural, geograficamente ancorado em polos institucionais centrais, com dinamismo temporal recente e agendas orientadas por digitalização, competitividade industrial e sus-

tentabilidade. Esses padrões informam as escolhas de especificação nos capítulos seguintes, notadamente a estratificação por perfis organizacionais, a construção de domínios temáticos e o controle para tendências temporais.

2.2 ANÁLISE DESCRITIVA DO TRATAMENTO

Esta seção apresenta uma análise descritiva sistemática dos dados utilizados para investigar os efeitos do lobbying na atividade parlamentar dos deputados do Parlamento Europeu. A abordagem adotada segue uma estratégia analítica multinível, iniciando com padrões agregados gerais e progredindo para análises desagregadas mais específicas. Esta progressão metodológica permite compreender tanto as tendências globais quanto os mecanismos específicos que operam no nível individual e temporal.

O conjunto de dados constitui um painel balanceado que combina informações sobre atividade parlamentar (perguntas) e intensidade de lobbying (reuniões) para 1.353 deputados ao longo de 63 meses, de julho de 2019 a novembro de 2024 em 9 domínios de política pública. Esta estrutura temporal permite capturar variações tanto na dimensão *cross-sectional* (entre deputados e domínios) quanto longitudinal (evolução temporal), fornecendo a base empírica necessária para estratégias de identificação causal robustas.

Considerando a unidade de análise a tríade MEP-domínio-mês, temos 767.151 observações com taxa de completude de 100%. Esta estrutura balanceada é metodologicamente vantajosa, pois elimina preocupações com viés de seleção decorrente de atrito amostral e garante que as estimativas não sejam distorcidas por padrões de observações ausentes.

A cobertura temporal de julho de 2019 a novembro de 2024 é particularmente relevante por abranger períodos de intensa atividade legislativa europeia, incluindo a transição entre legislaturas e eventos político-econômicos significativos. Destaca-se, nesse intervalo, o impacto da pandemia de COVID-19, que afetou profundamente tanto a dinâmica da atividade parlamentar quanto as estratégias de lobbying. A pandemia resultou em mudanças substanciais nos modos de trabalho do Parlamento Europeu, com a adoção de sessões remotas e restrições a reuniões presen-

ciais, o que pode ter alterado padrões de interação entre deputados e grupos de interesse. Assim, a análise cobre não apenas períodos de normalidade institucional, mas também um contexto de crise sanitária global, permitindo investigar como choques exógenos desse tipo influenciam o comportamento político e o lobbying.

A Figura 6 apresenta a evolução temporal das variáveis principais no nível mais agregado, revelando padrões que são fundamentais para compreender a dinâmica do sistema político europeu ao longo do período estudado.

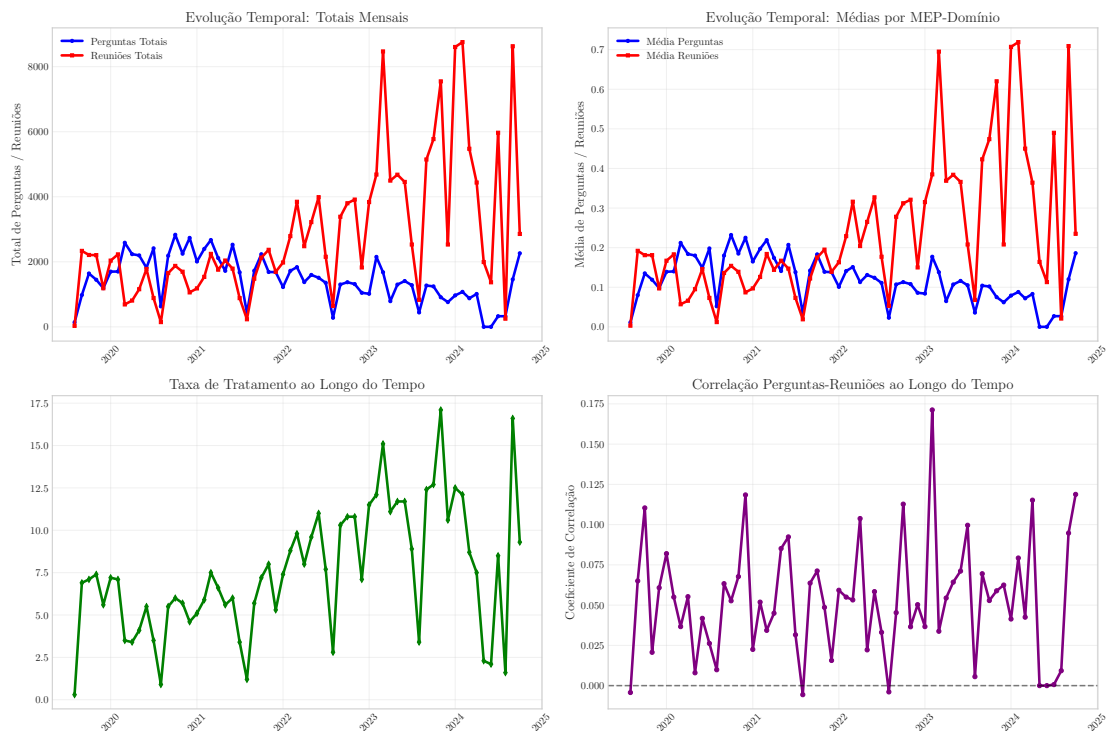


Figura 6 – Evolução temporal da atividade parlamentar e de lobbying

Nota: O painel superior esquerdo mostra os totais mensais agregados de perguntas e reuniões. O painel superior direito apresenta as médias mensais por observação MEP-domínio. O painel inferior esquerdo mostra a evolução da proporção de observações com atividade de lobbying. O painel inferior direito apresenta a estabilidade da correlação contemporânea entre as variáveis ao longo do tempo.

A análise temporal revela quatro padrões empiricamente relevantes. Primeiro, observa-se uma **tendência crescente** em ambas as variáveis ao longo do período, sugerindo intensificação tanto da atividade parlamentar quanto do lobbying.

Segundo, existe clara **sazonalidade** relacionada ao calendário parlamentar, com reduções sistemáticas durante períodos de recesso. Terceiro, identificam-se **picos de atividade** que coincidem com discussões de legislação relevante em domínios específicos, indicando resposta coordenada do sistema político. Quarto, a **correlação contemporânea** entre perguntas e reuniões permanece relativamente estável ao longo do tempo, sugerindo estabilidade estrutural na relação entre as variáveis.

Estes padrões temporais têm implicações metodológicas importantes. A presença de tendências temporais justifica a inclusão de efeitos fixos de tempo nas especificações econométricas para controlar choques temporais comuns. A sazonalidade observada valida a escolha da frequência mensal como unidade temporal, capturando variações de curto prazo sem introduzir ruído excessivo. A estabilidade da correlação fornece evidência preliminar contra quebras estruturais que poderiam comprometer a validade das estimativas.

Complementando a análise temporal, é fundamental examinar os padrões de participação no nível individual dos deputados. Esta perspectiva agregada revela a distribuição da atividade de lobbying entre os parlamentares e fornece insights sobre a concentração e heterogeneidade dos fenômenos estudados.

2.2.0.1 DISTRIBUIÇÃO DO TRATAMENTO ENTRE DEPUTADOS

A Figura 7 apresenta uma análise abrangente dos padrões de participação agregados por deputado, revelando aspectos fundamentais da distribuição da atividade de lobbying no Parlamento Europeu.

A análise revela três características fundamentais da distribuição de tratamento. Primeiro, existe **participação substancial mas não universal**: 46,3% dos deputados (627 de 1.353) receberam pelo menos uma reunião de lobbying durante o período estudado. Esta proporção indica que o lobbying é um fenômeno disseminado mas não ubíquo no sistema parlamentar europeu.

Segundo, observa-se **concentração extrema** na intensidade de tratamento. Entre os deputados que receberam lobbying, a distribuição é altamente assimétrica: enquanto a mediana é de 105 reuniões por deputado, a média é de 288,2 reuniões, indicando que uma minoria de parlamentares concentra uma proporção

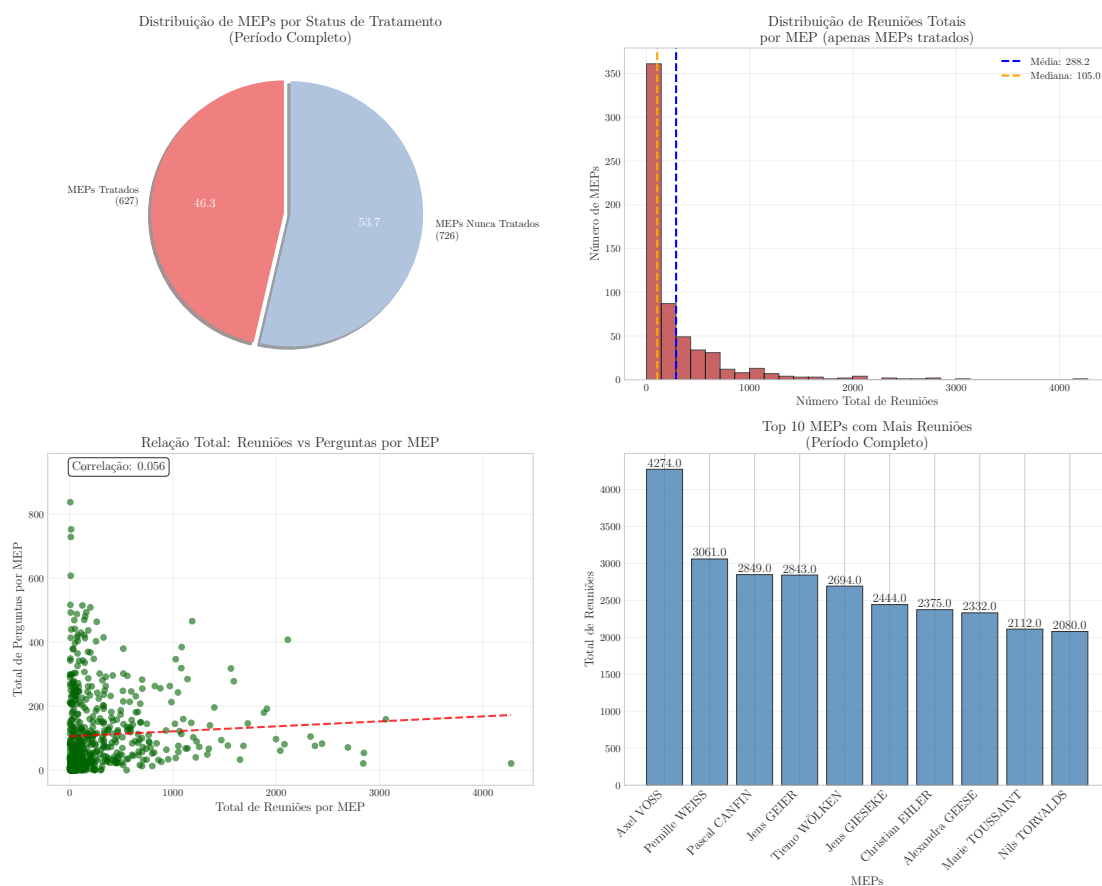


Figura 7 – Análise agregada por MEP: distribuição e intensidade do tratamento

Nota: O painel superior esquerdo mostra a proporção de deputados que receberam pelo menos uma reunião de lobbying durante todo o período. O painel superior direito apresenta a distribuição da intensidade total de reuniões entre deputados tratados. O painel inferior esquerdo examina a relação agregada entre reuniões e perguntas totais por deputado. O painel inferior direito identifica os deputados mais ativos em termos de recepção de lobbying.

desproporcional da atividade lobista. O caso extremo de um deputado com 4.274 reuniões ilustra esta concentração.

Terceiro, a **correlação agregada** entre reuniões e perguntas totais por deputado é surpreendentemente baixa (0,056), contrastando com correlações mais elevadas observadas no nível temporal. Este padrão sugere que os efeitos do lobbying podem ser mais evidentes em frequências temporais específicas do que em padrões de atividade agregados de longo prazo.

Estes padrões agregados têm implicações importantes para a identificação causal. A concentração do tratamento em uma minoria de deputados sugere que estratégias de identificação baseadas em variação cross-sectional podem sofrer de poder estatístico limitado. Simultaneamente, a variação substancial na intensidade de tratamento entre deputados tratados fornece fonte valiosa de identificação para estimativas de dose-resposta.

A baixa correlação agregada, combinada com correlações temporais mais elevadas, indica que a identificação causal pode beneficiar-se de estratégias que explorem variação temporal within-individual rather than cross-sectional between-individual. Esta evidência preliminar orienta a especificação de modelos com efeitos fixos de deputado para controlar heterogeneidade não observada time-invariant.

Tabela 2 – Estatísticas agregadas de tratamento por deputado

Estatística	Valor
Total de deputados únicos	1,353
Deputados que receberam tratamento	627
Taxa de tratamento por deputado (%)	46,3%
Entre deputados tratados:	
Reuniões médias por deputado	288,2
Reuniões medianas por deputado	105,0
Desvio padrão	468,9
Deputado mais ativo (reuniões)	4,274
Correlação agregada:	
Correlação reuniões-perguntas	0,056

2.2.1 HETEROGENEIDADE ENTRE DOMÍNIOS DE POLÍTICA PÚBLICA

A terceira dimensão da análise agregada examina a variação entre domínios de política pública. Esta heterogeneidade setorial é teoricamente relevante porque diferentes áreas de política podem apresentar características distintas em termos de complexidade técnica, interesse econômico e organização de grupos de pressão, afetando tanto a demanda por lobbying quanto a responsividade parlamentar.

A Figura 8 apresenta uma análise sistemática da variação inter-domínios em múltiplas dimensões: penetração, volume e intensidade do tratamento, bem como padrões temporais de iniciação.

Tabela 3 – Taxa de tratamento por domínio: deputados únicos que receberam lobbying

Domínio	Deputados Tratados	Total Deputados	Taxa (%)
Economia e Comércio	615	1,353	45,5
Tecnologia	615	1,353	45,5
Política Externa e Segurança	611	1,353	45,2
Infraestrutura e Indústria	610	1,353	45,1
Meio Ambiente e Clima	607	1,353	44,9
Saúde	599	1,353	44,3
Educação	578	1,353	42,7
Direitos Humanos	564	1,353	41,7
Agricultura	554	1,353	40,9

A análise revela heterogeneidade sistemática mas moderada entre domínios. Em termos de **penetração**, as taxas variam de 40,9% (agricultura) a 45,5% (economia e tecnologia), uma amplitude de apenas 4,6 pontos percentuais. Esta variação relativamente pequena sugere que o lobbying possui caráter transversal, não concentrando-se drasticamente em setores específicos.

Contudo, emergem padrões teoricamente consistentes. Domínios relacionados à **regulação econômica** (economia e comércio, tecnologia, infraestrutura) apresentam sistematicamente maiores taxas de penetração, refletindo os elevados interesses econômicos e a complexidade regulatória que incentivam investimento em

atividades de lobbying. Este padrão é especialmente relevante considerando que a promoção da integração econômica é o foco principal de um bloco como a União Europeia, o que naturalmente direciona maior atenção e mobilização de grupos de interesse para essas áreas. Inversamente, domínios com características de bem público (agricultura, direitos humanos) apresentam menores taxas, consistente com problemas de ação coletiva e menor capacidade organizacional de grupos difusos.

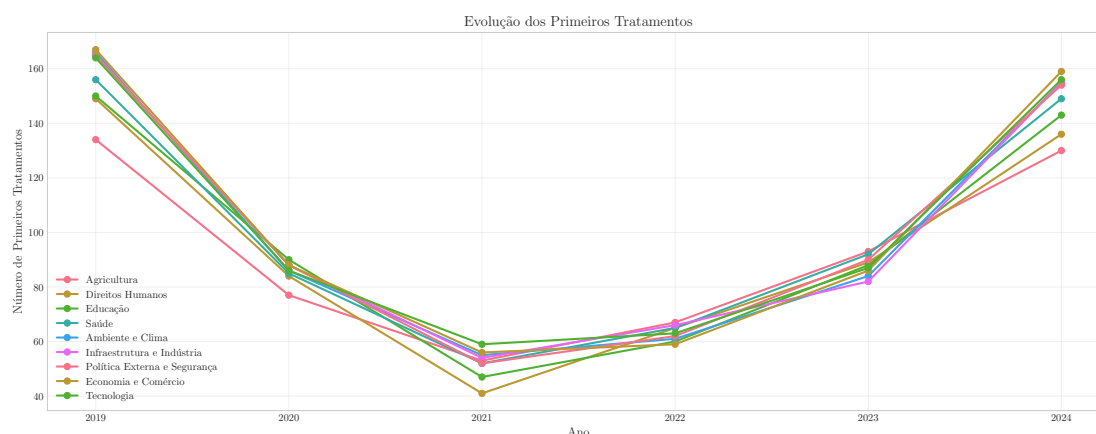


Figura 8 – Análise detalhada de tratamento por domínio de política pública

Nota: O painel superior esquerdo mostra a taxa de penetração (percentual de deputados únicos que receberam pelo menos uma reunião em cada domínio). O painel superior direito apresenta o volume absoluto de deputados tratados. O painel inferior esquerdo mostra a intensidade média condicional de tratamento. O painel inferior direito apresenta a evolução temporal dos primeiros tratamentos para os três domínios mais ativos.

A heterogeneidade observada entre domínios tem duas implicações metodológicas importantes. Primeiro, a variação sistemática sugere que estimativas de efeito médio podem mascarar diferenças substantivas entre setores, justificando análises de heterogeneidade de efeitos por domínio. Segundo, a ordenação consistente dos domínios por múltiplas métricas (penetração, volume, intensidade) sugere que esta heterogeneidade reflete características estruturais dos setores ao invés de variação aleatória, aumentando a credibilidade de interpretações causais diferenciadas.

A análise da inflação de zeros requer cuidado metodológico particular, pois a unidade de observação MEP-domínio-mês pode gerar inflação **artificial** de zeros. Como documentado na literatura sobre comportamento parlamentar (??),

deputados tendem a especializar-se tematicamente, concentrando atividade em subconjuntos específicos de domínios. Consequentemente, grupos de interesse, cientes desta especialização, direcionam esforços de lobbying apenas para deputados ativos em suas áreas de interesse.

Evidência de especialização temática: A análise da atividade parlamentar agregada por deputado revela que 97,6% dos MEPs são **generalistas** (Índice Herfindahl $< 0,4$), atuando em média em 7,47 dos 9 domínios disponíveis. Contudo, identificam-se 22 MEPs altamente especializados (HHI $> 0,8$) que concentram perguntas em domínios únicos, e 26 moderadamente especializados (HHI 0,4-0,8), demonstrando que a especialização, embora limitada, é empiricamente relevante.

2.2.1.0.1 ANÁLISE CORRIGIDA DA INFLAÇÃO DE ZEROS

Para evitar viés na interpretação, a análise da inflação de zeros deve considerar níveis de agregação teoricamente apropriados. A Tabela 4 apresenta uma comparação sistemática entre diferentes níveis de agregação.

Tabela 4 – Inflação de zeros por nível de agregação

Nível de Agregação	Observações	Zeros Perguntas	Zeros Reuniões
MEP-Domínio-Mês (original)	767.151	92,2%	92,5%
MEP-Mês (domínios ativos)	54.117	70,9%	85,2%
Domínio-Mês (agregado)	567	3,2%	0,0%

Os resultados revelam que a inflação de zeros é **sensível ao nível de agregação** e parcialmente **artificial** quando consideramos especializações temáticas:

1. **Nível MEP-domínio-mês:** A inflação aparentemente extrema ($>92\%$) reflete em grande parte combinações MEP-domínio onde não se espera atividade sistemática devido à especialização.

2. **Nível MEP-mês** (agregando apenas domínios onde o MEP demonstra atividade parlamentar): A inflação reduz substancialmente para 70,9% (perguntas) e 85,2% (reuniões), revelando que a atividade é *episódica* mas não *ausente*.
3. **Nível domínio-mês** (agregando todos os MEPs ativos): A inflação torna-se negligível (3,2% para perguntas, 0% para reuniões), indicando que há atividade consistente em todos os domínios quando consideramos o conjunto de deputados relevantes.

Esta análise corrigida tem implicações fundamentais para a estratégia econométrica:

1. **Falso problema de inflação de zeros**: A aparente inflação extrema ($>92\%$) é em grande parte **artificial**, resultante da inclusão de combinações MEP-domínio teoricamente implausíveis. A inflação **real** no nível MEP-mês é substancialmente menor (70,9%-85,2%).
2. **Justificativa para efeitos fixos**: A especialização temática documentada justifica ainda mais fortemente o uso de efeitos fixos $\text{MEP} \times \text{domínio}$, pois captura heterogeneidade não observada na propensão à atividade em áreas específicas.
3. **Modelos econométricos apropriados**: Enquanto a inflação moderada (70,9%) ainda sugere benefícios de modelos de contagem (PPML), a inflação não é tão extrema a ponto de requerer modelos zero-inflated especializados.
4. **Unidade de análise**: A evidência sustenta a escolha da unidade MEP-domínio-mês para análise causal, pois captura a granularidade necessária para identificação, desde que acompanhada de controles apropriados para especialização.
5. **Interpretação de resultados**: Efeitos estimados devem ser interpretados condicionalmente à especialização temática, com atenção particular às margens extensiva (entrada em novos domínios) versus intensiva (aumento de atividade em domínios existentes).

2.2.2 SÍNTESE E IMPLICAÇÕES METODOLÓGICAS

A análise descritiva multinível revela um conjunto coerente de *stylized facts* que informam tanto a compreensão teórica quanto as escolhas metodológicas para a análise causal subsequente.

Primeiro, o **lobbying é um fenômeno disseminado mas episódico**. Enquanto 46,3% dos deputados recebem lobbying durante o período estudado, a atividade temporal é concentrada: considerando apenas domínios onde MEPs demonstram atividade parlamentar, 70,9% das observações MEP-mês não apresentam perguntas e 85,2% não apresentam reuniões, indicando que a influência opera através de interações concentradas temporalmente.

Segundo, existe **concentração extrema** em múltiplas dimensões. No nível individual, a distribuição de reuniões é altamente assimétrica (mediana 105, média 288, máximo 4.274). Crucialmente, a aparente inflação extrema de zeros (>92%) no nível MEP-domínio-mês é em grande parte artificial, refletindo combinações onde não se espera atividade devido à especialização temática.

Terceiro, observa-se **heterogeneidade sistemática entre domínios** em todas as métricas analisadas. Domínios de regulação econômica apresentam consistentemente maior atividade de lobbying, refletindo diferenças estruturais em stakes econômicos e capacidade organizacional. Esta heterogeneidade é consistente com a especialização temática documentada.

Quarto, a **especialização temática é limitada mas empiricamente relevante**: enquanto 97,6% dos MEPs atuam como generalistas ($HHI < 0,4$), existem 22 deputados altamente especializados e 26 moderadamente especializados, com padrões claros de concentração que informam estratégias de lobbying e justificam controles econométricos específicos.

Quinto, a **correlação contemporânea entre lobbying e atividade parlamentar é baixa** no nível temporal mensal, mas padrões agregados sugerem relações mais complexas que podem envolver defasagens temporais ou mecanismos indiretos.

Estas características empíricas têm implicações diretas para a escolha da estra-

tégia econométrica:

1. **Especificação funcional:** A inflação moderada de zeros (70,9%-85,2% após correção) e natureza de contagem das variáveis justificam o uso de estimadores Poisson Pseudo-Maximum Likelihood (PPML), mas não requerem modelos zero-inflated especializados.
2. **Estrutura de efeitos fixos:** A evidência de especialização temática justifica fortemente efeitos fixos $\text{MEP} \times \text{domínio}$ para controlar heterogeneidade não observada na propensão à atividade em áreas específicas, complementados por efeitos fixos temporais.
3. **Correção de viés de seleção:** A especialização temática implica que observações MEP -domínio-mês com probabilidade zero de atividade podem distorcer estimativas. Controles ou exclusões baseadas em atividade histórica podem ser apropriados.
4. **Estrutura de erros:** A concentração temporal da atividade e especialização justificam erros-padrão agrupados no nível $\text{MEP} \times \text{domínio}$ para capturar correlação serial específica por área de atuação.
5. **Análise de heterogeneidade:** A variação sistemática entre domínios, combinada com especialização, justifica análises diferenciadas por setor e investigação de efeitos nas margens extensiva (entrada em novos domínios) versus intensiva.
6. **Interpretação causal:** Efeitos estimados devem ser interpretados condicionalmente à especialização temática existente, distinguindo entre expansão de atividade em domínios familiares versus diversificação para novas áreas.

Finalmente, a **estrutura balanceada do painel** e a **cobertura temporal substancial** fornecem condições ideais para estratégias de identificação baseadas em variação temporal within-individual, maximizando o poder estatístico while minimizando preocupações com confounding não observado time-invariant.

Esta análise descritiva abrangente estabelece as bases empíricas sólidas para a estratégia de identificação causal apresentada na seção seguinte, demonstrando que os dados possuem as características necessárias para investigar rigorosamente os efeitos do lobbying na atividade parlamentar dos deputados europeus.

2.3 ANÁLISE DE EFEITOS DO LOBBY

Optamos por estimar modelos de contagem via PPML com *link* logaritmo por três razões principais. Primeiro, as variáveis de interesse (perguntas e reuniões) são contagens, com forte assimetria e alta incidência de zeros no nível MEP–domínio–mês. O PPML lida naturalmente com zeros sem exigir transformações logarítmicas ad hoc que descartam observações. Segundo, o PPML é consistente sob especificação correta da média condicional mesmo na presença de sobredispersão e heterocedasticidade não especificada, fornecendo erros-padrão robustos quando combinados com *clustering*. Terceiro, a implementação com efeitos fixos de alta dimensão é estável e amplamente utilizada na literatura aplicada (estimador `fepois` do pacote `fixest`).

No PPML com *link* log, a expectativa condicional é $\mathbb{E}[y \mid X] = \exp(X\beta)$. Assim, para um regressor contínuo x_k (por exemplo, *meetings* em nível), o coeficiente β_k tem interpretação multiplicativa: um aumento de uma unidade em x_k está associado a uma variação percentual de $100 \times (e^{\beta_k} - 1)\%$ na média de y , *ceteris paribus*.

A especificação segue o *framework* analítico delineado no capítulo: controlamos por heterogeneidade não observada ao nível do membro e por choques comuns estruturados por partido e por país ao longo do tempo. Concretamente, estimamos modelos com efeitos fixos de membro (`member_id`) e efeitos fixos tempo-variantes por país (`country_time`) e por partido (`party_time`). Em especificações mais restritivas, adicionamos efeitos fixos *domínio*×*tempo* (`domain_time`) para absorver choques setoriais mensais. Os erros-padrão são agrupados em *domínio*×*tempo*, capturando correlação serial e choques idiossincráticos nesse nível, conforme implementado nos scripts empíricos.

Essa estrutura assegura: (i) comparação *within* do mesmo MEP ao longo do

tempo (controlando preferências, capital político e produtividade inobservados e invariantes no tempo); (ii) controle por variação comum a todos os MEPs de um mesmo país ou partido em cada mês; e (iii) robustez a choques setoriais específicos quando incluímos `domain_time`.

Para testar a Hipótese 1, utilizamos o painel agregado MEP–domínio–mês em amostra combinada (*pooled*) e estimamos PPML com a estrutura de efeitos fixos descrita acima. O coeficiente associado às reuniões (*meetings*) é **positivo**, indicando que aumentos na intensidade de lobbying estão associados a maior atividade parlamentar em termos de perguntas. Esse resultado é consistente em especificações alternativas, incluindo a versão com termo quadrático para capturar possíveis não linearidades e a inclusão de efeitos fixos *domínio*×*tempo*, sugerindo robustez do sinal e da magnitude qualitativa do efeito.

Em termos de interpretação, mantidos constantes os efeitos fixos, um incremento marginal em reuniões está associado a um aumento proporcional na média de perguntas dado por $e^{\hat{\beta}} - 1$. Reportamos os efeitos como variações percentuais estimadas na seção de tabelas de resultados, com intervalos de confiança baseados em erros-padrão agrupados.

Tabela 5 – Resumo dos modelos PPML para a Hipótese 1

	PPML	PPML (Quad.)
Reuniões	0.013*** (0.002)	0.037*** (0.005)
Reuniões ²		-0.001*** (0.000)
Observações	711,855	711,855
Efeitos fixos	país×tempo; partido×tempo; domínio×tempo	
Cluster	domínio×tempo; membro	

Nota: A coluna “DDD PPML” reporta o modelo principal com efeito linear em *meetings*. A coluna “DDD PPML (Quadrático)” adiciona *meetings*² para capturar retornos marginais decrescentes. Efeitos fixos: membro; país×tempo; partido×tempo. Erros-padrão agrupados em domínio×tempo e membro.

Tabela 5 mostra que o coeficiente de *meetings* no modelo PPML linear é positivo e estatisticamente significativo, evidenciando que aumentos na intensidade de lobbying associam-se a maior número de perguntas, *ceteris paribus*. Na especificação quadrática, o termo linear permanece positivo enquanto o termo quadrático é

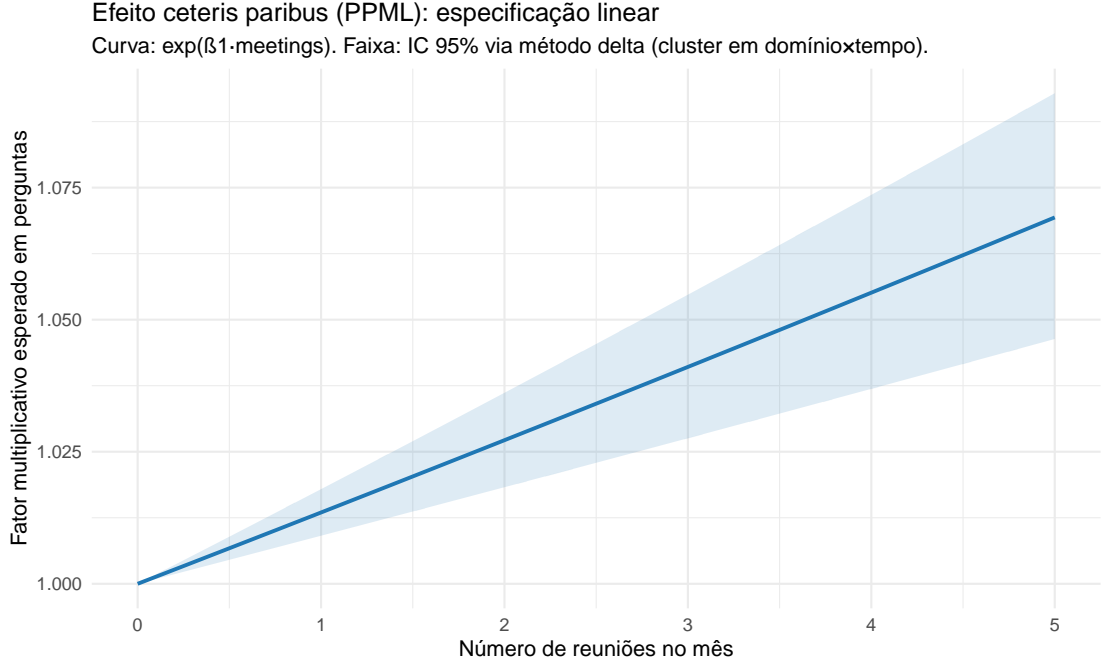


Figura 9 – Efeito esperado ceteris paribus: especificação linear (PPML)

Nota: A curva apresenta o fator multiplicativo esperado em perguntas como função do número de reuniões no mês, mantendo constantes os efeitos fixos ($\exp(\hat{\beta}_1 \cdot \text{meetings})$). A faixa sombreada corresponde ao intervalo de 95% obtido via método delta com erros-padrão agrupados em domínio \times tempo.

negativo, indicando retornos marginais decrescentes: o impacto adicional de reuniões sobre perguntas diminui à medida que o volume de reuniões cresce.

Essa interpretação decorre da forma funcional do PPML ($\mathbb{E}[y \mid X] = \exp(X\beta)$). No modelo linear, um acréscimo de uma unidade em *meetings* altera a média condicional de perguntas em $100 \times (e^{\hat{\beta}_1} - 1)\%$. No modelo quadrático, o efeito marginal em log-média é $\partial \log \mathbb{E}[y \mid X] / \partial \text{meetings} = \hat{\beta}_1 + 2\hat{\beta}_2 \text{meetings}$. Com $\hat{\beta}_2 < 0$, esse efeito declina com o nível de *meetings*, isto é, há retornos marginais decrescentes.

Em particular, a magnitude do termo quadrático é muito inferior ao efeito linear (0,002 vs. 0,066), o que indica retornos marginais decrescentes pequenos na faixa observada. Isso implica que atores com maior disponibilidade de recursos enfrentam pouca perda de eficácia ao intensificar o número de reuniões e, portanto,

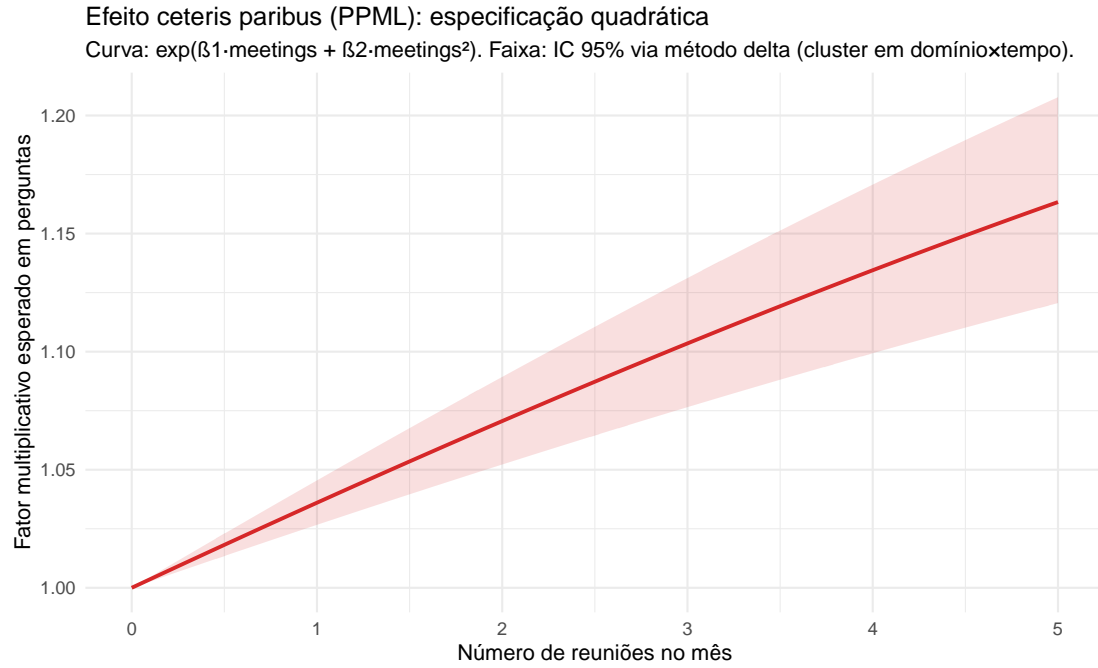


Figura 10 – Efeito esperado ceteris paribus: especificação quadrática (PPML)

Nota: A curva apresenta o fator multiplicativo esperado em perguntas como função do número de reuniões, permitindo retornos marginais decrescentes ($\exp(\hat{\beta}_1 \cdot \text{meetings} + \hat{\beta}_2 \cdot \text{meetings}^2)$). A faixa sombreada representa o IC de 95% via método delta com a matriz de variância-covariância agrupada.

podem sustentar níveis muito mais altos de lobbying; tal padrão é consistente com a hipótese de que grandes players conseguem alavancar sua capacidade financeira para obter influência relativamente maior, mesmo diante de retornos marginais decrescentes.

As curvas em Figura 9 e Figura 10 tornam essa dinâmica visual. A primeira apresenta um efeito multiplicativo crescente de forma monotônica ($\exp(\hat{\beta}_1 \cdot \text{meetings})$), com faixas de incerteza (IC 95%) obtidas por método delta usando a matriz de variância-covariância com *clustering* em domínio×tempo e membro. A segunda permite curvatura ($\exp(\hat{\beta}_1 \cdot \text{meetings} + \hat{\beta}_2 \cdot \text{meetings}^2)$) e revela concavidade compatível com saturação de agenda: para níveis altos de *meetings*, o ganho marginal em perguntas é menor. Em ambas as figuras, o eixo horizontal é mantido dentro do suporte observado dos dados para evitar extrapolações.

Do ponto de vista de identificação, os efeitos fixos por membro, país \times tempo e partido \times tempo controlam heterogeneidade não observada invariável e choques comuns, permitindo comparação *within* do mesmo MEP ao longo do tempo. A inferência usa erros-padrão agrupados em duas dimensões (domínio \times tempo; membro), acomodando dependência serial e seções cruzadas.

Em síntese, os resultados corroboram a Hipótese 1: há associação positiva entre lobbying e atividade de fiscalização medida por perguntas parlamentares, com evidência de retornos marginais decrescentes em níveis mais altos de *meetings*. Essa conclusão é robusta a especificações alternativas consideradas.

REFERÊNCIAS

AUSTEN-SMITH, D.; WRIGHT, J. R. Theory and evidence for counteractive lobbying. *American Journal of Political Science*, JSTOR, v. 40, n. 2, p. 543–564, 1996. 1

BACHRACH, P.; BARATZ, M. Two faces of power, in “American Political Science Review”, 56. *American Political Science Review*, 1962. 1

BERTRAND, M.; BOMBARDINI, M.; TREBBI, F. Is it whom you know or what you know? an empirical assessment of the lobbying process. *American Economic Review*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, v. 104, n. 12, p. 3885–3920, 2014. 1

CALDEIRA, G. A.; HOJNACKI, M.; WRIGHT, J. R. The lobbying activities of organized interests in federal judicial nominations. *Journal of Politics*, Wiley Online Library, v. 62, n. 1, p. 51–69, 2000. 1

DÜR, A. Measuring interest group influence in the EU: A note on methodology. v. 9, n. 4, p. 559–576, 2008. ISSN 1465-1165, 1741-2757. Disponível em: <<http://journals.sagepub.com/doi/10.1177/1465116508095151>>. 2

FIGUEIREDO, J. M. D.; RICHTER, B. K. Advancing the empirical research on lobbying. v. 17, n. 1, p. 163–185, 2014. ISSN 1094-2939, 1545-1577. Disponível em: <<https://www.annualreviews.org/doi/10.1146/annurev-polisci-100711-135308>>. 1, 2

FORDHAM, B. O.; MCKEOWN, T. J. Selection and influence: Interest groups and congressional voting on trade policy. *International Organization*, Cambridge University Press, v. 57, n. 3, p. 519–549, 2003. 2

GAWANDE, B. K.; KRISHNA, P.; OLARREAGA, M. Lobbying competition over trade policy. *International Economic Review*, Wiley Online Library, v. 53, n. 1, p. 115–132, 2012. 1

HOJNACKI, M.; KIMBALL, D. C. Pac contributions and lobbying contacts in congressional committees. *Political Research Quarterly*, University of Utah 260 SO. Central Campus Drive, RM. 252, Salt Lake City ..., v. 54, n. 1, p. 161–180, 2001. 1

HOLYOKE, T. T. Choosing battlegrounds: Interest group lobbying across multiple venues. *Political Research Quarterly*, Sage Publications Sage CA: Thousand Oaks, CA, v. 56, n. 3, p. 325–336, 2003. 1

KELLEHER, C. A.; YACKEE, S. W. A political consequence of contracting: Organized interests and state agency decision making. *Journal of public administration research and theory*, Oxford University Press, v. 19, n. 3, p. 579–602, 2009. 1

KOLLMAN, K. *Outside lobbying: Public opinion and interest group strategies*. [S.l.]: Princeton University Press, 1998. 1, 2

MARSHALL, D. Who to lobby and when: Institutional determinants of interest group strategies in european parliament committees. *European Union Politics*, SAGE Publications Sage UK: London, England, v. 11, n. 4, p. 553–575, 2010. 1