

SUMÁRIO

1	METODOLOGIA E ESTRATÉGIA EMPÍRICA	3
1.1	Desafios Metodológicos na Literatura	3
1.2	Estratégia de identificação e de operacionalização	5
1.3	Validação e Testes de Robustez	13
2	RESULTADOS	15
2.1	Análise descritiva dos lobistas	15
2.2	Análise descritiva do tratamento	23
2.3	Análise de efeitos do lobby	31
2.3.1	Teste da hipótese 2: Uma Análise Decomposta do Efeito do Lobby .	36
2.3.2	Teste da hipótese 3: O Efeito do Lobby em Temas Salientes	41
A	OPERACIONALIZAÇÃO	45
B	TESTES DE ROBUSTEZ	51
	REFERÊNCIAS	57

1 METODOLOGIA E ESTRATÉGIA EMPÍRICA

Este capítulo apresenta a proposta metodológica e estratégia empírica da tese, abordando desde a operacionalização dos dados até os métodos de estimação e validação dos resultados. Como discutido no capítulo anterior, estudar os efeitos do lobby apresenta desafios metodológicos importantes que devem ser cuidadosamente endereçados para assegurar identificação causal crível.

1.1 DESAFIOS METODOLÓGICOS NA LITERATURA

A identificação dos efeitos do lobby enfrenta seis principais complicações metodológicas que orientam a estratégia empírica desta pesquisa:

O primeiro desafio, denominado "**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, métodos de análise de dados em painel que exigem variação temporal nas variáveis de interesse podem não obter a variabilidade necessária para identificação causal.

A segunda complicação, ou "**problema do viés de seleção**", refere-se ao fato de que a escolha dos alvos pelos 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, denominado "**problema do lobby contra-ativo**", diz respeito à possibilidade de que, mesmo quando um grupo não obtém sucesso visível, isso não significa ausência de influência. Pode significar que houve um esforço de lobby no sentido contrário que obteve sucesso (AUSTEN-SMITH; WRIGHT,

1996).

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

A quinta dificuldade, ou "**problema das variáveis omitidas**" (FIGUEIREDO; RICHTER, 2014), está relacionada à possibilidade de que variáveis importantes podem não ser observadas, como a habilidade de negociação de cada lobista ou esforços de pressão não transparentes.

Por fim, a sexta complicação, denominada "**problema dos canais de influência**", refere-se à capacidade dos lobistas de agir através de diferentes canais (DÜR, 2008), desde lobby direto até campanhas de pressão perante a opinião pública (*outside lobbying*) (KOLLMAN, 1998), podendo inclusive influenciar a própria seleção dos tomadores de decisão (FORDHAM; MCKEOWN, 2003).

Para endereçar esses desafios, esta tese adota uma estratégia metodológica multifacetada que combina:

(1) **Operacionalização dos dados:** Processo rigoroso de extração, tratamento e harmonização de múltiplas fontes de dados públicos, com controles de qualidade extensivos (Apêndice A).

(2) **Identificação por efeitos fixos de alta dimensão:** Modelo econométrico que explora variação temporal intra-parlamentar após controlar por choques comuns em múltiplas dimensões.

(3) **Testes de robustez abrangentes:** Validação dos resultados através de especificações alternativas, amostras restringidas, e testes placebo.

(4) **Análise de heterogeneidade:** Investigação de efeitos diferenciados por tipo de ator, área temática, e contexto político.

Os detalhes da operacionalização dos dados e testes de robustez foram expostos no Apêndice A e B, respectivamente. Assim, as próximas sessões tratarão da estratégia de identificação e métodos de estimação específicos para cada hipótese testada.

1.2 ESTRATÉGIA DE IDENTIFICAÇÃO E DE OPERACIONALIZAÇÃO

A análise causal dos efeitos do lobby enfrenta desafios significativos, conforme discutido. Para superá-los, esta tese adota uma estratégia de identificação que combina um foco analítico restrito com um modelo empírico robusto, fundamentado no *framework* de análise do comportamento parlamentar (??).

A variável dependente utilizada, portanto, é a Atividade Legislativa (AL), operacionalizada como o número de perguntas parlamentares que um MPE (i) apresenta em um determinado domínio temático (d) e período (t). A escolha por focar no comportamento parlamentar, em vez de resultados de políticas públicas, reduz a complexidade da cadeia causal e aproxima a análise da ação individual do legislador.

Esta escolha é amplamente respaldada pela literatura sobre o Parlamento Europeu, que demonstra que as perguntas parlamentares são um instrumento multifacetado. Uma primeira vertente teórica, baseada na literatura de agente-principal, enxerga as perguntas como um mecanismo de fiscalização e responsabilização (JENSEN; PROKSCH; SLAPIN, 2013; MARICUT-AKBIK, 2020; MARTIN, 2013). Nessa perspectiva, os legisladores (o principal) delegam poder à burocracia (o agente), mas utilizam instrumentos como as perguntas para monitorar suas ações, especialmente em temas de alta saliência política (MCCUBBIN; SCHWARTZ, 1984; SAALFELD, 2000; STRØM, 2000; KOOP, 2011).

Uma segunda vertente foca nas perguntas como um mecanismo de sinalização e posicionamento político. Nessa visão, a atividade parlamentar é menos sobre a fiscalização e mais sobre estratégia: os parlamentares utilizam as perguntas para expressar preferências, sinalizar seus interesses para a liderança partidária, eleitores e grupos de interesse, sendo uma ferramenta valiosa especialmente para a oposição (MARTIN, 2013; OTJES; LOUWERSE, 2017; PROKSCH; SLAPIN, 2010; BEVAN; BORGHETTO; SEEBERG, 2023; NAVARRO; KELBEL, 2022). Por serem uma forma de baixo custo para demonstrar atividade e engajamento em temas específicos, as perguntas tornam-se um indicador sensível da alocação de

atenção e esforço de um parlamentar, funcionando como uma excelente proxy para medir mudanças no ativismo legislativo em resposta a estímulos externos, como o lobby.

A variável de tratamento, ou esforço de lobby (L), é mensurada pelo número de reuniões que o MPE i teve com lobistas com interesse no domínio d no período t . Os dados são extraídos dos registros públicos do Parlamento Europeu. A escolha por essa operacionalização, embora não capture a totalidade dos canais de influência (DÜR, 2008), se justifica pela centralidade do contato direto na atividade de lobby. A literatura define o lobby como a transferência de informações por meio de encontros privados (FIGUEIREDO; RICHTER, 2014) e, no contexto da UE, as táticas que envolvem engajamento face a face são consideradas particularmente eficazes (HUWYLER; MARTIN, 2022). Com o fortalecimento do PE como arena decisória, garantir acesso direto a parlamentares influentes, como relatores, tornou-se uma estratégia prioritária para os grupos de interesse (KLÜVER; BRAUN; BEYERS, 2015; MARSHALL, 2010). Portanto, entendemos que o número de reuniões funciona como uma proxy robusta e observável para a intensidade do esforço de lobby direcionado a legisladores individuais, refletindo o investimento em um dos recursos mais valiosos para os lobistas: o acesso e as conexões relacionais.

A relação entre o esforço de lobby e a AL é estimada por meio de um modelo de Poisson Pseudo Maximum Likelihood (PPML). O método, cuja fundamentação teórica remonta a Anderson (1979), é uma forma de Generalized Linear Model (GLM) robusto para dados de contagem, especialmente na presença de muitos zeros. O modelo é estimado por meio de uma distribuição quasi-Poisson com uma função de ligação logarítmica, o que evita os vieses que podem surgir da transformação logarítmica da variável dependente em modelos de Mínimos Quadrados Ordinários (MQO), uma questão comum com dados de contagem. Para a estimação, será utilizada a linguagem de programação estatística R, especificamente a função `fepois` do pacote `fixest`, que é eficiente para estimar modelos GLM com efeitos fixos de alta dimensão. A especificação do modelo, baseada no *framework* teórico, é a seguinte:

$$AL_{idt} = \beta L_{idt} + \gamma_{ct} + \lambda_{pt} + \theta_{dt} + X'_{it}\delta + \epsilon_{idt} \quad (1.1)$$

Onde:

- AL_{idt} é a contagem de perguntas do eurodeputado i no domínio d no mês t ;
- L_{idt} é a contagem de reuniões de lobby do eurodeputado i no domínio d no mês t ;
- β é o coeficiente de interesse, que captura o efeito médio de uma reunião adicional sobre a produção de perguntas;
- γ_{ct} , λ_{pt} e θ_{dt} são efeitos fixos de alta dimensão;
- X'_{it} é um vetor de controles individuais;
- ϵ_{idt} é o termo de erro.

O componente central da estratégia de identificação reside nos efeitos fixos, que absorvem uma vasta gama de fatores de confusão observáveis e não observáveis, alinhando-se diretamente às três dimensões do comportamento parlamentar identificadas no *framework* teórico:

- **Efeito fixo de país-tempo (γ_{ct}):** Captura qualquer choque ou tendência comum a todos os parlamentares de um mesmo país c em um dado período t . Isso controla por fatores como eleições nacionais, mudanças na opinião pública do país ou estratégias de política externa que poderiam afetar o ativismo de seus representantes.
- **Efeito fixo de partido-tempo (λ_{pt}):** Absorve choques comuns a todos os membros de um partido político europeu p no tempo t , como mudanças na liderança do partido, alterações na plataforma ideológica ou estratégias coordenadas de atuação.
- **Efeito fixo de domínio-tempo (θ_{dt}):** Controla por fatores que afetam a saliência de um domínio temático d no tempo t para todos os parlamentares, como crises políticas, novas diretivas da Comissão Europeia ou eventos de grande repercussão midiática.

Ao incluir essa estrutura de efeitos fixos, a identificação do coeficiente β se dá a partir da variação do número de reuniões *dentro* de um mesmo parlamentar ao longo do tempo, após controlar por todas as fontes de variação comuns ao seu país, partido e aos temas em que atua.

A especificação econométrica adotada permite mitigar os principais desafios metodológicos da literatura. Primeiramente, a abordagem lida com o **viés de seleção e o problema das variáveis omitidas**, considerados os desafios mais críticos. Os lobistas não escolhem seus alvos aleatoriamente; a seleção baseia-se em alinhamento prévio, posição em comitês ou influência. A estratégia de efeitos fixos controla grande parte desses critérios de seleção: fatores não observáveis e estáveis do parlamentar (como sua ideologia ou competência intrínseca) são absorvidos, enquanto fatores variantes no tempo são capturados pelos efeitos fixos de país-tempo, partido-tempo e domínio-tempo. A premissa de identificação é que, uma vez controlados esses fatores, a variação residual no número de reuniões que um parlamentar recebe é exógena à sua produção legislativa.

Para além dos efeitos fixos, o modelo inclui um vetor de variáveis de controle (X'_{it}) que capturam características individuais e variantes no tempo do parlamentar, com especial atenção para os cargos que ocupa. Variáveis *dummy* indicam se o eurodeputado exerce funções de liderança, como a presidência de comissões, delegações ou grupos de trabalho. A inclusão desses controles é crucial para a estratégia de identificação, pois tais posições de poder são um critério central na seleção de alvos pelos lobistas e, ao mesmo tempo, podem influenciar diretamente o ativismo legislativo do parlamentar. Ao controlar por essas posições, isolamos o efeito do lobby de um importante fator de confusão, fortalecendo a premissa de que a variação residual no número de reuniões é exógena.

Adicionalmente, o modelo contorna o problema da **persistência temporal**. Como a estimação utiliza apenas a variação intra-parlamentar, o fato de alguns parlamentares receberem consistentemente mais lobby do que outros não envies a estimativa de β . O efeito é identificado a partir de parlamentares que alteram seu nível de engajamento com lobistas ao longo do tempo. A questão do **lobby contra-ativo** também é endereçada, uma vez que a variável de tratamento (L_{idt}) representa o volume total de reuniões, sem distinguir a direção da pressão. Assim,

o coeficiente β estimado representa o efeito *líquido* de uma reunião adicional. Para explorar a heterogeneidade desse efeito, o modelo é também estimado separadamente para diferentes tipos de lobistas (empresariais, ONGs e outros), permitindo analisar se a natureza do grupo de interesse altera o resultado.

Finalmente, em relação às múltiplas **faces do poder e canais de influência**, esta pesquisa delimita seu escopo para um canal (lobby direto via reuniões) e um tipo de resultado (ativismo legislativo). Essa delimitação é uma escolha metodológica deliberada para obter maior validade interna e uma identificação causal mais crível, ainda que se reconheça que o lobby opera por múltiplos canais e visa a diferentes resultados.

Em suma, a estratégia de identificação se baseia em um modelo de painel com efeitos fixos de alta dimensão que exploram a estrutura de variação dos dados (parlamentar, país, partido, domínio e tempo) para isolar o efeito do lobby de inúmeros fatores de confusão, conforme fundamentado pelo *framework* de análise do comportamento parlamentar.

Para testar a primeira hipótese (H1) — de que MPEs sujeitos a maior pressão de lobby exibem maior AL — foi empregada a estratégia de identificação detalhada na Seção 1.2. O teste empírico baseia-se na estimação do modelo de PPML apresentado na Equação 1.1, que utiliza o número de perguntas parlamentares como variável dependente e o número de reuniões com lobistas como a principal variável de tratamento.

A hipótese H1 é corroborada se o coeficiente de interesse, β , associado à variável de reuniões de lobby (L_{idt}), for positivo e estatisticamente significativo. Os erros padrão foram clusterizados no nível de domínio-tempo para corrigir a possível autocorrelação nos resíduos dentro de cada área temática ao longo do tempo.

Adicionalmente, para investigar a possibilidade de retornos marginais decrescentes, foi estimado um modelo quadrático, adicionando o termo de reuniões ao quadrado (L_{idt}^2) à especificação. Este modelo, apresentado na Equação 1.2, permite verificar se o efeito do lobby se atenua em níveis mais altos de intensidade.

$$AL_{idt} = \beta_1 L_{idt} + \beta_2 L_{idt}^2 + \gamma_{ct} + \lambda_{pt} + \theta_{dt} + X'_{it} \delta + \epsilon_{idt} \quad (1.2)$$

Um coeficiente β_2 negativo e significativo indicaria que, embora reuniões adicionais aumentem a atividade legislativa, o impacto de cada nova reunião é progressivamente menor. Isso sugere a existência de um ponto de saturação, a partir do qual o esforço de lobby adicional perde eficácia, ou efeitos marginais decrescentes.

Por fim, para explorar a heterogeneidade do efeito entre diferentes áreas de política pública, o modelo PPML principal foi estimado separadamente para cada um dos domínios temáticos. Essa análise permite avaliar se a influência do lobby é um fenômeno homogêneo ou se sua magnitude e significância variam conforme o contexto temático em que o MPE atua.

Para testar a segunda hipótese (H2) — de que as empresas exercem influência agregada superior sobre a AL quando comparadas a outras categorias de atores — adotamos uma estratégia em duas etapas que espelha dois mecanismos centrais do lobby: (i) *persuasão por reunião* e (ii) *acesso via volume de reuniões*. Essa decomposição permite separar a eficácia marginal de cada encontro do poder de garantir encontros em maior quantidade.

Primeiro, estimamos a eficácia marginal de uma reunião por tipo de ator com o modelo PPML, compatível com a especificação base (Seção 1.2). Depois, modelamos o volume de acesso com uma regressão Binomial Negativa sobre dados de contagem de reuniões por lobista. Por fim, aproximamos o efeito agregado multiplicando o efeito marginal por reunião previsto pelo PPML pelo número esperado de reuniões previsto pelo modelo de acesso.

Reutilizamos a especificação central do PPML, incluindo efeitos fixos de alta dimensão de país-tempo, partido-tempo e domínio-tempo, e o mesmo vetor de controles individuais X'_{it} utilizado em H1. Para comparar a eficácia por tipo de ator, substituímos a variável de tratamento L_{idt} por versões alternativas que isolam o número de reuniões atribuíveis a cada categoria: Empresas, ONGs e Outros. A especificação estimada é

$$AL_{idt} = \beta^{(k)} L_{idt}^{(k)} + \gamma_{ct} + \lambda_{pt} + \theta_{dt} + X'_{it} \delta + \epsilon_{idt}, \quad k \in \{\text{Empresas, ONGs, Outros}\} \quad (1.3)$$

onde $\beta^{(k)}$ é o coeficiente de interesse para o tratamento k . Os erros padrão

são clusterizados em domínio-tempo, mitigando correlações intracluster ao longo do tempo dentro de cada área temática. A evidência a favor de maior eficácia marginal por reunião para um dado tipo de ator decorre de um $\beta^{(k)}$ maior e estatisticamente significativo.

Como dito anteriormente, o PPML é apropriado para dados de contagem com muitos zeros e evita vieses de transformações logarítmicas em MQO. A estrutura de efeitos fixos (país-tempo, partido-tempo e domínio-tempo) aborda choques comuns e heterogeneidade não observada correlacionada com a variável de tratamento, preservando identificação na variação intra-parlamentar ao longo do tempo, conforme discutido na Seção 1.2.

Para captar o componente de acesso, estimamos o número de reuniões por lobista com um modelo Binomial Negativo (`MASS::glm.nb`), apropriado para contagens com sobredispersão. A variável dependente é o total de reuniões realizadas pelo lobista, explicada por: (i) categoria do ator (ONG, Empresa, Outros), (ii) orçamento máximo de lobby em log, (iii) interação entre categoria e orçamento, e (iv) controles setoriais de atuação do lobista e país da sede. Para interpretação limpa, recentramos o orçamento em seu valor médio. Em termos funcionais:

$$\mathbb{E}[\text{reuniões}_\ell | Z_\ell] = \exp(\alpha_0 + \alpha_1 C_\ell + \alpha_2 \ln B_\ell + \alpha_3 C_\ell \cdot \ln B_\ell + W'_\ell \eta), \quad (1.4)$$

onde ℓ indexa lobistas, B_ℓ é o orçamento máximo declarado, C_ℓ é a categoria do ator, W_ℓ agrega os controles e a interação ($C_\ell \cdot \ln B_\ell$) permite que o efeito do orçamento sobre o número de reuniões varie por categoria. Este modelo estima o número de reuniões esperadas por categoria e nível de orçamento.

Combinamos as duas etapas para aproximar o impacto total do lobby por categoria e orçamento:

$$\text{Efeito total}_{(k,B)} \approx \underbrace{\mathbb{E}[\text{reuniões} | k, B]}_{\text{Acesso}} \times \underbrace{\beta^{(k)}}_{\text{Persuasão por reunião (PPML)}}. \quad (1.5)$$

Essa métrica captura a ideia de que influência política depende tanto da eficácia marginal de cada encontro quanto da capacidade de produzir muitos encontros. A interpretação requer a premissa de *separabilidade* entre os processos de acesso e

persuasão: condicionais às variáveis observadas (e aos efeitos fixos na Etapa 1), os fatores não observados que afetam a capacidade de agendar reuniões não devem enviesar sistematicamente a eficácia por reunião. Discutimos essa premissa e suas implicações na seção de resultados.

Para testar a terceira hipótese (H3) — de que, em temas mais salientes, o lobby não empresarial (ONGs) é relativamente mais eficaz do que o lobby empresarial — estendemos a especificação PPML com efeitos fixos (Seção 1.2) para incorporar uma medida de saliência do tema e suas interações com o tipo de ator.

A saliência é operacionalizada como o volume total de reuniões de lobby observadas em cada combinação domínio-tempo, transformado por $\log(1+x)$ para lidar com zeros e reduzir assimetria, e posteriormente padronizado para média zero e desvio-padrão unitário (`salienc_std`). Esta escolha alinha-se à literatura que utiliza intensidade de atividade como proxy de saliência agregada do tema e permite interpretar os coeficientes de interação como variações no efeito marginal ao longo do gradiente de saliência.

Mantemos os mesmos efeitos fixos de alta dimensão e o vetor de controles individuais utilizados em H1 e H2 (país-tempo, partido-tempo e domínio-tempo; ver 1.2). A novidade é a inclusão de termos de interação entre o número de reuniões atribuíveis a cada categoria de ator e a saliência padronizada. A forma geral estimada é:

$$AL_{idt} = \sum_{k \in \{\text{Business}, \text{NGOs}, \text{Other}\}} \left[\beta_k L_{idt}^{(k)} + \phi_k L_{idt}^{(k)} \cdot \text{Sal}_{dt} \right] + \gamma_{ct} + \lambda_{pt} + \theta_{dt} + X'_{it} \delta + \epsilon_{idt} \quad (1.6)$$

onde $L_{idt}^{(k)}$ é o número de reuniões do MPE i com a categoria k no domínio d e tempo t , e Sal_{dt} é a saliência (padronizada) do domínio d no tempo t . Os termos β_k capturam o efeito marginal por reunião quando a saliência está no seu valor médio, enquanto ϕ_k capturam como esse efeito varia com a saliência. Erros-padrão são clusterizados em domínio-tempo para acomodar correlação serial dentro de cada área temática.

A identificação repousa na variação intra-parlamentar ao longo do tempo, com choques comuns de país, partido e domínio absorvidos pelos efeitos fixos. Um

ϕ_{NGOs} menos negativo (ou mais positivo) do que ϕ_{Business} indica que, à medida que a saliência aumenta, o efeito marginal das ONGs se deteriora mais lentamente ou permanece mais resiliente que o das empresas, implicando uma vantagem relativa em temas sob maior escrutínio. Passemos, então para a estratégia de validação e testes de robustez.

1.3 VALIDAÇÃO E TESTES DE ROBUSTEZ

A credibilidade das inferências causais desta pesquisa repousa não apenas na adequação da estratégia de identificação, mas também na robustez dos resultados a diferentes especificações metodológicas. Por essa razão, esta tese implementa uma bateria abrangente de testes de robustez que incluem: comparação entre PPML e MQO, variações na estrutura de efeitos fixos, diferentes estruturas de clustering dos erros padrão, especificações com e sem controles individuais, exclusão de observações extremas (*outliers*), análise por período legislativo, testes *jackknife* por país, estabilidade temporal dos efeitos, diferentes operacionalizações da variável de tratamento, transformações alternativas da variável dependente, especificações binárias e categóricas e testes placebo com tratamentos falsos.

Os resultados detalhados de todos esses testes são apresentados no Apêndice B, onde demonstramos que as conclusões principais da tese são robustas a uma ampla gama de decisões metodológicas e especificações alternativas.

Todos os códigos de análise, procedimentos de operacionalização dos dados e especificações econométricas estão totalmente documentados e disponíveis neste repositório¹. A estrutura modular dos scripts permite replicação completa de todos os resultados, desde a extração dos dados brutos até a geração das tabelas e figuras finais.

O processo de operacionalização foi desenhado para ser completamente reproduzível, com controles automatizados de qualidade e validações cruzadas entre diferentes fontes. As decisões metodológicas são explicitamente justificadas e, quando apropriado, comparadas com abordagens alternativas para demonstrar a robustez

¹ <https://github.com/AcacioTelechi/thesis-lobbying-effects>

das escolhas realizadas.

2 RESULTADOS

Este capítulo tem como objetivo apresentar e discutir os principais resultados empíricos da tese, respondendo às hipóteses formuladas e avaliando o impacto do lobby sobre a atividade legislativa no PE. Inicialmente, são descritas as características dos lobistas e dos parlamentares presentes na amostra, bem como padrões descritivos das reuniões e perguntas parlamentares. Em seguida, são apresentados os resultados dos testes das hipóteses centrais, com análise dos efeitos estimados, heterogeneidades e mecanismos identificados. Por fim, o capítulo discute as implicações dos achados para o debate sobre influência política e transparência institucional.

2.1 ANÁLISE DESCRITIVA DOS LOBISTAS

A Tabela 1 apresenta a distribuição de lobistas e sua atividade (mensurada pelo número de reuniões) por categoria organizacional. Os dados revelam uma predominância de interesses empresariais, que constituem o maior grupo tanto em número de organizações (2.409) quanto em volume de reuniões (22.427). Em média, cada organização empresarial realizou 9,31 reuniões, a maior taxa entre as categorias. As ONGs figuram como o segundo grupo mais ativo, seguidas pela categoria "Outros".

Tabela 1 – Distribuição de organizações, reuniões e reuniões por lobista, por categoria de organização

Categoria	Organizações	Reuniões	Reuniões por Organização
Empresarial	2.409	22.427	9,31
ONGs	1.271	10.966	8,63
Outros	1.042	6.950	6,67
Total	4.722	40.343	8,54

Esses resultados corroboram uma vasta literatura que aponta para a assimetria de recursos e representação no universo do lobby. Interesses empresariais, que

dispõem de maiores recursos financeiros, tendem a predominar na atividade de lobby (FIGUEIREDO; RICHTER, 2014). No contexto específico da União Europeia, estudos demonstram que grupos empresariais e associações nacionais são os mais presentes (DÜR; MATEO, 2012; EISING, 2007), o que levou alguns autores a caracterizarem o sistema como um "pluralismo de elite" (COEN, 1997; SCHMIDT, 2006).

A intensidade da atividade, refletida no elevado número total de reuniões (40.343), sinaliza uma forte competição pelo acesso aos tomadores de decisão, um recurso considerado escasso pela literatura (HALL; WAYMAN, 1990). Embora os atores empresariais detenham mais acesso, a presença expressiva de ONGs indica um ambiente competitivo, e não monolítico. Essa coexistência sugere que diferentes tipos de organizações mobilizam recursos distintos para influenciar o processo político. Enquanto interesses empresariais podem alavancar recursos financeiros e expertise técnica, ONGs e outros grupos da sociedade civil podem recorrer a recursos como legitimidade, representatividade e apoio público para defender suas causas (COEN; KATSAITIS, 2019a; DÜR, 2008).

Portanto, a análise descritiva inicial já delineia uma das tensões centrais investigadas neste trabalho: o desequilíbrio de poder e recursos entre interesses empresariais e outros setores da sociedade, e como essa assimetria se manifesta na competição por influência política no PE.

A análise do orçamento máximo declarado por categoria (Figura 1), em escala logarítmica, revela padrões distributivos distintos. O grupo empresarial, embora mais numeroso em quantidade de organizações, exibe a menor mediana de orçamento e uma distribuição notavelmente mais concentrada, o que sugere uma maior homogeneidade nos gastos com lobby declarados por essas entidades.

Em contrapartida, as ONGs e a categoria "Outros" apresentam medianas mais elevadas e uma variância substancialmente maior. Essa dispersão indica uma forte heterogeneidade interna, sugerindo a coexistência de organizações com recursos modestos e outras com orçamentos muito elevados, que influenciam a distribuição.

Este achado é particularmente interessante, pois parece tensionar a premissa da literatura de que interesses empresariais dominam em termos de recursos financeiros.

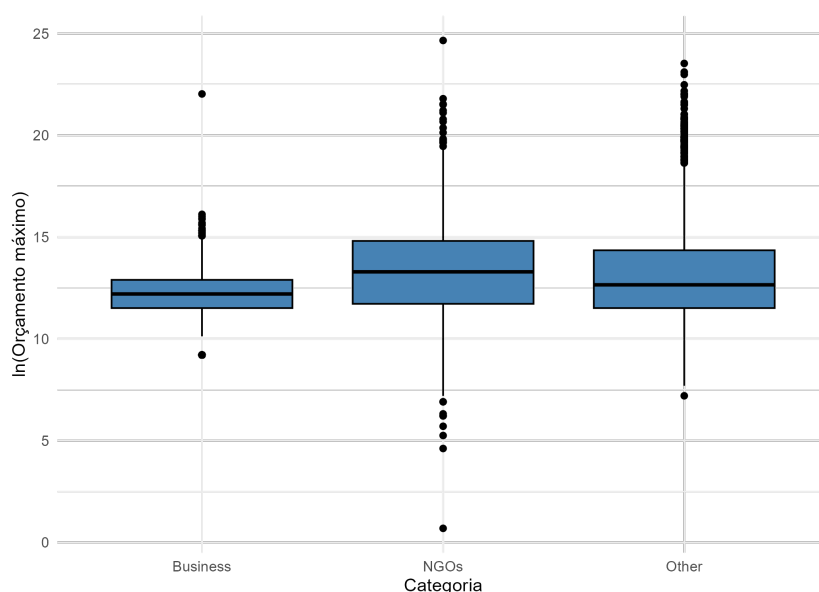


Figura 1 – Distribuição de orçamento máximo declarado ($\ln(\text{Orçamento Máximo Declarado})$) por categoria

ros (FIGUEIREDO; RICHTER, 2014; DÜR; MATEO, 2012). Uma interpretação possível é que a categoria empresarial seja composta por um grande número de atores com gastos mais padronizados, enquanto as categorias de ONGs e "Outros" contêm algumas organizações muito grandes, cujos orçamentos massivos podem refletir campanhas de alto custo ou estruturas de financiamento distintas. A alta variância pode, ainda, ser um reflexo de diferentes estratégias: enquanto empresas podem ter um dispêndio mais constante, ONGs podem mobilizar grandes somas para temas de alta saliência, onde a opinião pública é um fator crucial (MAHONEY, 2007).

Adicionalmente, a concentração no grupo empresarial pode estar relacionada ao problema da ação coletiva (OLSON, 1971), onde muitas empresas optam por atuar por meio de associações — que podem estar classificadas na categoria "Outros". De todo modo, os dados reforçam que a relação entre recursos financeiros e influência não é direta (SIMON, 1953), e que a capacidade de mobilizar diferentes tipos de recursos — não apenas financeiros, mas também de legitimidade e informação (COEN; KATSAITIS, 2019a) — é central para a dinâmica do lobby na UE.

A análise da distribuição geográfica dos lobistas (Figura 2) revela uma acentuada concentração em polos institucionais da União Europeia, com Bélgica (24,2%), Alemanha (15,3%) e França (10,1%) abrigando a maioria das organizações. Este padrão corrobora a tese da vantagem da proximidade institucional: a presença massiva na Bélgica, sede da Comissão Europeia e de grande parte das atividades do Parlamento, e na França, que sedia as sessões plenárias do Parlamento em Estrasburgo, é uma resposta estratégica à complexa governança multinível da UE (RICHARDSON, 2000).

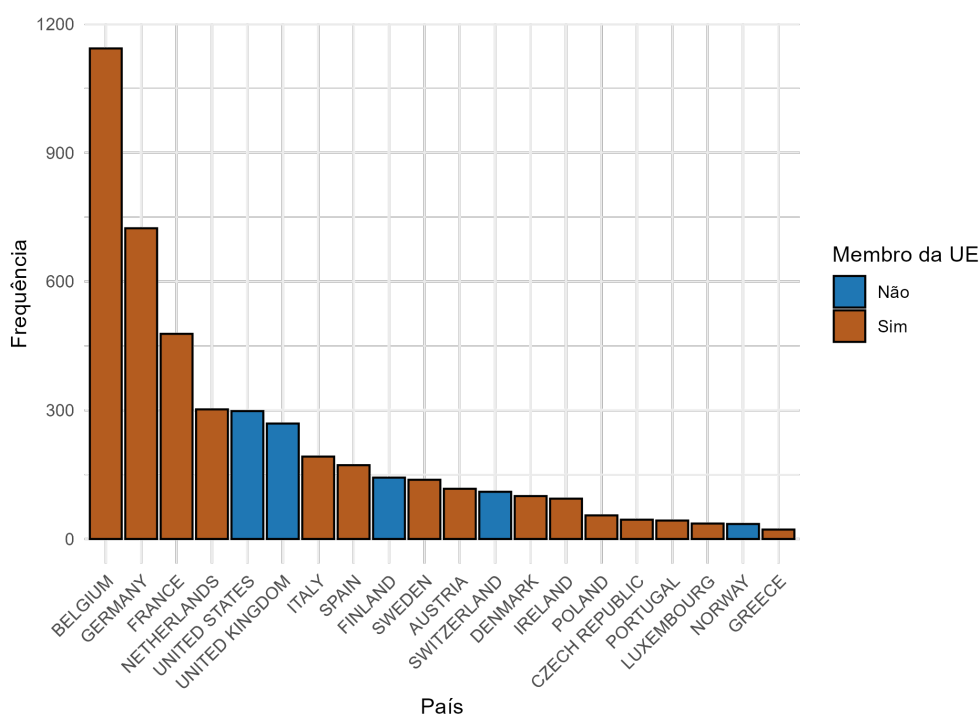


Figura 2 – Top 20 países-sede (n = 4.516 organizações, totalizando 95,6% do total de organizações)

O fenômeno do *venue-shopping* — a escolha estratégica do fórum mais favorável para exercer pressão — é central para entender essa distribuição (KLÜVER; BRAUN; BEYERS, 2015). A estrutura institucional da UE, com poder decisório partilhado entre a Comissão, o Conselho e o Parlamento, incentiva os grupos de interesse a se estabelecerem nos locais onde podem monitorar e intervir mais eficazmente no ciclo político. A proeminência de países como Alemanha e França

reflete não apenas a proximidade, mas também o peso político e econômico que exercem dentro da UE.

Adicionalmente, a presença significativa de atores extracomunitários, que representam cerca de 20% do total de organizações com sedes em 70 países (com destaque para os Estados Unidos, com 6,3%), evidencia a importância da UE como arena regulatória global. A atratividade do mercado europeu e o impacto de suas decisões incentivam atores internacionais a investir em uma presença local para influenciar a formulação de políticas que afetarão seus interesses.

O recorte do top 20, que mostra uma cauda longa com muitos países de baixa frequência, reforça a ideia de que, embora o sistema seja permeável, os altos custos de manter uma operação de lobby eficaz na Europa centralizam a influência em atores com maiores recursos e localização estratégica.

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.

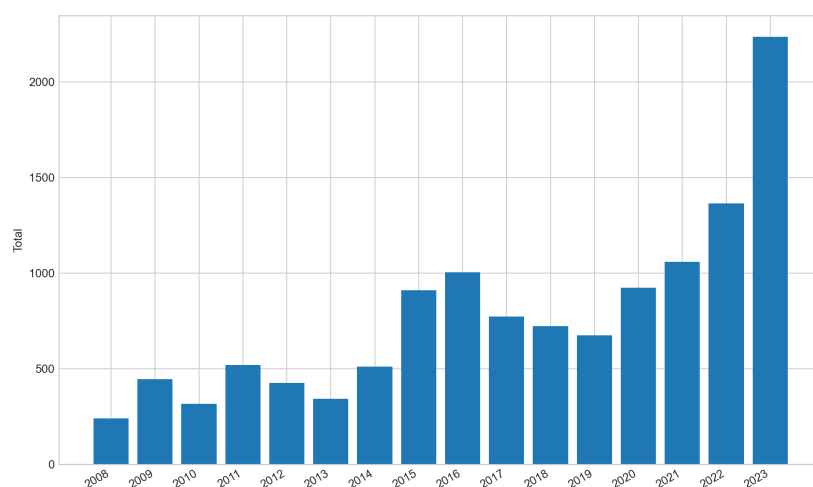


Figura 3 – Ano de registo

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 (Figura 4) destacam *Economia e Comércio* (15%), *Tecnologia* (14,9%) e *Infraestrutura e Indústria* (14,4%), seguidas por *Tecnologia* (14,2%) e *Meio Ambiente e Clima* (13,8%). Temas como *Saúde* (9,8%) e *Educação* (8,1%) são intermediários; *Agricultura* (7,2%) e *Direitos Humanos* (6,2%) têm menor incidência relativa.

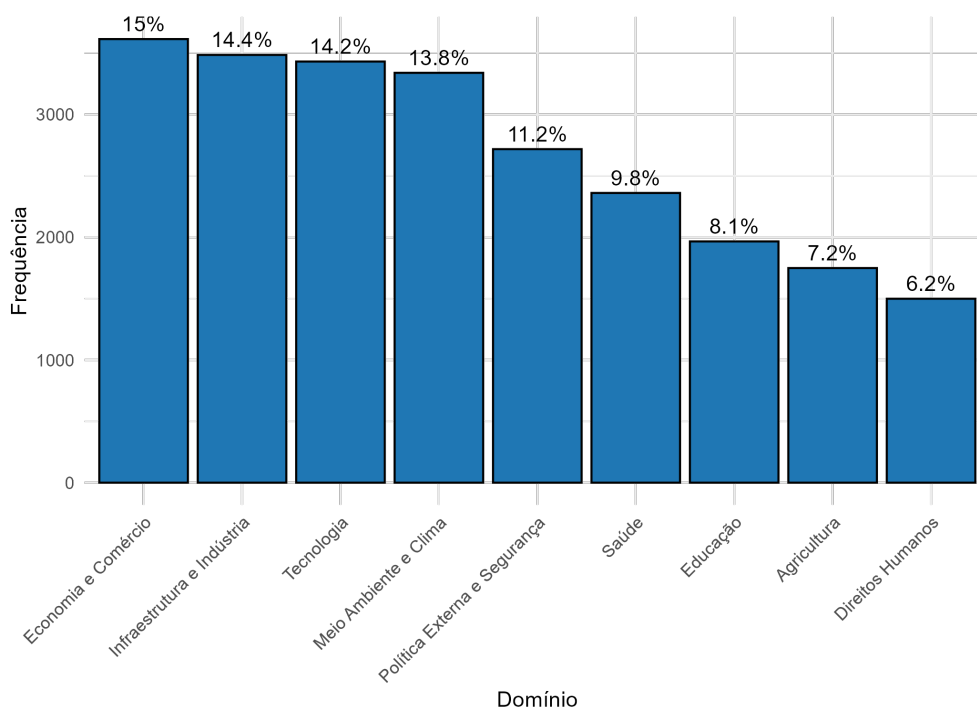


Figura 4 – Cobertura temática (proporção de entidades)

A distribuição temática (Figura 4) reflete uma concentração estratégica de esforços de lobby nas arenas de maior peso econômico e regulatório da União Europeia. A proeminência de domínios como "Economia e Comércio", "Infraestrutura e Indústria" e "Tecnologia" está alinhada com a literatura que demonstra uma correlação positiva entre a atividade de lobby e a saliência de um tema (CALDEIRA; HOJNACKI; WRIGHT, 2000; BAUMGARTNER; JONES, 2010). São áreas onde as decisões políticas implicam altas consequências financeiras, mobilizando um volume expressivo de atores.

Adicionalmente, esses temas são caracterizados por uma elevada complexidade técnica. Este fator aumenta a demanda por informação especializada por parte

dos decisores políticos, criando um vácuo que os lobistas buscam preencher para ganhar acesso e exercer influência, conforme sugere a literatura sobre o papel informacional do lobby (KLÜVER, 2012). A própria agenda política da UE, com iniciativas centrais como o Mercado Único Digital e o Pacto Ecológico Europeu, transforma esses domínios em focos de intensa atividade, corroborando a ideia de que os grupos de interesse são mais ativos nos temas em que o Estado também é mais atuante (MAHONEY, 2008).

Em contraste, temas como "Direitos Humanos" e "Agricultura", embora relevantes, apresentam menor incidência relativa. Isso pode indicar que os interesses difusos ou setoriais dessas áreas enfrentam maiores desafios de ação coletiva ou utilizam estratégias de influência distintas, que não se refletem com a mesma intensidade no lobby direto. O padrão observado, portanto, sugere um ecossistema de lobby que responde de forma racional tanto aos incentivos econômicos quanto às necessidades informacionais geradas pela agenda regulatória da UE.

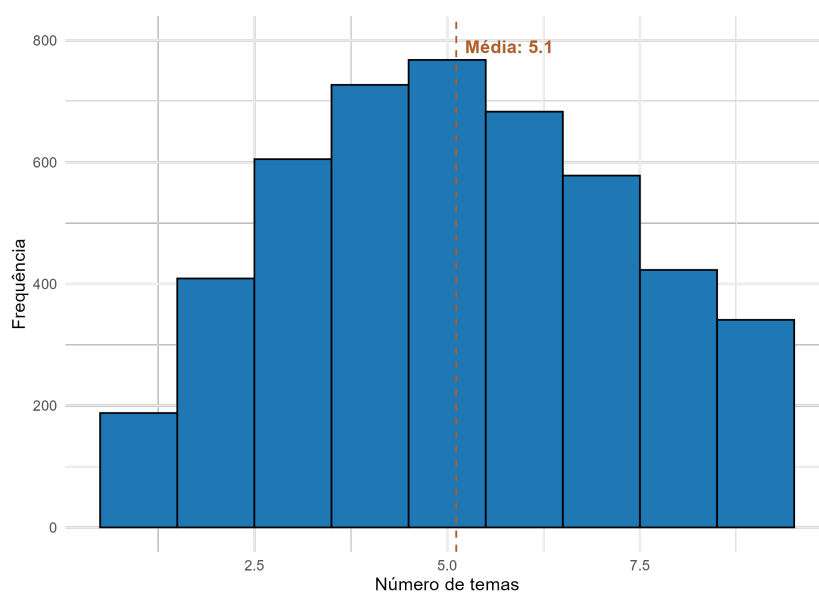


Figura 5 – Número de temas por lobista

A análise da diversidade temática das organizações (Figura 5) revela que a maioria dos lobistas atua em múltiplas frentes. A distribuição, que se assemelha a uma curva normal com média de 5,1 temas por organização, demonstra a coexistência de estratégias especializadas (focadas em poucos temas) e generalistas. Contudo,

o padrão dominante é o de uma abordagem multi-temática, onde as organizações cobrem um portfólio de 3 a 7 áreas de interesse.

Este achado dialoga diretamente com o debate na literatura sobre os recursos mais valiosos para o lobby: expertise ("o que você sabe") versus conexões ("quem você conhece") (BERTRAND; BOMBARDINI; TREBBI, 2014). A presença de atores especializados sugere uma aposta na expertise técnica como via de influência, especialmente em domínios de alta complexidade (KLÜVER, 2012). Por outro lado, a predominância de atores multi-temáticos indica que uma estratégia mais ampla, possivelmente baseada em redes de contatos e na capacidade de adaptação a diferentes arenas políticas, é a norma no complexo ambiente institucional da UE. A governança multinível e a interconexão de políticas na Europa podem incentivar as organizações a não se limitarem a um único nicho, buscando pontos de acesso em diferentes comissões e agências (COEN; KATSAITIS, 2019b).

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.

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 (*Empresas*, *ONGs* e *Outros*), 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 UE, sugerindo um campo de ação onde interesses difusos e concentrados buscam simultaneamente agenda e influência.

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 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.

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 PE. 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 1.727 deputados ao longo de 63 meses, de agosto de 2019 a outubro 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. A tabela 2 apresenta as estatísticas agregadas do painel de dados.

Tabela 2 – Estatísticas agregadas de tratamento por deputado

Estatística	Valor
Total de reuniões	235.638
Total de perguntas	118.238
Total de deputados únicos	1,727
Deputados que receberam tratamento	804
Taxa de tratamento por deputado (%)	46,5%
Entre deputados tratados:	
Reuniões médias por deputado	293,1
Reuniões medianas por deputado	109,0
Desvio padrão	468,9
Correlação agregada:	
Correlação reuniões-perguntas	0,05

Considerando a unidade de análise a tríade MEP-domínio-mês, temos 979.209 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 agosto de 2019 a outubro 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 presenciais, 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.

A análise da evolução temporal (Figura 6) revela uma dinâmica complexa na interação entre a atividade de lobbying (reuniões) e a atividade parlamentar (perguntas). O padrão mais saliente é a **divergência de tendências** a partir do início de 2022. Enquanto o volume de perguntas parlamentares (linha laranja) permanece relativamente estável ao longo de todo o período, oscilando dentro de uma faixa consistente - com leve tendência de queda -, o número de reuniões de lobby (linha azul) apresenta um crescimento acentuado e um aumento expressivo da volatilidade a partir de 2022.

A estabilidade no número de perguntas parlamentares sugere que este instrumento funciona mais como uma ferramenta de rotina da fiscalização e posicionamento político, menos elástica às flutuações de curto prazo da agenda legislativa mais contenciosa. Além disso, ambas as séries exibem uma clara **sazonalidade**,

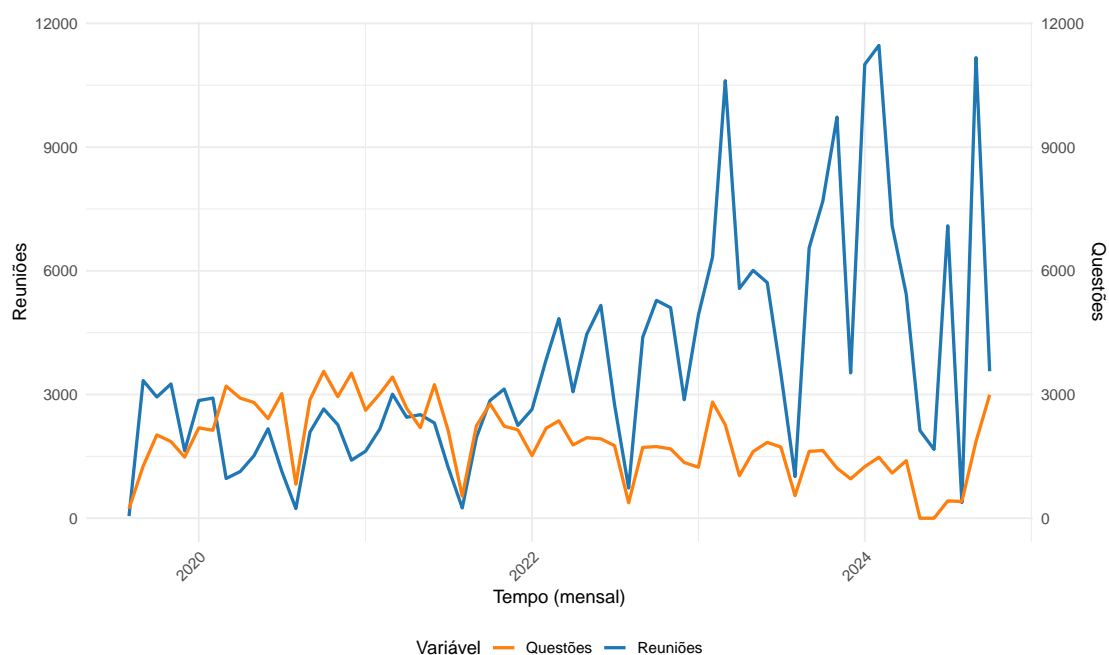


Figura 6 – Evolução temporal da atividade parlamentar e de lobbying

com quedas de atividade que correspondem aos períodos de recesso do calendário parlamentar, um fator institucional que molda o ritmo do trabalho político.

Do ponto de vista metodológico, estes padrões são cruciais. A tendência de crescimento no lobby e a sazonalidade em ambas as séries justificam a inclusão de efeitos fixos de tempo (ex: mês-ano) nos modelos econométricos, para controlar choques temporais comuns e variações sazonais que poderiam confundir a estimação dos efeitos. A divergência entre as séries a partir de 2022 também sugere a importância de se testar a estabilidade dos parâmetros do modelo ao longo do tempo.

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 informações sobre a concentração e heterogeneidade dos fenômenos estudados.

As Figura 7, Figura 8 e Figura 9 apresentam uma análise dos padrões de participação agregados por deputado, revelando aspectos da distribuição da atividade

de lobbying no PE que impactam a identificação causal.

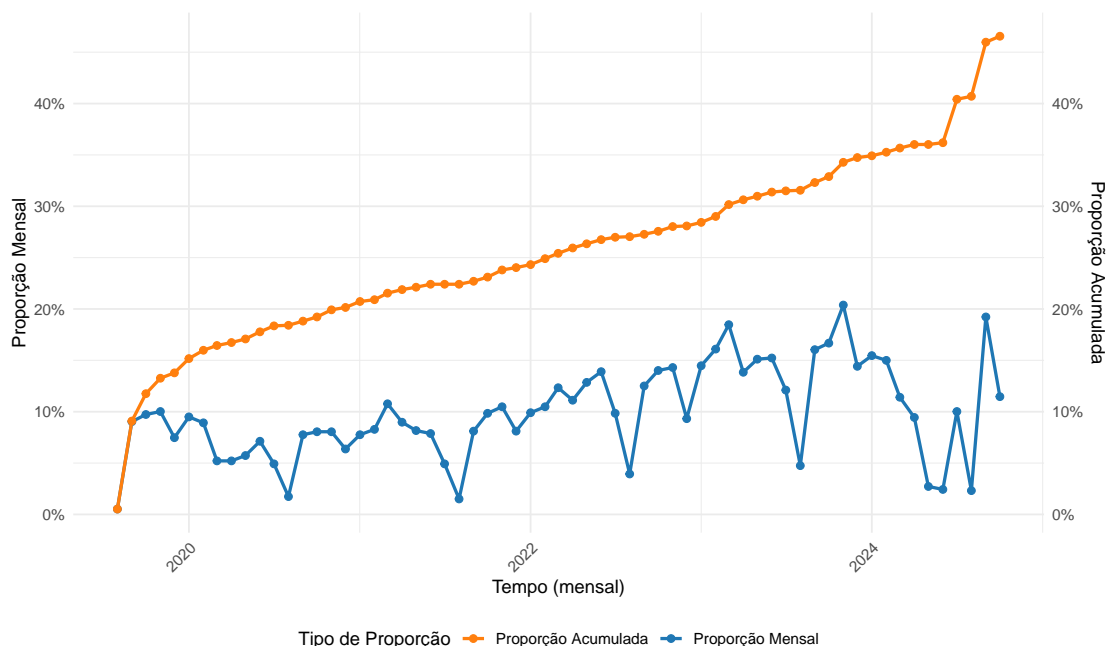


Figura 7 – Evolução temporal da proporção de MPEs que participaram de reuniões de lobbying

A análise revela três características fundamentais da distribuição de tratamento. Primeiro, existe **participação substancial mas não universal** (Figura 7): 46,5% dos deputados (804 de 1.727) 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 (Figura 8). Entre os deputados que receberam lobbying, a distribuição é altamente assimétrica: enquanto a mediana é de 109 reuniões por deputado, a média é de 293,1 reuniões, indicando que uma minoria de parlamentares concentra uma quantidade 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 (0,05), contrastando com correlações mais elevadas observadas no nível temporal (Figura 9). Este padrão sugere que os efeitos

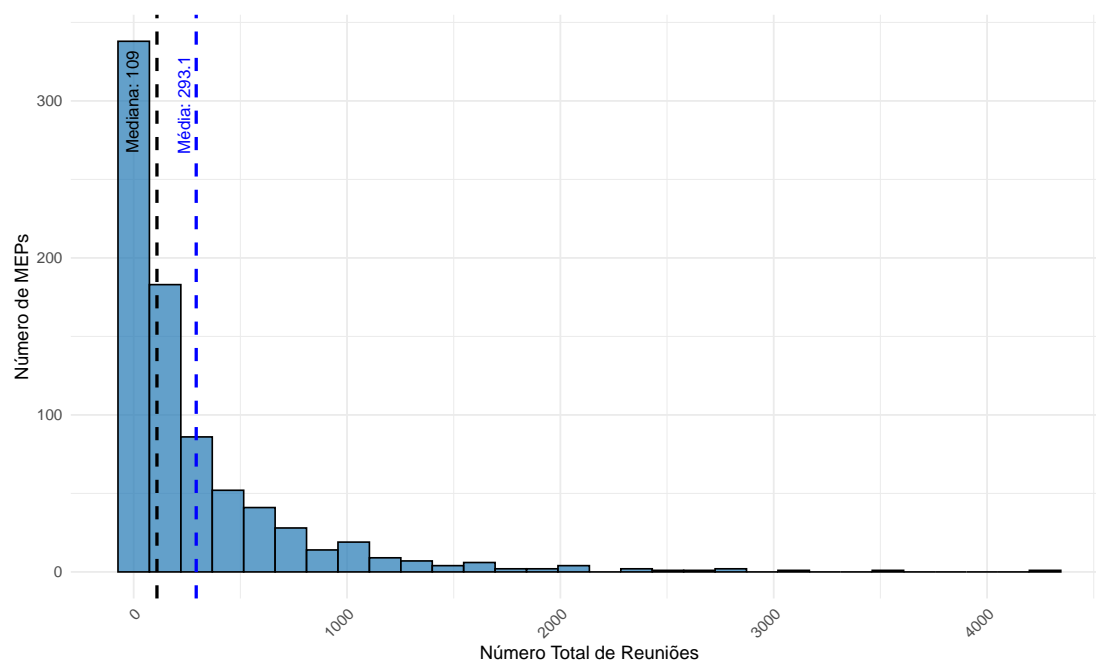


Figura 8 – Distribuição de reuniões por MPE

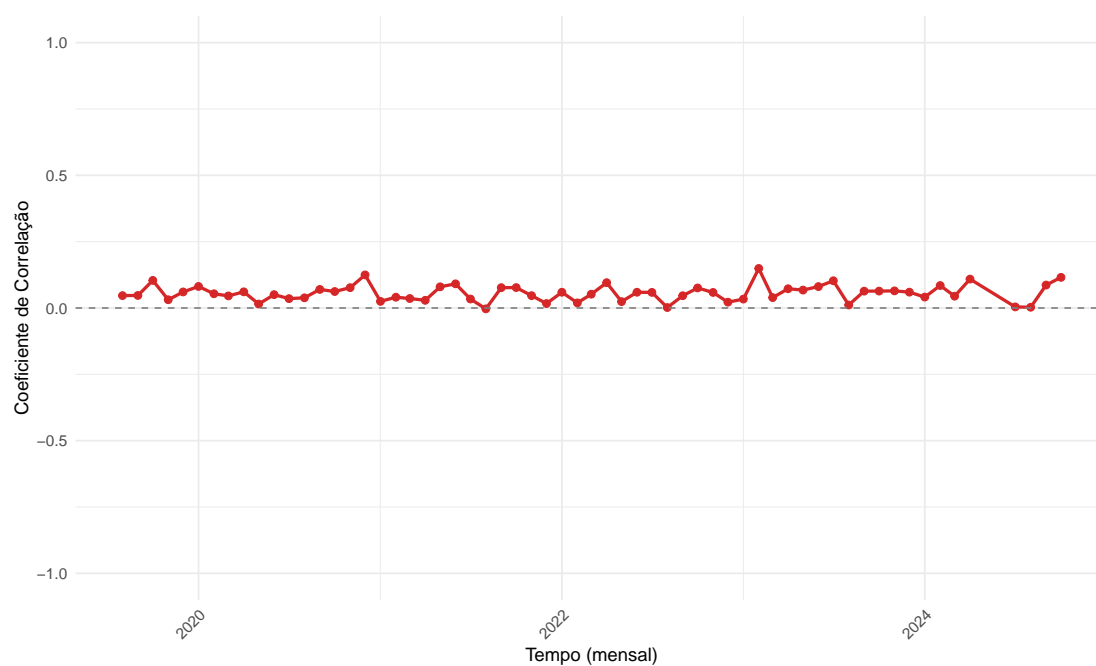


Figura 9 – Evolução temporal da correlação entre reuniões e perguntas

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* ao invés de *cross-sectional between-individual*. Esta evidência preliminar orienta a especificação de modelos com efeitos fixos de deputado para controlar heterogeneidade não observada invariante no tempo.

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 análise da Tabela 3 revela uma heterogeneidade sistemática entre os domínios de política pública, que se manifesta de forma distinta na penetração e na intensidade do lobby.

Em termos de **penetração** (a proporção de deputados que recebem ao menos uma reunião), a variação entre os domínios é moderada, oscilando entre 41,2% (Agricultura) e 45,7% (Tecnologia). Essa pequena amplitude sugere que o lobby, como prática, é uma atividade transversal e disseminada por todo o Parlamento Europeu, não se restringindo a nichos específicos. Ainda assim, o padrão é teoricamente consistente: domínios ligados à regulação econômica e de alta complexidade (Tecnologia, Economia e Comércio, Infraestrutura) apresentam as maiores taxas, refletindo os elevados interesses em jogo (MAHONEY, 2007).

A heterogeneidade torna-se muito mais acentuada quando se analisa a **intensidade média do tratamento** (o número médio de reuniões por deputado "tra-

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

Domínio	MEPs (A)	MEPs Tratados (B)	Reuniões (C)	Perguntas	Penetração (B/A %)	Intensidade Média (C/B)
Tecnologia	1727	789	32.067	9.918	45,69	40,64
Economia e Comércio	1727	788	34.762	13.884	45,63	44,11
Assuntos Externos e Segurança	1727	783	28.041	13.462	45,34	35,81
Infraestrutura e Indústria	1727	782	32.795	10.544	45,28	41,94
Meio Ambiente e Clima	1727	777	32.038	14.665	44,99	41,23
Saúde	1727	769	23.868	20.602	44,53	31,04
Educação	1727	738	18.670	2.949	42,73	25,30
Direitos Humanos	1727	727	17.194	26.482	42,10	23,65
Agricultura	1727	711	16.203	5.732	41,17	22,79

tado"). Aqui, a variação é extrema: um deputado ativo em "Economia e Comércio" recebe, em média, 44 reuniões, quase o dobro das 23 reuniões de um deputado ativo em "Agricultura". Esta é a dimensão onde a alocação estratégica de recursos se torna mais evidente. Domínios com altas consequências econômicas e complexidade técnica atraem não apenas um pouco mais de lobistas, mas um esforço de influência massivamente mais concentrado, validando a tese de que os recursos de lobby fluem para as arenas de maior saliência e complexidade regulatória (KLÜVER, 2012).

É notável também o contraste entre a intensidade do lobby e o volume de atividade parlamentar. O domínio de "Direitos Humanos", por exemplo, apresenta o maior número de perguntas parlamentares (26.482), mas uma das menores taxas de penetração e a segunda menor intensidade de lobby. Este padrão sugere que a dinâmica política varia conforme o tema. Enquanto domínios econômicos são caracterizados por um lobby intenso e técnico, temas de grande apelo público como "Direitos Humanos" podem ser mais influenciados por estratégias de *advocacy* e pela sinalização política dos parlamentares através de instrumentos como as perguntas, em vez do lobby direto medido por reuniões.

A unidade de análise adotada — deputado-domínio-mês — revela uma alta proporção de observações com valor zero ($>92\%$) tanto para reuniões quanto para perguntas. Contudo, esta "inflação de zeros" não representa um problema estatístico, mas sim uma característica substantiva do comportamento parlamentar: a **especialização temática**. Como documentado na literatura (SCHILLER, 1995; BURDEN, 2015), parlamentares concentram sua atividade em um subconjunto de domínios, resultando em zero atividade na maioria das combinações deputado-domínio. A atividade de lobby, por sua vez, segue essa especialização, direcionando-se aos parlamentares já ativos em cada tema.

A natureza "artificial" dessa inflação de zeros é confirmada quando os dados são agregados. Ao considerar apenas os domínios em que um deputado demonstra alguma atividade, a proporção de zeros cai para 47,1% para perguntas e 55,8% para reuniões. Em um nível agregado por domínio-mês, a inflação de zeros torna-se negligível (3,2% e 0%, respectivamente), indicando que há atividade constante em todos os domínios quando se considera o conjunto de parlamentares relevantes. Portanto, a escolha do modelo econométrico PPML justifica-se pela natureza da variável dependente (dados de contagem com excesso de zeros), evitando os vieses de modelos lineares com transformações logarítmicas.

A análise descritiva delineia um ecossistema de lobbying no PE caracterizado por (i) dinamismo temporal acentuado, (ii) concentração da atividade em um subconjunto de deputados e (iii) heterogeneidade entre os domínios de política pública. Longe de serem meramente contextuais, esses padrões empíricos são a principal justificativa para a estratégia econométrica adotada nesta tese, detalhada no Capítulo 1.

Primeiramente, a combinação de uma correlação agregada próxima de zero entre reuniões e perguntas com uma alta concentração da atividade de lobby em poucos deputados demonstra a inadequação de abordagens puramente *cross-sectional*. A fonte de identificação mais promissora reside na variação *within-individual*, ou seja, em como a atividade de um mesmo deputado muda ao longo do tempo em resposta a variações no lobby que recebe.

Em segundo lugar, a presença de tendências temporais claras, sazonalidade e

choques de atividade específicos por domínio (como o aumento expressivo do lobby a partir de 2022) torna imperativo o uso de uma estrutura robusta de efeitos fixos. A escolha por efeitos fixos de alta dimensão — **país**×**tempo**, **partido**×**tempo** e **domínio**×**tempo** — é uma resposta direta a esses padrões, permitindo controlar uma vasta gama de fatores de confusão não observados e isolar de forma mais crível o efeito causal de interesse.

Terceiro, a heterogeneidade da intensidade do lobby entre os domínios valida a decisão de não se limitar a um único efeito médio, motivando as análises de heterogeneidade que exploram como o impacto do lobby pode variar conforme o contexto temático.

Finalmente, a natureza da variável dependente — um dado de contagem com uma alta proporção de zeros explicada pela especialização temática — fundamenta a escolha do estimador PPML. Esta abordagem modela adequadamente a estrutura dos dados, evitando os vieses de modelos lineares e reconhecendo a especialização como um comportamento substantivo, e não como um mero problema estatístico.

Em suma, a análise descritiva não apenas caracteriza o fenômeno em estudo, mas também estabelece as bases empíricas que guiam e validam a estratégia de identificação causal empregada nos capítulos seguintes para testar as hipóteses centrais desta pesquisa.

2.3 ANÁLISE DE EFEITOS DO LOBBY

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 descritas no Capítulo 1. O coeficiente estimado 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 incre-

mento 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 Tabela 4, com intervalos de confiança baseados em erros-padrão agrupados.

Tabela 4 – Resumo dos modelos PPML para a Hipótese 1

	PPML	PPML (Quad.)
Reuniões	0,025*** (0,002)	0,098*** (0,007)
Reuniões ²		-0,004*** (0,001)
Observações	600.237	600.237
Efeitos fixos	país×tempo; partido×tempo; domínio×tempo	
Cluster	domínio×tempo; membro	

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

Tabela 4 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 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 (0,098 vs. 0,004), 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, podem sustentar níveis muito mais altos de lobbying; tal padrão é consistente com

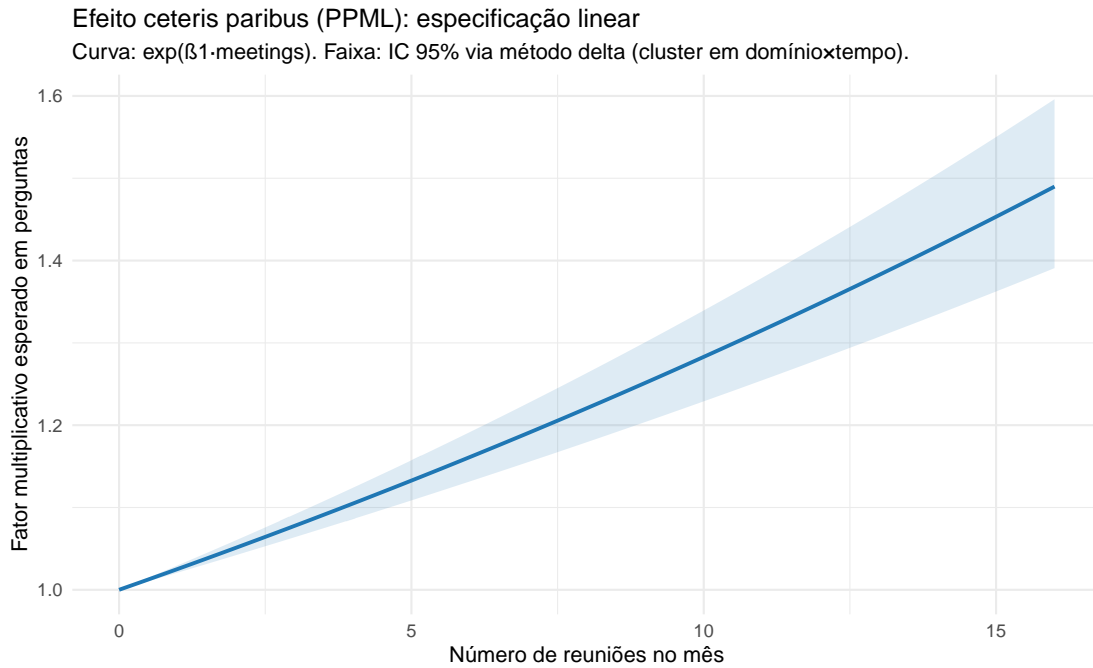


Figura 10 – Efeito esperado ceteris paribus: especificação linear (PPML)

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

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 10 e Figura 11 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×tempo

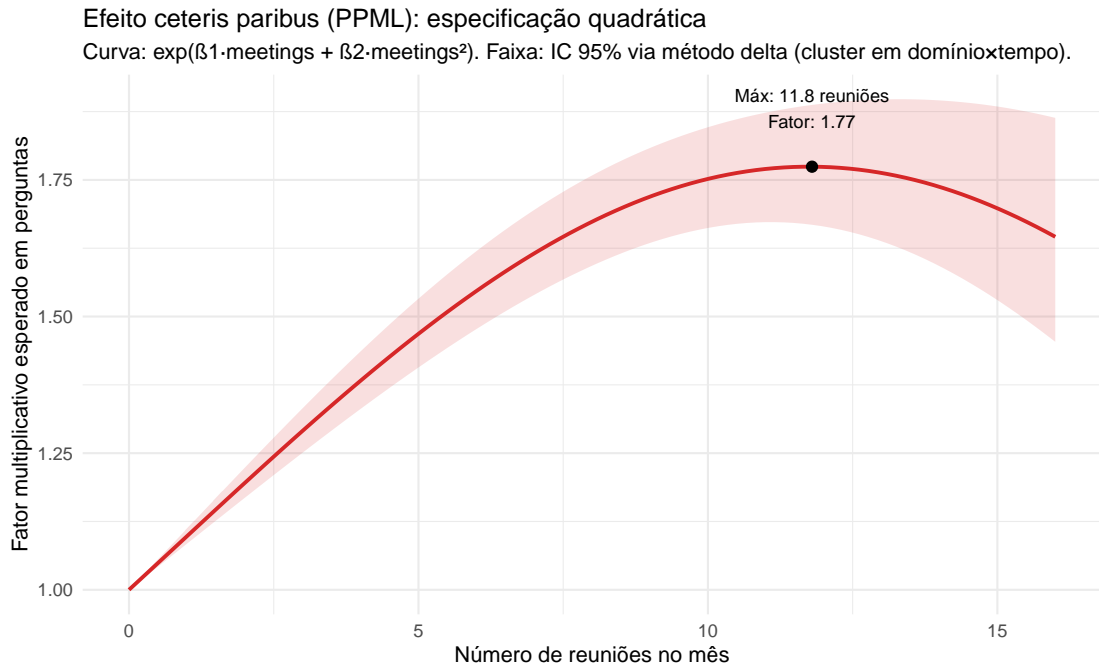


Figura 11 – Efeito esperado ceteris paribus: especificação quadrática (PPML)

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.

e partido×tempo controlam heterogeneidade não observada invariável e choques comuns, permitindo comparação *within* do mesmo MPE ao longo do tempo. A inferência usa erros-padrão agrupados em duas dimensões (domínio×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.

A análise desagregada por domínios de políticas públicas revela que o efeito positivo das reuniões sobre a atividade parlamentar é consistente em praticamente todas as áreas temáticas consideradas. Conforme ilustrado na Figura 12, a estimativa do coeficiente associado a *meetings* permanece positiva em todos os domínios,

ainda que a magnitude do efeito varie entre eles. Por exemplo, domínios como "Agricultura" e "Educação" apresentam efeitos mais pronunciados, sugerindo que nesses setores o lobbying pode ser particularmente eficaz em estimular a apresentação de perguntas parlamentares. Já em áreas como "Economia e Comércio" ou "Tecnologia", embora o efeito também seja positivo, sua magnitude é ligeiramente inferior, o que pode refletir diferenças na dinâmica de atuação dos grupos de interesse ou na agenda dos parlamentares nesses temas.

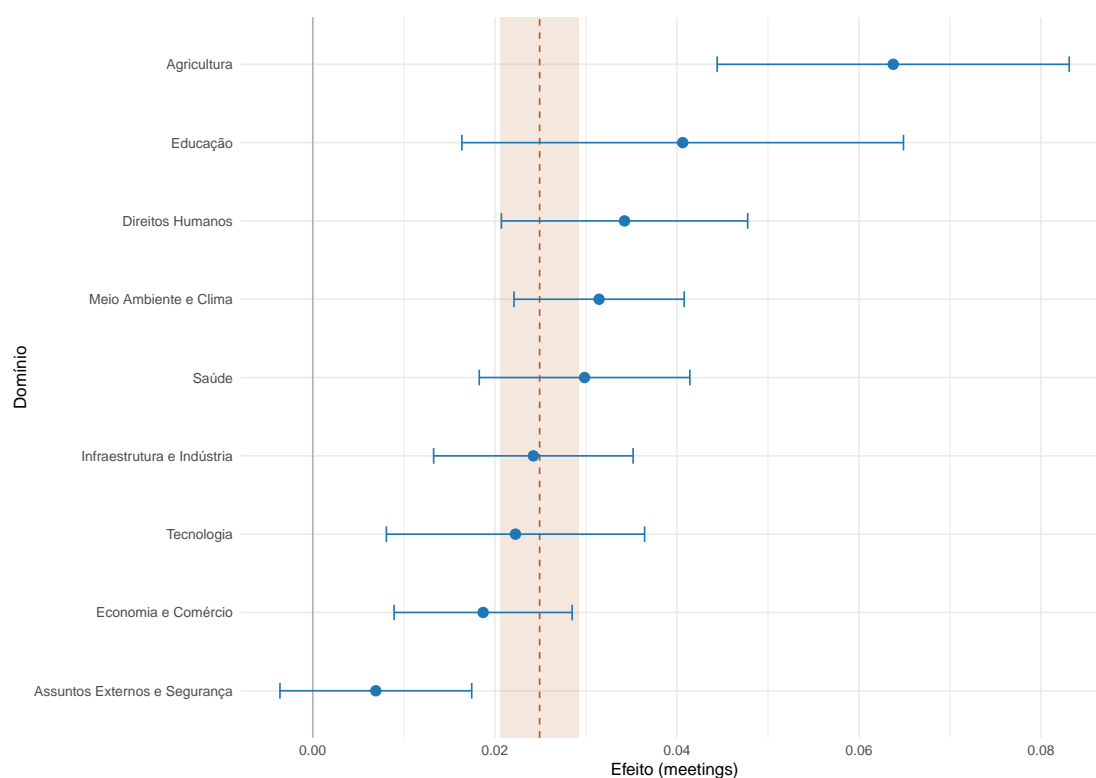


Figura 12 – Efeito esperado *ceteris paribus*: especificação linear (PPML) para cada domínio

Nota: Cada ponto azul representa a estimativa do coeficiente associado a *meetings* para um domínio específico de políticas públicas, refletindo o efeito marginal esperado de reuniões sobre o número de perguntas parlamentares naquele domínio *ceteris paribus*. As linhas horizontais correspondem aos intervalos de confiança de 95% para cada estimativa, indicando a incerteza estatística. A linha tracejada laranja indica o efeito médio estimado para todos os domínios, servindo como referência para comparação entre áreas temáticas.

Além disso, os intervalos de confiança indicam que, apesar de variações na

precisão das estimativas entre domínios, o sinal positivo do efeito é robusto e estatisticamente distinto de zero na maioria dos casos. Isso reforça a conclusão de que a associação entre intensidade de lobbying e atividade de fiscalização parlamentar não se restringe a um setor específico, mas se manifesta de forma generalizada no Parlamento Europeu, ainda que com intensidades distintas conforme o contexto temático.

De modo geral, esses resultados sugerem que o impacto do lobbying sobre a produção de perguntas parlamentares é um fenômeno transversal aos diferentes campos de políticas públicas, evidenciando a relevância desse mecanismo de influência em múltiplas agendas legislativas.

2.3.1 TESTE DA HIPÓTESE 2: UMA ANÁLISE DECOMPOSTA DO EFEITO DO LOBBY

A avaliação da Hipótese 2, que postula uma maior influência das empresas sobre a atividade parlamentar em comparação com outros atores, exige uma análise que transcenda a simples contagem de reuniões. Uma análise preliminar do efeito marginal por reunião, apresentada na Figura 13, sugere que as ONGs, paradoxalmente, exercem uma influência maior por encontro. Este resultado, embora contraintuitivo, destaca a necessidade de um modelo mais completo que considere a heterogeneidade dos atores de lobby.

O efeito marginal isolado, contudo, é insuficiente para um teste robusto da hipótese. A influência total de um grupo de interesse não depende apenas da *eficácia* de cada reunião, mas também da sua capacidade de assegurar *acesso* – isto é, o volume de reuniões que consegue realizar. Argumentamos que o impacto total é uma função dessas duas componentes: a frequência do acesso e a eficácia da persuasão em cada encontro.

Para capturar essa dualidade, adotamos uma estratégia de modelagem em duas etapas que decompõe o processo de lobby da seguinte forma:

- **Acesso (Frequência):** O processo pelo qual um lobista garante reuniões com os MPEs. Esta etapa, modelada com uma regressão Binomial Negativa,

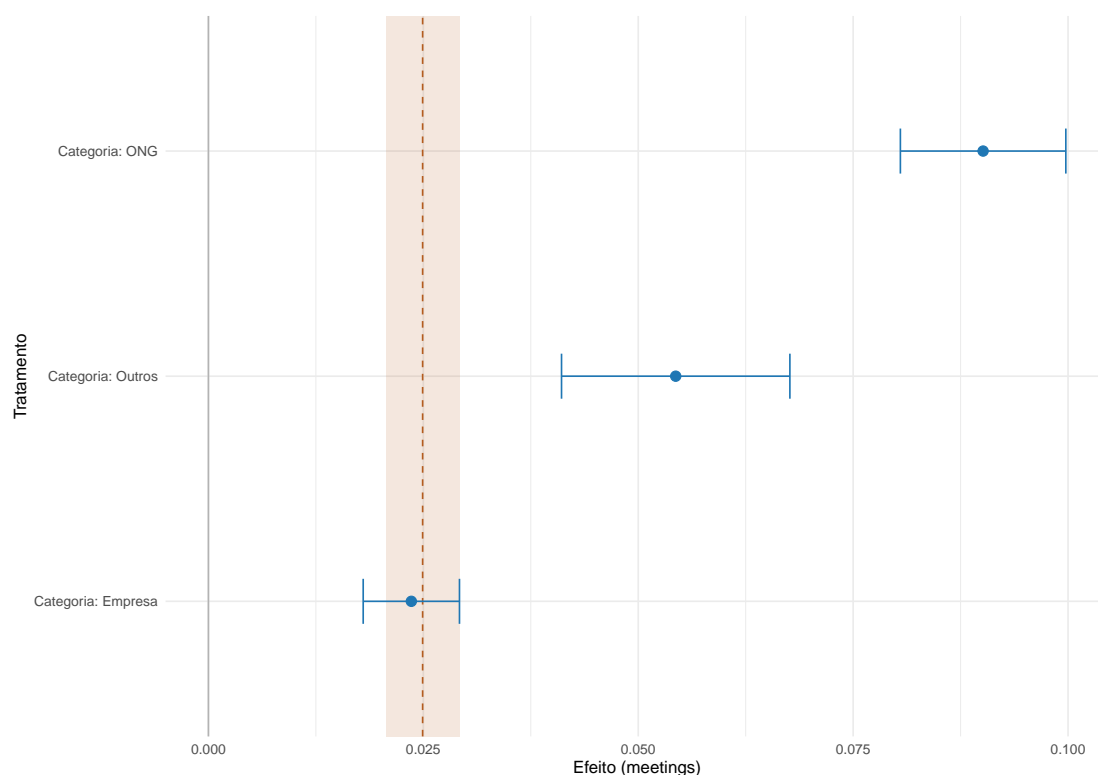


Figura 13 – Efeito marginal por tipo de ator: especificação PPML
 Nota: Cada ponto representa a estimativa do coeficiente para *reuniões*, associado a um tipo de ator (tratamento), refletindo o efeito marginal esperado de uma única reunião sobre o número de perguntas parlamentares, mantendo os efeitos fixos constantes. As linhas horizontais indicam os intervalos de confiança de 95%. A faixa vertical sombreada representa o intervalo de confiança do efeito médio geral, servindo como referência.

responde à pergunta: "*Quantas reuniões um determinado lobista consegue obter?*"

- **Persuasão (Eficácia):** O processo pelo qual um lobista utiliza uma reunião para influenciar a atividade de um MPE. Esta etapa, modelada com PPML, responde à pergunta: "*Quão eficaz é uma única reunião para gerar atividade parlamentar?*"

Esta abordagem permite-nos decompor e compreender os mecanismos através dos quais diferentes atores exercem influência.

Para estimar o volume de reuniões (Acesso) - Etapa 1 -, utilizamos um modelo de regressão Binomial Negativo, apropriado para dados de contagem com sobre-dispersão. A variável dependente é o número de reuniões que um lobista realiza, e as variáveis explicativas incluem o orçamento de lobby, a categoria do ator (ONG, Empresa, Outros) e um termo de interação entre orçamento e categoria, além de controles setoriais e geográficos.

O coeficiente de interação entre ser uma empresa e o orçamento de lobby é particularmente revelador. Um resultado positivo e estatisticamente significativo para este termo indica que as empresas não só tendem a realizar mais reuniões em média, mas também demonstram uma "eficiência alocativa" superior: cada aumento percentual no seu orçamento se traduz num aumento maior no número de reuniões em comparação com as ONGs. A Figura 14 ilustra essa dinâmica, mostrando o número esperado de reuniões em função do orçamento.

Observa-se um ponto de inflexão: abaixo de um orçamento de aproximadamente \$27.000 ($\approx e^{10.2}$), as ONGs tendem a realizar mais reuniões. Acima desse limiar, a capacidade das empresas de converter recursos financeiros em acesso torna-se proeminente, e a disparidade cresce exponencialmente com o orçamento.

Para estimar o efeito agregado (Etapa 2), combinamos os resultados das duas etapas, multiplicando o número esperado de reuniões (o *Acesso*, da Etapa 1) pelo efeito marginal por reunião (a *Persuasão*, da Figura 13) para cada categoria de ator e nível de orçamento. O resultado, ilustrado na Figura 15, representa uma aproximação de primeira ordem do impacto total do lobby.

A análise do efeito total revela uma relação complexa e não-linear. Embora uma única reunião com uma ONG seja, em média, mais influente, a superioridade das empresas em garantir um grande volume de acesso, especialmente quando dispõem de orçamentos elevados, reverte essa vantagem. Para orçamentos abaixo de \$8,8 milhões ($\approx e^{16}$), o efeito total das ONGs permanece superior. No entanto, acima de \$40 milhões ($\approx e^{17.5}$), o efeito agregado das empresas torna-se substancialmente maior.

Estes resultados oferecem um suporte nuançado à Hipótese 2 e dialogam diretamente com a literatura sobre os mecanismos de influência do lobby. A influência

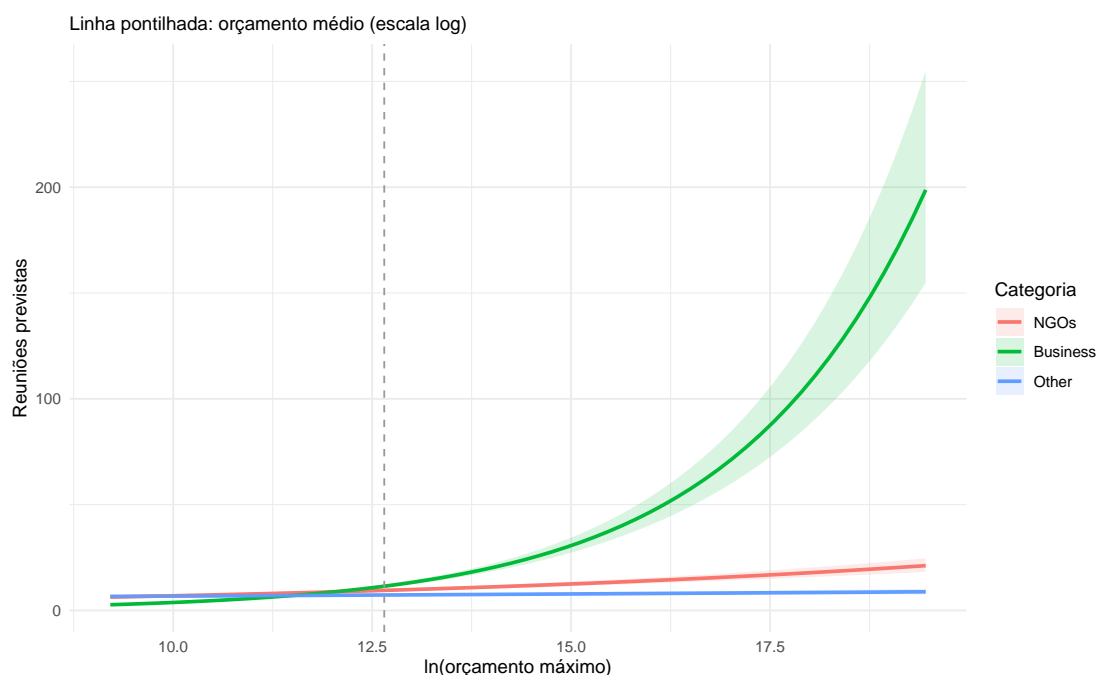


Figura 14 – Previsão do número de reuniões por categoria e orçamento
 Nota: O gráfico exibe o número esperado de reuniões (eixo Y) em função do logaritmo do orçamento de lobby (eixo X), com base no modelo Binomial Negativo. As curvas representam a previsão para cada categoria de ator, mantendo as demais variáveis em seus valores médios ou modais. A linha tracejada vertical indica o orçamento médio na amostra. As áreas sombreadas correspondem aos intervalos de confiança de 95%.

das empresas não é incondicionalmente superior, mas torna-se dominante quando alavancada por vastos recursos financeiros, um achado que se alinha com a discussão sobre a desigualdade de representação (MAHONEY, 2007). A decomposição do efeito em "acesso" e "persuasão" permite-nos explorar as diferentes naturezas dos recursos mobilizados pelos atores, conforme aponta a literatura (FIGUEIREDO; RICHTER, 2014; POP, 2013).

A maior eficácia marginal por reunião das ONGs pode ser interpretada à luz do seu capital reputacional e da sua legitimidade percebida (BUNEA, 2018). Do ponto de vista do comportamento parlamentar, interagir com ONGs pode ser uma estratégia de *vote-seeking* para os MPEs, que buscam sinalizar alinhamento com causas de interesse público e, assim, aumentar seu apelo eleitoral (IBENSKAS;

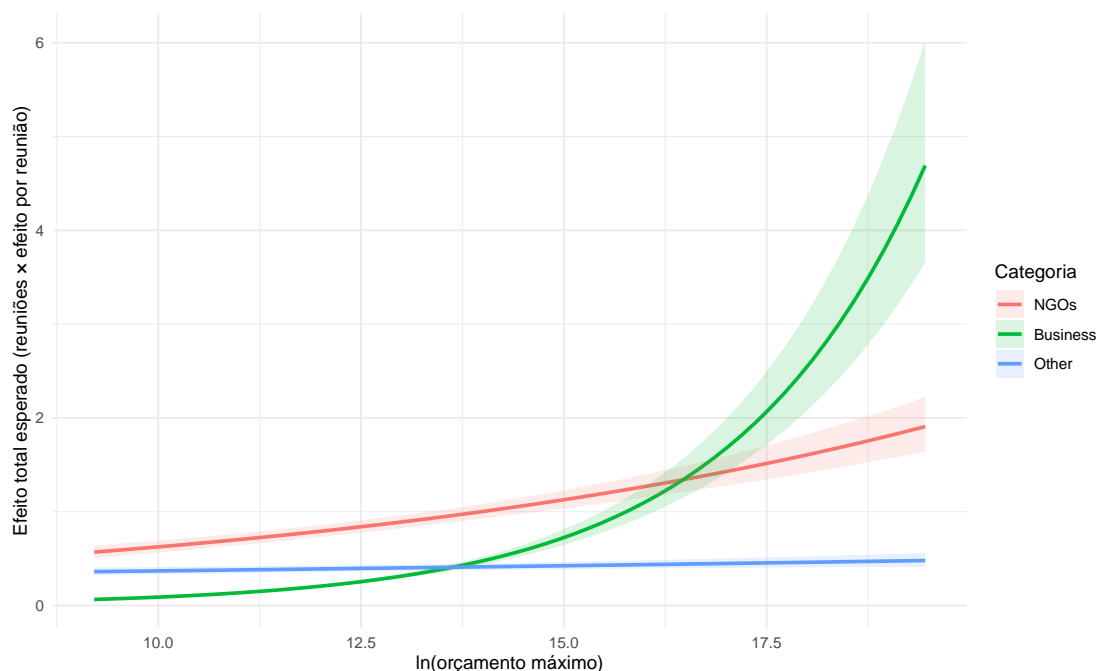


Figura 15 – Efeito total estimado por categoria e orçamento

Nota: O gráfico representa o efeito total esperado, calculado como o produto do número previsto de reuniões (da Figura 14) e do coeficiente de efeito marginal por reunião (da Figura 13). O eixo Y representa o aumento esperado no número de perguntas parlamentares.

BUNEA, 2021; MAYHEW, 2004).

Por outro lado, a capacidade das grandes corporações de converter recursos financeiros em um volume massivo de acesso aponta para outro mecanismo de influência: a subsidiação de informação. Para atingir seus objetivos de carreira (*career-seeking*) e de formulação de políticas (*policy-seeking*), os parlamentares necessitam de informação técnica e especializada (DANIEL, 2015; KLÜVER, 2012). As grandes empresas, com seus recursos, estão em posição privilegiada para fornecer este subsídio informacional, estabelecendo uma relação de troca (HUWYLER; TURNER-ZWINKELS; BAILER, 2023) que lhes garante um acesso privilegiado e contínuo. Assim, a capacidade de "saturar" o ambiente informacional com interações frequentes parece ser um fator decisivo para a sua influência agregada.

É importante notar que esta abordagem metodológica assenta na premissa de

separabilidade entre os processos de "acesso" e "persuasão". Esta premissa implica que, após controlarmos pelas variáveis observáveis (como o orçamento), os fatores não observados que tornam um lobista eficaz em garantir reuniões são estatisticamente independentes dos fatores não observados que o tornam influente durante essas reuniões. A suposição de separabilidade poderia ser violada se, por exemplo, uma "qualidade" ou "reputação" intrínseca do lobista, não capturada pelo modelo, afetasse simultaneamente a sua capacidade de agendar reuniões e a receptividade dos parlamentares às suas propostas. Nesse cenário, a multiplicação dos efeitos poderia levar a uma estimativa enviesada do impacto total.

Em suma, os resultados indicam que, embora o discurso das ONGs possa ter maior ressonância por interação, a capacidade financeira das grandes empresas permite-lhes superar essa desvantagem através de uma presença quantitativamente esmagadora, confirmando a importância crítica dos recursos na determinação da influência política.

2.3.2 TESTE DA HIPÓTESE 3: O EFEITO DO LOBBY EM TEMAS SALIENTES

A Hipótese 3 postula que, em temas de maior saliência, o lobby exercido por organizações não empresariais (como as ONGs) tem uma maior probabilidade de influenciar a atividade legislativa dos MPEs em comparação com o lobby de organizações empresariais. A lógica subjacente é que, quando um tema está sob intenso escrutínio público, os parlamentares se tornam mais sensíveis a argumentos que ressoam com a opinião pública e a interesses difusos, frequentemente representados por ONGs.

Para testar esta hipótese, mantivemos a estrutura do modelo PPML com efeitos fixos, garantindo a consistência com as análises anteriores. A principal diferença metodológica foi a introdução de uma variável para capturar a saliência de um tema e a sua interação com os diferentes tipos de lobistas.

A saliência foi operacionalizada como uma proxy baseada na intensidade da atividade de lobby, uma abordagem que encontra respaldo na literatura (BAUMGARTNER; JONES, 2010). Especificamente, criamos uma variável (*saliencia_std*)

PPML com Interação (H3)	
Empresa (base)	0,035*** (0,006)
ONG (base)	0,090*** (0,006)
Outros (base)	0,032** (0,010)
Empresa x Saliência	−0,022*** (0,005)
ONG x Saliência	−0,016* (0,007)
Outros x Saliência	−0,024* (0,011)
Num.Obs.	600 237
R2	0,253
RMSE	0,56
Std.Errors	by: cl_dt
FE: fe_ct	X
FE: fe_pt	X
FE: fe_dt	X
+ p <0,1, * p <0,05, ** p <0,01, *** p <0,001	

que mede o volume total de reuniões de lobby dentro de cada domínio temático para cada período mensal, padronizada para ter média zero e desvio padrão um. Um valor mais alto nesta variável indica que um tema atraiu mais atenção de todos os grupos de interesse, sendo, portanto, considerado mais saliente.

O modelo econométrico foi então especificado para incluir termos de interação entre cada categoria de lobista (Empresa, ONG, Outros) e a variável de saliência, apresentado na Equação 1.6. Esta especificação permite-nos estimar como o efeito marginal de uma reunião de cada tipo de ator varia em função do nível de saliência do tema. Os resultados da regressão estão sumarizados na Tabela ?? e visualizados no gráfico de efeitos marginais na Figura 16.

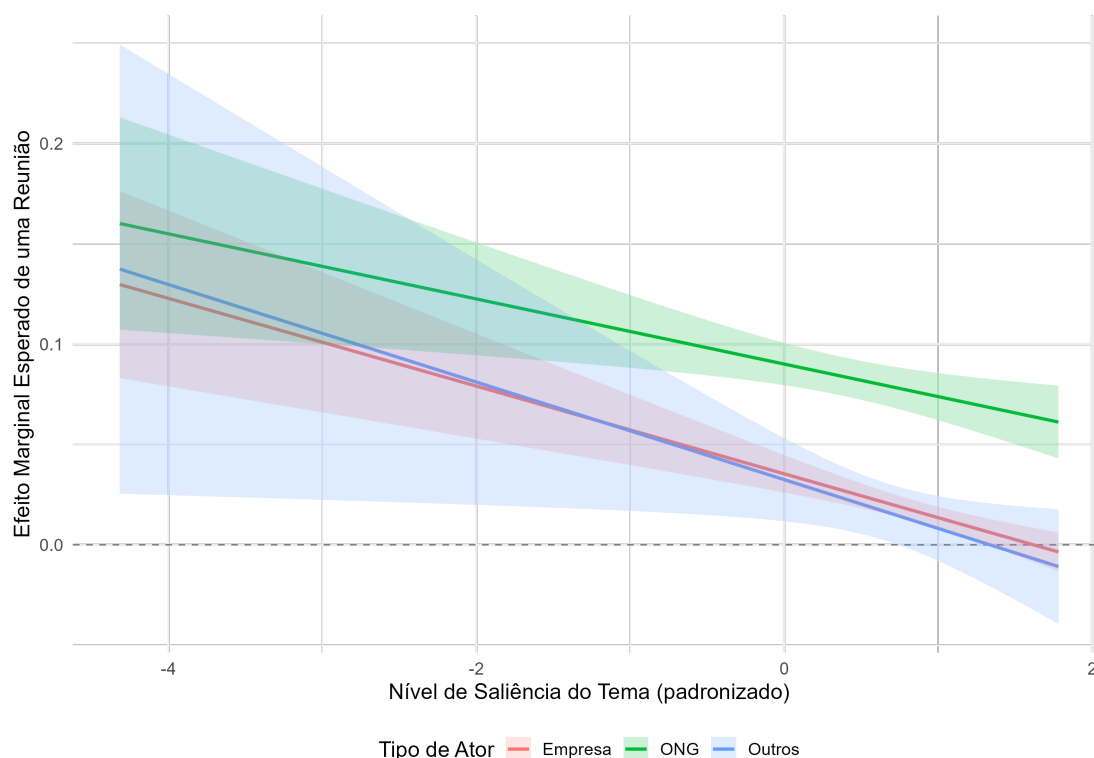


Figura 16 – Efeito do Lobby Condicional à Saliência do Tema

Nota: O gráfico exibe o efeito marginal esperado de uma única reunião sobre o número de perguntas parlamentares (eixo Y) em diferentes níveis de saliência do tema (eixo X). As linhas representam a estimativa para cada categoria de ator, e as áreas sombreadas correspondem aos intervalos de confiança de 95%, calculados via bootstrap.

A análise revela um padrão complexo que contradiz parcialmente, mas também enriquece, a Hipótese 3. Contrariamente à expectativa de que o efeito das ONGs aumentaria com a saliência, observamos que o efeito marginal de uma reunião **diminui** para todos os grupos à medida que um tema se torna mais saliente (coeficiente negativo para todos os grupos nas variáveis de interação). Este achado está em forte alinhamento com a literatura, que sugere que a influência do lobby direto decresce quando a opinião pública e a atenção da mídia se intensificam, forçando os parlamentares a se alinharem a considerações eleitorais mais amplas (MAHONEY, 2007; KOLLMAN, 1998).

No entanto, a análise revela uma heterogeneidade crucial na taxa dessa diminui-

ção. Três pontos principais se destacam na Figura 16. Em temas de baixa saliência (à esquerda do gráfico), o efeito das ONGs é similar estatisticamente ao de empresas e outros atores. Isso pode ser observado pela intersecção das áreas sombreadas das linhas, que indicam o intervalo de confiança de 95% das estimativas.

À medida que a saliência aumenta (movendo-se para a direita no gráfico), a vantagem comparativa das ONGs se acentua significativamente. O efeito do lobby de empresas e de outros atores decai rapidamente, enquanto o efeito das ONGs se mostra muito mais resiliente, diminuindo a uma taxa consideravelmente menor.

Em temas de alta saliência, onde a influência de empresas e outros grupos se torna estatisticamente indistinguível de zero (seus intervalos de confiança cruzam a linha pontilhada), o efeito das ONGs permanece positivo, robusto e estatisticamente significativo. É precisamente neste contexto de maior escrutínio público que a sua influência relativa se torna mais pronunciada.

Os resultados validam a Hipótese 3. Em temas de maior saliência, o lobby de organizações não empresariais é, de fato, mais eficaz em aumentar a atividade parlamentar em comparação com o lobby empresarial. A nuance importante é que essa maior eficácia não se manifesta como um aumento absoluto do efeito, mas sim como uma resiliência superior à pressão do escrutínio público, o que amplia a sua vantagem comparativa.

Esta descoberta dialoga diretamente com a teoria sobre os recursos do lobby e os incentivos parlamentares. Em temas de baixa saliência, os parlamentares, focados em seus objetivos de formulação de políticas (*policy-seeking*), podem valorizar o subsídio informacional técnico fornecido por empresas. Contudo, quando um tema ganha visibilidade, os incentivos de reeleição (*vote-seeking*) tornam-se dominantes (MAYHEW, 2004). Nesse contexto, alinhar-se a interesses empresariais pode ter um custo político elevado, enquanto responder a ONGs, que detêm maior capital de legitimidade (BUNEA, 2018), reforça a imagem pública do parlamentar.

Os achados confirmam que a influência do lobby é altamente contextual e que, sob o escrutínio público, a vantagem se desloca para os atores percebidos como representantes de interesses mais amplos e difusos.

A OPERACIONALIZAÇÃO

Esta seção descreve o processo de operacionalização dos dados utilizados na presente pesquisa, desde a extração das fontes primárias até a construção final dos painéis de dados utilizados nas análises econométricas. A transparência e replicabilidade deste processo são fundamentais para assegurar a confiabilidade dos resultados obtidos.

A pesquisa utiliza três principais fontes de dados, todas públicas e oficiais:

(1) API do Parlamento Europeu: Utilizada para extrair dados sobre perguntas parlamentares e reuniões com representantes de interesse. A API fornece dados estruturados em formato JSON sobre atividades parlamentares desde 2009, incluindo metadados como datas, participantes, e identificadores únicos.

(2) Registro de Transparência da União Europeia: Base de dados em formato Excel contendo informações sobre organizações registradas como representantes de interesse, incluindo categoria, orçamento, área de atuação e data de registro.

(3) Base de Dados de Membros do Parlamento Europeu (MEPs): Informações sobre os parlamentares europeus, incluindo filiação partidária, país de origem, comissões de participação e períodos de mandato.

O processo de extração de dados sobre reuniões foi implementado através de um sistema de *web scraping* paralelo (*script 1_extraction.ipynb*). O sistema utiliza a classe `ParliamentDataHandler` para realizar requisições à API do Parlamento Europeu com controle de taxa de requisições para evitar sobrecarga do servidor.

A extração cobriu o período de janeiro de 2024 a outubro de 2024 inicialmente, sendo posteriormente expandida para o período completo de 2009 a 2024. O sistema implementa:

- Processamento paralelo com controle de concorrência (máximo de 10 *threads*)
- Controle de taxa de requisições (500 requisições por janela de 5 minutos)

- Sistema de *cache* para evitar requisições duplicadas
- Tratamento robusto de erros com *retry* automático

Cada reunião extraída contém informações sobre: data, parlamentar(es) envolvido(s), representantes presentes, capacidade do parlamentar na reunião, e identificadores únicos.

A extração de perguntas parlamentares seguiu processo similar, mas com etapas adicionais devido à necessidade de acessar o conteúdo textual completo (*script 1.2_extraction_questions_data.ipynb*). O processo envolveu:

Fase 1 - Metadados: Extração dos metadados das perguntas através da API, obtendo identificadores únicos, autores, datas de publicação e tipos de pergunta.

Fase 2 - Conteúdo Textual: Download dos documentos PDF correspondentes a cada pergunta a partir do repositório documental do Parlamento Europeu. Esta fase processou 72.154 documentos, implementando sistema de *retry* para documentos inacessíveis.

Fase 3 - Processamento Textual: Extração e limpeza do conteúdo textual dos PDFs utilizando a biblioteca PyPDF2, com foco na seção "Subject" de cada pergunta para análise temática posterior.

Um dos desafios centrais foi o *matching* entre os nomes das organizações presentes nas reuniões e aquelas registradas no Registro de Transparência (*script 2.0_treat_fuzzy.ipynb*). Como, na grande maioria dos casos, não havia um identificador único do lobista na base de dados de reuniões, mas apenas o nome e organização representada, foi preciso uma estratégia de parear o nome da organização dessa base de dados com a do Registro de Transparência. Este processo utilizou técnicas de correspondência *fuzzy* devido a variações na grafia e denominações oficiais.

O algoritmo implementado utiliza múltiplas métricas de similaridade:

- *Ratio*: Similaridade geral entre strings
- *Partial ratio*: Melhor correspondência de subsequências

- *Token sort ratio*: Similaridade após ordenação de tokens
- *Token set ratio*: Similaridade de conjuntos de tokens únicos

O processo foi paralelizado em lotes de 10 organizações, processando 99.370 correspondências potenciais em 5 horas e 40 minutos. Foram consideradas correspondências de alta confiança aquelas com $\text{score} \geq 85\%$, resultando em 2.166 matches confirmados.

A classificação temática das perguntas parlamentares foi realizada utilizando técnicas de processamento de linguagem natural (*script 2.6_analyze_questions_docs.ipynb*). O processo empregou o modelo `facebook/bart-large-mnli` para classificação *zero-shot* multi-rótulo. Após iterações exploratórias com diferentes conjuntos de categorias baseados nos interesses declarados pelos lobistas no Registro de Transparência, definiu-se uma taxonomia final de 9 domínios temáticos:

1. Assuntos Externos e Segurança
2. Economia e Comércio
3. Tecnologia
4. Infraestrutura e Indústria
5. Meio Ambiente e Clima
6. Saúde
7. Direitos Humanos
8. Educação
9. Agricultura

O processo classificou 72.155 perguntas em lotes de 8 documentos, executado em GPU durante 28 horas e 58 minutos. Cada pergunta recebeu scores de probabilidade para todos os domínios, permitindo análises tanto de classificação única (domínio mais provável) quanto multi-rótulo.

O tratamento dos dados dos parlamentares envolveu a criação de um *time-line* temporal para capturar mudanças em filiações partidárias, participação em comissões e outros aspectos institucionais (*script 2.3_treat_meps.ipynb*).

Para cada parlamentar e data relevante, o sistema determina:

- Grupo político de filiação
- País de origem
- Participação em comissões (tipos e funções)
- Outras funções institucionais

Este processo gerou um dataset temporal com 1.353 parlamentares únicos ao longo de todas as datas relevantes do período de análise.

Os lobistas registrados foram categorizados em três grupos principais com base em sua natureza organizacional (*script 2.1_treatment_meetings_orgs.ipynb*):

Business: Empresas, grupos empresariais e associações comerciais e de negócios.

NGOs: Organizações não-governamentais, plataformas e redes similares.

Other: Demais categorias incluindo sindicatos, think tanks, instituições acadêmicas, consultorias, entidades públicas mistas, e representações de terceiros países.

Adicionalmente, foram criadas variáveis de classificação orçamentária (baixo, médio, alto) baseadas na distribuição dos orçamentos declarados e de tempo desde registro (em dias). Para ambas as variáveis, a categorização foi feita com base em quartis. As organizações no primeiro quartil (os 25% menores valores) foram classificadas como 'baixo'. As que se encontravam entre o primeiro e o terceiro quartil (valores entre 25% e 75% da distribuição) foram classificadas como 'médio', e as no quartil superior (os 25% maiores valores) como 'alto'. Para a variável de orçamento, foi utilizado o logaritmo natural do valor máximo do orçamento declarado para suavizar a distribuição antes da categorização.

A construção do painel de dados envolveu decisões importantes sobre a agregação temporal. Testaram-se três níveis: diário, semanal e mensal. A agregação

mensal foi selecionada como padrão por (i) reduzir esparsidade excessiva dos dados diários, (ii) manter granularidade suficiente para capturar variações temporais, e (iii) facilitar a identificação econométrica com maior variabilidade.

O painel final possui estrutura tridimensional (parlamentar \times tempo \times domínio temático), permitindo análises dos efeitos heterogêneos por área temática. A construção envolveu (*scripts* 2.4_panel_data.ipynb e 2.7_panel_data_w_topics.ipynb):

Passo 1 - Agregação Temporal: Contagem de perguntas e reuniões por parlamentar-mês-domínio, considerando múltiplas classificações temáticas quando aplicável.

Passo 2 - Junção com Dados Institucionais: Para cada observação parlamentar-tempo, incorporação das características institucionais vigentes (partido, país, comissões).

Passo 3 - Preenchimento de Missings: Criação de observações com valor zero para todas as combinações parlamentar-tempo-domínio não observadas, assegurando painel balanceado.

Passo 4 - Construção de Variáveis de Controle: Criação de variáveis *dummy* para grupos políticos, países, e funções em comissões, utilizando categorias de referência (França para país, não especificado).

O dataset final (*df_long_v2.csv*) contém:

- **Observações:** Painel tridimensional parlamentar \times mês \times domínio
- **Período:** 2014-2019 (8^a legislatura) e 2019-2024 (9^a legislatura)
- **Variáveis principais:** *questions* (contagem de perguntas), *meetings* (contagem de reuniões)
- **Controles institucionais:** Dummies para grupos políticos, países, funções em comissões
- **Identificadores:** *member_id*, *Y.m* (ano-mês), *domain* (domínio temático)

Durante todo o processo de operacionalização, foram implementados múltiplos controles de qualidade:

Validação de Extração: Comparação de amostras extraídas com dados disponíveis publicamente no site do Parlamento Europeu.

Validação de Matching: Revisão manual de amostras aleatórias de matches de organizações, confirmando precisão superior a 90% para matches com score $\geq 85\%$.

Validação Temática: Avaliação qualitativa de classificações temáticas através de amostras aleatórias, confirmando adequação das categorias desenvolvidas.

Verificações de Consistência: Testes de integridade referencial entre datasets, verificação de períodos de mandato dos parlamentares, e consistência temporal das observações.

B TESTES DE ROBUSTEZ

A credibilidade das inferências causais desta pesquisa repousa não apenas na adequação da estratégia de identificação, mas também na robustez dos resultados a diferentes especificações metodológicas. Por essa razão, esta tese implementa uma bateria abrangente de testes de robustez. Esta seção detalha os principais testes realizados para validar a H1, cujos resultados estão sumarizados na Tabela 5.

A Tabela 5 apresenta os resultados de uma bateria de 15 testes de robustez que avaliam a estabilidade do efeito positivo do lobby sobre a atividade parlamentar, conforme estimado no modelo PPML de base. A análise sistemática destes testes, detalhada a seguir, demonstra a consistência do achado principal em diferentes estimadores, especificações, sub-amostras e estruturas de erro, ao mesmo tempo que revela nuances importantes sobre a dinâmica da relação estudada.

Tabela 6 – Resultados dos testes de robustez

	Baseline PPML	MQO	PPML (sem domino x tempo)	PPML (sem país x tempo)	PPML (sem partido x tempo)	PPML (somente FE individuais)	Sem outliers	Período recente (2019-2024)	Período inicial (2014-2019)	Tratamento binário	Cluster: membro	Cluster: bidirecio- nal	Robust SE	Placebo: lead	Placebo: aleatório
meetings	0,025*** (0,002)	0,007*** (0,001)	0,025*** (0,002)	0,025*** (0,002)	0,024*** (0,002)	0,014*** (0,003)	0,052*** (0,003)	0,025*** (0,002)	0,112*** (0,012)		0,025*** (0,005)	0,025*** (0,006)	0,025*** (0,005)		
meetings_binary										0,327*** (0,013)					
meetings_lead														0,024*** (0,002)	
meetings_random															0,000 (0,003)
Num.Obs.	600 237	979 209	600 237	615 609	600 237	625 401	599 594	600 237	40 320	600 237	600 237	600 237	600 237	592 110	600 237
R2	0,253	0,194	0,199	0,239	0,243	0,257	0,254	0,253	0,211	0,254	0,253	0,253	0,253	0,254	0,252
RMSE	0,56	0,45	0,58	0,56	0,56	0,56	0,56	0,56	0,52	0,56	0,56	0,56	0,56	0,55	0,56
Std.Errors	by: cl_dt	by: cl_dt	by: cl_dt	by: cl_dt	by: cl_dt	by: member_id	by: cl_dt	by: cl_dt	by: cl_dt	by: cl_dt	by: cl_dt	by: mem- ber_id	by: cl_dt & mem- ber_id	by: fe_ct	by: cl_dt
FE: fe_dt	X	X		X	X		X	X	X	X	X	X	X	X	X
FE: fe_ct	X	X	X		X		X	X	X	X	X	X	X	X	X
FE: fe_pt	X	X	X	X			X	X	X	X	X	X	X	X	X
FE: fe_i						X									

+ p <0,1, * p <0,05, ** p <0,01, *** p <0,001

Tabela 5 – Resultados dos testes de robustez

A primeira série de testes valida a escolha do modelo. A estimação por **Mínimos Quadrados Ordinários (MQO)** (coluna 2) produz um coeficiente positivo e significativo, confirmando que o resultado não é um artefato do estimador Poisson. As colunas 3 a 6 examinam a sensibilidade à estrutura de efeitos fixos. A remoção sequencial dos efeitos fixos de **domínio×tempo**, **país×tempo** e **partido×tempo** (colunas 3-5) resulta em coeficientes notavelmente estáveis e estatisticamente indistinguíveis do modelo base. Isso indica que a identificação não é impulsionada por um único conjunto de efeitos fixos, mas sim pelo controle robusto de múltiplas fontes de heterogeneidade. O modelo com **apenas efeitos fixos individuais** (coluna 6), que não controla por choques variantes no tempo, resulta em um coeficiente menor (0.014), sublinhando a importância dos efeitos fixos de alta dimensão para mitigar o viés de variáveis omitidas.

O efeito se mantém robusto em diferentes recortes da amostra. A exclusão de **outliers** (coluna 7) — deputados com atividade de lobby extrema — mais do que duplica o coeficiente (0.052), sugerindo que o resultado base é conservador. A análise por período revela um efeito significativamente maior no **período legislativo anterior (2014-2019)** (coluna 9) em comparação com o período recente, indicando uma possível mudança na dinâmica da influência do lobby ao longo do tempo. A especificação com **tratamento binário** (coluna 10) mostra que a margem extensiva é particularmente importante: passar de zero para pelo menos uma reunião está associado a um aumento de aproximadamente 38% ($e^{0.327} - 1$) na atividade parlamentar.

As colunas 11 a 13 demonstram que a inferência não é sensível à estrutura de erros-padrão. A utilização de **clusters no nível do deputado** ou **clusters bidirecionais** (deputado e domínio×tempo) aumenta, como esperado, a incerteza da estimativa, mas o coeficiente permanece altamente significativo, validando a robustez estatística do achado.

Os dois últimos testes avaliam a premissa de exogeneidade. O **placebo com tratamento aleatório** (coluna 15) resulta, como esperado, em um coeficiente nulo e não significativo, fortalecendo a confiança de que o resultado não é espúrio. Contudo, o **placebo com o lead da variável de tratamento** (coluna 14) — ou seja, o lobby do mês seguinte — apresenta um coeficiente positivo e significativo,

quase idêntico ao do modelo base. Este resultado viola a premissa de exogeneidade estrita e sugere a presença de dinâmicas temporais complexas, como causalidade reversa (parlamentares mais ativos sinalizam que serão alvos futuros de lobby) ou antecipação (lobistas e parlamentares se preparam para atividades legislativas futuras). Embora este achado complice uma interpretação estritamente causal do coeficiente como o efeito de uma reunião *exógena*, ele não invalida a forte e robusta associação positiva documentada. Em conjunto, os testes demonstram que a relação entre lobby e atividade parlamentar é consistente e não-aleatória, ainda que a sua dinâmica causal seja complexa e potencialmente bidirecional.

A Figura 17 apresenta os resultados da exclusão sequencial de cada país da amostra. A estabilidade dos coeficientes estimados indica que nenhum país individual influencia desproporcionalmente os resultados.

A bateria de testes de robustez apresentada nesta seção oferece um suporte abrangente à validade da hipótese central (H1). Os resultados demonstram que a associação positiva e estatisticamente significativa entre a intensidade do lobby (medida pelo número de reuniões) e a atividade parlamentar (medida pelo número de perguntas) não é um artefato de decisões metodológicas específicas. O efeito persiste sob diferentes estimadores (PPML e MQO), especificações de efeitos fixos, estruturas de erros-padrão, sub-amostras (excluindo outliers e por período legislativo) e definições da variável de tratamento.

O conjunto de testes revela três achados centrais. Primeiro, a relação é **estatisticamente robusta**: a significância do coeficiente se mantém mesmo sob as especificações mais exigentes, como a clusterização bidirecional dos erros-padrão. Segundo, o efeito estimado no modelo base é **conservador**, pois a exclusão de outliers resulta em um coeficiente de magnitude consideravelmente maior. Terceiro, o efeito não é **espúrio**, como demonstrado pelo teste placebo com tratamento aleatório, que produz um resultado nulo.

Apesar da forte evidência, os testes também apontam para a complexidade da relação causal. O resultado significativo do teste placebo com o *lead* da variável de tratamento sugere a presença de dinâmicas temporais, como causalidade reversa ou comportamento antecipatório por parte de lobistas e parlamentares. Embora

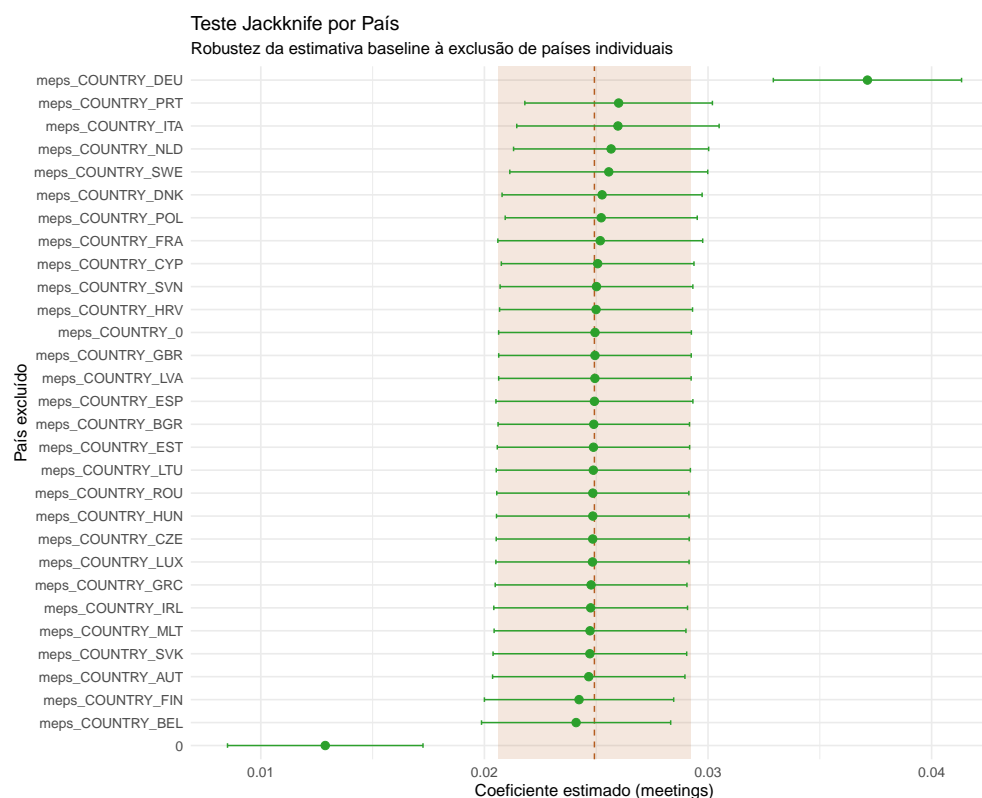


Figura 17 – Teste Jackknife por País

Nota: Cada ponto representa o coeficiente estimado do efeito do lobby (reuniões) sobre a atividade parlamentar (perguntas) ao excluir o país correspondente da amostra. As barras de erro indicam o intervalo de confiança de 95%. A linha vertical tracejada marca a estimativa do modelo base (com todos os países), e a área sombreada representa seu intervalo de confiança. A estabilidade dos coeficientes demonstra que o resultado não é impulsionado por nenhum país em particular.

isso complique uma interpretação estritamente causal do coeficiente como o efeito de uma reunião exógena, não invalida a conclusão principal. A evidência, em sua totalidade, aponta para uma relação sistêmica, robusta e não-aleatória, na qual a atividade de lobby e a atividade parlamentar estão intrinsecamente interligadas.

REFERÊNCIAS

ANDERSON, J. E. A theoretical foundation for the gravity equation. *The American Economic Review*, JSTOR, v. 69, n. 1, p. 106–116, 1979. ISSN 00028282. 6

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. 4

BACHRACH, P.; BARATZ, M. Two faces of power, in "american political science review", 56. *American Political Science Review*, 1962. 4

BAUMGARTNER, F. R.; JONES, B. D. *Agendas and instability in American politics*. [S.l.]: University of Chicago Press, 2010. 20, 41

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. 3, 22

BEVAN, S.; BORGHETTO, E.; SEEBERG, H. Do different parties respond to different problems? a comparative study of parliamentary questions across multiple countries. *Journal of European Public Policy*, Taylor & Francis, v. 31, n. 7, p. 1856–1878, 2023. 5

BUNEA, A. Legitimacy through targeted transparency? regulatory effectiveness and sustainability of lobbying regulation in the european union. *European Journal of Political Research*, Wiley Online Library, v. 57, n. 2, p. 378–403, 2018. 39, 44

BURDEN, B. C. *Personal roots of representation*. [S.l.]: Princeton University Press, 2015. 30

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. 3, 20

COEN, D. The evolution of the large firm as a political actor in the european union. *Journal of European Public Policy*, Taylor & Francis, v. 4, n. 1, p. 91–108, 1997. 16

COEN, D.; KATSAITIS, A. Legislative efficiency and political inclusiveness: The effect of procedures on interest group mobilization in the european parliament. *Journal of Legislative Studies*, v. 25, p. 278–294, 4 2019. ISSN 1357-2334. 16, 17

COEN, D.; KATSAITIS, A. Legislative efficiency and political inclusiveness: The effect of procedures on interest group mobilization in the european parliament. *The Journal of Legislative Studies*, Taylor & Francis, v. 25, n. 2, p. 278–294, 2019. 22

DANIEL, W. T. *Career behaviour and the European parliament: All roads lead through Brussels?* [S.l.]: Oxford University Press, 2015. 40

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>>. 4, 6, 16

DÜR, A.; MATEO, G. Who lobbies the european union? national interest groups in a multilevel polity. *Journal of European Public Policy*, ROUTLEDGE JOURNALS, TAYLOR & FRANCIS LTD, 2-4 PARK SQUARE, MILTON PARK, ABINGDON OX14 4RN, OXON, ENGLAND, v. 19, n. 7, p. 969–987, 2012. ISSN 1350-1763. 16, 17

EISING, R. Institutional context, organizational resources and strategic choices: Explaining interest group access in the european union. *European Union Politics*, Sage Publications Sage UK: London, England, v. 8, n. 3, p. 329–362, 2007. 16

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>>. 3, 4, 6, 16, 17, 39

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. 4

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. 3

HALL, R. L.; WAYMAN, F. W. Buying time: Moneyed interests and the mobilization of bias in congressional committees. *American political science review*, Cambridge University Press, v. 84, n. 3, p. 797–820, 1990. 16

- 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. 3
- 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. 3
- HUWYLER, O.; MARTIN, S. Interest group tactics and legislative behaviour: how the mode of communication matters. *Journal of European Public Policy*, v. 29, n. 8, p. 1268–1287, AUG 3 2022. ISSN 1350-1763. 6
- HUWYLER, O.; TURNER-ZWINKELS, T.; BAILER, S. No representation without compensation: The effect of interest groups on legislators' policy area focus. v. 76, n. 3, p. 1388–1402, 2023. ISSN 1065-9129, 1938-274X. Disponível em: <<http://journals.sagepub.com/doi/10.1177/10659129221137035>>. 40
- IBENSKAS, R.; BUNEA, A. Legislators, organizations and ties: Understanding interest group recognition in the european parliament. *European Journal of Political Research*, v. 60, p. 560–582, 8 2021. ISSN 0304-4130. 40
- JENSEN, C. B.; PROKSCH, S.-O.; SLAPIN, J. B. Parliamentary questions, oversight, and national opposition status in the european parliament. *Legislative Studies Quarterly*, Wiley Online Library, v. 38, n. 2, p. 259–282, 2013. 5
- 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. 3
- KLÜVER, H. Informational lobbying in the european union: The effect of organisational characteristics. v. 35, n. 3, p. 491–510, 2012. ISSN 0140-2382, 1743-9655. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/01402382.2012.665737>>. 21, 22, 29, 40
- KLÜVER, H.; BRAUN, C.; BEYERS, J. Legislative lobbying in context: towards a conceptual framework of interest group lobbying in the european union. *Journal of European Public Policy*, Taylor & Francis, v. 22, n. 4, p. 447–461, 2015. 6, 18
- KOLLMAN, K. *Outside lobbying: Public opinion and interest group strategies*. [S.l.]: Princeton University Press, 1998. 3, 4, 43

KOOP, C. Explaining accountability of independent agencies: The importance of political salience. *Journal of Public Policy*, Cambridge University Press, v. 31, n. 2, p. 209–234, 2011. 5

MAHONEY, C. Lobbying success in the united states and the european union. v. 27, n. 1, p. 35–56, 2007. ISSN 0143-814X, 1469-7815. Disponível em: <https://www.cambridge.org/core/product/identifier/S0143814X07000608/type/journal_article>. 17, 28, 39, 43

MAHONEY, C. *Brussels versus the beltway: Advocacy in the United States and the European Union*. [S.l.]: Georgetown University Press, 2008. 21

MARICUT-AKBIK, A. Q&a in legislative oversight: A framework for analysis. *European Journal of Political Research*, Wiley Online Library, v. 60, n. 3, p. 539–559, 2020. 5

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. 3, 6

MARTIN, S. *Parliamentary questions (Comparative Assessment of Parliaments Note)*. [S.l.], 2013. 5

MAYHEW, D. R. *Congress: The electoral connection*. [S.l.]: Yale university press, 2004. 40, 44

MCCUBBIN, M. D.; SCHWARTZ, T. Congressional oversight overlooked: Police patrols and fire alarms. *American Journal of Political Science*, JSTOR, v. 28, n. 1, p. 165–179, 1984. 5

NAVARRO, J.; KELBEL, C. Banking on meps? the strategic and partisan motivations of the parliamentary scrutiny of the european central bank. *International Journal of Parliamentary Studies*, v. 2, n. 2, p. 176–201, 2022. 5

OLSON, M. *The Logic of Collective Action: Public Goods and the Theory of Groups, with a new preface and appendix*. [S.l.]: Harvard University Press, 1971. v. 124. 17

OTJES, S.; LOUWERSE, T. Parliamentary questions as strategic party tools. *West European Politics*, Taylor & Francis, v. 41, n. 2, p. 496–516, 2017. 5

POP, A.-L. Lobbying the european union: Institutions, actors and issues. *Romanian Journal of European Affairs*, v. 13, p. 83, 2013. 39

PROKSCH, S.-O.; SLAPIN, J. B. Parliamentary questions and oversight in the european union. *European Journal of Political Research*, Wiley Online Library, v. 50, n. 1, p. 53–79, 2010. 5

RICHARDSON, J. Government, interest groups and policy change. *Political studies*, Wiley Online Library, v. 48, n. 5, p. 1006–1025, 2000. 18

SAALFELD, T. Members of parliament and governments in western europe: Agency relations and problems of oversight. *European Journal of Political Research*, Wiley Online Library, v. 37, n. 3, p. 353–376, 2000. 5

SCHILLER, W. J. Senators as political entrepreneurs: Using bill sponsorship to shape legislative agendas. *American journal of political science*, JSTOR, p. 186–203, 1995. 30

SCHMIDT, V. A. Procedural democracy in the eu: the europeanization of national and sectoral policy-making processes. *Journal of European Public Policy*, Taylor & Francis, v. 13, n. 5, p. 670–691, 2006. 16

SIMON, H. A. Notes on the observation and measurement of political power. v. 15, n. 4, p. 500–516, 1953. ISSN 0022-3816, 1468-2508. Disponível em: <<https://www.journals.uchicago.edu/doi/10.2307/2126538>>. 17

STRØM, K. Delegation and accountability in parliamentary democracies. *European Journal of Political Research*, Wiley Online Library, v. 37, n. 3, p. 261–289, 2000. 5