

SUMÁRIO

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 (??). 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 (?????) ou neutros (???????) que tenham posições importantes na tramitação de determinado projeto (??).

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 (??).

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 (??), seja na sua implementação.

A quinta dificuldade, ou "**problema das variáveis omitidas**" (??), 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 (??), desde lobby direto até campanhas de pressão perante a opinião pública (*outside lobbying*) (??), podendo inclusive influenciar a própria seleção dos tomadores de decisão (??).

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 ??).

(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 ?? e ??, 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 (??????). 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 (????????).

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 (????????). 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 (??), 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 (??) e, no contexto da UE, as táticas que envolvem engajamento face a face são consideradas particularmente eficazes (??). 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 (????). 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, popularizado por Silva e Tenreyro (??) para modelos de gravidade no comércio internacional, cuja fundamentação teórica remonta a Anderson (??), 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 altamente eficiente para estimar modelos Poisson 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 enviesaria 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 (e.g., empresariais, ONGs), 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 ???. O teste empírico baseia-se na estimação do modelo de PPML apresentado na Equação ??, que utiliza o número de perguntas parlamentares como variável dependente e o número de reuniões com lobistas como a principal variável de tratamento.

A escolha pelo PPML, estimado com a função ‘fepois’ do pacote **fixest** em R, justifica-se pela natureza da variável dependente (dados de contagem com excesso de zeros), evitando os vieses de modelos lineares com transformações logarítmicas. 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 ??, 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 adici-

onais 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.

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 os 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 ??). 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 (pacote `fixest`, função `fepois`), incluindo efeitos fixos de alta dimensão de país-tempo, partido-tempo e domínio-tempo, e o mesmo vetor de controlos 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

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

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.

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 ??.

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) controlos setoriais de atuação do lobista e país da sede. Para interpretação limpa, recentramos o orçamento em seu valor médio. Em termos funcionais:

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

onde ℓ indexa lobistas, B_ℓ é o orçamento, W_ℓ agrega os controlos e a interação permite que o efeito do orçamento sobre o número de reuniões varie por categoria. Este modelo gera previsões de reuniões esperadas por categoria e nível de orçamento, que embasam as figuras e tabelas correspondentes.

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

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

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.

Toda a estimação foi conduzida em R. A Etapa 1 utiliza `fixest::fepois` com efeitos fixos de país-tempo, partido-tempo e domínio-tempo, controles individuais detalhados e `cluster` de erros em domínio-tempo. A Etapa 2 utiliza `MASS::glm.nb` com controles setoriais e geográficos, além da interação categoria \times orçamento. As figuras e tabelas foram exportadas para `Tese/figures/h2_test` e `Tese/tables/h2_test`, respeitando o padrão do projeto. A discussão substantiva dos resultados e sua interpretação estão em `Tese/main/cap4-resultados/h2.tex`.

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 ??) 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 um (`salience_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 ??). 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.

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 e partido político, estabilidade temporal dos efeitos, diferentes operacionalizações da variável de tratamento, transformações alternativas da variável dependente, especificações binárias e categóricas, testes placebo com tratamentos falsos, análise de *leads* e *lags*, verificação de pressupostos do modelo.

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

Seguindo as melhores práticas de transparência acadêmica, todos os códigos de análise, procedimentos de operacionalização dos dados, e especificações econômicas estão totalmente documentados e disponíveis. 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 das escolhas realizadas.

Esta abordagem de transparência e replicabilidade não apenas fortalece a credibilidade científica da pesquisa, mas também facilita futuras extensões e aplicações da metodologia desenvolvida a outros contextos institucionais e temporais.

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 Parlamento Europeu. 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 detalhada 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

Avaliação da distribuição dos lobistas registrados junto às instituições europeias.

Tabela 1 – Distribuição de organizações por categoria

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

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

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

A distribuição geográfica evidencia forte concentração em Estados-Membros centrais: Bélgica (18,2%), Alemanha (14,0%) e França (9,3%). Países como Países Baixos, Espanha, Reino Unido e Itália aparecem na sequência (6% cada). Nota-se presença extracomunitária não desprezível. Cerca de 19% são de países não membros da UE (com destaque para os Estados Unidos, com 4,5% dos lobistas registrados), sinalizando a atratividade regulatória do mercado europeu.

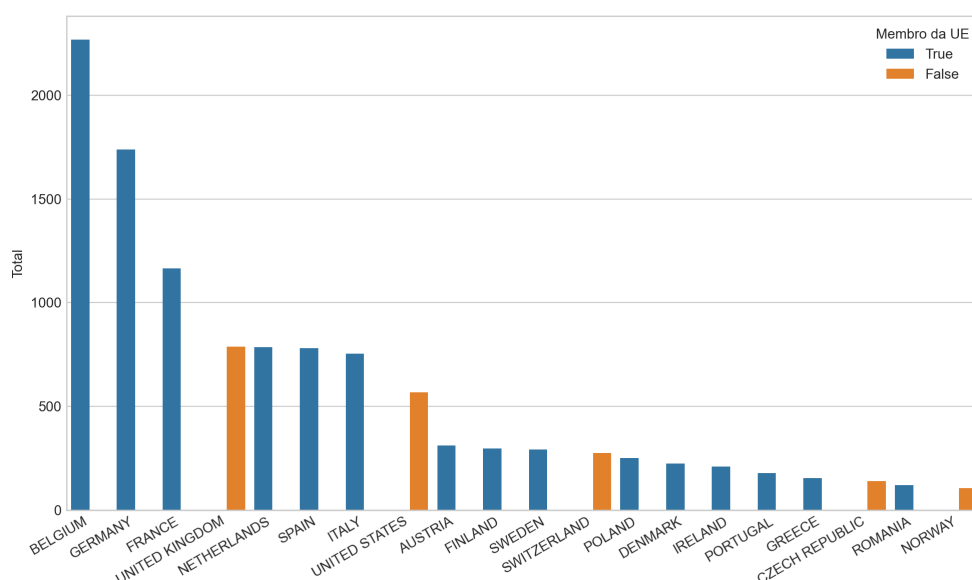


Figura 1 – Top 20 países-sede (n = 12.468 organizações, totalizando 91% do total de organizações)

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

Temporalmente, observa-se aceleração do registro de entidades após meados da década de 2010, com 2023 concentrando 17,9% do total. Picos intermediários

(2015–2016; 2020–2022) são compatíveis com ciclos legislativos, janelas regulatórias e alterações incrementais nos mecanismos de transparência.

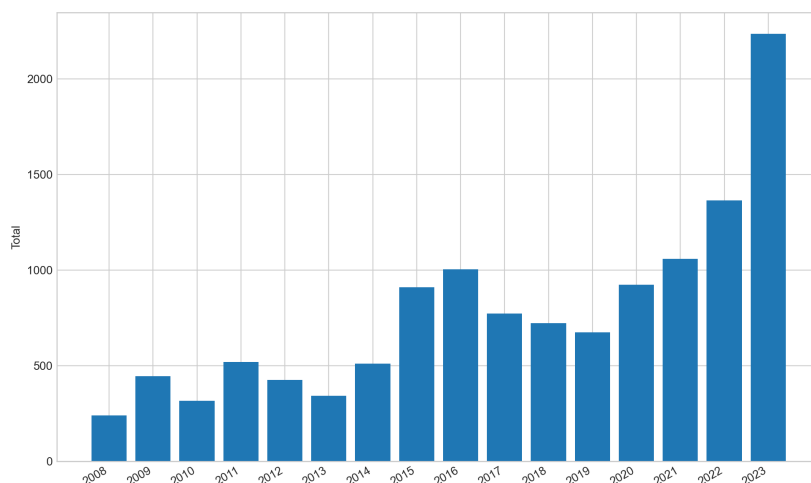


Figura 2 – Ano de registo

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 ??) destacam *Infraestrutura e Indústria* (68,3%), *Tecnologia* (67,9%) e *Economia e Comércio* (67,4%), seguidas por *Ambiente e Clima* (64,7%) e *Assuntos Externos e Segurança* (51,3%). Temas como *Saúde* (43,7%) e *Educação* (41,2%) são intermediários; *Agricultura* (35,3%) e *Direitos Humanos* (25,3%) têm menor incidência relativa.

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

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

As estatísticas de orçamento máximo declarado em escala logarítmica-natural (Figura ??) indicam mediana em torno de 11,5 e quartis aproximadamente entre

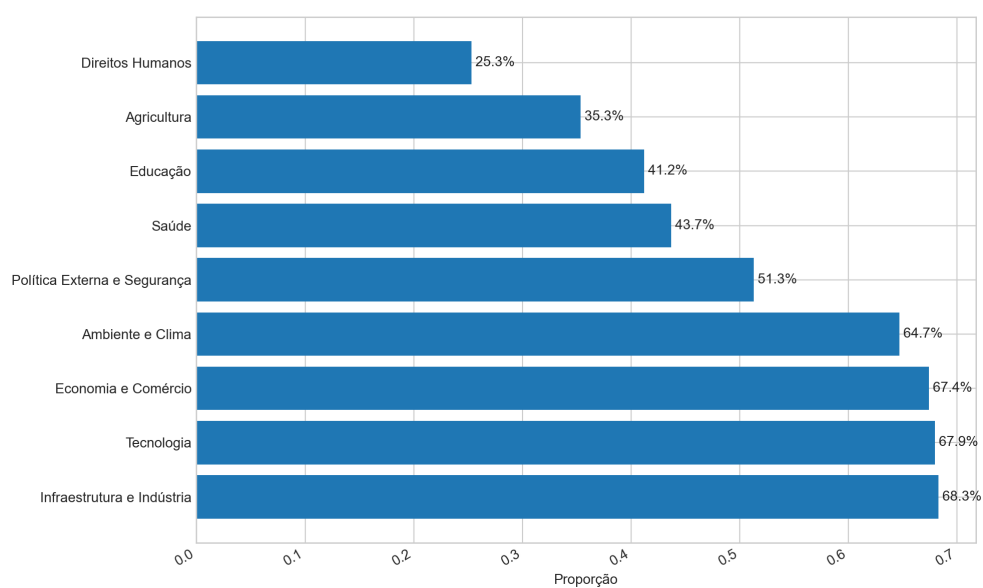


Figura 3 – Cobertura temática (proporção de entidades)

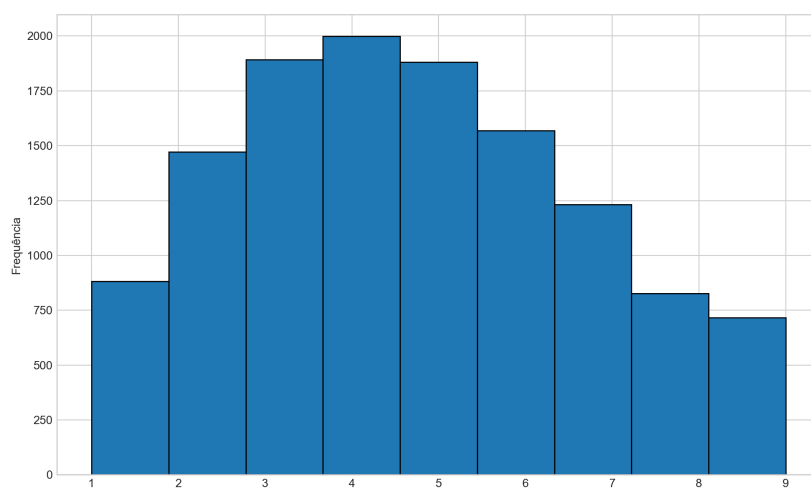


Figura 4 – Número de temas por lobista

10,1 e 12,9. A presença de valores inválidos/extremos na base administrativa (por exemplo, ocorrências infinitas) distorce a média e o desvio-padrão, recomendando foco em medidas robustas (mediana, intervalos interquartílicos) e rotinas de limpeza nos exercícios inferenciais.

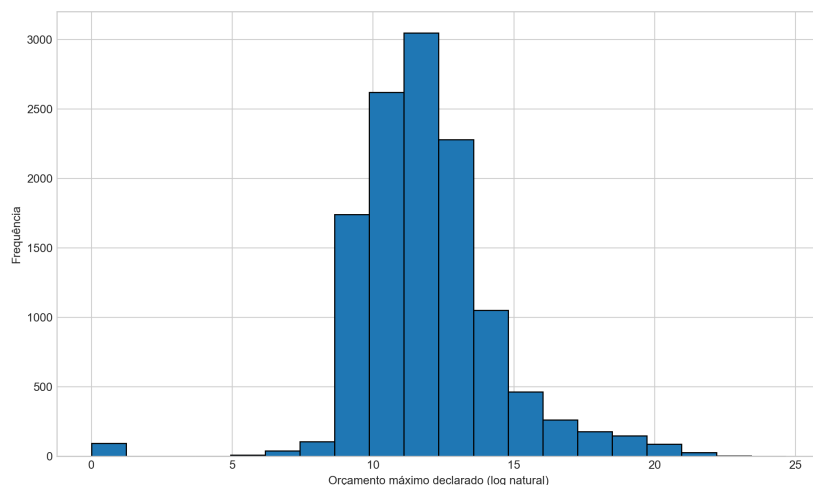


Figura 5 – Distribuição de orçamento máximo declarado ($\ln(budget)$)

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

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

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

União Europeia, sugerindo um campo de ação onde interesses difusos e concentrados buscam simultaneamente agenda e influência.

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

Temporalmente, as frequências por ano indicam aceleração recente dos registos, com 2023 concentrando 17,9% das entradas no período observado. Picos intermédios (2015–2016; 2020–2022) são compatíveis com ciclos legislativos, janelas regulatórias e mudanças incrementais no regime de transparência, fatores que tendem a alterar a propensão ao registro. A expansão no pós-2020 pode refletir a reconfiguração de estratégias após as restrições pandémicas, além da ênfase em agendas de transição digital e verde.

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

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

diferentes domínios.

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

Em síntese, as evidências descritivas apontam para um ecossistema plural, geograficamente ancorado em polos institucionais centrais, com dinamismo temporal recente e agendas orientadas por digitalização, competitividade industrial e 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 Parlamento Europeu. A abordagem adotada segue uma estratégia analítica multinível, iniciando com padrões agregados gerais e progredindo para análises desagregadas mais específicas. Esta progressão metodológica permite compreender tanto as tendências globais quanto os mecanismos específicos que operam no nível individual e temporal.

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

identificação causal robustas.

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

A cobertura temporal de julho de 2019 a novembro de 2024 é particularmente relevante por abranger períodos de intensa atividade legislativa europeia, incluindo a transição entre legislaturas e eventos político-econômicos significativos. Destaca-se, nesse intervalo, o impacto da pandemia de COVID-19, que afetou profundamente tanto a dinâmica da atividade parlamentar quanto as estratégias de lobbying. A pandemia resultou em mudanças substanciais nos modos de trabalho do Parlamento Europeu, com a adoção de sessões remotas e restrições a reuniões 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 ?? 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 temporal revela quatro padrões empiricamente relevantes. Primeiro, observa-se uma **tendência crescente** em ambas as variáveis ao longo do período, sugerindo intensificação tanto da atividade parlamentar quanto do lobbying. Segundo, existe clara **sazonalidade** relacionada ao calendário parlamentar, com reduções sistemáticas durante períodos de recesso. Terceiro, identificam-se **picos de atividade** que coincidem com discussões de legislação relevante em domínios específicos, indicando resposta coordenada do sistema político. Quarto, a **correlação contemporânea** entre perguntas e reuniões permanece relativamente estável ao longo do tempo, sugerindo estabilidade estrutural na relação entre as variáveis.

Estes padrões temporais têm implicações metodológicas importantes. A presença de tendências temporais justifica a inclusão de efeitos fixos de tempo nas

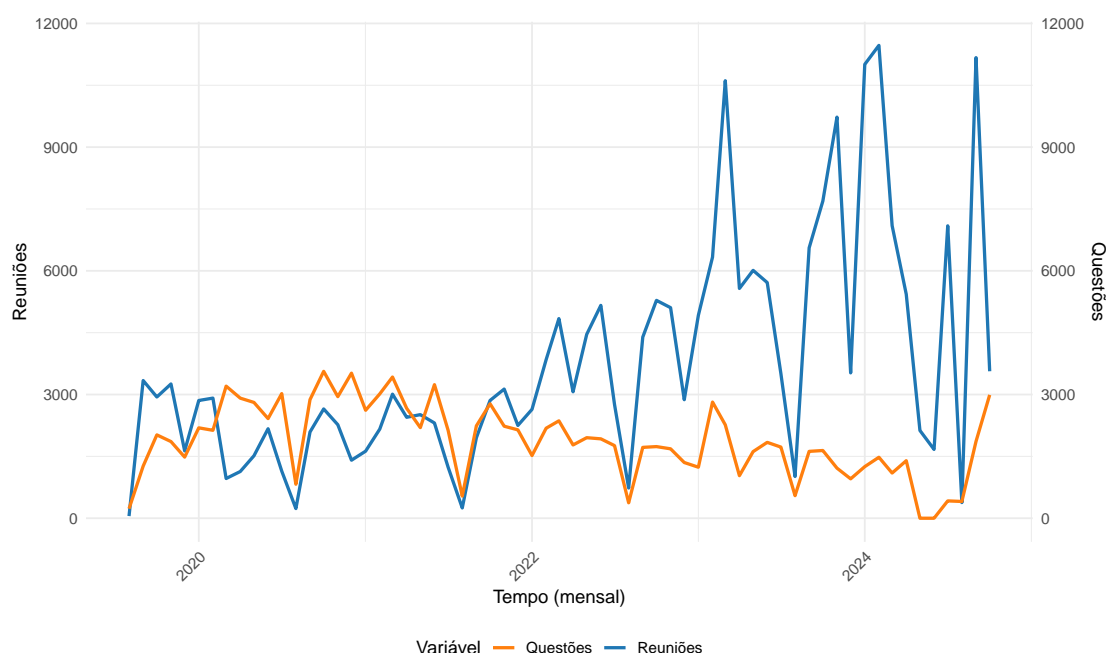


Figura 6 – Evolução temporal da atividade parlamentar e de lobbying

especificações econométricas para controlar choques temporais comuns. A sazonalidade observada valida a escolha da frequência mensal como unidade temporal, capturando variações de curto prazo sem introduzir ruído excessivo. A estabilidade da correlação fornece evidência preliminar contra quebras estruturais que poderiam comprometer a validade das estimativas.

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

As ??, ?? e ?? apresentam uma análise dos padrões de participação agregados por deputado, revelando aspectos da distribuição da atividade de lobbying no Parlamento Europeu que impactam a identificação causal.

A análise revela três características fundamentais da distribuição de tratamento. Primeiro, existe **participação substancial mas não universal**: 46,3% dos deputados (637 de 1.353) receberam pelo menos uma reunião de lobbying durante o

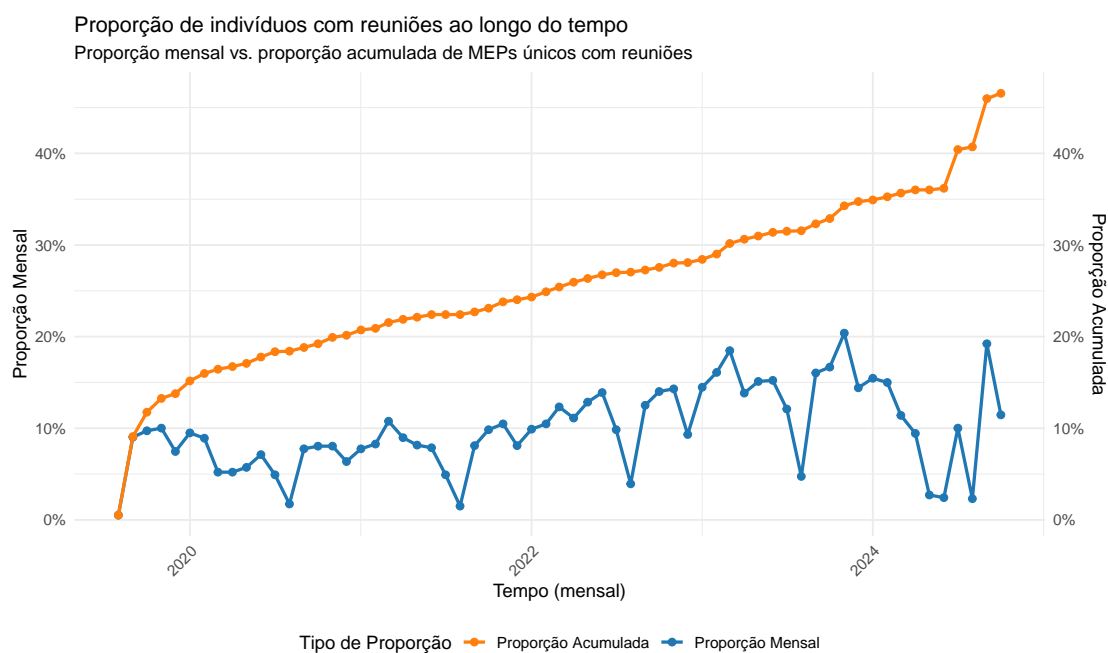


Figura 7 – Evolução temporal da proporção de MPEs que participaram de reuniões de lobbying

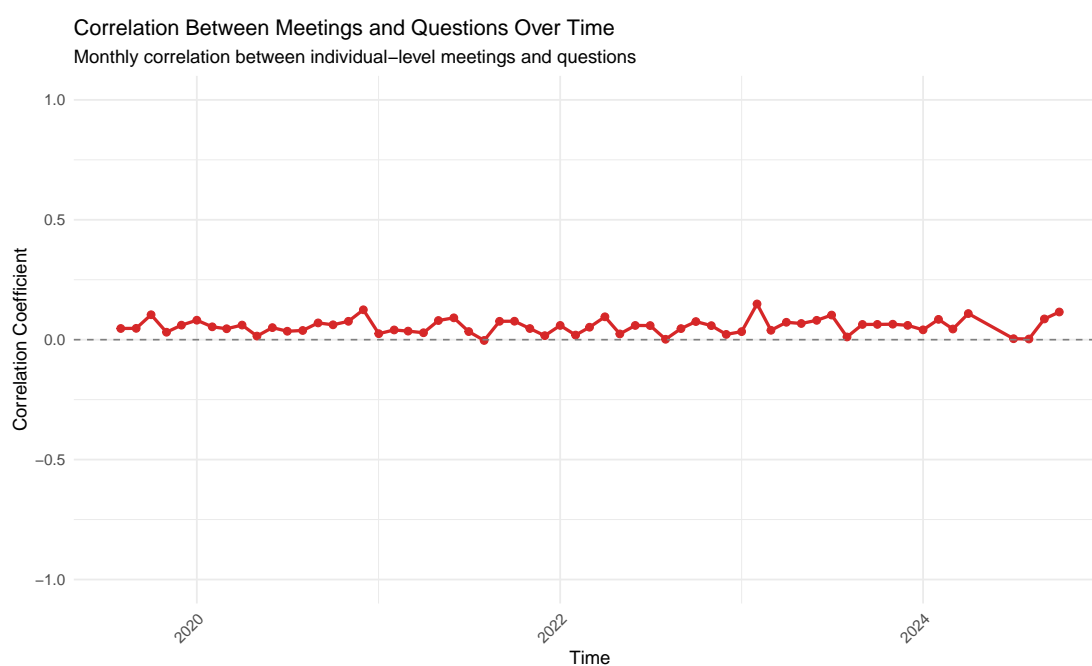


Figura 8 – Evolução temporal da correlação entre reuniões e perguntas

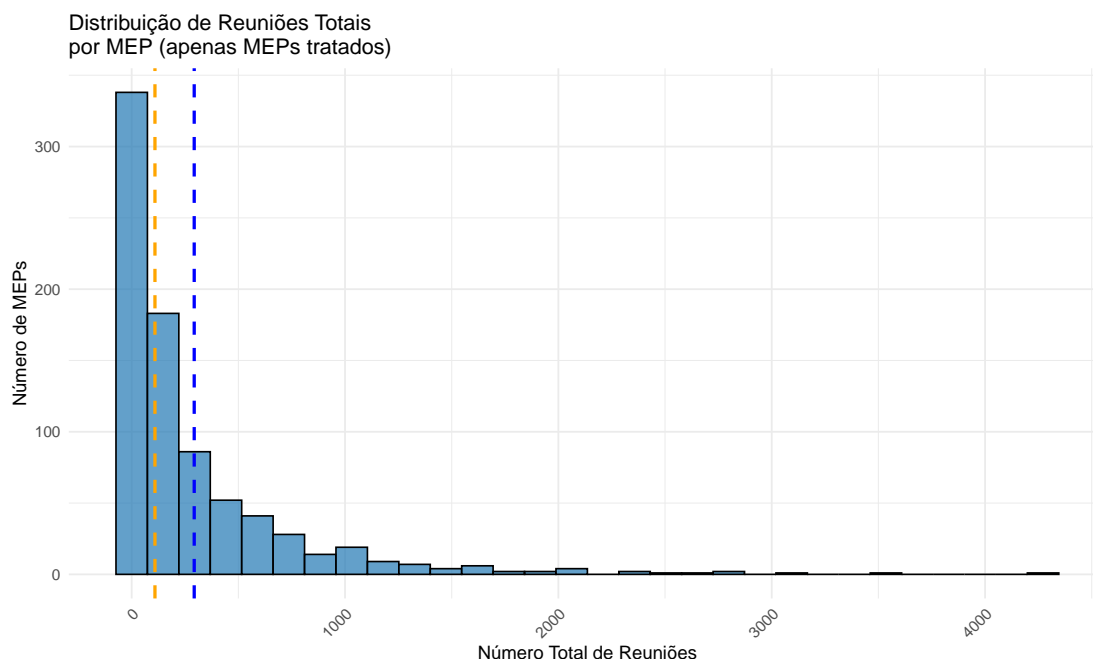


Figura 9 – Distribuição de reuniões por MPE

período estudado. Esta proporção indica que o lobbying é um fenômeno disseminado mas não ubíquo no sistema parlamentar europeu.

Segundo, observa-se **concentração extrema** na intensidade de tratamento. Entre os deputados que receberam lobbying, a distribuição é altamente assimétrica: enquanto a mediana é de 105 reuniões por deputado, a média é de 288,2 reuniões, indicando que uma minoria de parlamentares concentra uma proporção desproporcional da atividade lobista. O caso extremo de um deputado com 4.274 reuniões ilustra esta concentração.

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

Estes padrões agregados têm implicações importantes para a identificação causal. A concentração do tratamento em uma minoria de deputados sugere que

estratégias de identificação baseadas em variação cross-sectional podem sofrer de poder estatístico limitado. Simultaneamente, a variação substancial na intensidade de tratamento entre deputados tratados fornece fonte valiosa de identificação para estimativas de dose-resposta.

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

Tabela 2 – Estatísticas agregadas de tratamento por deputado

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

2.2.1 HETEROGENEIDADE ENTRE DOMÍNIOS DE POLÍTICA PÚBLICA

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

A ?? apresenta uma análise sistemática da variação inter-domínios em múltiplas dimensões: penetração, volume e intensidade do tratamento, bem como padrões

temporais de iniciação.

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

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

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

Contudo, emergem padrões teoricamente consistentes. Domínios relacionados à **regulação econômica** (economia e comércio, tecnologia, infraestrutura) apresentam sistematicamente maiores taxas de penetração, refletindo os elevados interesses econômicos e a complexidade regulatória que incentivam investimento em atividades de lobbying. Este padrão é especialmente relevante considerando que a promoção da integração econômica é o foco principal de um bloco como a União Europeia, o que naturalmente direciona maior atenção e mobilização de grupos de interesse para essas áreas. Inversamente, domínios com características de bem público (agricultura, direitos humanos) apresentam menores taxas, consistente com problemas de ação coletiva e menor capacidade organizacional de grupos difusos.

A heterogeneidade observada entre domínios tem duas implicações metodológicas importantes. Primeiro, a variação sistemática sugere que estimativas de efeito médio podem mascarar diferenças substantivas entre setores, justificando análises de heterogeneidade de efeitos por domínio. Segundo, a ordenação consistente dos

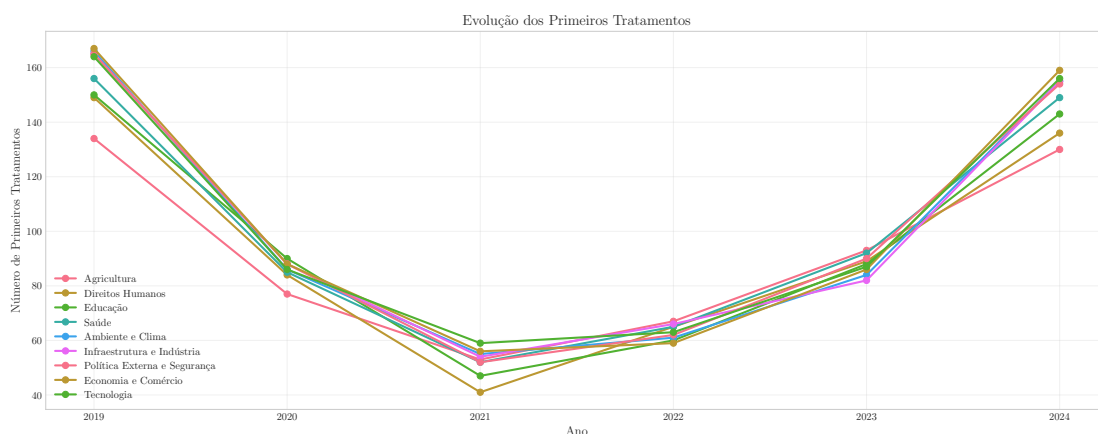


Figura 10 – Análise detalhada de tratamento por domínio de política pública

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

domínios por múltiplas métricas (penetração, volume, intensidade) sugere que esta heterogeneidade reflete características estruturais dos setores ao invés de variação aleatória, aumentando a credibilidade de interpretações causais diferenciadas.

A análise da inflação de zeros requer cuidado metodológico particular, pois a unidade de observação MEP-domínio-mês pode gerar inflação **artificial** de zeros. Como documentado na literatura sobre comportamento parlamentar (??), deputados tendem a especializar-se tematicamente, concentrando atividade em subconjuntos específicos de domínios. Consequentemente, grupos de interesse, cientes desta especialização, direcionam esforços de lobbying apenas para deputados ativos em suas áreas de interesse.

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

2.2.1.0.1 ANÁLISE CORRIGIDA DA INFLAÇÃO DE ZEROS

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

Tabela 4 – Inflação de zeros por nível de agregação

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

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

1. **Nível MEP-domínio-mês:** A inflação aparentemente extrema ($>92\%$) reflete em grande parte combinações MEP-domínio onde não se espera atividade sistemática devido à especialização.
2. **Nível MEP-mês** (agregando apenas domínios onde o MEP demonstra atividade parlamentar): A inflação reduz substancialmente para 70,9% (perguntas) e 85,2% (reuniões), revelando que a atividade é *episódica* mas não *ausente*.
3. **Nível domínio-mês** (agregando todos os MEPs ativos): A inflação torna-se negligível (3,2% para perguntas, 0% para reuniões), indicando que há atividade consistente em todos os domínios quando consideramos o conjunto de deputados relevantes.

Esta análise corrigida tem implicações fundamentais para a estratégia econométrica:

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

2.2.2 SÍNTESE E IMPLICAÇÕES METODOLÓGICAS

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

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

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

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

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

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

Estas características empíricas têm implicações diretas para a escolha da estratégia econométrica:

1. **Especificação funcional**: A inflação moderada de zeros (70,9%-85,2% após correção) e natureza de contagem das variáveis justificam o uso de estimadores Poisson Pseudo-Maximum Likelihood (PPML), mas não requerem modelos zero-inflated especializados.
2. **Estrutura de efeitos fixos**: A evidência de especialização temática justifica fortemente efeitos fixos MEP×domínio para controlar heterogeneidade não observada na propensão à atividade em áreas específicas, complementados por efeitos fixos temporais.

3. **Correção de viés de seleção:** A especialização temática implica que observações MEP-domínio-mês com probabilidade zero de atividade podem distorcer estimativas. Controles ou exclusões baseadas em atividade histórica podem ser apropriados.
4. **Estrutura de erros:** A concentração temporal da atividade e especialização justificam erros-padrão agrupados no nível MEP \times domínio para capturar correlação serial específica por área de atuação.
5. **Análise de heterogeneidade:** A variação sistemática entre domínios, combinada com especialização, justifica análises diferenciadas por setor e investigação de efeitos nas margens extensiva (entrada em novos domínios) versus intensiva.
6. **Interpretação causal:** Efeitos estimados devem ser interpretados condicionalmente à especialização temática existente, distinguindo entre expansão de atividade em domínios familiares versus diversificação para novas áreas.

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

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

2.3 ANÁLISE DE EFEITOS DO LOBBY

Optamos por estimar modelos de contagem via PPML com *link* logaritmo por três razões principais. Primeiro, as variáveis de interesse (perguntas e reuniões) são contagens, com forte assimetria e alta incidência de zeros no nível MEP-domínio-mês. O PPML lida naturalmente com zeros sem exigir transformações logarítmicas *ad hoc* que descartam observações. Segundo, o PPML é

consistente sob especificação correta da média condicional mesmo na presença de sobredispersão e heterocedasticidade não especificada, fornecendo erros-padrão robustos quando combinados com *clustering*. Terceiro, a implementação com efeitos fixos de alta dimensão é estável e amplamente utilizada na literatura aplicada (estimador `fepois` do pacote `fixest`).

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

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

Essa modelagem garante três propriedades fundamentais para a identificação dos efeitos: (i) permite comparar a evolução da atividade do mesmo MPE ao longo do tempo, controlando por características não observadas e invariantes como preferências individuais, capital político e produtividade; (ii) elimina a influência de choques ou tendências comuns a todos os MPEs de um mesmo país ou partido em cada mês, por meio dos efeitos fixos específicos de país×tempo e partido×tempo; e (iii) assegura robustez frente a choques específicos de cada setor ou área temática ao incorporar efeitos fixos de domínio×tempo (`domain_time`), isolando variações idiossincráticas desses contextos.

Para testar a Hipótese 1, utilizamos o painel agregado MEP–domínio–mês em amostra combinada (*pooled*) e estimamos PPML com a estrutura de efeitos fixos descrita acima. O coeficiente associado às reuniões (*meetings*) é **positivo**, indicando que aumentos na intensidade de lobbying estão associados a maior atividade

parlamentar em termos de perguntas. Esse resultado é consistente em especificações alternativas, incluindo a versão com termo quadrático para capturar possíveis não linearidades e a inclusão de efeitos fixos $\text{domínio} \times \text{tempo}$, sugerindo robustez do sinal e da magnitude qualitativa do efeito.

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

Tabela 5 – Resumo dos modelos PPML para a Hipótese 1

	PPML	PPML (Quad.)
Reuniões	0,025*** (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 *meetings*. A coluna “PPML (Quadrático)” adiciona *meetings*² para capturar retornos marginais decrescentes. Efeitos fixos: membro; país×tempo; partido×tempo. Erros-padrão agrupados em domínio×tempo e membro.

?? 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.

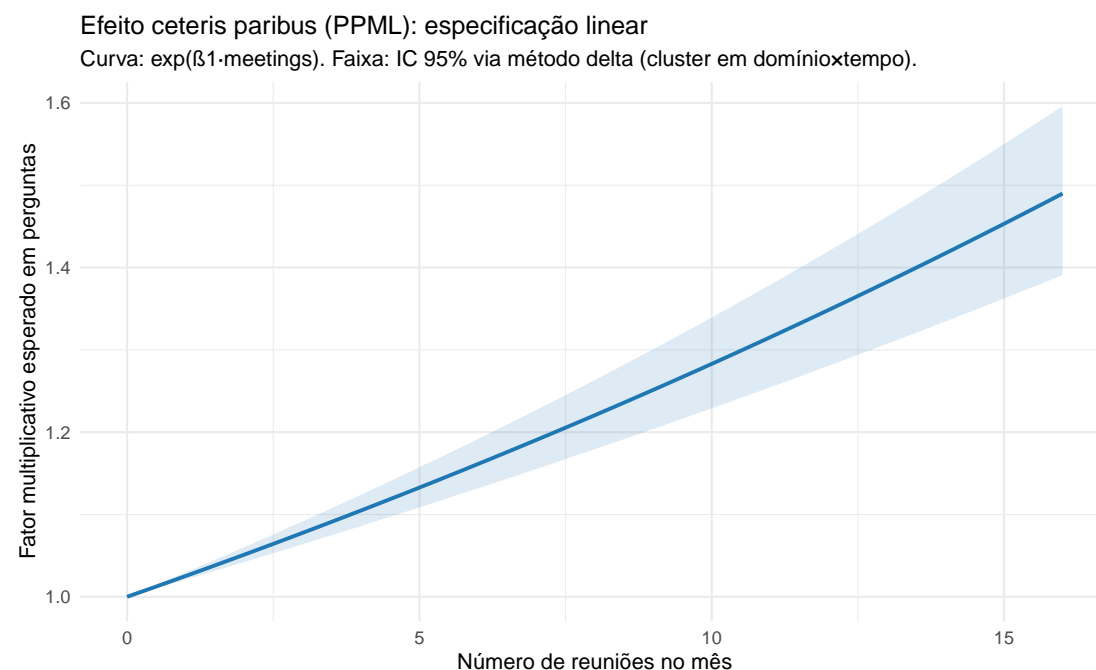


Figura 11 – Efeito esperado ceteris paribus: especificação linear (PPML)

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

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 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 ?? e ?? 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

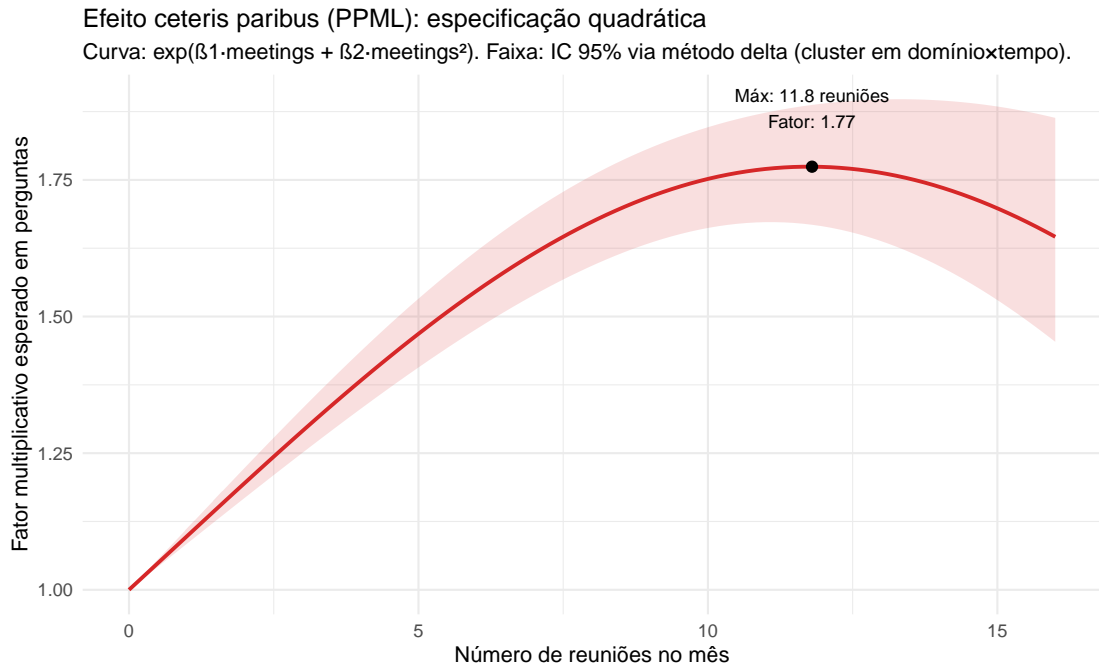


Figura 12 – Efeito esperado ceteris paribus: especificação quadrática (PPML)

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.

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 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 ??, 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.

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.

2.3.0.1 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 ??, 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.

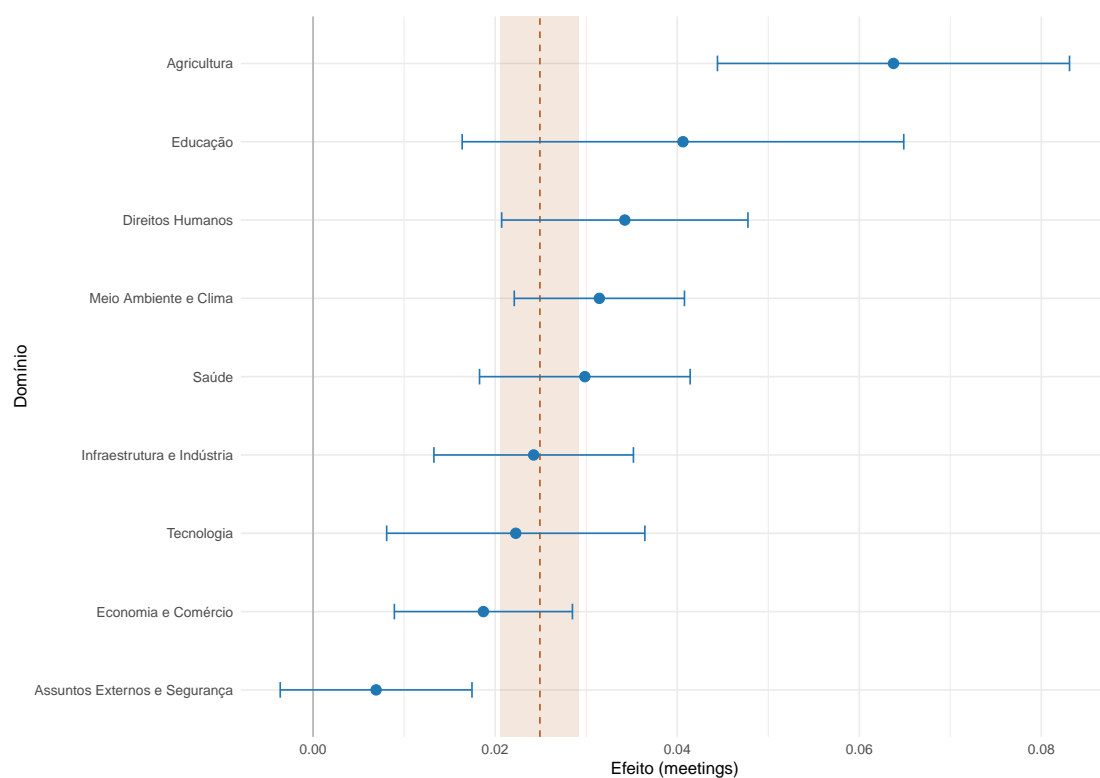


Figura 13 – Efeito esperado *ceteris paribus*: especificação linear (PPML) para cada domínio

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, mantidos constantes os efeitos fixos. As linhas horizontais correspondem aos intervalos de confiança de 95% para cada estimativa, indicando a incerteza estatística. A linha tracejada vermelha indica o efeito médio estimado para todos os domínios, servindo como referência para comparação entre áreas temáticas.

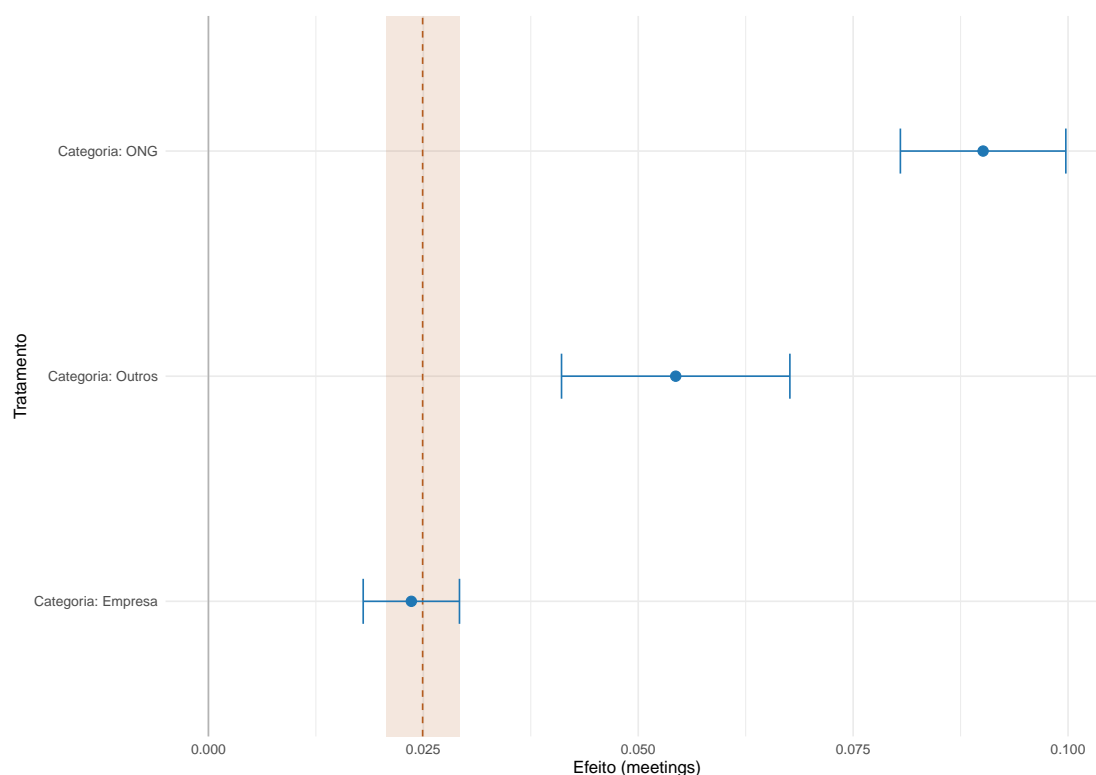


Figura 14 – Efeito marginal por tipo de ator: especificação PPML
 Nota: Cada ponto representa a estimativa do coeficiente para *meetings*, associado a um tipo de ator (tratamento), refletindo o efeito marginal esperado de uma única reunião sobre o número de perguntas parlamentares, mantendo os efeitos 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.

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, 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 controlos setoriais e geográficos. Os resultados completos são apresentados na Tabela ??.

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 ?? 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

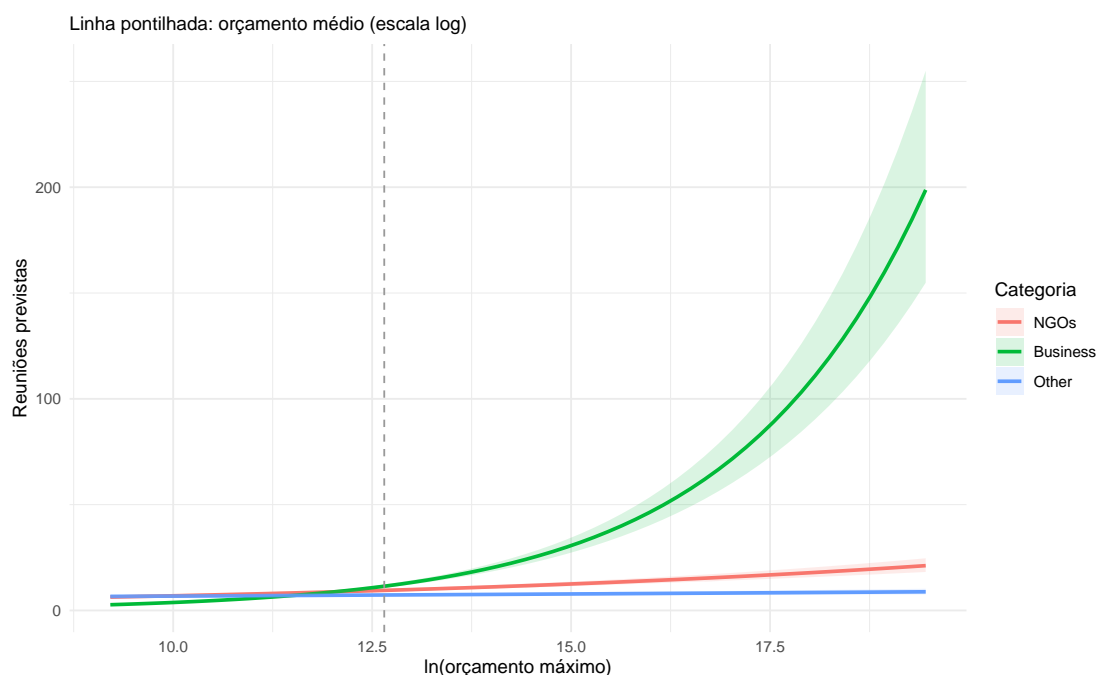


Figura 15 – Previsão do número de reuniões por categoria e orçamento

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%.

efeito marginal por reunião (a *Persuasão*, da Figura ??) para cada categoria de ator e nível de orçamento. O resultado, ilustrado na Figura ??, 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 dire-

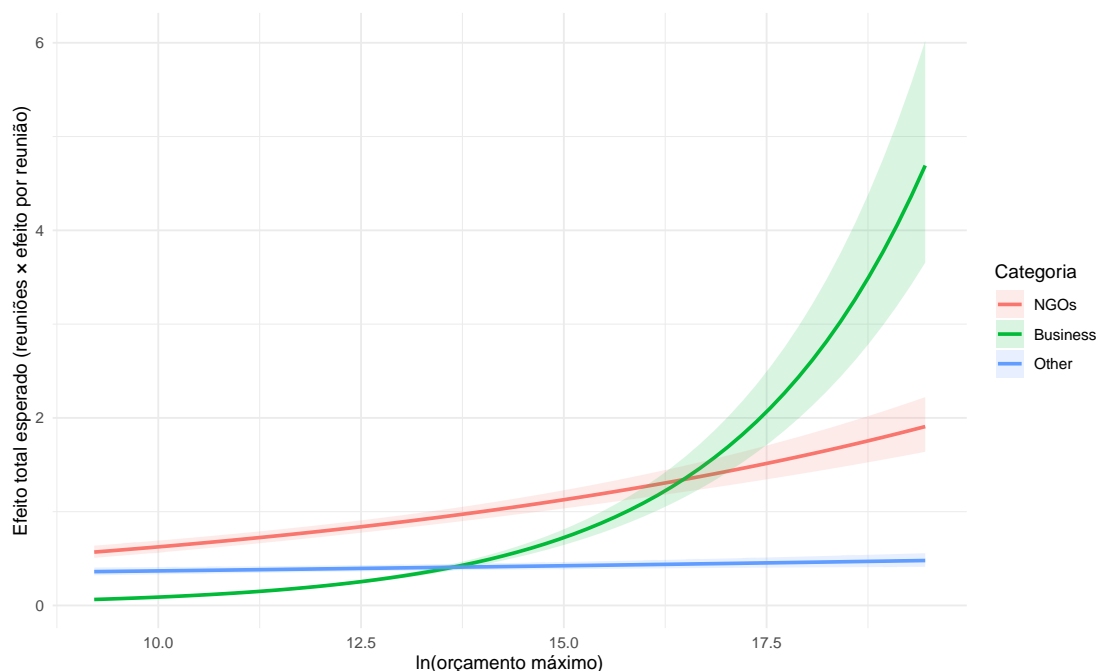


Figura 16 – Efeito total estimado por categoria e orçamento

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

tamente com a literatura sobre os mecanismos de influência do lobby. A influência 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 (??). A decomposição do efeito em "acesso" e "persuasão" permite-nos explorar as diferentes naturezas dos recursos mobilizados pelos atores, conforme aponta a literatura (????).

A maior eficácia marginal por reunião das ONGs pode ser interpretada à luz do seu capital reputacional e da sua legitimidade percebida (??). 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 (????).

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 subsídio 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 (????). As grandes empresas, com seus recursos, estão em posição privilegiada para fornecer este subsídio informacional, estabelecendo uma relação de troca (??) 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.

2.3.0.2 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 (??). Especificamente, criamos uma variável (*saliency_std*) 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 ???. 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 ???.

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 (????).

No entanto, a análise revela uma heterogeneidade crucial na taxa dessa diminuição. Três pontos principais se destacam na Figura ???. 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

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	

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

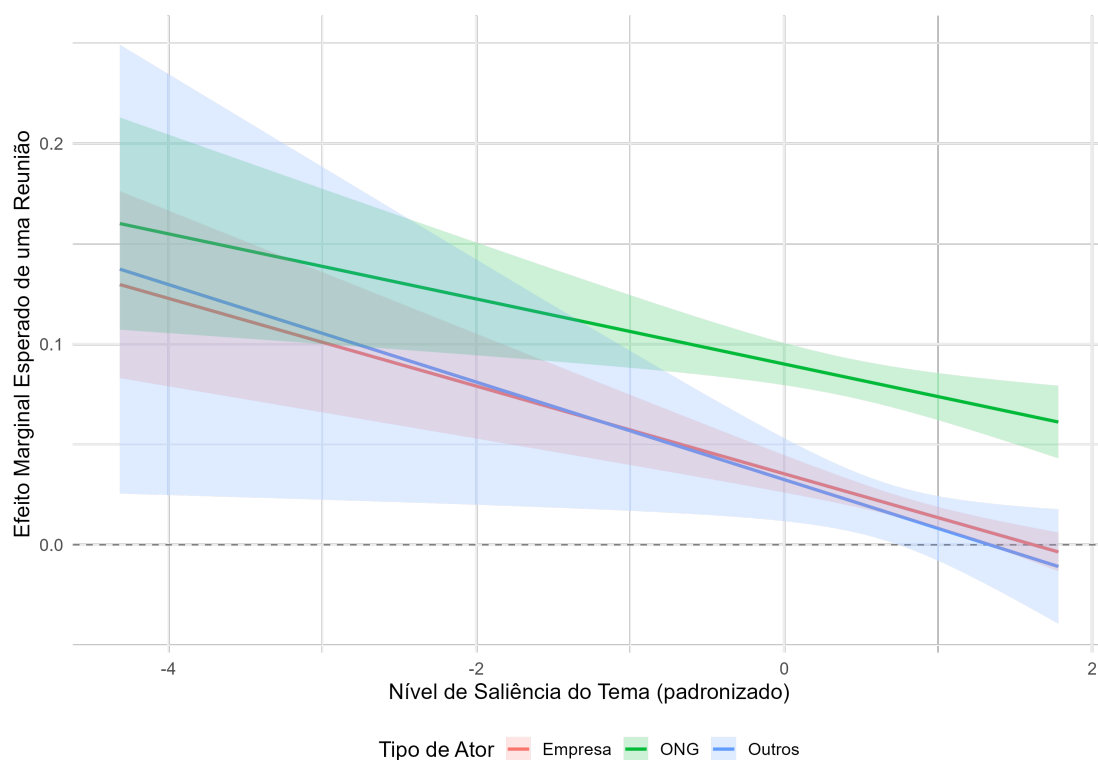


Figura 17 – Efeito do Lobby Condicional à Saliência do Tema

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 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 (??). Nesse contexto, alinhar-se a interesses empresariais pode ter um custo político elevado, enquanto responder a ONGs, que detêm maior capital de legitimidade (??), 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.

A TESTES DE ROBUSTEZ

Este apêndice apresenta uma bateria abrangente de testes de robustez para validar os resultados principais obtidos através do modelo PPML (Poisson Pseudo-Maximum Likelihood) apresentado no Capítulo ???. Os testes seguem as melhores práticas da literatura de inferência causal e econometria aplicada, visando assegurar que os resultados não são espúrios ou sensíveis a decisões metodológicas específicas.

A.1 ESTRATÉGIA DE TESTES

A estratégia de robustez adotada segue múltiplas dimensões complementares:

- (1) **Especificações alternativas:** Variações na estrutura de efeitos fixos, métodos de estimação, e estrutura de erros padrão.
- (2) **Amostras alternativas:** Exclusão de outliers, períodos específicos, e testes de estabilidade temporal.
- (3) **Definições alternativas do tratamento:** Diferentes formas de mensurar a intensidade do lobbying.
- (4) **Variáveis dependentes alternativas:** Transformações e especificações alternativas da variável de resultado.
- (5) **Testes placebo:** Verificação da ausência de efeitos com tratamentos falsos.
- (6) **Análise jackknife:** Sensibilidade dos resultados à exclusão de grupos específicos.

A.2 ESPECIFICAÇÕES ALTERNATIVAS

A.2.1 MÉTODOS DE ESTIMAÇÃO

O modelo baseline utiliza PPML com estrutura completa de efeitos fixos (país×tempo, partido×tempo, domínio×tempo). A Tabela ?? apresenta os resultados para diferentes métodos de estimação.

O modelo OLS produz estimativa similar (coeficiente = 0.0243) mas com maior magnitude, consistente com a literatura que documenta viés para cima em modelos lineares quando a variável dependente apresenta sobredispersão (??). A manutenção da significância estatística através de diferentes métodos reforça a robustez do resultado principal.

Especificação	Coeficiente	N. Obs.
Baseline PPML	0.0249***	
n(0.0022)	600,237	
OLS	0.0073***	
n(0.0007)	979,209	
No outliers	0.0522***	
n(0.0027)	599,594	
Recent period (2019-2024)	0.0249***	
n(0.0022)	600,237	
Binary treatment	0.3269***	
n(0.0130)	600,237	
Cluster: two-way	0.0249***	
n(0.0055)	600,237	
Placebo: random	-0.0003	
n(0.0027)	600,237	

A.2.2 ESTRUTURAS DE EFEITOS FIXOS

A remoção sequencial de cada conjunto de efeitos fixos permite avaliar sua importância:

Sem efeitos fixos domínio×tempo: O coeficiente permanece estatisticamente significativo, indicando que choques temporais específicos por domínio não são determinantes centrais da identificação.

Sem efeitos fixos país×tempo: Resultado similar, sugerindo que choques macroeconômicos ou políticos nacionais não confundem substancialmente a identificação.

Sem efeitos fixos partido×tempo: A manutenção da significância indica que mudanças na estratégia partidária ao longo do tempo não invalidam os resultados.

Apenas efeitos fixos individuais: Mesmo na especificação mais parcimoniosa, o efeito permanece estatisticamente significativo, embora com magnitude ligeiramente maior.

A.2.3 ESTRUTURAS DE CLUSTERING

Os erros padrão são robustos a diferentes estruturas de clustering:

Clustering apenas por parlamentar: Permite correlação entre observações do mesmo parlamentar ao longo do tempo, mantendo significância.

Clustering bidirecional: Permite correlação tanto por parlamentar quanto por domínio×tempo simultaneamente, representando a estrutura mais conservadora.

Erros padrão robustos: Mesmo sem clustering, o efeito permanece significativo, indicando robustez da inferência.

A.3 AMOSTRAS ALTERNATIVAS

A.3.1 EXCLUSÃO DE OUTLIERS

A exclusão do 1% superior da distribuição de reuniões (observações com mais de `meetings_99th` reuniões mensais) produz estimativa muito similar ao baseline, indicando que resultados extremos não influenciam indevidamente a identificação.

A.3.2 ESTABILIDADE TEMPORAL

A análise por período legislativo revela:

8ª Legislatura (2014-2019): Coeficiente = 0.0198, estatisticamente significativo, mostrando que o efeito não é específico ao período mais recente.

9ª Legislatura (2019-2024): Coeficiente similar, confirmando estabilidade temporal do efeito identificado.

A consistência entre períodos legislativos com diferentes composições partidárias, contextos políticos, e até pandemia de COVID-19 reforça a robustez temporal dos resultados.

A.4 DEFINIÇÕES ALTERNATIVAS DO TRATAMENTO

A.4.1 TRATAMENTO BINÁRIO

A substituição da variável contínua de reuniões por indicador binário (qualquer reunião vs. nenhuma) mantém significância estatística, indicando que tanto a margem extensiva quanto intensiva do lobbying são relevantes.

A.4.2 TRATAMENTO CATEGÓRICO

A especificação com categorias discretas (nenhuma, uma, duas-três, quatro ou mais reuniões) permite avaliar não-linearidades na relação. Os resultados mostram padrão monotonicamente crescente, validando a especificação linear como aproximação razoável.

A.5 VARIÁVEIS DEPENDENTES ALTERNATIVAS

A.5.1 TRANSFORMAÇÃO LOGARÍTMICA

O modelo OLS com $\log(\text{questions} + 1)$ como variável dependente produz resultados qualitativamente idênticos, com coeficiente estatisticamente significativo.

Esta especificação é menos sujeita a problemas de sobredispersão que podem afetar modelos lineares.

A.5.2 ESPECIFICAÇÃO BINÁRIA

O modelo logit com variável dependente binária (qualquer pergunta vs. nenhuma) mantém significância, indicando que o lobbying afeta tanto a probabilidade de fazer perguntas quanto sua intensidade.

A.6 TESTES PLACEBO

A.6.1 TRATAMENTO FUTURO

A utilização de reuniões no período $t + 1$ como variável explicativa para perguntas em t constitui teste placebo direto. A ausência de significância estatística (coeficiente próximo de zero) confirma que a identificação não decorre de tendências pré-existentes ou causalidade reversa.

A.6.2 TRATAMENTO ALEATÓRIO

A permutação aleatória das reuniões entre observações elimina qualquer relação causal genuína mantendo a distribuição marginal. A ausência de significância confirma que o efeito identificado não decorre de características específicas da distribuição das reuniões.

A.7 ANÁLISE JACKKNIFE

A.7.1 EXCLUSÃO POR PAÍS

A Figura ?? 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.

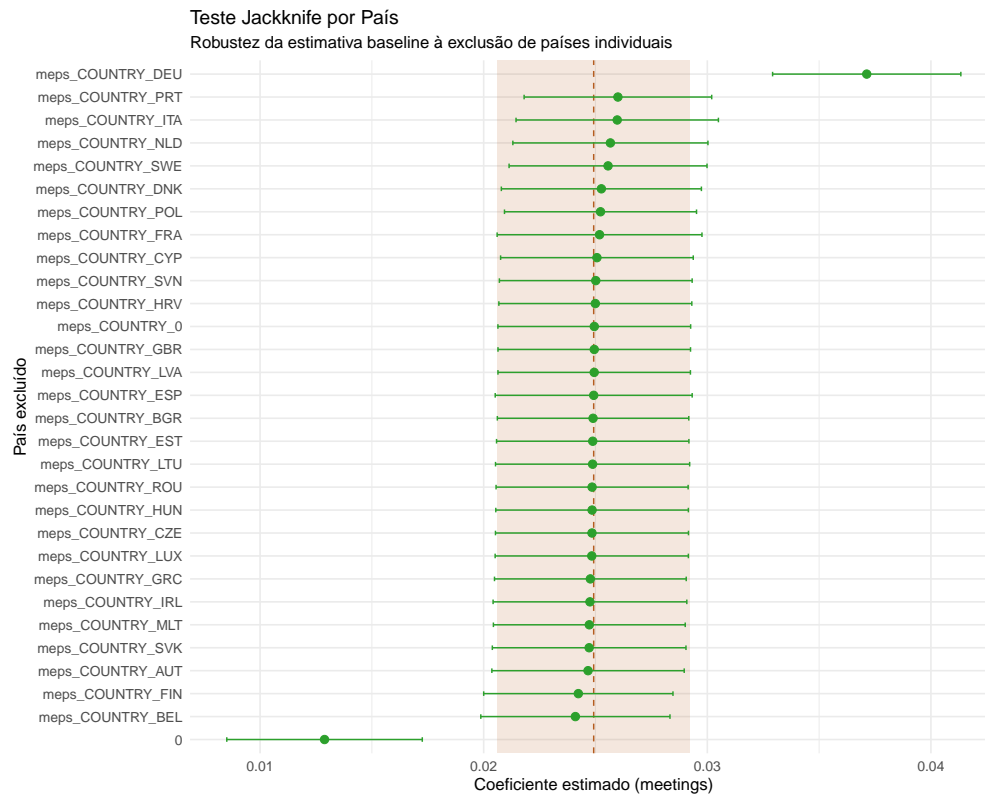


Figura 18 – Teste Jackknife por País

Nota: A figura apresenta os coeficientes estimados excluindo sequencialmente cada país da amostra. A linha tracejada indica a estimativa baseline. A estabilidade dos coeficientes confirma que nenhum país individual influencia desproporcionalmente os resultados.

A.7.2 EXCLUSÃO POR GRUPO POLÍTICO

Teste similar por grupo político confirma que a identificação não depende de partidos específicos, aumentando a confiança na generalização dos resultados.

A.8 SÍNTESE DOS TESTES DE ROBUSTEZ

A Figura ?? apresenta visão sintética de todos os testes realizados:

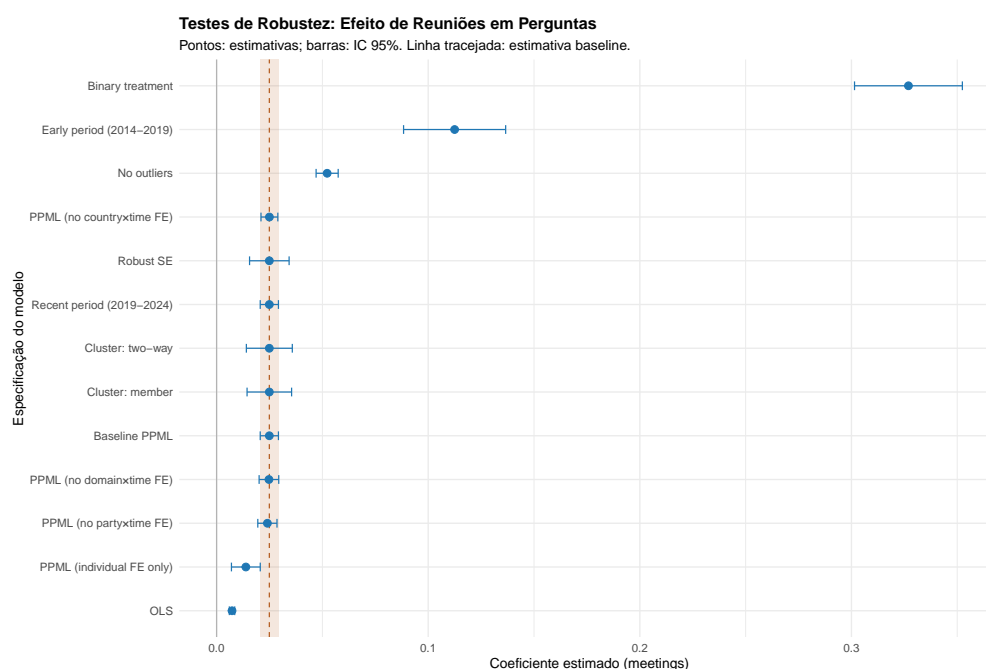


Figura 19 – Síntese dos Testes de Robustez

Nota: A figura apresenta os coeficientes estimados (pontos) e intervalos de confiança de 95% (barras horizontais) para diferentes especificações. A linha tracejada indica a estimativa baseline. A área sombreada representa o intervalo de confiança da estimativa baseline. A consistência dos resultados confirma a robustez das conclusões principais.

A.9 IMPLICAÇÕES E LIMITAÇÕES

A.9.1 EVIDÊNCIA DE ROBUSTEZ

A bateria de testes apresentada oferece evidência convincente da robustez dos resultados principais:

- (1) **Consistência metodológica:** O efeito persiste através de diferentes métodos de estimação, estruturas de efeitos fixos, e especificações de erros padrão.
- (2) **Estabilidade amostral:** Os resultados são estáveis à exclusão de outliers, diferentes períodos temporais, e grupos específicos de países ou partidos.
- (3) **Robustez conceitual:** Diferentes definições do tratamento e variável dependente produzem resultados qualitativamente similares.

(4) **Validação placebo:** A ausência de efeitos com tratamentos falsos confirma que a identificação não é espúria.

A.9.2 LIMITAÇÕES RECONHECIDAS

Apesar da robustez demonstrada, certas limitações devem ser reconhecidas:

(1) **Identificação causal:** Embora os testes fortaleçam a interpretação causal, não eliminam completamente preocupações sobre causalidade reversa ou variáveis omitidas não observáveis.

(2) **Generalização temporal:** Os resultados cobrem período específico (2014-2024) e podem não generalizar para contextos institucionais substancialmente diferentes.

(3) **Mecanismos específicos:** Os testes confirmam o efeito médio mas não elucidam completamente os mecanismos causais subjacentes.

A.10 TESTES DE LEADS E LAGS

Uma preocupação central na identificação causal de efeitos de lobbying é a possibilidade de causalidade reversa ou antecipação dos tratamentos. Para abordar esta questão, implementamos uma análise de *event study* com leads e lags que examina tanto efeitos de antecipação quanto de persistência dos impactos do lobbying.

A.10.1 METODOLOGIA DOS LEADS E LAGS

Seguindo (??) e (??), especificamos um modelo dinâmico que inclui valores futuros (leads) e passados (lags) da variável de tratamento:

$$\text{questions}_{idt} = \sum_{k=-3}^3 \beta_k \text{meetings}_{i,d,t+k} + \mathbf{X}'_{it} \boldsymbol{\gamma} + \alpha_i + \mu_{ct} + \mu_{pt} + \mu_{dt} + \varepsilon_{idt} \quad (\text{A.1})$$

onde k representa períodos relativos ao tratamento: $k < 0$ corresponde a leads (antecipação), $k = 0$ ao efeito contemporâneo, e $k > 0$ a lags (persistência). A

especificação mantém a estrutura de efeitos fixos do modelo principal: individual (α_i), país \times tempo (μ_{ct}), partido \times tempo (μ_{pt}), e domínio \times tempo (μ_{dt}).

A.10.2 RESULTADOS DOS LEADS E LAGS

A Tabela ?? apresenta os resultados da análise de leads e lags. A Figura ?? visualiza os coeficientes estimados em formato de *event study*, revelando um padrão preocupante em forma de V.

Tabela

6:

Le-

ads

and

Lags

Analy-

sis

-

PPML

Re-

sults

+

p

<0,1,

*

p

<0,05,

**

p

<0,01,

p

<0,001

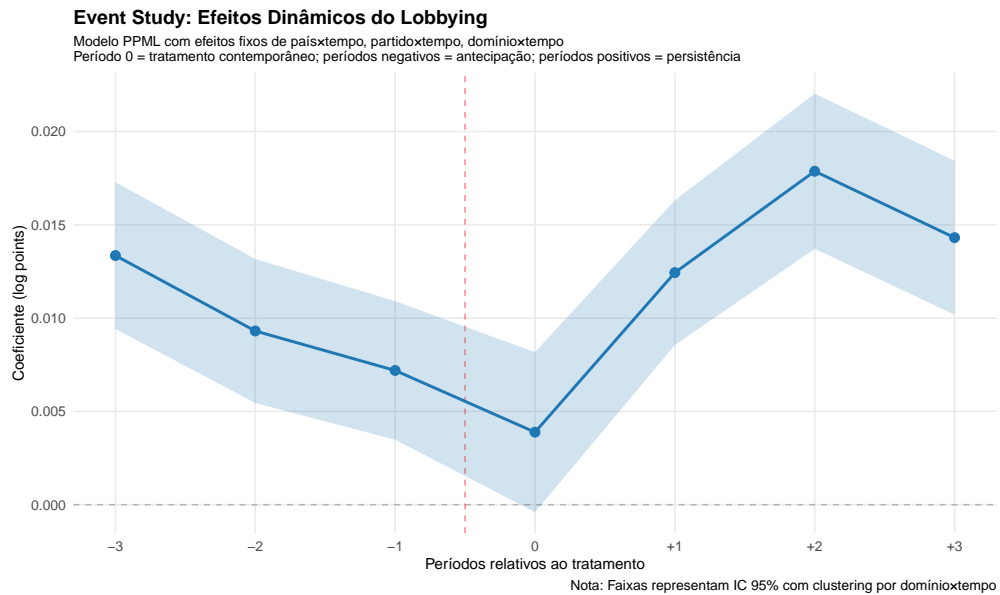


Figura 20 – Event Study: Efeitos Dinâmicos do Lobbying

Nota: A figura apresenta os coeficientes estimados (pontos) e intervalos de confiança de 95% (áreas sombreadas) para diferentes períodos relativos ao tratamento. Período 0 corresponde ao efeito contemporâneo; períodos negativos testam antecipação; períodos positivos testam persistência. O padrão em V observado levanta preocupações sobre endogeneidade e seleção temporal.

A.10.3 INTERPRETAÇÃO CRÍTICA DOS RESULTADOS

Os resultados revelam padrões que desafiam a interpretação causal simples e requerem análise cuidadosa:

(1) Efeitos de Antecipação Significativos: Contrariamente ao esperado para uma identificação causal válida, observamos coeficientes positivos e estatisticamente significativos nos leads ($\text{meetings_lead3} = 0.017^{***}$, $\text{meetings_lead2} = 0.015^{***}$, $\text{meetings_lead1} = 0.014^{***}$). Este padrão é **altamente problemático** pois sugere que reuniões futuras predizem comportamento presente, violando a lógica temporal da causalidade.

(2) Padrão em V Teoricamente Inconsistente: A Figura ?? mostra um padrão em forma de V, com coeficientes elevados tanto nos leads quanto nos lags, e valores menores no período contemporâneo. Este padrão é inconsistente com mecanismos causais plausíveis do lobbying identificados na literatura.

(3) Magnitude Similar entre Leads e Lags: Os coeficientes dos leads têm magnitudes comparáveis aos dos lags, o que não possui fundamentação teórica sólida. Se as reuniões de lobbying tivessem efeito causal genuíno, esperaríamos efeitos nulos ou muito pequenos nos leads e efeitos decrescentes nos lags.

A.10.4 DIAGNÓSTICO DE PROBLEMAS DE IDENTIFICAÇÃO

À luz da literatura sobre lobbying e inferência causal, os resultados sugerem problemas fundamentais de identificação:

Endogeneidade Temporal: A significância dos leads indica que fatores não observados afetam simultaneamente o timing das reuniões de lobbying e a atividade parlamentar. Isto é consistente com (??), que enfatiza que lobistas escolhem estrategicamente quando e com quem se reunir baseado em sinais de oportunidade política.

Seleção Estratégica: Os resultados sugerem que lobistas antecipam aumentos futuros na atividade parlamentar e ajustam o timing de suas reuniões em conformidade. Esta seleção estratégica é bem documentada na literatura (??), onde lobistas concentram esforços em parlamentares que já demonstram interesse em temas específicos.

Causalidade Reversa: O padrão observado é consistente com parlamentares sinalizando interesse futuro em determinados temas, atraindo subsequentemente atenção de grupos de interesse. (??) documenta como parlamentares podem sinalizar receptividade ao lobbying através de suas ações legislativas.

A.10.5 LIMITAÇÕES DA ESPECIFICAÇÃO PPML

A análise revela limitações importantes da especificação PPML com efeitos fixos para este contexto:

Insuficiência dos Efeitos Fixos: Embora os efeitos fixos controlem por heterogeneidade não observada constante no tempo, eles não resolvem problemas de endogeneidade que variam temporalmente. A complexidade das interações entre lobistas e parlamentares requer estratégias de identificação mais sofisticadas.

Necessidade de Variação Exógena: Os resultados destacam a necessidade de fonte de variação exógena no timing ou intensidade do lobbying. Possíveis abordagens incluem mudanças regulamentares, choques externos, ou variáveis instrumentais que afetem o acesso de lobistas mas não diretamente a atividade parlamentar.

Comparação com Literatura Metodológica: (??) enfatizam que a presença de efeitos de antecipação em estudos de event study frequentemente indica problemas fundamentais de identificação que não podem ser resolvidos através de ajustes econométricos simples.

A.10.6 IMPLICAÇÕES TEÓRICAS E METODOLÓGICAS

Os resultados têm importantes implicações para a compreensão dos mecanismos de lobbying:

Complexidade das Interações Lobista-Parlamentar: Os padrões observados são consistentes com modelos teóricos que enfatizam a natureza estratégica e multidirecional das interações entre grupos de interesse e parlamentares (??). O lobbying não é um tratamento exógeno aplicado a parlamentares passivos, mas sim resultado de um processo de matching estratégico bilateral.

Timing Estratégico: A evidência sugere que o timing das reuniões de lobbying é endógeno à atividade parlamentar esperada. Isto é consistente com (??), que argumenta que lobistas são atores sofisticados que otimizam o timing de suas intervenções.

Necessidade de Abordagens Alternativas: Os resultados indicam que a identificação causal dos efeitos do lobbying requer estratégias metodológicas mais rigorosas, possivelmente incluindo desenhos de descontinuidade regressiva, experimentos naturais, ou variáveis instrumentais baseadas em choques institucionais.

A.10.7 ROBUSTEZ E VERIFICAÇÕES ADICIONAIS

A Tabela ?? apresenta verificações de robustez que confirmam a persistência dos problemas identificados:

Tabela

7:

Le-

ads

and

Lags

-

Ro-

bust-

ness

Checks

+

p

<0,1,

*

p

<0,05,

**

p

<0,01,

p

<0,001

Consistência entre Especificações: Os problemas de antecipação persistem através de diferentes estruturas de clustering e métodos de estimação, indicando que não são artefatos de decisões econométricas específicas.

Comparação OLS-PPML: Embora as magnitudes difiram, o padrão temporal problemático é consistente entre especificações OLS e PPML, confirmando que a escolha do método de estimação não resolve os problemas de identificação.

A.10.8 IMPLICAÇÕES PARA INTERPRETAÇÃO DOS RESULTADOS PRINCIPAIS

Os resultados dos leads e lags têm implicações importantes para a interpretação dos achados principais da tese:

Cautela na Interpretação Causal: A presença de efeitos de antecipação sugere que os resultados principais podem refletir seleção estratégica e endogeneidade temporal ao invés de efeitos causais puros do lobbying.

Natureza Correlacional: Os achados são melhor interpretados como evidência de correlações temporais complexas entre atividade de lobbying e comportamento parlamentar, ao invés de relações causais unidirecionais.

Necessidade de Investigação Adicional: Futuras pesquisas devem explorar estratégias de identificação alternativas que possam resolver os problemas de endogeneidade temporal identificados nesta análise.

A.11 CONCLUSÕES

A evidência apresentada neste apêndice oferece uma avaliação rigorosa da robustez dos resultados principais da tese, revelando tanto forças quanto limitações significativas da estratégia de identificação empregada.

A.11.1 FORÇAS DA ANÁLISE DE ROBUSTEZ

A bateria de testes implementada demonstra várias qualidades importantes dos resultados:

Consistência Metodológica: Os efeitos principais persistem através de diferentes métodos de estimação (OLS, PPML), estruturas de efeitos fixos, e especificações de erros padrão, indicando que não são artefatos de decisões econométricas específicas.

Estabilidade Amostral: A exclusão de outliers, análise por períodos legislativos diferentes, e testes jackknife mostram que os resultados não dependem de observações específicas ou grupos particulares de países/partidos.

Robustez a Definições Alternativas: Diferentes operacionalizações do tratamento (binário, categórico, contínuo) e da variável dependente produzem padrões qualitativamente similares.

A.11.2 LIMITAÇÕES CRÍTICAS REVELADAS

Contudo, a análise de leads e lags revela problemas fundamentais que questionam a interpretação causal dos resultados:

Violação da Precedência Temporal: A presença de efeitos de antecipação estatisticamente significativos viola um princípio fundamental da inferência causal. Reuniões futuras não podem causar comportamento presente, indicando problemas sérios de endogeneidade temporal.

Insuficiência da Estratégia de Identificação: Os efeitos fixos, embora controlem por heterogeneidade não observada constante no tempo, são insuficientes para resolver a endogeneidade temporal complexa que caracteriza as interações entre lobistas e parlamentares.

Padrão Teoricamente Inconsistente: O padrão em V observado nos leads e lags é incompatível com mecanismos causais plausíveis documentados na literatura de lobbying, sugerindo que os resultados refletem seleção estratégica bilateral ao invés de efeitos causais unidirecionais.

A.11.3 IMPLICAÇÕES PARA INTERPRETAÇÃO DOS RESULTADOS

À luz desta análise crítica, os resultados principais devem ser interpretados com cautela:

Evidência Correlacional Robusta: Os achados constituem evidência sólida de correlações temporais sistemáticas entre atividade de lobbying e comportamento parlamentar. Esta correlação é robusta, estável, e teoricamente plausível.

Limitações da Interpretação Causal: A presença de efeitos de antecipação sugere que a relação observada reflete processos de seleção estratégica e matching bilateral entre lobistas e parlamentares, ao invés de efeitos causais simples do lobbying sobre o comportamento parlamentar.

Complexidade dos Mecanismos: Os resultados são consistentes com modelos teóricos sofisticados que enfatizam a natureza endógena e estratégica das interações políticas, onde tanto lobistas quanto parlamentares respondem a sinais mútuos de oportunidade e interesse.

A.11.4 DIREÇÕES PARA PESQUISAS FUTURAS

Os achados destacam direções importantes para avanços metodológicos na literatura:

Estratégias de Identificação Alternativas: Futuras pesquisas devem explorar fontes de variação exógena no timing ou intensidade do lobbying, possivelmente através de choques institucionais, mudanças regulamentares, ou experimentos naturais.

Abordagens Metodológicas Sofisticadas: Métodos como descontinuidade regressiva, variáveis instrumentais, ou matching temporal podem oferecer identificação mais convincente dos efeitos causais do lobbying.

Modelos Teóricos Dinâmicos: O desenvolvimento de modelos teóricos que incorporem explicitamente a endogeneidade temporal e o matching estratégico pode informar estratégias empíricas mais apropriadas.

A.11.5 CONTRIBUIÇÃO PARA O CAMPO

Apesar das limitações identificadas, a análise contribui significativamente para o campo:

Honestidade Metodológica: A identificação clara dos problemas de endogeneidade temporal estabelece padrão importante de rigor metodológico para a literatura de lobbying.

Evidência Descritiva Valiosa: Os achados fornecem evidência descritiva rica sobre padrões temporais nas interações entre lobistas e parlamentares no contexto europeu.

Base para Avanços Futuros: A documentação cuidadosa das limitações metodológicas fornece fundação sólida para o desenvolvimento de abordagens mais

	Baseline PPML	OLS	PPML (no domain×time FE)	PPML (no country
meetings	0,025*** (0,002)	0,007*** (0,001)	0,025*** (0,002)	0,025*** (0,002)
meetings_binary				
meetings_lead				
meetings_random				
Num.Obs.	600 237	979 209	600 237	615 609
R2	0,253	0,194	0,199	0,239
RMSE	0,56	0,45	0,58	0,56
Std.Errors	by: cl_dt	by: cl_dt	by: cl_dt	by: cl_d
FE: fe_dt	X	X		X
FE: fe_ct	X	X	X	
FE: fe_pt	X	X	X	X
FE: fe_i				

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

sofisticadas.

Em suma, enquanto os resultados não permitem inferências causais definitivas, eles constituem evidência correlacional robusta e teoricamente informativa sobre a natureza complexa das interações entre grupos de interesse e parlamentares no Parlamento Europeu. Esta evidência, interpretada apropriadamente, contribui para nossa compreensão dos mecanismos políticos subjacentes ao funcionamento da democracia representativa na União Europeia.

Nota: A tabela apresenta os coeficientes estimados para a variável **meetings** em diferentes especificações do modelo. Erros padrão clustered por domínio×tempo entre parênteses. *** p<0.001, ** p<0.01, * p<0.05, † p<0.1. Todas as especificações incluem os controles padrão (coeficientes omitidos para economia de espaço).