



AVALIAÇÃO DE IMPACTO: MÉTODO DE DIFERENÇA-EM- -DIFERENÇAS (DID)

Avaliação de impacto: Método de Diferença-em-Diferenças (DID)

Organizadores

André Portela Souza
Lycia Lima

Equipe técnica

Alei Fernandes Santos
Ana Luiza Camillo Mansano
Caio de Souza Castro
Michel Szklo
Murilo Sepulvida Cardoso
Pedro Molina Ogeda

APRESENTAÇÃO

Este guia integra a série de publicações *Avaliação na Prática*, desenvolvida pelo FGV CLEAR com o objetivo de ampliar o acesso a conhecimentos sobre monitoramento e avaliação com foco em políticas públicas.

Fundado em 2015, o FGV CLEAR tem se dedicado ao fortalecimento da cultura de gestão orientada por evidências no Brasil e em países lusófonos. Com sede na Escola de Economia de São Paulo da Fundação Getúlio Vargas (FGV EESP), o FGV CLEAR atua como centro regional da Iniciativa CLEAR (Centers for Learning on Evaluation and Results).

A Iniciativa CLEAR, criada em 2010, é um programa de desenvolvimento de capacidades em monitoramento e avaliação que congrega instituições acadêmicas e parceiros doadores de modo a contribuir para o uso de evidências na tomada de decisões em países em desenvolvimento. Ao todo, são seis centros regionais CLEAR, coordenados pela *Global Evaluation Initiative (GEI)*, um programa liderado pelo Banco Mundial e pelo Programa das Nações Unidas para o Desenvolvimento (PNUD).

O FGV CLEAR atua com governos, organizações multilaterais, sociedade civil e academia, oferecendo capacitação, assistência técnica, geração e disseminação de conhecimento científico na área.

Saiba mais sobre o FGV CLEAR e acesse outras publicações em: **www.fgvclear.org**

Avaliação de Impacto: Método de Diferença-em-Diferenças

Maio 2025

Resumo

Este é um guia prático e revisado para cientistas de dados que utilizam a metodologia de diferença-em-diferenças (DID) na análise de efeitos causais de intervenções não randomizadas. Com foco na abordagem mais recente de DID escalonado, o manual revisita o DID tradicional e com múltiplos períodos, além de apresentar um referencial teórico acompanhado de exemplos práticos extraídos de estudos científicos dos principais autores sobre o tema: : Goodman-Bacon (2021); Sun and Abraham (2021); Callaway and Sant'Anna (2021); Borusyak et al. (2024).

Sumário

1	Relembrando Diferença-em-Diferenças	3
1.1	Modelo de resultados potenciais e a importância de definirmos bem o que queremos estimar.	3
1.1.1	ATE vs ATT	5
1.2	DID como diferenças de médias	8
1.3	DID vs TWFE	10
1.4	DID com múltiplos períodos	11
1.4.1	Regressão DID com múltiplos períodos pós-tratamento	15
1.4.2	Regressão com Dummies de tempo	18
1.5	Regressão DID com múltiplos períodos pré-tratamento	22
2	Diferença-em-Diferenças escalonado	24
2.1	O que é um DID escalonado?	24
2.2	Aumentando o modelo de resultados potenciais	24
2.3	Quando e por que um DID escalonado apresenta problemas?	27
3	Como usar um DID escalonado?	34
3.1	Sun and Abraham (2021)	35
3.1.1	Estimador	35
3.1.2	Hipóteses	35
3.1.3	Código em R com exemplo	36
3.2	Callaway and Sant’Anna (2021)	43
3.2.1	Estimador	43
3.2.2	Hipóteses	45
3.2.3	Código em R com exemplo	46
3.3	De Chaisemartin and d’Haultfoeuille (2020)	52
3.3.1	Estimador	52
3.3.2	Hipóteses	53
3.3.3	Código em R com exemplo	55
3.4	Borusyak et al. (2024)	57
3.4.1	Estimador	57
3.4.2	Hipóteses	60
3.4.3	Código em R com exemplo	60
3.5	Diferença entre os estimadores	66
3.5.1	Hipóteses	66
3.5.2	Usabilidade	71
3.6	Como escolher os pesos	72
3.6.1	Caçando resultados	72
3.6.2	Definindo os Pesos	73
3.6.3	Pesos Mais Utilizados na Literatura	74
4	Conclusão	87

1 Relembrando Diferença-em-Diferenças

Parte importante do processo de avaliação de uma política pública ou intervenção é avaliar seu impacto naquele ambiente. No entanto, estimar esse impacto de forma **causal** nem sempre é simples. Essa análise é fundamental porque nos permite saber se a política está, de fato, alcançando seus objetivos — e, mais importante, distinguir o que é efeito da política do que é consequência de outros fatores.

Nosso objetivo, portanto, é analisar, com base em alguma métrica específica, a diferença entre o resultado de um indivíduo que recebe a intervenção e o que esse mesmo indivíduo apresentaria caso não a recebesse. No caso da avaliação da eficácia de uma vacina, por exemplo, queremos saber qual é a probabilidade de uma pessoa contrair a doença ao ser vacinada, em comparação com a probabilidade de essa mesma pessoa contrair a doença sem a vacina.

Na realidade, apenas uma dentre essas duas situações se concretizará: o indivíduo receberá a vacina ou não. É justamente essa limitação que dá origem ao conceito de modelos de **resultados potenciais**. Esses modelos propõem dois cenários possíveis para o mesmo indivíduo, que diferem apenas pela presença ou ausência da intervenção. No exemplo da vacina, seriam: (1) a probabilidade de contrair a doença sendo vacinado e (2) a probabilidade de contrair a doença sem a vacinação. A partir desses cenários, conseguimos estimar de forma mais precisa o impacto da intervenção — e é a esses cenários que damos o nome de **resultados potenciais**, conceito que será detalhado na próxima seção.

1.1 Modelo de resultados potenciais e a importância de definirmos bem o que queremos estimar.

Definimos

- $y_i(1)$ como o resultado potencial do indivíduo i caso ele tenha sido exposto à intervenção.
- $y_i(\infty)$ como o resultado potencial do indivíduo i caso ele **não** tenha sido exposto à intervenção.

o efeito da intervenção para o indivíduo i , δ_i , será, portanto,

$$\delta_i = y_i(1) - y_i(\infty).$$

Ou seja, no caso da vacina, trata-se da diferença entre a probabilidade de o indivíduo i contrair a doença ao receber a vacina e a probabilidade de contrair a doença sem recebê-la — exatamente o resultado que nos interessa estimar. Para inferir o impacto causal de uma intervenção sobre um indivíduo tratado, precisamos, portanto, estimar o resultado desse indivíduo em um cenário que não foi observado, conhecido como **contrafactual**.

Existem diferentes técnicas para realizar essa estimativa, e a escolha da melhor abordagem depende de hipóteses cuja razoabilidade varia de acordo com o contexto da intervenção. Uma estratégia possível é encontrar um indivíduo que **não tenha participado** da intervenção, mas que seja muito semelhante ao indivíduo tratado — ou seja, formar um grupo de controle comparável. No entanto, como é impossível encontrar indivíduos exatamente iguais, muitas vezes **não é viável estimar o efeito individual do tratamento**.

Por isso, trabalhamos com o conceito de **efeito médio do tratamento** (*Average Treatment Effect* — ATE), definido como:

$$ATE = E[y_i(1) - y_i(\infty)].$$

Em alguns casos, como será demonstrado neste guia, o contexto da intervenção não permite estimar o efeito médio do tratamento de forma geral. Nessas situações, podemos tentar estimar o efeito médio do tratamento apenas para o grupo de indivíduos que de fato recebeu a intervenção.¹ Esse conceito é chamado de efeito médio

¹Nem sempre $ATE=ATT$ e o mais provável é que sejam diferentes quando estamos olhando para intervenções que não foram randomizadas como nos modelos de Diferença-em-Diferenças. A Seção 1.1.1 discute essa diferença mais profundamente.

de tratamento nos tratados (*ATT - Average Treatment Effect on the Treated*) e é definido como:

$$\begin{aligned} ATT &= E[y_i(1) - y_i(\infty)|d_i = 1] \\ &= E[y_i(1)|d_i = 1] - E[y_i(\infty)|d_i = 1], \end{aligned}$$

onde d_i é uma variável indicadora igual a 1 se o indivíduo i é tratado e 0 caso contrário (Angrist and Pischke, 2009).

Nesse contexto, não é necessário criar um cenário contrafactual para cada indivíduo tratado, mas sim formar dois grupos comparáveis, um de tratamento e outro de controle. Em outras palavras, buscamos estabelecer um grupo de controle no qual a média do resultado potencial do grupo tratado, caso não tivesse recebido o tratamento (representada em vermelho na equação acima), seja igual à média do grupo de controle, caso este também não tivesse sido tratado. Ou seja, buscamos garantir que: $E[y_i(\infty)|d_i = \infty] = E[y_i(\infty)|d_i = 1]$.

O grande desafio, portanto, é construir um grupo de controle que seja verdadeiramente comparável ao grupo de tratamento. Se desejamos estimar o **ATE**, é necessário que os dois grupos tenham a mesma distribuição de todas as características que influenciam o efeito do tratamento. Por outro lado, se o objetivo é estimar o **ATT**, basta que a alocação do tratamento seja estatisticamente independente do resultado potencial no cenário em que o indivíduo não é tratado.²

Tanto na estimativa do ATE quanto do ATT, garantir que os grupos sejam comparáveis em termos de características observáveis — como a proporção de homens e mulheres, jovens e idosos, pessoas altas e baixas — não costuma ser um grande desafio.

No entanto, nem todas as características relevantes podem ser observadas. Retomando o exemplo da vacina, o ideal seria que o grupo tratado e o grupo de controle tivessem proporções semelhantes de pessoas com alto e baixo risco de contágio. Contudo, o risco de contágio depende de diversos fatores não observáveis, como estilo de vida, aversão ao risco, círculo social, genética, sistema imunológico, entre outros.

Uma forma de garantir que os grupos tenham, em média, características semelhantes é aleatorizar a atribuição do tratamento. No exemplo da vacina, isso significa sortear aleatoriamente quem receberá a vacina (grupo de tratamento) e quem não receberá (grupo de controle). No entanto, por motivos éticos, legais ou práticos, muitas vezes essa aleatorização não é viável. Surge então a pergunta: como estimar de forma consistente o efeito da intervenção quando a alocação não é aleatória?

O exemplo a seguir foi criado com informações fictícias e nos acompanhará ao longo de todos os capítulos, inclusive nos exemplos práticos.

Exemplo - Programa de sustentabilidade

Um programa governamental selecionou 1.200 municípios, distribuídos em 12 estados, para participar de um estudo sobre o potencial da sustentabilidade em prédios novos. Em 2012, parte dessas localidades foi beneficiada por uma iniciativa de estímulo à construção e adaptação de edifícios sustentáveis. O objetivo foi fomentar práticas mais ecológicas e eficientes no setor da construção civil, contribuindo para a promoção de uma sociedade mais sustentável.

O programa, por meio de incentivos financeiros e apoio técnico, permitiu que esses municípios adotassem medidas para tornar seus edifícios mais amigáveis ao meio ambiente.

Se desejarmos avaliar se o programa reduz o consumo de energia elétrica, $y_i(1)$ representará a quantidade de energia elétrica consumida se o município i for submetido à intervenção, enquanto $y_i(\infty)$ será a quantidade de energia elétrica consumida se o município i não for submetido à intervenção. A variável $\delta_i = y_i(1) - y_i(\infty)$ representará a variação (redução ou aumento)

²Para mais detalhes sobre a diferença entre ATT e ATE ver MaCurdy et al. (2011).

nos gastos com energia do município i devido à intervenção. Repare que só observamos um dos dois resultados para um dado município. Ou seja, se o município foi tratado $y_i(\infty)$ não é observado. Se o município não foi tratado, $y_i(1)$ não é observado.

1.1.1 ATE vs ATT

A diferença entre ATE e ATT passa, principalmente, pela independência entre a atribuição de tratamento e o resultado potencial, ou seja, $y_i(1), y_i(\infty) \perp d$. Neste caso, é possível estimar o ATE mesmo que um grupo de comparação esteja sendo usado para estimar o resultado potencial do grupo tratado caso não tivesse sido tratado, i.e., $E[y_i(\infty)|d = 1]$. Isso é possível porque, se $y_i(1), y_i(\infty) \perp d$, então, $ATE = ATT$. Para ver isso, podemos reescrever δ_i como:

$$\delta_i = y_i(1) - y_i(\infty) = g_i(X_i) \quad (1)$$

onde X é um vetor que contém todas as características observáveis e *não observáveis* que influenciam o efeito do tratamento. Como exemplo de características observáveis, temos: idade, gênero, raça, anos de estudo, renda e localização geográfica. Já para características não observáveis, podemos citar: habilidade, inteligência, genética, habilidade socioemocional e rede de contatos.

Ao definirmos δ_i como na equação (1), deixamos explícito que o efeito do tratamento depende de características observáveis e não observáveis. Usando nosso exemplo do programa de sustentabilidade, a redução, ainda que per capita, pode ser maior em cidades maiores que têm maior proporção de pessoas em edifícios, por exemplo. Cidades que não têm nenhuma outra política (pública ou privada) para diminuição ou uso consciente da energia também impactam no efeito total, entre várias outras características.

Ou seja, podemos definir o ATE como:

$$\begin{aligned} ATE &= E[\delta_i] = E[y_i(1) - y_i(\infty)] \\ &= E[g_i(X_i)] = \int E[y_i(1) - y_i(\infty)|X]f(X)dX \end{aligned}$$

onde $f(X)$ é a densidade conjunta do vetor X (MaCurdy et al., 2011). Portanto, ATE é a média do efeito do tratamento ponderada pela probabilidade de todas as combinações possíveis de características X na população estudada.

Já o ATT representa o efeito do tratamento apenas entre os indivíduos tratados:

$$\begin{aligned} ATT &= E[\delta_i|d = 1] = E[y_i(1) - y_i(\infty)|d = 1] \\ ATT &= E[g(X_i)|d = 1] = \int E[y_i(1) - y_i(\infty)|X, d = 1]f(X|d = 1)dX \end{aligned}$$

onde $f(X|d = 1)$ é a distribuição de probabilidade conjunta das características X condicional a ter sido tratado MaCurdy et al. (2011).

Ou seja, se $y_i(1), y_i(\infty) \perp d$, então $E[g(X_i)] = E[g(X_i)|d]$ e, portanto, o ATT é igual ao ATE. Uma forma de garantir isso é aleatorizar a atribuição do tratamento. Entretanto, nem sempre é possível ou ético aleatorizar o tratamento, e assumimos uma hipótese um pouco menos restritiva, que seria que $y_i(\infty) \perp d|X$, isto é, o resultado potencial caso o indivíduo não seja tratado é independente da atribuição do tratamento condicional às covariadas (observáveis ou não).

Caso o resultado potencial não seja independente do tratamento, teremos que:

$$E[y_i(1) - y_i(\infty)|d = 1] \neq E[y_i(1) - y_i(\infty)|d = 0].$$

Isso acontece, primeiro, pois não podemos garantir que a distribuição das características dos tratados, $X|d = 1$, seja igual à dos controles $X|d = 0$ e, segundo, não podemos garantir que o resultado potencial de quem é tratado seja afetado da

mesma maneira, mesmo que as características sejam iguais, ou seja, $y_i(1) \perp d|X$. Portanto, só conseguiremos estimar o resultado do tratamento no subgrupo tratado.

Abaixo temos dois exemplos: um onde $y_i(1), y_i(\infty) \perp d$, i.e., $ATT = ATE$, e outro onde $y_i(1), y_i(\infty) \perp d|X$ e, nesse caso, $ATT \neq ATE$.

Exemplo Numérico com independência entre resultado potencial e a atribuição do tratamento

Vamos considerar um exemplo simples onde o programa estudado são aulas de carpintaria para pessoas de baixa renda afim de promover a entrada no mercado de trabalho. Vamos supor por simplicidade que a única característica, observável ou não, que afeta o resultado potencial é a habilidade, denotada $h \in [0, 1]$. Ou seja $X = (h)$. Assumiremos que $y_i(1), y_i(\infty) \perp d$, condição necessária para termos o ATE. Retornaremos a esse exemplo relaxando essa hipótese quando tratarmos do ATT. Defina o salário potencial y_i medidos em salários mínimos como:

$$\begin{aligned} y_i(1) &= 1.5 + 1.2h_i + \varepsilon_i^1 \\ y_i(\infty) &= 1 + 1h_i + \varepsilon_i^\infty \end{aligned}$$

onde $\varepsilon_i \sim N(0, 1)$ e $\varepsilon_i^1, \varepsilon_i^\infty \perp h$. Além disso, vamos definir a densidade de probabilidade de h como sendo independente entre os indivíduos i e dada por:

$$f(X) = f(h) = 3h^2.$$

Observe que $f(h|d) = f(h)$ então $ATT = E[y_1(1) - y_i(\infty)|d = 1] = E[y_1(1) - y_i(\infty)] = ATE$. Calculando o ATE=ATT:

$$\begin{aligned} g(h) &= E[y_i(1) - y_i(\infty)|h_i] = E[0.5 - 0.2h_i + \varepsilon_i^1 - \varepsilon_i^\infty|h_i] \\ &= 0.5 - 0.2h_i \end{aligned}$$

e

$$\begin{aligned} ATE &= E[y_i(1) - y_i(\infty)] \\ &= \int E[y_i(1) - y_i(\infty)|X]f(X)dX \\ &= \int_0^1 E[y_i(1) - y_i(\infty)|h_i = h]f(h)dh \\ &= \int_0^1 (0.5 - 0.2h)3h^2dh \\ &= \int_0^1 1.5h^2 - 0.6h^3dh \\ &= \left. \frac{1.5}{3}h^3 - \frac{0.6}{4}h^4 \right|_0^1 \\ &= \frac{1}{2} - 0.15 \\ &= 0.35 \end{aligned}$$

Exemplo Numérico com independência apenas condicional entre resultado potencial e a atribuição do tratamento

Considere o mesmo modelo do exemplo anterior. No entanto, vamos supor agora que o resultado potencial é independente da atribuição do tra-

tamento apenas condicional ao vetor de características X , i.e, $y_i(\infty) \perp d|X$. Como a única característica que temos no exemplo é a habilidade h , temos que $y_i(\infty) \perp d|h_i$. Vamos supor que $f(h|d=1) \neq f(h|d=0)$ onde:

$$f(h) = \begin{cases} 3h^2 & \text{se } d = 0 \\ 2h & \text{se } d = 1 \end{cases}$$

Ou seja,

$$\begin{aligned} ATT &= E[y_i(1) - y_i(\infty)|d = 1] \\ &= \int E[y_i(1) - y_i(\infty)|X, d = 1]f(X|d = 1)dX \\ &= \int_0^1 E[y_i(1) - y_i(\infty)|h_i = h, d = 1]f(h|d = 1)dh \\ &= \int_0^1 (0.5 - 0.2h)2h dh \\ &= \int_0^1 h - 0.4h^2 dh \\ &= \left. \frac{h^2}{2} - \frac{0.4}{3}h^3 \right|_0^1 \\ &= \frac{1}{2} - 0.33333 \\ &= 0.26777 \end{aligned}$$

Agora, se quisermos calcular o ATE:

$$\begin{aligned} ATE &= E[y_i(1) - y_i(\infty)] \\ &= \int E[y_i(1) - y_i(\infty)|X, d]f(X, d)dx \\ &= \int_0^1 E[y_i(1) - y_i(\infty)|h_i = h, d = 0]f(h, d = 0) \\ &\quad + E[y_i(1) - y_i(\infty)|h_i = h, d = 1]f(h, d = 1)dh \\ &= \int_0^1 E[y_i(1) - y_i(\infty)|h_i = h, d = 0]f(h|d = 0)P(d = 0) + E[y_i(1) \\ &\quad - y_i(\infty)|h_i = h, d = 1]f(h|d = 1)P(d = 1)dh \end{aligned}$$

como $E[y_i(1) - y_i(\infty)|h_i = h, d = 0] = E[y_i(1) - y_i(\infty)|h_i = h, d = 1] = g(h) = 0.5 - 0.2h_i$,

$$\begin{aligned} ATE &= \int_0^1 \{f(h|d = 0)P(d = 0) + f(h|d = 1)P(d = 1)\}(0.5 - 0.2h_i)dh \\ &= \int_0^1 \{3h^2P(d = 0) + 2hP(d = 1)\}(0.5 - 0.2h)dh \\ &= \underbrace{\int_0^1 3h^2(0.5 - 0.2h) dh}_{0.35(ex1)} + \underbrace{P(d = 1) \int_0^1 2h(0.5 - 0.2h) dh}_{=ATT} \\ &= P(d = 0)0.35 + P(d = 1)ATT \\ &= P(d = 0)0.35 + P(d = 1)0.26777 \end{aligned}$$

$\int_0^1 3h^2(0.5 - 0.2h)$ é exatamente igual ao ATE calculado no exemplo 1, por isso substituímos diretamente por 0.35. Já $\int_0^1 2h(0.5 - 0.2h) = ATT$ calculado anteriormente. Veja que, nesse caso, $ATE \neq ATT$.

1.2 DID como diferenças de médias

Geralmente, é desafiador obter grupos de tratamento e controle suficientemente comparáveis na ausência de aleatorização na atribuição do tratamento. O modelo de diferença-em-diferenças (DiD) surge como uma abordagem que exige uma hipótese menos restritiva nesse sentido.

Nesse procedimento, avaliamos os efeitos da política antes e depois da intervenção. Para começar, vamos relembrar o caso mais simples, em que temos apenas dois períodos: um antes e um depois da intervenção. Denotaremos o período posterior à intervenção como $t=1$, e o período imediatamente anterior como $t=0$.

Agora, além dos resultados potenciais, é necessário indicar o período ao qual o resultado se refere. Por exemplo, o resultado potencial caso o indivíduo i seja tratado no período $t = 1$ será denotado por $y_{i1}(1)$. Ou seja:

- $y_{it}(1)$ representa o resultado potencial no período t caso o indivíduo i tenha sido exposto à intervenção, no período.
- $y_{it}(\infty)$ representa o resultado potencial no período t caso o indivíduo i **não** tenha sido exposto à intervenção.

Ainda estamos interessados em estimar o ATT; no entanto, não é necessário que $E[y_{i1}(\infty)|d = 1] = E[y_{i1}(\infty)|d = \infty]$. Isso porque, em vez de apenas calcular a diferença entre o grupo de tratamento e controle, realizaremos duas comparações: a diferença ao longo do tempo dentro de cada grupo e a diferença entre as diferenças entre o grupo de tratamento e controle.

$$\begin{aligned}\Delta_{Tr} &= E[y_{i1}(1)|d_i = 1] - E[y_{i0}(\infty)|d_i = 1] \\ \Delta_{Cr} &= E[y_{i1}(\infty)|d_i = \infty] - E[y_{i0}(\infty)|d_i = \infty]\end{aligned}$$

$$\begin{aligned}\gamma^{DID} &= \Delta_{Tr} - \Delta_{Cr} \\ &= (E[y_{i1}(1)|d_i = 1] - E[y_{i0}(\infty)|d_i = 1]) \\ &\quad - (E[y_{i1}(\infty)|d_i = \infty] - E[y_{i0}(\infty)|d_i = \infty])\end{aligned}$$

Somando e subtraindo $E[y_{i1}(\infty)|d_i = 1]$:

$$\begin{aligned}\gamma^{DID} &= \underbrace{E[y_{i1}(1) - y_{i1}(\infty)|d_i = 1]}_{ATT} \\ &\quad + E[y_{i1}(\infty) - y_{i0}(\infty)|d_i = 1] - E[y_{i1}(\infty) - y_{i0}(\infty)|d_i = \infty]\end{aligned}\tag{2}$$

Para que essas diferenças realmente representem o ATT, precisamos assumir que três condições, consideradas as principais hipóteses do modelo de diferença em diferenças, sejam satisfeitas:

Hipóteses

1. **SUTVA:** O resultado potencial no caso de não tratamento não é afetado pela alocação do tratamento sobre outros indivíduos.
2. **Não antecipação:** $\forall i, y_{i0} = y_{i0}(\infty)$.
3. **Tendências Paralelas:**

$$E[y_{i1}(\infty) - y_{i0}(\infty)|d_i = 1] = E[y_{i1}(\infty) - y_{i0}(\infty)|d_i = \infty]$$

As hipóteses 1 e 2 são necessárias para derivarmos a equação (2), enquanto a terceira hipótese implica que os últimos termos da equação (2) se anulam. Ou seja, sob a hipótese padrão de amostra aleatória e as hipóteses 1 a 3 temos que $\gamma^{DID} = ATT$.

A hipótese 3, Tendências Paralelas, estabelece que a variação nos resultados potenciais, na ausência de tratamento, para o grupo tratado deve ser igual à variação nos resultados potenciais, também na ausência de tratamento, para o grupo de controle. Em outras palavras, assume-se que, em média, os resultados potenciais evoluam de forma paralela ao longo do tempo nos dois grupos — o que dá nome à hipótese.

Nesse sentido, um estimador simples para o DID será:

$$\hat{\gamma}^{DID} = (\bar{y}_{11} - \bar{y}_{10}) - (\bar{y}_{\infty 1} - \bar{y}_{\infty 0}) \quad (3)$$

onde $\bar{y}_{gt} = \frac{1}{N_g} \sum_{i=1}^N y_{it} \mathbb{1}_{\{d=g\}}$, N_g é o número de indivíduos no grupo $g \in \{\infty, 1\}$ (Angrist and Pischke, 2009; Callaway and Sant'Anna, 2021).

Como mostrado por Angrist and Pischke (2009), podemos estimar o ATT acima utilizando uma regressão linear interagindo Dummies de tratamento e período:

$$y_{it} = \alpha + \theta d_i + \lambda T_t + \gamma(d_i * T_t) + \varepsilon_{it}, \quad (4)$$

onde $T_t = \begin{cases} 1 & t = 1 \\ 0 & t = 0 \end{cases}$, $d_i = \begin{cases} 1 & i \in \mathcal{D} \\ 0 & c.c. \end{cases}$ e \mathcal{D} representa o conjunto de todos os indivíduos tratados.

Portanto, podemos derivar que:

$$\begin{aligned} \alpha &= E[y_{it} \mid i \notin \mathcal{D}, t = 0] \\ \theta &= E[y_{it} \mid i \in \mathcal{D}, t = 0] - E[y_{it} \mid i \notin \mathcal{D}, t = 0] \\ \lambda &= E[y_{it} \mid i \notin \mathcal{D}, t = 1] - E[y_{it} \mid i \notin \mathcal{D}, t = 0] \\ \gamma &= (E[y_{it} \mid i \in \mathcal{D}, t = 1] - E[y_{it} \mid i \in \mathcal{D}, t = 0]) \\ &\quad - (E[y_{it} \mid i \notin \mathcal{D}, t = 1] - E[y_{it} \mid i \notin \mathcal{D}, t = 0]). \end{aligned}$$

Dado as hipóteses 1 e 2, temos que:

$$\begin{aligned} \gamma &= (E[y_{it} \mid i \in \mathcal{D}, t = 1] - E[y_{it} \mid i \in \mathcal{D}, t = 0]) \\ &\quad - (E[y_{it} \mid i \notin \mathcal{D}, t = 1] - E[y_{it} \mid i \notin \mathcal{D}, t = 0]) \\ &= (E[y_{i1}(1) \mid d = 1] - E[y_{i0}(\infty) \mid d = 1]) \\ &\quad - (E[y_{i1}(\infty) \mid d = \infty] - E[y_{i0}(\infty) \mid d = \infty]) \end{aligned}$$

Ou seja, o $\hat{\gamma}$ estimado pela equação (4) é numericamente equivalente ao $\hat{\gamma}^{DID}$ (equação (3)). Se a hipótese de tendência paralela for válida, ambas as estimativas serão consistentes para o ATT.

Além disso, conforme argumentado por Angrist and Pischke (2009), é relativamente simples estimar o DID condicional a um vetor de características X_{it} , por meio da regressão:

$$y_{it} = \alpha + \theta d_i + \lambda T_t + \gamma(d_i * T_t) + X'_{it}\beta + \varepsilon_{it} \quad (5)$$

e, neste caso, podemos mostrar que:

$$\begin{aligned} \gamma &= (E[y_{it} | i \in \mathcal{D}, t = 1, X_{it}] - E[y_{it} | i \in \mathcal{D}, t = 0, X_{it}]) \\ &\quad - (E[y_{it} | i \notin \mathcal{D}, t = 1, X_{it}] - E[y_{it} | i \notin \mathcal{D}, t = 0, X_{it}]) \\ &= (E[y_{i1}(1) | d = 1, X_{it}] - E[y_{i0}(\infty) | d = 1, X_{it}]) \\ &\quad - (E[y_{i1}(\infty) | d = \infty, X_{it}] - E[y_{i0}(\infty) | d = \infty, X_{it}]) \end{aligned}$$

Com a hipótese de tendências paralelas condicional sendo válida:

$$\gamma = E[y_{i1}(1) - y_{i1}(\infty) | d_i = 1, X_{it}] = \text{ATT condicional às covariadas } X.$$

1.3 DID vs TWFE

O modelo de **efeitos fixos em dois sentidos** (*two-way fixed effects* — TWFE) incorpora efeitos fixos tanto de tempo quanto de indivíduo. Ou seja:

$$y_{it} = \alpha_i + \lambda_t + \underbrace{\gamma^{TWFE} D_{it}}_{d_i * T_t} + \varepsilon_{it} \quad (6)$$

ou equivalentemente

$$\ddot{y}_{it} = \gamma^{TWFE} \ddot{D}_{it} + \ddot{\varepsilon}_{it}$$

onde

$$\begin{aligned} \ddot{h}_{it} &= h_{it} - \bar{h}_{i.} - \bar{h}_{.t} + \bar{h} \\ \bar{h}_{i.} &= \frac{1}{T} \sum_{t=1}^T h_{it}, \\ \bar{h}_{.t} &= \frac{1}{N} \sum_{i=1}^N h_{it} \text{ e} \\ \bar{h} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T h_{it}, \end{aligned}$$

para h sendo y_{it} , D_{it} ou ε_{it} .

Em geral, não estimamos todos os efeitos fixos dos indivíduos (α_i) mas apenas os efeitos fixos de cada grupo. Estimamos, portanto, a regressão:

$$y_{igt} = \alpha_g + \lambda_t + \underbrace{\gamma^{TWFE} D_{gt}}_{d_i * T_t} + \varepsilon_{it} \quad (7)$$

para $g \in \{\infty, 1\}$ e $t \in \{0, 1\}$. Calculando a expectativa de y_{igt} em cada um dos cenários possíveis — tratados no período pré-tratamento, $d_i = 1$ e $t = 0$; tratados no período pós-tratamento, $d_i = 1$ e $t = 1$; controle no período pré-tratamento, $d_i = \infty$ e $t = 0$; e controle no período pós-tratamento, $d_i = \infty$ e $t = 1$ — temos as seguintes relações:

$$\begin{aligned} E[y_{igt} | d_i = 1, t = 0] &= \alpha_1 + \lambda_0 \\ E[y_{igt} | d_i = 1, t = 1] &= \alpha_1 + \lambda_1 + \gamma^{TWFE} \\ E[y_{igt} | d_i = \infty, t = 0] &= \alpha_\infty + \lambda_0 \\ E[y_{igt} | d_i = \infty, t = 1] &= \alpha_\infty + \lambda_1 \end{aligned}$$

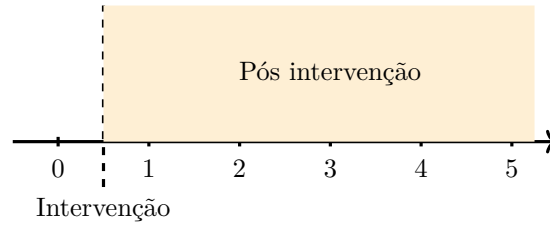
Assumindo as hipóteses 1-3:

$$\begin{aligned}
ATT &= (E[y_{igt}|d_i = 1, t = 1] - E[y_{igt}|d_i = 1, t = 0]) - (E[y_{igt}|d_i = \infty, t = 1] - E[y_{igt}|d_i = \infty, t = 0]) \\
&= (\alpha_1 + \lambda_1 + \gamma^{TWFE} - \alpha_1 - \lambda_0) - (\alpha_\infty + \lambda_1 - \alpha_\infty - \lambda_0) \\
&= \gamma^{TWFE}
\end{aligned}$$

Dessa forma, dadas as hipóteses de 1 a 3, o parâmetro γ^{TWFE} nas equações (6) e (7) equivale ao ATT. Além disso, como na equação (5), podemos estimar o ATT condicional a um vetor de características, X_{it} , utilizando o seguinte modelo: $y_{it} = \alpha_i + \lambda_t + \gamma D_{it} + \beta X_{it} + \varepsilon_{it}$ (Angrist and Pischke, 2009).

1.4 DID com múltiplos períodos

Consideramos agora o caso em que ainda temos apenas dois grupo de indivíduos, tratados e controle, mas adicionaremos múltiplos períodos após a intervenção. Vamos supor que todos os indivíduos do grupo de tratamento sejam expostos à intervenção simultaneamente no período $t = 1$ e permaneçam tratados depois disso. No final do capítulo, discutiremos as implicações de termos múltiplos períodos pré-tratamento.



Dessa forma, para cada período $t \geq 1$, é possível que o efeito da exposição ao tratamento seja diferente. Pode ocorrer, por exemplo, que o efeito do tratamento seja crescente na quantidade de exposição, resultando em um aumento do efeito do tratamento com o passar dos períodos. Também é possível que o efeito da intervenção seja constante, independentemente da duração da exposição. Ou ainda, o tratamento pode gerar um efeito positivo no curto prazo e nulo no longo prazo.

Portanto, é necessário definirmos o efeito da intervenção em cada período do tempo, dado por $ATT(t)$.

$$ATT(t) = E[y_{it}(1) - y_{it}(\infty)|d = 1].$$

Por exemplo, no primeiro período o $ATT(t = 1)$ será igual a $E[y_{i1}(1) - y_{i1}(\infty)|d = 1]$.

Como não observamos simultaneamente $y_{it}(1)$ e $y_{it}(\infty)$, precisamos seguir o mesmo procedimento do DID padrão, mas agora para cada período t . Ou seja, faremos as duas diferenças para cada período t .

Contudo, é fundamental ter cautela. Se simplesmente tomarmos a diferença entre os períodos t e $t - 1$, estaremos comparando o resultado potencial sob tratamento de indivíduos que foram tratados nos dois períodos t e $t - 1$. Isso não é equivalente à diferença na média dos resultados do grupo de controle em t e $t - 1$, mesmo se assumirmos tendências paralelas.

Para ilustrar esse ponto, considere o caso em que $t = 3$. Se escrevermos a diferença das diferenças, teremos:

$$\begin{aligned}
\tau(3) &= E[y_{i3} - y_{i2}|d = 1] - E[y_{i3} - y_{i2}|d = \infty] \\
&= E[y_{i3}(1) - y_{i2}(1)|d = 1] - E[y_{i3}(\infty) - y_{i2}(\infty)|d = \infty]
\end{aligned}$$

Isso ocorre porque, dado que $d = 1$ tanto em $t = 2$ quanto em $t = 3$, o indivíduo foi exposto à intervenção em ambos os períodos. Assim, o resultado potencial dele será o resultado caso ele tenha sido tratado. Portanto, somando e subtraindo $E[y_{i3}(\infty)|d = 1]$:

$$\begin{aligned}\tau(3) &= \underbrace{E[y_{i3}(1) - y_{i3}(\infty)|d=1]}_{ATT(3)} \\ &\quad + E[y_{i3}(\infty) - \textcolor{red}{y}_{i2}(1)|d=1] - E[y_{i3}(\infty) - y_{i2}(\infty)|d=\infty]\end{aligned}$$

Note que, mesmo que haja tendências paralelas ($E[y_{i3}(\infty) - y_{i2}(\infty)|d=1] = E[y_{i3}(\infty) - y_{i2}(\infty)|d=\infty]$), os dois últimos termos da expressão anterior não se anulam. Isso ocorre porque é possível que $y_{i2}(1) \neq y_{i2}(\infty)$. No caso do DID 2x2, a hipótese de não antecipação somada à de tendências paralelas garantia a anulação dos últimos termos pois como só temos 2 períodos sempre estaríamos fazendo a diferença em relação ao período antes de tratamento. No entanto, para que o $\tau(3)$ seja igual ao $ATT(3)$, precisaríamos assumir que $y_{i2}(1) = y_{i2}(\infty)$, que é o mesmo que assumir que o tratamento não tem efeito no período 2!

O grande problema aqui advém do fato de estarmos tirando a diferença em dois períodos nos quais o grupo de tratamento já foi tratado. No nosso exemplo, o grupo de tratamento já foi tratado nos períodos $t = 3$ e $t - 1 = 2$. Para resolver isso, vamos sempre calcular a diferença em relação ao período pré-tratamento, independentemente do período t . Em outras palavras, faremos sempre a média dos grupos em t menos a média dos grupos em $t = 0$.

$$\begin{aligned}\gamma^{DID}(t) &= E[y_{it} - y_{i0}|d=1] - E[y_{it} - y_{i0}|d=\infty] \\ &= \underbrace{E[y_{it}(1) - y_{it}(\infty)|d=1]}_{ATT(t)} \\ &\quad - E[y_{it}(\infty) - y_{i0}(\infty)|d=1] - E[y_{it}(\infty) - y_{i0}(\infty)|d=\infty]\end{aligned}$$

Se a hipótese de tendências paralelas ($E[y_{it}(\infty) - y_{i0}(\infty)|d=1] = E[y_{it}(\infty) - y_{i0}(\infty)|d=\infty]$) for satisfeita, teremos que $\gamma^{DID}(t) = ATT(t)$. Abaixo, listamos novamente as hipóteses do DID, agora para múltiplos períodos:

Hipóteses

1. **SUTVA:** O resultado potencial no caso de não tratamento não é afetado pela alocação do tratamento sobre outros indivíduos.
2. **Não antecipação:** $\forall i, y_{i0} = y_{i0}(\infty)$.
3. **Irreversibilidade do Tratamento:** $\forall t \geq 1, \forall i \in \mathcal{D}, y_{it} = y_{it}(1)$.
4. **Tendências Paralelas:**

$$\forall t, E[y_{it}(\infty) - y_{i0}(\infty)|d_i = 1] = E[y_{it}(\infty) - y_{i0}(\infty)|d_i = \infty]$$

E, assumindo que as hipóteses 1-4 são satisfeitas, o estimador abaixo é um estimador consistente para o $ATT(t)$:

$$\hat{\gamma}_t^{DID} = (\bar{y}_{1t} - \bar{y}_{10}) - (\bar{y}_{\infty t} - \bar{y}_{\infty 0}) \quad (8)$$

onde $\bar{y}_{gt} = \frac{1}{N_g} \sum_{i=1}^N y_{it} \mathbb{1}_{\{d_i=g\}}$.

Entretanto, diferentemente do caso simples onde temos apenas 2 grupos e 2 períodos (DID 2x2), aqui não é tão simples responder quais são os parâmetros de interesse. Por exemplo, estamos interessados em saber o valor de cada $ATT(t)$? Ou queremos calcular uma média simples dos $ATT(t)$? Ainda, seria mais apropriado considerar uma média ponderada dos $ATT(t)$? E, nesse caso, quais seriam os pesos adequados e por quê?

Além disso, surge uma questão importante: **qual parâmetro, de fato, pode ser estimado pelo modelo de TWFE?** Na próxima subseção, discutiremos o que o modelo de efeitos fixos em dois sentidos (TWFE) realmente nos retorna.

Estimando os $ATT(t)$ via diferença de médias

No cenário apresentado no exemplo do programa de sustentabilidade da seção 1, temos um programa governamental destinado a 1.200 municípios distribuídos por 12 estados. Esses municípios foram selecionados para participar de um estudo sobre o potencial da sustentabilidade em prédios novos. Em 2012, uma parte dessas localidades foi contemplada por um programa governamental que visava incentivar a construção e adaptação de edifícios sustentáveis. O objetivo era promover práticas mais ecológicas e eficientes no setor de construção, contribuindo para uma sociedade mais sustentável.

Por meio de incentivos financeiros e apoio técnico, o programa possibilitou que esses municípios adotassem medidas para tornar seus edifícios mais sustentáveis e ambientalmente responsáveis.

Os dados relativos a esse programa estão disponíveis na base "base_MULT_did.csv". Nessa base, a variável d indica se o município é tratado ou controle, enquanto a variável y representa a variação percentual na quantidade de energia elétrica utilizada durante o ano no município de 2011 até 2020.

```
library(dplyr)
library(stargazer)

df <- read.csv("base_MULT_did.csv")
```

Apos abrir a base, vamos calcular a média de cada grupo para cada período do tempo, i.e., $\bar{y}_{\infty t}$ e \bar{y}_{1t}

```
Bary_inf = c()
Bary_1 = c()
for(t in 2011:2020){

  Bary_inf[as.character(t)] = mean(df$y[df$d==0 & df$ano == t])
  Bary_1[as.character(t)] = mean(df$y[df$d==1 & df$ano == t])

}

round(Bary_inf, 3)
```

```
## 2011 2012 2013 2014 2015 2016 2017
## 9.937 8.792 7.442 6.226 5.014 3.702 2.419
## 2018 2019 2020
## 1.178 0.045 -1.240
```

```
round(Bary_1, 3)
```

```
## 2011 2012 2013 2014 2015 2016 2017
## 16.185 14.480 12.951 11.481 10.004 8.511 6.928
## 2018 2019 2020
## 5.520 3.976 2.559
```

E vamos calcular $\hat{ATT}(t)$. Pela formula abaixo:

$$\hat{ATT}(t) = (\bar{y}_{1t} - \bar{y}_{10}) - (\bar{y}_{\infty t} - \bar{y}_{\infty 0})$$


```

ATT= c()
for( t in 2012:2020){
  t = as.character(t)
  ATT[t] = (Bary_1[t] - Bary_1["2011"])
           - (Bary_inf[t] - Bary_inf["2011"])
}

round(ATT, 3)

```

```

## 2012 2013 2014 2015 2016 2017
## -0.561 -0.740 -0.993 -1.258 -1.439 -1.740
## 2018 2019 2020
## -1.907 -2.318 -2.450

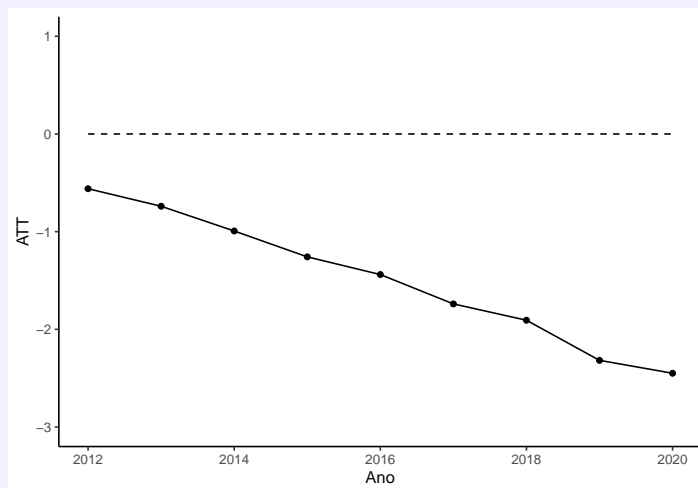
```

Agora vamos plotar o efeito da intervenção ao longo dos anos:

```

library(ggplot2)
ggplot(data = data.frame(x=2012:2020, y=ATT),
aes(x=x, y=y)) +
geom_line() + geom_point() +
  geom_line(y = 0, linetype = "dashed") +
  ylim(-3, 1) + ylab("ATT") + xlab("Ano") +
  theme_classic()

```



Ao plotar o gráfico, torna-se evidente como a intervenção tem um efeito negativo e crescente, em valor absoluto, ao longo do tempo, resultando em um efeito de -2.45% no 8º ano pós intervenção. Se desejarmos calcular a média no período, podemos fazê-lo somando os efeitos médios estimados para cada ano:

$$\gamma = \frac{1}{T} \sum_{t=1}^T \hat{ATT}(t)$$

```
mean(ATT)
```

```
## [1] -1.489463
```

Portanto, o efeito médio ao longo dos 8 anos foi de 1.489%.

1.4.1 Regressão DID com múltiplos períodos pós-tratamento

O que acontece se rodarmos a regressão padrão do DID representada pela equação (4) no caso de múltiplos períodos de tratamento? Suponha, portanto, que estamos estimando a regressão abaixo:

$$y_{it} = \alpha + \theta d_i + \lambda T_t + \gamma(d_i * T_t) + \varepsilon_{it} \quad (9)$$

onde $T_t = \begin{cases} 1 & t > 0 \\ 0 & t = 0 \end{cases}$, $d_i = \begin{cases} 1 & i \in \mathcal{D} \\ 0 & c.c. \end{cases}$ e \mathcal{D} representa o conjunto de todos os indivíduos tratados.

Como derivado para o DID 2x2, podemos escrever γ como:

$$\begin{aligned} \gamma = & E[y_{it} \mid d = 1, t > 0] - E[y_{it} \mid d = 1, t = 0] \\ & - (E[y_{it} \mid d = \infty, t > 0] - E[y_{it} \mid d = \infty, t = 0]) \end{aligned}$$

E pela Lei das Expectativas Iteradas:

$$\begin{aligned} E[y_{it} \mid d = 1, t > 0] &= E_t[y_{it} \mid d = 1, t = t] \\ &= \sum_{\tau=1}^T P(t = \tau) E[y_{i\tau} \mid d = 1] \\ &= \sum_{\tau=1}^T P(t = \tau) E[y_{i\tau}(1) \mid d = 1] \end{aligned}$$

com $P(t=0)=0$ pois estamos tirando a esperança condicional de $t > 0$.³

Vamos assumir inicialmente que temos um painel balanceado. Portanto, $P(t = \tau) = 1/T$. Dessa forma, podemos escrever:

$$\begin{aligned} E[y_{it} \mid d = 1, t > 0] &= \frac{1}{T} \sum_{\tau=1}^T E[y_{i\tau} \mid d = 1] \\ E[y_{it} \mid d = 1, t = 0] &= E[y_{i0} \mid d = 1] \\ E[y_{it} \mid d = 0, t > 0] &= \frac{1}{T} \sum_{\tau=1}^T E[y_{i\tau} \mid d = 0] \\ E[y_{it} \mid d = \infty, t = 0] &= E[y_{i0} \mid d = \infty] \end{aligned}$$

e γ será uma função dos $ATT(t)$:

$$\begin{aligned} \gamma &= E[y_{it} \mid d = 1, t > 0] - E[y_{it} \mid d = 1, t = 0] - (E[y_{it} \mid d = \infty, t > 0] - E[y_{it} \mid d = \infty, t = 0]) \\ &= \frac{1}{T} \sum_{t=1}^T E[y_{it} \mid d = 1] - E[y_{i0} \mid d = 1] - \left(\frac{1}{T} \sum_{t=1}^T E[y_{it} \mid d = \infty] - E[y_{i0} \mid d = \infty] \right) \\ &= \frac{\sum_{t=1}^T \{E[y_{it} \mid d = 1] - E[y_{i0} \mid d = 1]\}}{T} - \left(\frac{\sum_{t=1}^T \{E[y_{it} \mid d = \infty] - E[y_{i0} \mid d = \infty]\}}{T} \right) \\ &= \frac{1}{T} \sum_{t=1}^T E[y_{it} - y_{i0} \mid d = 1] - E[y_{it} - y_{i0} \mid d = \infty] = \frac{1}{T} \sum_{t=1}^T \gamma_t^{DID} \end{aligned}$$

Se estivermos lidando com um painel balanceado, o parâmetro γ da equação (9) representará a média simples dos efeitos médios do tratamento em cada período de tempo. Agora, o que acontecerá se não tivermos um painel balanceado? Nesse caso, a probabilidade de o indivíduo i aparecer no tempo t não será constante (e igual a $1/T$).

³ $E[y_{it}, d = 1, t = t] = E[y_{it}, d = 1]$ é um abuso de notação, no fundo estamos definindo que $E[y_{it}] = E[y_i | t = t]$.

Como consequência, a regressão estimará uma média ponderada dos ATT(t), na qual os pesos dependerão da quantidade de indivíduos observados naquele ano (N_{gt}) em relação ao total de indivíduos-ano ($NT = \sum_{g \in G} \sum_{t=1}^T N_{gt}$). Você pode demonstrar esse resultado como exercício.

Exemplo - Programa de sustentabilidade

Estimando o efeito tratamento com o modelo de DID via MQO

Voltamos ao exemplo do programa de sustentabilidade cujos dados estão disponíveis na base de dados "base_MULT_did.csv". A variável d representa se o município é tratado ou controle e y representa a variação, em porcentagem, da quantidade de energia elétrica consumida durante o ano no município.

Assumiremos que a hipótese de tendências paralelas incondicional é válida (sem a necessidade de nenhum controle). Agora, vamos estimar γ da equação (9). Para isso, vamos criar nossa variável *dummy*:

$$T_t = \begin{cases} 1 & \text{ano} \geq 2012 \\ 0 & \text{c.c} \end{cases}.$$

```
df <- df %>% mutate(Tt = ifelse(ano>=2012, 1, 0))
```

Agora, vamos estimar via MQO a equação abaixo:

$$y_{it} = \alpha + \theta d_i + \lambda T_t + \gamma(d_i * T_t) + \varepsilon_{it}$$

```
Reg_did <- lm(y ~ d*Tt, data = df)
```

E, finalmente, calcular os desvios-padrão robustos à heteroscedasticidade:

```
# Load libraries
library("lmtest")
library("sandwich")

# Robust t test
cov1 <- sandwich::vcovHC(Reg_did, type = "HCO")
sd = sqrt(diag(cov1))
stargazer(Reg_did, type = "latex", se = list(sd),
           title = "MQO com Erros Padrão Robustos")
```

<i>Dependent variable:</i>	
	y
d	6.249*** (0.075)
Tt	-6.206*** (0.080)
d:Tt	-1.489*** (0.106)
Constant	9.937*** (0.059)
Observations	12,000
R ²	0.424
Adjusted R ²	0.424
Residual Std. Error	3.750 (df = 11996)
F Statistic	2,949.144*** (df = 3; 11996)
Note:	*p<0.1; **p<0.05; ***p<0.01

Ou seja:

$$\begin{aligned}\alpha &= 9.937 & \theta &= 6.249 \\ \lambda &= -6.206 & \gamma &= -1.489\end{aligned}$$

A intervenção teve um efeito médio ao longo dos anos de 2012 a 2020 de -1.489%, i.e., $\gamma = \frac{1}{T} \sum_{t=2012}^{2020} ATT(t) = -1.489$. Ou seja, o programa reduziu o consumo de energia elétrica em aproximadamente 1.49%, em média, nas cidades expostas à intervenção.

Vale ressaltar que podemos obter o mesmo resultado da seção anterior quando calculamos a média simples de todos os $ATT(t)$. Isto ocorre exatamente porque estamos lidando com um painel balanceado e porque estamos assumindo que não é necessário controlar por nenhuma variável adicional. Entretanto, como argumenta Angrist and Pischke (2009), poderíamos facilmente estimar a regressão com variáveis de controle, se necessário. Além disso, é possível calcular os erros-padrão robustos à heteroscedasticidade e ajustados por clusters, a fim de realizar os testes de hipóteses corretamente.

Em um modelo de DID com múltiplos períodos pós-tratamento e **todos os indivíduos do grupo de tratamento sendo tratados ao mesmo tempo**, a regressão nos retornará o efeito médio do tratamento ao longo dos anos pós-tratamento. Se tivermos diante de um painel balanceado, esse efeito corresponderá a uma média simples dos $ATT(t)$. Em painéis **desbalanceados**, o resultado será uma **média ponderada**, cujos pesos refletem o tamanho relativo de cada grupo em cada período. Em ambos os casos — diferentemente do DID com tratamento escalonado (que será estudado nas seções seguintes) —, não temos o problema de estimar consistentemente uma média dos $ATT(t)$ informativa.

No entanto, mesmo estimando consistentemente essa média dos efeitos ao longo do tempo, surge a questão: isso é realmente o que desejamos estimar? Embora a abordagem mais óbvia seja calcular a média simples dos efeitos em cada período de tempo, essa medida pode "esconder" informações relevantes para a análise da intervenção, já que ela não nos revela informações sobre a dinâmica do efeito do tratamento ao longo do tempo.

A Figura 1 representa três exemplos de dinâmicas para o $ATT(t)$ com a mesma

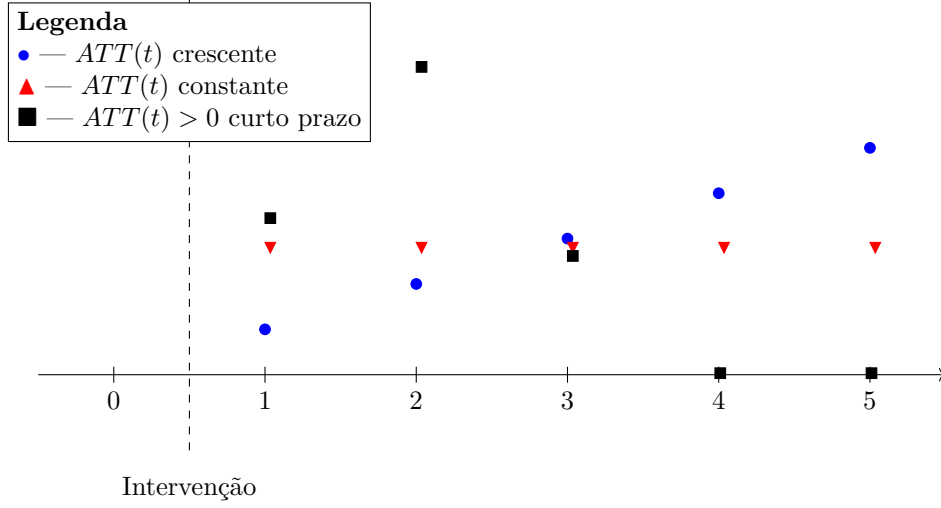


Figura 1: $ATT(t)$

média. A primeira, em azul, mostra um efeito crescente ao longo do tempo de exposição ao tratamento, enquanto a segunda (em vermelho) exibe um efeito constante. A última (em preto) ilustra um efeito significativo no curto prazo, mas nenhum efeito no longo prazo. Ao estimarmos γ para cada uma dessas intervenções por MQO, obteríamos a mesma estimativa para cada caso. Para contornarmos o problema, podemos estimar de forma simples os $ATT(t)$ interagindo nossa *dummy* de tratamento com uma *dummy* para cada ano.

1.4.2 Regressão com Dummies de tempo

Para conseguirmos estimar cada um dos $ATT(t)$, que serão iguais a cada um dos γ_t na equação (10), definiremos um conjunto de *dummies* de tempo:

$$\mathbb{1}_\tau = \begin{cases} 1 & t = \tau \\ 0 & c.c. \end{cases}$$

Considere o exemplo de uma intervenção feita no ano de 2012. Nesse caso, $t = 0$ seria o ano de 2011 e, em 2011, $\mathbb{1}_0 = 1$ e $\mathbb{1}_1 = \mathbb{1}_2 = \dots \mathbb{1}_T = 0$. Temos, então:

$$y_{igt} = \alpha_g + \theta_t + \sum_{t=t_0+1}^T \gamma_t \mathbb{1}_\tau d * T_t + \varepsilon_{igt} \quad (10)$$

onde $g \in \{\infty, 1\}$ determina se o indivíduo é do grupo de controle (i.e., $g = \infty$) ou do grupo de tratamento (i.e., $g = 1$). Dessa forma, como mostrado por Angrist and Pischke (2009), se as hipóteses 1-4 da seção anterior forem satisfeitas:

$$\gamma_t = ATT(t)$$

e, do mesmo modo que na regressão sem dummies, conseguimos facilmente calcular os erros-padrão robustos à heterocedasticidade e clusters.

Isso pode ser demonstrado pelo mesmo procedimento da seção anterior. Taremos:

$$\begin{aligned} \gamma_t &= E[y_{it} - y_{i0} | d = 1] - E[y_{it} - y_{i0} | d = \infty] \\ &= \underbrace{E[y_{it}(1) - y_{it}(\infty) | d = 1]}_{ATT(t)} \\ &\quad - (E[y_{it}(\infty) - y_{i0}(\infty) | d = 1] - E[y_{it}(\infty) - y_{i0}(\infty) | d = \infty]) \end{aligned}$$

onde, sob tendências paralelas, $E[y_{it}(\infty) - y_{i0}(\infty) | d = 1] - E[y_{it}(\infty) - y_{i0}(\infty) | d = \infty] = 0$.

Além disso, como estimamos consistentemente cada $ATT(t)$, podemos calcular qualquer média ponderada dos $ATT(t)$ que seja interessante para a análise da intervenção. Isto é:

$$\Gamma = \sum_{t>0} w(t)ATT(t)$$

onde $w(t)$ é um peso arbitrariamente escolhido tal que $\sum_{t>0} w(t) = 1$ e $w(t) \geq 0$.

Entretanto, qual média ponderada seria interessante? Como argumentado por Callaway and Sant'Anna (2021), isso depende muito das características específicas da intervenção avaliada e dos objetivos da pesquisa. Por exemplo:

- Se quisermos avaliar o efeito médio ao longo de todos os anos da análise, poderíamos determinar um peso igual para cada ano da amostra, isto é, $w(t) = 1/T$ (exatamente como na regressão sem dummies).
- Se quisermos ver o efeito acumulado durante todos os anos da intervenção, definiríamos: $w(t) = 1$ para todo t .
- Agora, se estivermos interessados em avaliar o efeito do tratamento no curto prazo, poderíamos definir que o curto prazo seria até 4 anos após a intervenção e $w(t) = \begin{cases} 1/4 & t \leq 4 \\ 0 & t > 4 \end{cases}$.
- Ou, ainda, se quisermos avaliar se o efeito de curto prazo (≤ 4 anos) é maior que o efeito de longo prazo (> 4 anos), definiríamos $w(t) = \begin{cases} -\frac{1}{4} & t \leq 4 \\ \frac{1}{T-4} & t > 4 \end{cases}$.

A inferência sobre o parâmetro Γ também não é difícil, uma vez que os erros-padrão do estimador podem ser calculados rapidamente com o *Delta method*.

Exemplo - Programa de sustentabilidade

Estimando $ATT(t)$ com DID via MQO

Primeiro, vamos transformar a variável *ano* em *factors*, para que possamos gerar *dummies* pra cada ano com a função *lm*.

```
df <- df %>% mutate(Tt = ifelse(ano >= 2012, 1, 0),
                           ano = as.factor(ano) #criar Dummies
                           )
```

Agora, vamos estimar a regressão (10), isto é:

$$y_{ist} = \alpha_s + \theta_t + \sum_{t=t_0+1}^T \gamma_t \mathbb{1}_\tau d * T_t + \varepsilon_{ist}$$

```
Reg_did <- lm(y ~ as.factor(d) + ano + ano:d:Tt,
              data = df)
```

E, como sempre, calcular os erros-padrão robustos e gerar a tabela.

```

library("lmtest")
library("sandwich")

# Robust t test
cov1 <- sandwich::vcovHC(Reg_did, type = "HCO")
sd_robust = sqrt(diag(cov1))
# Especifique as variáveis que você deseja incluir na tabela

# Extraia os coeficientes relevantes
coef_names <- names(coef(Reg_did))
vars_to_include <- grep(paste0("ano\\d{4}:d:Tt"),
                        coef_names, value = TRUE)

# Gere a tabela usando stargazer
stargazer(Reg_did, type = "latex", se = list(sd_robust),
          title = "MQO com Erros Padrão Robustos")

```

Como usamos a fórmula "ano:d:Tt", γ_{2012} = será mostrado como "ano2012:d:Tt", γ_{2013} = será como "ano2013:d:Tt", e assim por diante. Além disso, vamos ocultar as estimativas dos efeitos fixos de tempo e de indivíduo, dado o alto número de regressores.

Dependent variable:	
	y
⋮	⋮
ano2012:d:Tt	−0.561*** (0.103)
ano2013:d:Tt	−0.740*** (0.105)
ano2014:d:Tt	−0.993*** (0.108)
ano2015:d:Tt	−1.258*** (0.112)
ano2016:d:Tt	−1.439*** (0.116)
ano2017:d:Tt	−1.740*** (0.120)
ano2018:d:Tt	−1.907*** (0.123)
ano2019:d:Tt	−2.318*** (0.132)
ano2020:d:Tt	−2.450 (0.132)
Constant	9.937*** (0.059)
Observations	12,000
R ²	0.907
Adjusted R ²	0.907
Residual Std. Error	1.509 (df = 11980)
F Statistic	6,141.730*** (df = 19; 11980)
Note:	*p<0.1; **p<0.05; ***p<0.01

Ou seja, temos que:

$$\begin{aligned}\gamma_{2012} &= -0.561, \gamma_{2013} = -0.740, \gamma_{2014} = 0.993, \gamma_{2015} = -1.258 \\ \gamma_{2016} &= -1.439, \gamma_{2017} = -1.740, \gamma_{2018} = 1.907, \gamma_{2019} = -2.318 \\ \gamma_{2020} &= -2.45\end{aligned}$$

Agora, podemos estimar qualquer combinação dos $ATT(t)$ (γ_t).

$$\Gamma = \sum_{t=t_0+1}^T w_t \gamma_t$$

Primeiro, vamos armazenar todos os coeficientes que representam os $ATT(t)$ em um vetor.

```
ATT = Reg_did$coefficients[paste0("ano", 2012:2020, ":d:Tt")]
```

E vamos calcular Γ e seus desvios-padrão usando o método delta. ^a Para começar, podemos calcular a média simples, ou seja $w(t) = \frac{1}{T} = \frac{1}{9}$. Logo:

$$\Gamma = \frac{1}{9} \sum_{t=2012}^{2020} \gamma_t$$

```
wt= 1/9 #(1/T)

Gamma= sum(wt*ATT)

#calculado Sd via Delta method
Grad = matrix(c(rep(0, 12), rep(wt,8)), ncol = 1)
Sd_Gamma = sqrt(t(Grad)%*%cov1%*%Grad)
# Calcula o p_valor
valor_t <- (Gamma)*sqrt(12000) /Sd_Gamma
gl <- Reg_did$df.residual
valor_p <- 2 * pt(abs(valor_t), df=gl, lower.tail=FALSE)

print(c(Gamma, Sd_Gamma, valor_p))

## [1] "Gamma" "Sd" "P_valor"
## [1] -1.48946344 0.0804382 0.00000000
```

Como esperado, estimamos o mesmo valor que o obtido com a regressão DID padrão. O desvio-padrão também foi muito semelhante, e rejeitamos a hipótese de efeito zero.

Entretanto, como discutido, pode ser interessante analisar outros fatores. Por exemplo, será que o efeito de médio a longo prazo é maior do que o efeito de curto prazo (até o quarto período após o tratamento)? Para investigar essa possibilidade, vamos definir:

$$w_t = \begin{cases} -1/4 & t \leq 4 \\ 1/4 & t > 4 \end{cases}$$

E, como os $ATT(t)$ são negativos, se $\Gamma > 0$, o efeito de longo prazo será menor, em valor absoluto, do que o efeito de curto prazo. Se $\Gamma < 0$, os efeitos de longo prazo serão maiores. Por fim, se $\Gamma = 0$, o efeito, em média, será igual no longo e no curto prazo.


```
wt = c(rep(-1/4, 4), rep(1/4,5))

Gamma = sum(wt*ATT)
#calculado Sd via Delta method
Grad = matrix(c(rep(0, 12), wt), ncol = 1)
Sd_Gamma = sqrt(t(Grad)%*%cov1%*%Grad)
# Calcula o p_valor
valor_t <- (Gamma)*sqrt(12000) /Sd_Gamma
gl <- Reg_did$df.residual
valor_p <- 2 * pt(abs(valor_t), df=gl, lower.tail=FALSE)

print(c(Gamma, Sd_Gamma, valor_p))

## [1]      "Gamma"      "Sd"      "P_valor"
## [1] -1.57538221  0.07025662  0.00000000
```

Portanto, como $\Gamma < 0$, o efeito de longo prazo tem maior intensidade do que o efeito de curto prazo. Também poderíamos ter chegado a essa conclusão apenas plotando o gráfico dos $ATT(t)$, como feito na seção anterior. No entanto, com esse procedimento, além de obtermos facilmente os erros-padrão para realizar testes de hipóteses, podemos usar o mesmo raciocínio para avaliar outras formas interessantes. Por exemplo, podemos definir $w(t) = 1$ para todo t e calcular o efeito acumulado ao longo dos anos.

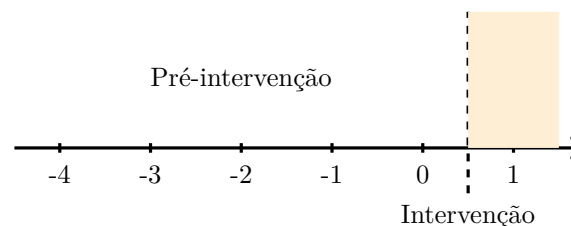
Também poderíamos ter chegado a essa conclusão apenas analisando o gráfico dos $ATT(t)ATT(t)ATT(t)$, como feito na seção anterior. No entanto, com esse procedimento, além de obtermos facilmente os **erros-padrão** para realizar testes de hipóteses, podemos aplicar o mesmo raciocínio para avaliar outras formas de agregação dos efeitos.

Por exemplo, podemos definir $w(t)=1w(t)=1$ para todo t e calcular o **efeito acumulado** ao longo dos anos.

^aPara mais informações sobre o método delta e como usar veja: Ver Hoef (2012).

1.5 Regressão DID com múltiplos períodos pré-tratamento

Nesta seção, será estudado o modelo de diferença-em-diferenças com mais de um período pré-tratamento. Os períodos pré-tratamento serão definidos com índices menores ou iguais a zero, ou seja, $0, -1, -2, \dots$, e o único período pós-tratamento será definido com índice 1. A figura abaixo representa graficamente os períodos antes e depois da intervenção.



Diferentemente do modelo com múltiplos períodos pós-intervenção, a definição do efeito do tratamento (ATT) não dependerá do tempo, pois há apenas um período pós-intervenção. Portanto, o ATT se mantém igual a:

$$ATT = E[y_i(1) - y_i(\infty) | d = 1]$$

A principal diferença está na hipótese de tendências paralelas, que agora deverá ser válida para todos os períodos pré-tratamento. Semelhante ao que foi feito na seção 1.2, derivaremos a condição suficiente da tendência paralela no caso com mais de um período pré-tratamento. Como temos mais de um período pré-tratamento,

a diferença será calculada entre o período pós-tratamento e a média de todos os períodos pré-tratamento. Defina, então, $N_{t \leq 0}$ como a quantidade total de períodos pré-intervenção. Teremos, sob a hipótese de não antecipação (i.e. $\forall t \leq 0, y_{it} = y_{it}(\infty)$):

$$\begin{aligned}\Delta_{Tr} &= E[y_{i1}(1)|d_i = 1] - E\left[\frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i = 1\right] \\ \Delta_{Cr} &= E[y_{i1}(\infty)|d_i = 0] - E\left[\frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i = 0\right].\end{aligned}$$

A segunda diferença continua a mesma, ou seja:

$$\begin{aligned}\gamma^{DID} &= \Delta_{Tr} - \Delta_{Cr} \\ &= \left(E[y_{i1}(1)|d_i = 1] - E\left[\frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i = 1\right]\right) \\ &\quad - \left(E[y_{i1}(\infty)|d_i = 0] - E\left[\frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i = 0\right]\right)\end{aligned}$$

Somando e subtraindo $E[y_{i1}(\infty)|d_i = 1]$:

$$\begin{aligned}\gamma^{DID} &= \underbrace{E[y_{i1}(1) - y_{i1}(\infty)|d_i = 1]}_{ATT} \\ &\quad + E\left[y_{i1}(\infty) - \frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i = 1\right] - E\left[y_{i1}(\infty) - \frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i = 0\right].\end{aligned}$$

Portanto, para que γ^{DID} seja igual ao ATT , é necessário que:

$$E\left[y_{i1}(\infty) - \frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i = 1\right] = E\left[y_{i1}(\infty) - \frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i = 0\right] \quad (11)$$

Usando álgebra, podemos mostrar que essa hipótese é equivalente à hipótese de tendências paralelas convencional em todos os períodos pré-tratamento. Primeiro, vamos mostrar que a diferença $E\left[y_{i1}(\infty) - \frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i\right]$ é equivalente à média das diferenças para todos os períodos pré-tratamento:

$$\begin{aligned}E\left[y_{i1}(\infty) - \frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i\right] &= E\left[\frac{N_{t \leq 0}}{N_{t \leq 0}} y_{i1}(\infty) - \frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i\right] \\ &= E\left[\frac{\sum_{t \leq 0} y_{i1}(\infty)}{N_{t \leq 0}} - \frac{\sum_{t \leq 0} y_{it}(\infty)}{N_{t \leq 0}}|d_i\right] \\ &= \frac{\sum_{t \leq 0} E[y_{i1}(\infty) - y_{it}(\infty)|d_i]}{N_{t \leq 0}}\end{aligned}$$

Ou seja, usando a igualdade acima e a igualdade dada na equação (11):

$$\begin{aligned}\frac{\sum_{t \leq 0} E[y_{i1}(\infty) - y_{it}(\infty)|d_i = 1]}{N_{t \leq 0}} - \frac{\sum_{t \leq 0} E[y_{i1}(\infty) - y_{it}(\infty)|d_i = 0]}{N_{t \leq 0}} &= 0 \\ \frac{1}{N_{t \leq 0}} \sum_{t \leq 0} [E[y_{i1}(\infty) - y_{it}(\infty)|d_i = 1] - E[y_{i1}(\infty) - y_{it}(\infty)|d_i = 0]] &= 0\end{aligned}$$

Portanto, para que um DID funcione com múltiplos períodos pré-tratamento, a média simples da diferença entre as tendências do grupo de controle e tratamento deve ser igual a zero. Essa hipótese pode soar um pouco estranha, pois não é necessário que a tendência paralela seja válida para cada período, apenas que a média simples delas em relação a todos os períodos seja igual para tratados e controles.

Dada a dificuldade de interpretar essa hipótese, muitos trabalhos assumem uma hipótese suficiente, porém mais restritiva, de que para todos os períodos de pré-tratamento, as tendências dos grupos de tratamento e controle são paralelas:

$$\forall t \leq 0, E[y_{i1}(\infty) - y_{it}(\infty)|d_i = 1] = E[y_{i1}(\infty) - y_{it}(\infty)|d_i = 0]$$

2 Diferença-em-Diferenças escalonado

2.1 O que é um DID escalonado?

O DID escalonado assemelha-se ao DID múltiplo discutido anteriormente. Contudo, em vez de todos os indivíduos serem tratados simultaneamente, os grupos são tratados em diferentes momentos no tempo. Considere o estudo do efeito de uma política econômica, como um aumento no salário mínimo, em diferentes estados de um país. Em vez de implementar o aumento do salário mínimo de uma só vez em todos os estados, ele é aplicado em estados diferentes em momentos distintos. Por exemplo, é implementado em 2018 em um grupo de estados, em 2019 em outro grupo, e o restante em 2020. Neste contexto, após 2018, o tratamento é escalonado até que todos (ou parte) dos estados sejam tratados.

Os períodos anteriores ao momento em que o primeiro grupo recebe o tratamento, nos quais nenhum grupo ainda foi tratado, serão chamados de **período pré-tratamento**. Os períodos seguintes, a partir do momento em que algum grupo passa a ser tratado, serão chamados de **períodos pós-tratamento**. Os demais períodos serão períodos pós-tratamento. Os indivíduos tratados pela primeira vez no períodos $t = g$ constituirão o grupo de tratamento g . Já o grupo não tratado em nenhum período chamaremos de *never treated* (ou nunca tratados) e serão denotados por ∞ . Denotaremos por G o conjunto de todos os grupos tratados. Assim, se $i \in g$ e $g \in G$ significa que o indivíduo i pertence ao grupo que foi tratado pela primeira vez no período $t = g$.

2.2 Aumentando o modelo de resultados potenciais

Assim como para o DID com múltiplos períodos, aqui será necessário definir ATT para cada período no tempo. No entanto, aqui o resultado potencial pode depender do momento em que o indivíduo recebe o tratamento: ser exposto à intervenção antes ou depois pode ter efeitos diferentes. Por isso, definiremos o resultado potencial de um indivíduo considerando especificamente o cenário em que ele teria sido tratado pela primeira vez no período g , o que denotaremos por $y_{it}(g)$.

Por exemplo, considere uma intervenção onde cada estado do Brasil tem seu salário mínimo aumentado. A exposição foi escalonada entre os anos de 2012 a 2014. O estado de São Paulo ($i = SP$), por exemplo, terá quatro resultados potenciais diferentes para cada período t : um para o caso em que ele é tratado pela primeira vez em 2012, ou seja, $y_{SP,t}(2012)$; outro para o caso em que é tratado em 2013, $y_{SP,t}(2013)$; um para o caso em que é tratado pela primeira vez em 2014, $y_{SP,t}(2014)$; e, por fim, para caso não tenha mudado o salário mínimo em nenhum desses anos, $y_{SP,t}(\infty)$.

Novamente, apenas observamos um dos resultados potenciais para cada indivíduo. Isso significa que, no nosso exemplo, se SP mudar seu salário mínimo em 2014, apenas $y_{(SP,t)}(2014)$ será observado.

Dessa forma, passamos a considerar não apenas um ATT para cada período no tempo t , mas, para cada t um ATT específico para cada período de início de tratamento g .

Denotaremos o ATT como:

$$ATT(g, t) = E[y_{it}(g) - y_{it}(\infty)|d = g].$$

que representa o efeito médio do tratamento no período t para o grupo tratado pela primeira vez no período g .

Todos as soluções da literatura, de alguma forma, passarão pela estimação de cada $ATT(g, t)$ de forma análoga a um DID padrão (2x2). Ou seja, as hipóteses

e, principalmente, a interpretação dos resultados, se assemelham muito aos casos visitados nas seções anteriores. Todas essas soluções requerem que tenhamos tendências paralelas, com algumas poucas variações entre as hipóteses. Porém, o mais importante é que, em todas as soluções, precisamos ter tendências paralelas para todos os resultados potenciais em todos os períodos de tempo. Em outras palavras, precisaremos de tendências paralelas para $y_{it}(g)$ para todo t e para todo $g \in G \cup \{\infty\}$.

Nesta seção, definiremos as hipóteses para um DID escalonado sem controles, de forma que os estimadores apresentados na literatura acabem assumindo hipóteses bastante semelhantes. Quando estivermos destrinchando cada um dos diferentes estimadores, comentaremos as diferenças de hipóteses entre eles e os casos em que necessitaremos de hipóteses mais restritivas do que as apresentadas nesse momento. Essas diferenças estarão relacionadas, principalmente, às hipóteses de tendências paralelas condicionais que cada estimador pressupõe.

Hipóteses

1. **SUTVA:** O resultado potencial no caso de não tratamento não é afetado pela alocação do tratamento sobre outros indivíduos.
2. **Não antecipação:** $\forall i \text{ e } g \in \{1, 2, \dots, G\} \cup \{\infty\}, y_{i0} = y_{i0}(g) = y_{i0}(\infty)$.
3. **Irreversibilidade do Tratamento:** $\forall t \geq g, y_{it} = y_{it}(g)$.
4. **Tendências Paralelas para todos os grupos:**

$$\forall t \text{ e } \forall g, E[y_{it}(\infty) - y_{i0}(\infty)|d = g] = E[y_{it}(\infty) - y_{i0}(\infty)|d_i = \infty]$$

Podemos estimar os $ATT(g, t)$ para cada período de forma semelhante ao DID múltiplo, mas separadamente para cada um dos grupos. Ou seja, para o grupo tratado pela primeira vez em $t = g$:

$$\begin{aligned} \gamma_{gt}^{DID} &= (E[y_{it}|d = g] - E[y_{i0}|d = g]) - (E[y_{it}|d = \infty] - E[y_{i0}|d = \infty]) \\ &= E[y_{it}(g) - y_{i0}(\infty)|d = g] - E[y_{it}(\infty) - y_{i0}(\infty)|d_i = \infty] \\ &= \underbrace{E[y_{it}(g) - y_{it}(\infty)|d = g]}_{ATT(g,t)} \\ &\quad + E[y_{it}(\infty) - y_{i0}(\infty)|d = g] - E[y_{it}(\infty) - y_{i0}(\infty)|d_i = \infty] \end{aligned}$$

onde da segunda para terceira linha somamos e subtraímos $E[y_{it}(\infty)|d = g]$. Sob as hipóteses 1-4, temos que $\gamma_{gt}^{DID} = ATT(g, t)$. Observe que, para estimarmos o $ATT(g, t)$ consistentemente sob a hipótese de tendências paralelas não condicionais, o procedimento é exatamente igual ao discutido nas seções anteriores, mas realizamos a estimativa separadamente para cada grupo de exposição.

Portanto, um estimador consistente seria:

$$\hat{\gamma}_{gt}^{DID} = (\bar{y}_{gt} - \bar{y}_{g0}) - (\bar{y}_{\infty t} - \bar{y}_{\infty 0})$$

Note que, sob as hipóteses 1-4, temos que $\hat{\gamma}_{gt}^{DID} \xrightarrow{p} ATT(g, t)$. E, dado que temos um estimador consistente para cada $ATT(g, t)$, podemos, como fizemos para o DID múltiplo, estimar qualquer média ponderada deles. Mas, afinal, qual é o problema de um DiD escalonado? A dificuldade aparece quando tentamos estimá-lo usando um modelo com efeitos fixos de duas vias (TWFE), como fizemos nos casos anteriores.

Na próxima seção, vamos detalhar os problemas que surgem ao aplicar um modelo TWFE em um contexto de DiD escalonado.

Exemplo - Programa de sustentabilidade

Vamos continuar com nosso exemplo sobre o programa de sustentabilidade nos municípios brasileiros. Desta vez, porém, o programa será imple-

mentado de maneira escalonada. Ou seja, parte dos municípios será exposta à intervenção em 2012; outra parte em 2014; uma terceira parte em 2016 e o restante em 2018. Assumiremos que alguns municípios não serão tratados em nenhum momento.

A base com os dados fictícios encontra-se no arquivo "base_stagger_did.csv". Vamos, primeiro, observar o tamanho de cada grupo de tratamento G :

```
table(df$G, df$ano)
```

```
##
##      2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
## 0      480 480 480 480 480 480 480 480 480 480
## 2012 240 240 240 240 240 240 240 240 240 240
## 2014 120 120 120 120 120 120 120 120 120 120
## 2016 180 180 180 180 180 180 180 180 180 180
## 2018 180 180 180 180 180 180 180 180 180 180
```

Agora, vamos usar a equação

$$\hat{\gamma}_{gt}^{DID} = (\bar{y}_{gt} - \bar{y}_{g0}) - (\bar{y}_{\infty t} - \bar{y}_{\infty 0})$$

para estimarmos cada $ATT(g, t)$ consistentemente.

```
All_G = sort(unique(df$G))[-1]
All_t = sort(unique(df$ano))[-1]

ATT <- matrix(ncol = length(All_G),
              nrow = length(All_t))

rownames(ATT) = All_t
colnames(ATT) = All_G

for(g in 1:length(All_G)){
  for(t in 1:length(All_t)){
    gg = All_G[g]
    tt = All_t[t]
    if(tt >= gg){
      ATT[t, g] = (mean(df$y[df$G==gg & df$ano == tt])
                  - mean(df$y[df$G==gg & df$ano == 2011]))
                  - (mean(df$y[df$G==0 & df$ano == tt])
                    - mean(df$y[df$G==0 & df$ano == 2011]))
    }
  }
}
```

Vamos plotar os $ATT(g, t)$ numa tabela onde as colunas denotam o momento de primeira exposição de cada grupo (g), e as linhas, o ano para o qual o impacto está sendo avaliado (t).

ATT

##		2012	2014	2016	2018
##	2012	-0.3961984	NA	NA	NA
##	2013	-0.3214912	NA	NA	NA
##	2014	-0.6649661	-0.2423996	NA	NA
##	2015	-0.9659459	-0.3577052	NA	NA
##	2016	-1.0635417	-0.5236565	-0.007105557	NA

```
## 2017 -1.1597543 -1.0455578 -0.262119340 NA
## 2018 -1.2528935 -1.0309909 -0.480166057 -0.2141823
## 2019 -1.6833693 -1.2518550 -0.697514637 -0.5321645
## 2020 -1.7334163 -1.3034411 -0.804441748 -0.7870462
```

Ou seja, o ATT do grupo que foi tratado pela primeira vez em 2014 avaliado em 2018, i.e, $ATT(g = 2014, t = 2018) = -1,031$.

2.3 Quando e por que um DID escalonado apresenta problemas?

O problema na estimação de diferença-em-diferenças escalonado surge quando tentamos realizá-la via regressão linear. Em geral, neste contexto, colocamos efeitos fixos de grupo (α_g) e de tempo (θ_t) num TWFE. Vamos definir $g(i)$ como o ano que o indivíduo i é tratado pela primeira vez.

$$y_{igt} = \alpha_g + \theta_t + \gamma D_{it} + \varepsilon_{igt} \quad (12)$$

onde $D_{it} = \begin{cases} 1 & t \geq g(i) \\ 0 & \text{c.c.} \end{cases}$.

A grande questão é: o que γ representa?

Ao contrário de modelos que envolvem apenas dois grupos, a interpretação de γ torna-se mais complexa. Para simplificar, vamos considerar um cenário específico com 3 grupos e 3 períodos. Um grupo permanece sem intervenção ao longo dos 3 períodos analisados, sendo identificado como nosso grupo de controle $g = \infty$. Outro grupo recebe tratamento no primeiro período pós-intervenção, enquanto o último é tratado no segundo período pós-intervenção. Definimos o período em que nenhum grupo foi tratado como $t = 0$, o primeiro período após a intervenção como $t = 1$, e o segundo como $t = 2$.

Na representação da Figura 2, observamos as tendências dos grupos $g = 1$, $g = 2$ e $g = \infty$. Embora os três grupos apresentem tendências paralelas neste exemplo, é importante destacar que, ao estimar a equação 12, o coeficiente γ pode não refletir adequadamente o que desejamos analisar, conforme discutido por Goodman-Bacon (2021).

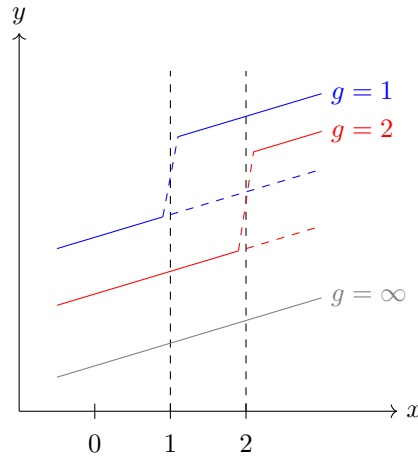


Figura 2: DID 3x3

O ponto crítico é que, em alguns casos, todos os grupos podem exibir um efeito positivo do tratamento, enquanto o γ da equação 12 resulta em zero ou em um valor negativo. Isso sugere que, mesmo que a intervenção tenha um impacto em todos os grupos, podemos erroneamente concluir que ela não tem efeito, ou ainda pior, inferir incorretamente que o efeito é oposto ao real.

Como mostrado por Goodman-Bacon (2021), o γ da regressão será uma média ponderada das comparações dois a dois de cada um dos grupos e dos períodos analisados. Podemos expressá-la da seguinte forma:

$$\gamma = VWATT + VWCT - \Delta ATT \quad (13)$$

onde, $VWATT$ corresponde a uma média ponderada dos $ATT(g, t)$, em que a ponderação é determinada pela variância relativa de cada grupo g no período t . Em outras palavras, atribuímos maior peso às comparações que envolvem grupos maiores e àquelas em que a **variância relativa do tratamento** na subamostra composta pelos grupos g e controle no período t é maior. $VWCT$ representa a tendência paralela para todos os grupos: se a hipótese de tendências paralelas for satisfeita para todos os grupos, $VWCT = 0$. Por fim, $\Delta ATT = \sum_{g=1}^G \sum_{k \neq u} \sum_{l > k} \sigma_{ku}^g [ATT(k, l) - ATT(k, l)]$ onde, σ_{kl}^l é o peso relativo do grupo g entre

Veja que, se o ATT for constante ao longo do tempo, então $\Delta ATT = 0$ e, $\gamma = VWATT$ (sob tendências paralelas para todos os grupos e períodos).

Portanto, agora podemos resumir o que um TWFE (eq. 12) nos retornará. Assumindo que as tendências paralelas são satisfeitas pra todos os grupos ($VWCT = 0$):

1. Se ATT for constante para todos os grupos e todos os períodos, i.e., $\forall g, t, ATT(g, t) = ATT(g', t') = ATT$:

$$\begin{aligned} \gamma &= VWATT - \underbrace{\Delta ATT}_{=0} \\ &= VWATT = ATT \end{aligned}$$

2. Se ATT for constante ao longo do tempo, mas diferente entre grupos, i.e., $\forall t, t', ATT(g, t) = ATT(g, t') = ATT(g)$ mas $ATT(g) \neq ATT(g')$:

$$\begin{aligned} \gamma &= VWATT - \underbrace{\Delta ATT}_{=0} \\ &= VWATT = \sum_{t=1}^T \sum_{g=1}^G w_{MQO}(g, t) ATT(g, t) \end{aligned}$$

Nesse caso, temos uma média ponderada dos $ATT(g, t)$, em que o peso depende das variâncias relativas de cada combinação 2x2. Isso, provavelmente, não reflete o que estamos interessados em estimar (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021).

3. Se ATT variar ao longo do tempo (e, possivelmente, ao longo dos grupos):

$$\begin{aligned} \gamma &= VWATT - \underbrace{\Delta ATT}_{\neq 0} \\ &= VWATT - \Delta ATT \\ &= \sum_{t=1}^T \sum_{g=1}^G w_{MQO}(g, t) ATT(g, t) - \Delta ATT \end{aligned}$$

Nessa situação, além de estimarmos uma média ponderada que provavelmente não é a de interesse, ainda introduzimos um termo de viés representado por ΔATT . Além disso, a depender das magnitudes dos dois termos, poderemos observar um resultado com sinal trocado.

Em resumo, apenas quando o efeito da intervenção é homogêneo entre os grupos e ao longo do tempo, conseguimos estimar o ATT sem viés. Se o efeito é heterogêneo entre os grupos, mas constante ao longo do tempo, estimaremos consistentemente uma média ponderada dos ATT . No entanto, em geral, essa média ponderada pode ter uma interpretação difícil e pouco relevante para a análise. Caso o ATT seja heterogêneo, pelo menos ao longo do tempo, estimaremos um parâmetro enviesado e, possivelmente, com sinal inverso do verdadeiro.

Para ilustrar a aplicação prática do modelo de Diferença-em-Diferenças escalonado via TWFE, reproduzimos nosso exemplo do programa de sustentabilidade com três tipos de efeito, todos com média simples dos ATTs é igual a 1, ou seja, $\frac{1}{NG} \sum_t \sum_g ATT(g, t) = 1$.

No primeiro cenário, o efeito é constante ao longo do tempo e entre os grupos, resultando em uma estimativa consistente do ATT médio. No segundo cenário, o efeito é constante ao longo do tempo, mas varia entre os grupos, levando a uma estimativa consistente, porém de uma média ponderada pouco informativa. No terceiro cenário, a heterogeneidade ocorre em todas as dimensões, resultando em uma estimativa com viés.

Em todos os cenários, temos 4 grupos de tratamento e um grupo de controle no período de 2011 a 2020. O primeiro grupo recebeu tratamento em 2012, o segundo em 2014, o terceiro em 2016 e o último em 2018. Além disso, assumimos que, **condicionalmente à escolaridade e à população**, a hipótese de **tendências paralelas** é satisfeita. Ou seja, $VWCT = 0$ na equação (13).

CASO 1: Tratamento Homogêneo

$$\forall t, g \text{ } ATT(g, t) = 1$$

```
rm(list=ls())
library(dplyr)

df <- read.csv("base_stagger_did_const.csv")

#TWFE regressão
reg_TWFE = lm(y ~ as.factor(G) + as.factor(ano) + d_x_t +
  escolaridade + populacao, data = df)

library("lmtest")
library("sandwich")
# t test Robusto
cov1 <- sandwich::vcovHC(reg_TWFE, type = "HCO")
sd_robust = sqrt(diag(cov1))

# Gere a tabela usando stargazer
library(stargazer)
stargazer(reg_TWFE, type = "latex", se = list(sd_robust),
title = "MQO com Erros Padrao Robustos")
```

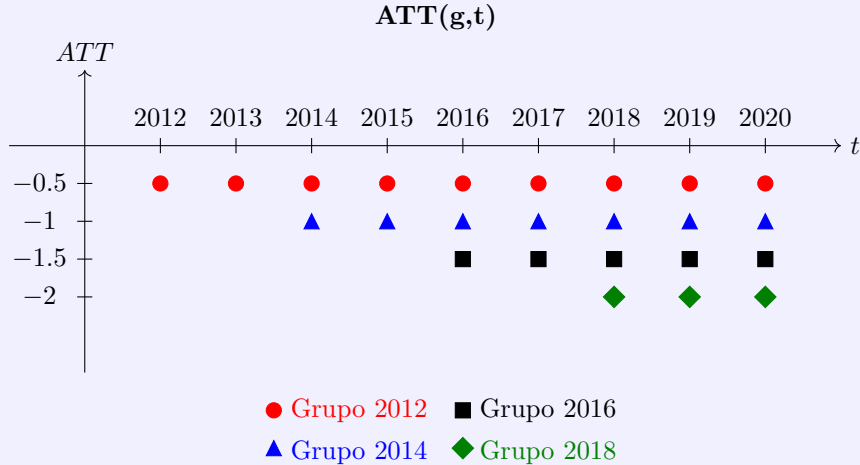

Dependent variable:	
y	
⋮	⋮
d_x_t	-1.038*** (0.051)
escolaridade	-0.064*** (0.014)
populacao	-0.00000 (0.00000)
Constant	10.500*** (0.124)
Observations	12,000
R ²	0.907
Adjusted R ²	0.907
Residual Std. Error	1.499 (df = 11983)
F Statistic	7,282.051*** (df = 16; 11983)

Note: *p<0.1; **p<0.05; ***p<0.01

A variável d_x_t representa uma indicadora de tratamento interagida com uma indicadora de períodos em que o grupo ao qual o indivíduo pertence está sendo tratado. Observa-se que, como $ATT(g, t) = -1$ para todos os grupos e períodos, estimamos um efeito médio (ATT) de -1 , conforme esperado. Nesse cenário, por não haver heterogeneidade no efeito do tratamento, estimamos o ATT de forma consistente, e qualquer média ponderada será igual ao ATT , que é -1 . O desafio significativo dessa abordagem reside na improbabilidade de que o efeito do tratamento seja uniforme ao longo dos períodos e grupos, especialmente quando se lida com um grande número de períodos ou grupos.

CASO 2: Tratamento Homogêneo ao longo do tempo

Neste cenário, geramos dados com o efeito do tratamento variando entre os grupos, mas permanecendo constante ao longo do tempo. A figura abaixo representa os $ATT(g, t)$ de cada grupo, para cada período. O grupo tratado pela primeira vez em 2012 ($g = 2012$) tem um efeito de -0.5 em todos os períodos; o grupo tratado pela primeira vez em 2014 ($g = 2014$) tem um efeito de -1 ; o grupo tratado pela primeira vez em 2016 ($g = 2016$) tem um efeito de -1.5 ; e, por fim, o grupo tratado pela primeira vez em 2018 ($g = 2018$) tem um efeito de -2 .



Mais uma vez, a média simples dos $ATTs$ é -1 , ou seja, $\frac{1}{NG} \sum_i \sum_g ATT(g) = -1$. E já que o ATT de cada grupo não varia ao longo do tempo, temos $\Delta ATT = 0$. Assim, podemos estimar uma média ponderada dos $ATTs$ de forma consistente. No entanto, é importante destacar que essa média ponderada é de difícil interpretação, como argumentado por Goodman-Bacon (2021) e Callaway and Sant'Anna (2021).

Vamos, portanto, estimar um DID via TWFE nesse caso:

```
rm(list=ls())
library(dplyr)

df <- read.csv("base_stagger_did_const_group.csv")

#TWFE regressão
reg_TWFE = lm(y ~ as.factor(G) + as.factor(ano) + d_x_t +
  escolaridade + populacao, data = df)

library("lmtest")
library("sandwich")
# t test Robusto
cov1 <- sandwich::vcovHC(reg_TWFE, type = "HCO")
sd_robust = sqrt(diag(cov1))

# Gere a tabela usando stargazer
library(stargazer)
stargazer(reg_TWFE, type = "latex", se = list(sd_robust),
  title = "MQO com Erros Padrao Robustos")
```

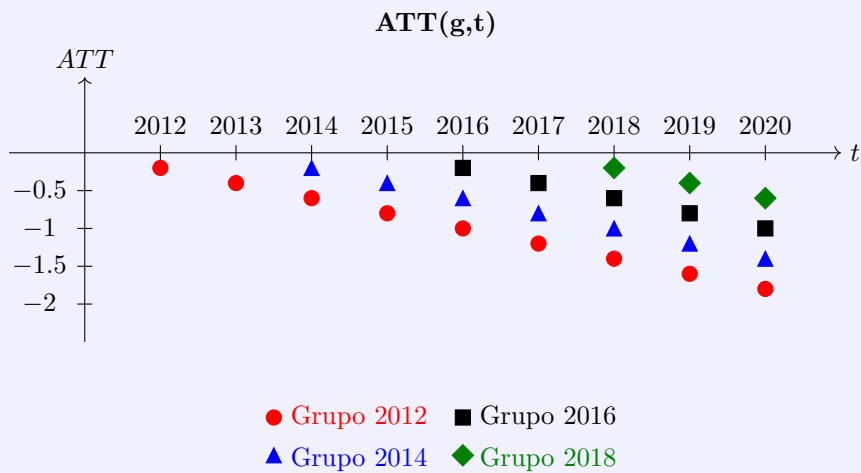
Dependent variable:	
	y
⋮	⋮
d_x_t	-1.439*** (0.052)
escolaridade	-0.064*** (0.014)
populacao	-0.00000 (0.00000)
Constant	10.344*** (0.125)
Observations	12,000
R ²	0.907
Adjusted R ²	0.907
Residual Std. Error	1.499 (df = 11983)
F Statistic	7,282.051*** (df = 16; 11983)
Note: *p<0.1; **p<0.05; ***p<0.01	

A estimativa do efeito é de -1.439 . Isso reflete uma média ponderada dos $ATTs$ apresentados na figura anterior. No entanto, os pesos associados carecem de intuição e são de difícil interpretação. O valor -1.439 não representa o efeito médio da redução do gasto de energia nos grupos tratados durante o período analisado, e também não é equivalente à média simples dos $ATTs$ (que é -1). Além disso, questões comuns em uma análise de impacto, tais como "Qual o efeito médio ao longo dos anos da análise?", "Qual o efeito médio, caso a política continue no próximo ano?" e "Qual o custo-benefício da

política?", não podem ser respondidas. Aqui, mesmo estimando uma média ponderada consistente, ela nos é pouco informativa (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021).

CASO 3: Tratamento heterogêneo em todas as dimensões

Agora, geramos uma base de dados onde o tratamento varia nas duas dimensões, tanto ao longo do tempo quanto entre os grupos. Novamente, a média simples dos ATTs é igual a 1, i.e., $\frac{1}{NG} \sum_g \sum_t ATT(g, t) = 1$. A figura abaixo representa o $ATT(g, t)$ de todos os grupos em todos os períodos de tratamento.



O efeito do tratamento aumenta, em valor absoluto, à medida que o tempo de exposição se prolonga. Para cada ano exposto ao tratamento, o efeito médio naquele ano é incrementado em 0.25. Por exemplo, considerando o grupo tratado pela primeira vez em 2012; no período de 2016, ele estará sob tratamento por 5 anos, resultando em um efeito naquele ano de $-1.25 = -(5 * 0.25)$. Neste cenário, temos o tratamento variando ao longo do tempo, o que significa que, além dos problemas do caso anterior, agora $\Delta ATT \neq 0$. Em outras palavras, não estaremos estimando consistentemente uma média ponderada dos *ATTs*.

Novamente, vamos estimar um DID via TWFE:

```
rm(list=ls())
library(dplyr)

df <- read.csv("base_stagger_did.csv")

#TWFE regressão
reg_TWFE = lm(y ~ as.factor(G) + as.factor(ano) + d_x_t +
  escolaridade + populacao, data = df)

library("lmtest")
library("sandwich")
# t test Robusto
cov1 <- sandwich::vcovHC(reg_TWFE, type = "HCO")
sd_robust = sqrt(diag(cov1))

# Gere a tabela usando stargazer
library(stargazer)
stargazer(reg_TWFE, type = "latex", se = list(sd_robust),
  title = "MQO com Erros Padrao Robustos")
```

<i>Dependent variable:</i>	
	y
⋮	⋮
d_x_t	-0.410*** (0.051)
escolaridade	-0.064*** (0.014)
populacao	-0.00000 (0.00000)
Constant	10.664*** (0.125)
Observations	12,000
R ²	0.908
Adjusted R ²	0.908
Residual Std. Error	1.513 (df = 11983)
F Statistic	7,364.357*** (df = 16; 11983)

Note: *p<0.1; **p<0.05; ***p<0.01

Observe que, mesmo quando a média simples é igual a 1, estimamos um coeficiente menor que 1 em valor absoluto (-0.41). Em outras palavras, além dos problemas mencionados anteriormente, estamos subestimando os resultados, pois $\Delta|ATT| < 0$.

3 Como usar um DID escalonado?

Relembrando algumas definições importantes

- $y_{it}(g)$ é o resultado potencial do indivíduo i , no período t , caso ele tenha sido tratado pela primeira vez no período g .
- $y_{it}(\infty)$ é o resultado potencial do indivíduo i , no período t , quando o indivíduo i não é tratado.
- Assumiremos sempre que $\forall t < g, y_{it}(g) = y_{it}(\infty)$
- $ATT(g, t) := E[y_{it}(g) - y_{it}(\infty)]$

O principal desafio do DID escalonado reside no fato de que os modelos de TWFE provavelmente fornecerão um estimador enviesado para qualquer média dos $ATT(g, t)$ ou, na melhor das hipóteses, uma média ponderada com pesos que não serão necessariamente informativos para a análise e/ou de difícil interpretação. Apenas no caso em que o efeito é homogêneo para todos os grupos e períodos, o estimador de TWFE será consistente e informativo; no entanto, isso raramente ocorrerá na prática.

A solução, portanto, passa por estimarmos de forma consistente cada $ATT(g, t)$. Com isso, conseguiremos construir qualquer combinação linear que seja relevante para a análise da intervenção. Todos os estimadores da literatura de alguma forma seguirão esse caminho, direta ou indiretamente.

Além disso, todos os trabalhos discutidos nesta seção utilizam, em essência, as mesmas hipóteses, desde que consideremos que a hipótese de tendências paralelas

é satisfeita de forma incondicional, ou seja, sem a necessidade de controlar por um vetor de características X_{it} . Dessa forma, nesta seção, serão estudados os modelos que assumem a hipótese mais restritiva de tendências paralelas condicionais. Caso contrário, a escolha de qual modelo utilizar dependerá mais da facilidade de implementação para o caso específico do que das hipóteses assumidas por cada abordagem.

3.1 Sun and Abraham (2021)

3.1.1 Estimador

O estimador proposto por Sun and Abraham (2021) representa uma alteração no modelo padrão de TWFE. Basicamente, calcularemos cada $ATT(g, t)$ por meio de uma regressão TWFE. Definimos:

$$D_{i,t}^l = \begin{cases} 1 & \text{se } t - g(i) = l \\ 0 & \text{caso contrário} \end{cases}$$

Ou seja, $D_{i,t}^l$ é igual a 1 se estivermos l períodos após o primeiro período em que o indivíduo i foi tratado ($g(i)$) e zero, caso contrário. A princípio, vamos considerar $l > 0$, mas podemos definir $l < 0$ para testarmos se a hipótese de tendências paralelas pré-intervenção é válida.

Podemos, então, realizar a seguinte regressão:

$$y_{it} = \alpha_i + \gamma_t + \sum_{l \neq 0} \sum_g \delta_{l,g} D_{i,t}^l \mathbb{1}_{\{g(i)=g\}} + X_{it}'\beta + \varepsilon_{it}. \quad (14)$$

Sob algumas hipóteses que serão mencionadas na próxima sub-seção, temos que $\delta_{l,g} = ATT(g, t')$, onde $t' = g + l$.⁴ Para entender isso, definimos:

$$D_{i,t} = \begin{cases} 1 & \text{se } i \text{ é tratado no período } t \\ 0 & \text{caso contrário} \end{cases}$$

Observe que o terceiro termo da especificação 14 equivale a interagir nossa variável de tratamento com dummies de tempo e de grupo de tratamento (cohorte).⁵

$$D_{i,t}^l \mathbb{1}_{\{g(i)=g\}} = D_{i,t} \mathbb{1}_{\{g(i)=g\}} \mathbb{1}_{\{t\}}$$

Ou seja, semelhante ao que fizemos no capítulo 2:

$$\delta_{l,g} = E[y_{it} - y_{i0} | G = g, X] - E[y_{it} - y_{i0} | G = \infty, X]$$

onde $t = g + l$. Portanto, sob a hipótese de tendências paralelas condicionais, $\gamma_{l,t} = ATT(g, t)$. Isso é equivalente a calcularmos um DID 2x2 padrão em uma subamostra apenas com os grupos g e ∞ (controle) para os períodos 0 (pre-tratamento) e t (pós-tratamento).

3.1.2 Hipóteses

Como estamos estimando $ATT(g, t)$ via um TWFE, as hipóteses são basicamente as mesmas dos DID 2x2. Porém, agora as hipóteses devem ser satisfeitas para todos os períodos e grupos. Ou seja:

⁴No artigo, as autoras definem o período pré-tratamento como -1, mas, para manter a notação das seções anteriores, consideraremos o período pré-tratamento como 0.

⁵Os efeitos fixos a nível de indivíduo podem ser substituídos por efeitos fixos a nível de grupo g

Hipóteses

1. (Amostragem aleatória): Todos os indivíduos i foram amostrados de forma aleatória da população
2. (SUTVA) O resultado potencial no caso de não tratamento não é afetado pela alocação do tratamento sobre outros indivíduos.
3. (Irreversibilidade do Tratamento) Se $t \geq g$, $y_{it} = y_{it}(g)$.
4. (Não antecipação) Para todo indivíduo i , se $t < g$ então $y_{it} = y_{it}(\infty)$
5. (Tendências paralelas condicionais) **Para todo i , g , t :**

$$\begin{aligned} E[y_{it}(\infty) - y_{i0}(\infty)|G = g, X] &= E[y_{it}(\infty) - y_{i0}(\infty)|G = \infty, X] \\ &= E[y_{it}(\infty) - y_{i0}(\infty)|X] \end{aligned}$$

Basicamente, mantemos as mesmas hipóteses do DID padrão; no entanto, agora a tendência paralela precisa ser válida para todos os períodos e grupos (hipótese 5). O ponto principal argumentado por Marcus and Sant'Anna (2021) (citado por Callaway and Sant'Anna (2021)) é, que a primeira igualdade implica que tenhamos unidades nunca tratadas (*Never-Treated*), e a segunda, implica que o tratamento deve ser independente das tendências paralelas.⁶

Na literatura aplicada

Carrillo and Feres (2019) analisam o impacto do Programa Mais Médicos (PMM), uma política que visava aumentar a presença de médicos em regiões carentes, especialmente nas Unidades Básicas de Saúde (UBS). O objetivo principal era avaliar se a maior oferta de médicos resultou em melhorias na utilização de cuidados de saúde e em indicadores de saúde infantil, como mortalidade, peso ao nascer e duração da gestação. Para isso, foram utilizados dados administrativos bimestrais de saúde entre 2008 e 2016, abrangendo mais de 5.500 municípios brasileiros, que são a unidade de implementação da política. Ainda que o artigo tenha sido publicado antes de Sun and Abraham (2021), uma das especificações adotadas é bastante semelhante à metodologia proposta por esses autores, podendo, portanto, servir como exemplo de aplicação.

A análise mostrou que a implementação do PMM aumentou significativamente o número de médicos por 1.000 habitantes nas regiões tratadas, representando um incremento médio — ao longo do tempo e entre os grupos de municípios — de 18% em relação à média inicial. O impacto foi mais expressivo na disponibilidade de médicos nas UBSs, especialmente médicos da família. Além disso, observou-se um aumento de 5% a 8% nas visitas médicas, incluindo consultas pré-natais. No entanto, essas consultas substituíram atendimentos previamente realizados por enfermeiros, sem provocar um aumento no número total de atendimentos.

Apesar do aumento no acesso a médicos, os resultados não indicaram melhorias significativas nos principais indicadores de saúde infantil. Não foram observados efeitos estatisticamente significantes em métricas como peso ao nascer, duração da gestação ou mortalidade infantil. Isso sugere que o aumento da disponibilidade de médicos pode ter efeitos reduzidos se seus serviços substituírem serviços de outros profissionais como enfermeiros, em vez de ofertar serviços anteriormente indisponíveis.

3.1.3 Código em R com exemplo

Para estimar os efeitos de tratamento, usaremos um TWFE interagindo dummies de tempo e grupo com a dummy de tratamento, conforme a equação (14). Utilizaremos

⁶O tratamento ser independente das tendências paralelas implica que, por exemplo, não é possível o tratamento depender das tendências paralelas passadas.

a mesma base de dados do caso 3 da seção 2.3.

Primeiro, vamos relembrar como está estruturada a base de dados:

Código R

```
rm(list=ls())
library(dplyr)

df <- read.csv("base_stagger_did.csv")

head(df)

##   id_estado id_municipio   setor  ano renda escolaridade
## 1         1         1001 Serviços 2011  55.9           8.9
## 2         1         1001 ServiC'os 2012  60.0           8.8
## 3         1         1001 ServiC'os 2013  65.0           8.8
## populacao      y      d G      Tt d_x_t
## 117002      9.8040061  0 0      0  0
## 117510     11.145862  0 2013  1  0
## 117754      5.578302  0 0      1  0
```

Ou seja, G é a variável de grupo, e ano é a variável de tempo. Ao usarmos ‘as.factor(G)’ e ‘as.factor(ano)’ dentro do ambiente ‘lm’, criamos uma dummy para cada um dos grupos e anos da amostra. Portanto, para rodarmos a equação (14):

Código R

```
#TWFE regressão
reg_TWFE = lm(y ~-1+ as.factor(G) + as.factor(ano)
+ d_x_t : as.factor(G):as.factor(ano) + escolaridade
+ populacao, data = df)

library("lmtest")
library("sandwich")
# t test Robusto
cov1 <- sandwich::vcovHC(reg_TWFE, type = "HC0")
sd_robust = sqrt(diag(cov1))

# Gere a tabela usando stargazer
library(stargazer)
stargazer(reg_TWFE, se = list(sd_robust),
title = "MQO com Erros Padrao Robustos")
```

	Dependent variable:
	y
⋮	⋮
as.factor(G)2012:as.factor(ano)2012:d_x_t	-0.424*** (0.119)
as.factor(G)2012:as.factor(ano)2013:d_x_t	-0.284** (0.126)
as.factor(G)2012:as.factor(ano)2014:d_x_t	-0.690*** (0.132)
as.factor(G)2014:as.factor(ano)2014:d_x_t	-0.142 (0.147)

as.factor(G)2012:as.factor(ano)2015:d_x_t	-1.024*** (0.139)
as.factor(G)2014:as.factor(ano)2015:d_x_t	-0.289* (0.151)
as.factor(G)2012:as.factor(ano)2016:d_x_t	-1.047 (0.147)
as.factor(G)2014:as.factor(ano)2016:d_x_t	-0.380 (0.157)
as.factor(G)2016:as.factor(ano)2016:d_x_t	-0.093 (0.140)
as.factor(G)2012:as.factor(ano)2017:d_x_t	-1.182 (0.152)
as.factor(G)2014:as.factor(ano)2017:d_x_t	-0.942 (0.170)
as.factor(G)2016:as.factor(ano)2017:d_x_t	-0.388 (0.138)
as.factor(G)2012:as.factor(ano)2018:d_x_t	-1.263 (0.159)
as.factor(G)2014:as.factor(ano)2018:d_x_t	-0.916 (0.196)
as.factor(G)2016:as.factor(ano)2018:d_x_t	-0.594 (0.154)
as.factor(G)2018:as.factor(ano)2018:d_x_t	-0.237 (0.146)
as.factor(G)2012:as.factor(ano)2019:d_x_t	-1.694 (0.171)
as.factor(G)2014:as.factor(ano)2019:d_x_t	-1.138 (0.190)
as.factor(G)2016:as.factor(ano)2019:d_x_t	-0.812 (0.161)
as.factor(G)2018:as.factor(ano)2019:d_x_t	-0.556 (0.169)
as.factor(G)2012:as.factor(ano)2020:d_x_t	-1.743 (0.171)
as.factor(G)2014:as.factor(ano)2020:d_x_t	-1.189 (0.196)
as.factor(G)2016:as.factor(ano)2020:d_x_t	-0.918 (0.167)
as.factor(G)2018:as.factor(ano)2020:d_x_t	-0.810 (0.163)
escolaridade	-0.064*** (0.014)
populacao	-0.00000 (0.00000)
Observations	12,000
R ²	0.908
Adjusted R ²	0.908
Residual Std. Error	1.513 (df = 11983)
F Statistic	7,364.357*** (df = 16; 11983)
Note: *p<0.1; **p<0.05; ***p<0.01	

Cada interação de ‘as.factor(G)’ e ‘as.factor(ano)’ representa um estimador consistente do $ATT(g, t)$; ou seja, ‘as.factor(G)2014:as.factor(ano)2016’ representa o estimador para o $ATT(g = 2014, t = 2016)$. Como estamos estimando um número muito grande de parâmetros em subgrupos reduzidos da amostra, é provável que nosso poder seja baixo. Por isso, podemos observar que vários dos nossos parâmetros não são significantes. Entretanto, como argumentado por Callaway and Sant’Anna

(2021), ao calcular médias ponderadas dos $ATT(g, t)$, conseguimos realizar testes de hipóteses com o tamanho correto e com poder estatístico suficiente, mesmo que para cada $ATT(g, t)$ individualmente não tenhamos poder suficiente.

Agora que possuímos estimadores consistentes para cada $ATT(g, t)$, podemos responder a algumas perguntas interessantes, tais como: qual a média simples dos ATT ? Qual a média do efeito em cada ano da amostra? E em cada grupo? Ou ainda, qual o efeito de ser tratado por 1, 2, ..., n anos?

Para isso, vamos renomear os $ATT(g, t)$ estimados via TWFE e armazená-los em um vetor para facilitar sua manipulação:

Código R

```
library(stringr)

#Selecionando todos os ATT(g,t) [todos os parâmetros que
#contêm d_x_t no nome]
nomes_betas = names(reg_TWFE$coefficients)
nomes_atts = nomes_betas[str_detect(nomes_betas,
  pattern = 'd_x_t')]

#selecionando somente os ATTs e excluindo os nulos
(excluindo as interações com G=0)
ATT_gt = reg_TWFE$coefficients[nomes_atts]
ATT_gt = ATT_gt[!is.na(ATT_gt)]

#Renomeando para ATT(g,t) para facilitar o uso:
grupo <- sub(".*\\(G\\)(\\d{4}).*", "\\1", names(ATT_gt))
ano <- sub(".*\\(ano\\)(\\d{4}).*", "\\1", names(ATT_gt))
names(ATT_gt) <- paste("ATT(", grupo, ",", ano, ")")
, sep = ",")

#printando ATT(g,t)
ATT_gt

## ATT(2012,2012) ATT(2012,2013) ATT(2012,2014)
## -0.424 -0.284 -0.690
## ATT(2014,2014) ATT(2012,2015) ATT(2014,2015)
## -0.142 -1.024 -0.289
## ATT(2012,2016) ATT(2014,2016) ATT(2016,2016)
## -1.047 -0.380 -0.09
## ATT(2012,2017) ATT(2014,2017) ATT(2016,2017)
## -1.182 -0.942 -0.38
## ATT(2012,2018) ATT(2014,2018) ATT(2016,2018)
## -1.263 -0.916 -0.594
## ATT(2018,2018) ATT(2012,2019) ATT(2014,2019)
## -0.237 -1.694 -1.13
## ATT(2016,2019) ATT(2018,2019) ATT(2012,2020)
## -0.812 -0.556 -1.743
## ATT(2014,2020) ATT(2016,2020) ATT(2018,2020)
## -1.189 -0.918 -0.810
```

Em geral, desejamos alguma média ponderada dos $ATT(g, t)$. Um exemplo seria a média do ATT com pesos iguais para todos os grupos e períodos, ou seja, uma média simples dos ATT s. Para realizar testes de hipóteses, como estamos essencialmente realizando uma transformação dos parâmetros estimados via TWFE, podemos usar o método delta.⁷

Dessa forma, podemos construir uma função e aplicar o método delta para qual-

⁷Para mais informações sobre o método delta e como usar veja: Ver Hoef (2012).

quer média ponderada dos $ATTs$,⁸ ou seja,

$$\Gamma = \sum_g \sum_t w(g,t) ATT(g,t)$$

Código R

```
#delta metodo para media ponderada
delta_metodo = function(w, ATT_gt, VAR){
  W = matrix(w, ncol=1)
  Coef = t(W[17:40])%*%ATT_gt
  Grad = W #linear function
  VAR = t(Grad)%*%VAR%*%Grad
  dp = VAR^(1/2)
  t_value = (Coef)/dp
  p_value = 2*(1 - pnorm(abs(t_value)))

  res = c(Coef,dp,t_value, p_value)
  names(res) = c('Coef', 'dp', 't', 'P-Valor')
  return(res)
}
```

Agora, podemos calcular as médias ponderadas dos $ATTs$ para quaisquer conjuntos de pesos, além de realizar testes de hipóteses sobre essas médias. Para responder, por exemplo, à pergunta sobre qual é a média simples dos $ATTs$ e testar a hipótese de que ela é diferente de zero, basta atribuir **pesos iguais** a cada um dos 24 $ATTs$, ou seja, $w(g,t) = 1/24$. Como estamos utilizando o método delta, precisamos calcular o gradiente para todos os parâmetros estimados no TWFE. Os pesos para os 16 parâmetros iniciais serão sempre zero, pois representam os controles e os efeitos fixos, enquanto os 24 últimos parâmetros representarão cada $ATT(g,t)$.

Qual é a média dos $ATTs$ atribuindo peso igual para todos os períodos e grupos?

Como queremos a média simples, definiremos $w = c(rep(16,0), rep(1,24)/24)$ que representa $w = (\underbrace{0, \dots, 0}_{16 \text{ vezes}}, \underbrace{\frac{1}{24}, \dots, \frac{1}{24}}_{24 \text{ vezes}})$:

Código R

```
# Simple mean
delta_metodo(w=c(rep(0,16),rep(1,24)/24), ATT_gt, cov1)

##          Coef          dp          t          P-Valor
## -0.7815036   0.0574231 -13.6095679   0.0000000
```

Ou seja, a média simples dos $ATTs$ é de -0.78% de redução no consumo de água e é estatisticamente significativa.

Qual o efeito médio do tratamento em cada ano da amostra?

Como desejamos calcular o efeito médio em cada ano, devemos atribuir pesos iguais para cada grupo tratado naquele ano específico e zero para todos os outros períodos. Ou seja:

$$w(g,t) = \begin{cases} \frac{1}{\sum_{g'} \mathbb{1}_{g' \leq t}} & g \leq t \\ 0 & c.c. \end{cases}$$

onde $\sum_{g'} \mathbb{1}_{g' \leq t}$ representa exatamente a quantidade de grupos tratados no período t . Se, por exemplo, temos 3 grupos tratados em 2013, atribuiremos peso de $1/3$

⁸Para mais informações sobre o método delta e como usar veja: Ver Hoef (2012).

para cada um deles e zero para todos os outros *ATT*s. Este processo será repetido para todos os períodos.

Código R

```
# Time effect
ATT_Time = matrix(NA, nrow = length(2012:2020), ncol = 4)
rownames(ATT_Time) = 2012:2020
colnames(ATT_Time) = c('Coef', 'sd', 't', 'P-Valor')
for(tt in 2012:2020){

  #Selecionando o ano pelo nome de cada att
  time_att = str_sub(names(ATT_gt), start = 10, end = 13)
  #1 se ATT for p periodos apos a intervencao, 0 caso
  contrario.
  weights = ifelse(time_att %in% as.character(tt), 1, 0)
  #dividindo pelo número de ATTs daquele grupo
  weights = weights/sum(weights)

  #Calculando o teste de hipotese via metodo delta e salvando
  em uma matrix
  ATT_Time[as.character(tt),] = delta_metodo(w=c(rep(0,16),weights),
    ATT_gt, cov1)

}

stargazer(ATT_Time, title = "Time Effect")
```

	Efeito médio do tratamento por ano			
	ATT(t) - Time effect			
	Coef	sd	t	P-Valor
2012	-0.424	0.119	-3.549	0.0004
2013	-0.284	0.126	-2.250	0.024
2014	-0.416	0.106	-3.939	0.0001
2015	-0.656	0.110	-5.986	0
2016	-0.507	0.100	-5.084	0.00000
2017	-0.838	0.104	-8.050	0
2018	-0.753	0.108	-6.969	0
2019	-1.050	0.116	-9.078	0
2020	-1.165	0.117	-9.968	0

Em média, portanto, o efeito foi de -0.424% no ano de 2012, -0.284% em 2013 e assim por diante.

Qual é o efeito médio do tratamento para cada ano adicional de tratamento?

No ano de 2018, por exemplo, temos grupos tratados por 1 período, outros por 3, e outro por 5 períodos. A média dos efeitos por ano agrega todos esses grupos com diferentes tempo de tratamento. Entretanto, uma questão geralmente crucial em análises de políticas públicas é avaliar o efeito gerado por cada período adicional de tratamento. Portanto, inúmeras vezes, desejamos estimar o efeito médio do tratamento ao ser tratado por 1 período, por 2 períodos e assim por diante.

Construímos uma variável que representa a quantidade de períodos desde o tratamento, ou seja, $d = t - g$. Por exemplo, se estivermos no ano de 2016 e o

grupo foi tratado pela primeira vez em 2012 ($g = 2012$), estamos no quarto ano de tratamento, $d = 2016 - 2012 = 4$. Para calcular o efeito médio de ser tratado por 1 período, atribuiremos peso igual para todos os $ATT(g, t)$ tais que $d = t - g = 1$. Em outras palavras, $ATT(2012, 2013)$, $ATT(2013, 2014)$, e assim por diante. Para todos os outros $ATTs$, daremos peso 0.

Código R

```
#Mudando ATT names para 'dynamic'
ATT_names = names(ATT_gt)
g = str_sub(ATT_names, start= 5, end = 8)
tt = str_sub(ATT_names, start= 10, end = 13)
#calculando quantidade de periodos apos a intervencao
d = as.numeric(tt)-as.numeric(g)
names(ATT_gt) = d

# Dynamic effect
ATT_dynamic = matrix(NA, nrow = max(d)+1, ncol = 4)
rownames(ATT_dynamic) = unique(d)
colnames(ATT_dynamic) = c('Coef', 'sd', 't', 'P-Valor')
for(p in unique(d)){

dynamic_att = names(ATT_gt)
#1 se ATT for p periodos apos a intervencao, 0 caso contrário.
weights = ifelse(dynamic_att %in% as.character(d), 1, 0)
#dividindo pelo número de ATTs daquele grupo
weights = weights/sum(weights)

#Calculando o teste de hipotese via metodo delta e salvando
em uma matrix
ATT_dynamic[as.character(p),] = delta_metodo(w=c(rep(0,16)
,weights), ATT_gt, cov1)

}

stargazer(ATT_dynamic, title = "Dynamic effect")
```

Efeito médio por quantidade de anos tratado

	ATT(d) - Dynamic effect			
	Coef	sd	t	P-Valor
0	-0.224	0.069	-3.228	0.001
1	-0.379	0.074	-5.148	0.00000
2	-0.619	0.077	-7.995	0
3	-0.926	0.092	-10.043	0
4	-0.960	0.101	-9.513	0
5	-1.160	0.123	-9.412	0
6	-1.226	0.128	-9.582	0
7	-1.694	0.171	-9.917	0
8	-1.743	0.171	-10.221	0

Assim, observamos que o efeito do tratamento é crescente (em valor absoluto) conforme aumenta o número de períodos de exposição ao tratamento, sendo estatisticamente significante para cada d .

3.2 Callaway and Sant’Anna (2021)

3.2.1 Estimador

Novamente, o objetivo é estimar de forma consistente cada um dos $ATT(g, t)$. Entretanto, utilizar a hipótese de tendências paralelas condicionais não é tão simples para estimar consistentemente o $ATT(g, t)$ incondicional. Para ver isso, defina o $ATT(g, t)$ condicional como:

$$ATT_x(g, t) = E[y_{it}(g) - y_{ig-1}(\infty) | X, d = g]$$

Assumindo que temos tendências paralelas condicionais, como na seção a seguir, conseguiremos estimar de forma consistente esses $ATTs$. Entretanto, o que desejamos é o ATT incondicional, que é uma média ponderada dos $ATTs$ acima.⁹ Ou seja, desejamos estimar:

$$ATT(g, t) = E_x[ATT_x(g, t)].$$

Para conseguirmos recuperar o $ATT(g, t)$, Callaway and Sant’Anna (2021) propuseram três métodos não paramétricos igualmente consistentes. O primeiro é o *outcome regression* (OR), que ajusta o fato de que o ATT incondicional é uma combinação linear dos $ATTs$ condicionais às covariáveis (X), especificando essa relação por meio de uma regressão linear sobre essas covariáveis. O segundo representa o *inverse probability weighting*, que, como o nome sugere, constrói o ATT não condicional usando uma média ponderada dos $ATTs$ condicionais, sendo o peso dado pelo inverso do *propensity score*. Por fim, o *double robust*, que utiliza os dois métodos simultaneamente.

Como argumentado por Callaway and Sant’Anna (2021), apesar de os três métodos serem consistentes, o *double robust* terá melhores propriedades de robustez, principalmente quando há erro na especificação das funções do *propensity score* ou das covariáveis na regressão.

Para simplificar, daqui em diante utilizaremos sempre o grupo *never treated* (aquele que não é tratado em nenhum período) como grupo de controle. Caso optássemos por definir o grupo de controle como os indivíduos pertencentes a grupos que serão tratados em algum momento, mas ainda não tratado até o período específico (*not yet treated*), o raciocínio seria praticamente o mesmo. Além disso, suporemos que não há antecipação.¹⁰ Caso haja, a única diferença é que o período pré-tratamento utilizado será $g - \delta - 1$ em vez de $g - 1$, onde δ é o número de períodos de antecipação.

Outcome Regression (OR)

Neste estimador, utilizaremos uma regressão linear para ajustar a possível heterogeneidade do ATT ao longo das covariadas X . Para isso, usaremos a lei das expectativas iteradas, pela qual o ATT incondicional será uma média ponderada dos $ATTs$ condicionais.

Definimos $G_g = \begin{cases} 1 & d = g \\ 0 & c.c. \end{cases}$. Do Teorema 1 de Callaway and Sant’Anna (2021):

⁹O resultado vem do fato de que o ATT condicional é uma função de X , ou seja, podemos aplicar a lei das expectativas iteradas.

¹⁰Antecipação refere-se ao fato de que os indivíduos conseguem antecipar que serão tratados no futuro. Isso torna-se um problema se eles mudarem seus comportamentos antecipadamente visando o tratamento futuro, enviesando a diferença de médias subjacente no DID.

$$\begin{aligned}
ATT(g, t) &= E_x[ATT_x(g, t)|d = g] \\
&= E_x[(E[y_t - y_{g-1}|X, d = g] - E[y_t - y_{g-1}|X, d = \infty])|d = g] \\
&= E_x[E[y_t - y_{g-1}|X, d = g]|d = g] - E_x[E[y_t - y_{g-1}|X, d = \infty]|d = g] \\
&= E[y_t - y_{g-1}|d = g] - E_x \left[\underbrace{E[y_t - y_{g-1}|X, d = \infty]}_{m_{g,t}(X)} |d = g \right] \\
&= E \left[\frac{G_g}{E[G_g]} (y_t - y_{g-1} - m_{g,t}(X)) \right]
\end{aligned}$$

Portanto, precisaremos estimar o termo de ajuste $m_{g,t}(X)$. Para estimá-lo, Callaway and Sant'Anna (2021) propõem uma regressão linear, ou seja, estimaremos a seguinte regressão na subamostra com somente os controles ($g = \infty$):

$$y_{i,t} - y_{i,g-1} = X'_{i,t}\beta_{g,t} + \varepsilon_t$$

Isso nos permite estimar $m_{g,t}(X) = E[y_t - y_{g-1}|X, d = \infty] = X'_{i,t}\beta_{g,t}$. Com isso, conseguimos estimar o $ATT(g, t)$ dado por:

$$ATT_{or}(g, t) = E \left[\frac{G_g}{E[G_g]} (y_t - y_{g-1} - X'_{i,t}\beta_{g,t}) \right] \quad (15)$$

substituindo a equação acima pelos seus equivalentes amostrais. No entanto, é crucial que $m_{g,t}(X)$ esteja bem especificada, como mencionam os autores:

A abordagem OR exige que os pesquisadores modelem corretamente a evolução dos resultados do grupo de comparação para estimar os efeitos médios do tratamento no tempo para o grupo. Esta abordagem está explicitamente conectada com a suposição de tendências paralelas condicionais necessária na análise de Diferença em Diferenças (DiD), pois essa condição geralmente é expressa em termos de expectativas condicionais (Callaway and Sant'Anna, 2021).¹¹

Inverse Probability Weighting (IPW)

O estimador IPW recupera o ATT incondicional por meio dos $ATTs$ condicionais. Em vez de ajustarmos o ATT pela evolução dos resultados dos controles ($m_{g,t}(X)$), construímos a média ponderada dos $ATTs$ condicionais usando o *propensity score*. A intuição aqui é semelhante ao estimador OR, uma vez que utilizamos os controles para ajustar a heterogeneidade do ATT condicional. Entretanto, em vez de estimarmos uma forma funcional para esse ajuste, usamos os controles ponderados, dando mais peso àqueles que têm uma maior probabilidade de serem tratados dadas suas covariáveis.

$$ATT_{ipw}(g, t) = E \left[\left(\frac{G_g}{E[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\frac{E[p_g(X)C]}{E[p_g(X)C-1]}} \right) \cdot (y_t - y_{g-1}) \right] \quad (16)$$

onde, C é uma dummy igual a 1 se o indivíduo é do grupo de controle (i.e. $g = \infty$) e zero caso contrário. $p_g(X)$ será estimado via *propensity score*.

Neste caso, o *propensity score* deve ser corretamente especificado, como mencionado por Callaway and Sant'Anna (2021):

Por outro lado, a abordagem IPW evita modelar explicitamente a evolução dos resultados do grupo de comparação e, portanto, não depende de o modelo paramétrico ser válido. Em vez disso, a abordagem IPW exige que se modele corretamente a probabilidade condicional de a unidade i estar no grupo g , dados suas covariadas X e que estejam ou no grupo g ou em um grupo de comparação apropriado.

¹¹Tradução independente do original Callaway and Sant'Anna (2021)

Doubly Robust (DR)

O estimador Doubly Robust (DR) combina os métodos IPW e OR. Ele estima o ATT incondicional da seguinte forma:

$$ATT_{dr}(g, t) = E \left[\left(\frac{G_g}{E[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{\frac{E[p_g(X)C]}{E[p_g(X)C-1]}} \right) \cdot (y_t - y_{g-1} - m_{g,t}(X)) \right] \quad (17)$$

A grande vantagem do DR é que, neste caso, se um dos dois métodos estiver bem especificado, mesmo que o outro não esteja, o estimador ainda será consistente para o $ATT(g, t)$. Em outras palavras, se apenas o IPW estiver bem especificado, mas o OR não (ou vice-versa), o DR ainda será um estimador consistente para o ATT (Callaway and Sant'Anna, 2021).

Agrupamento dos $ATT(g, t)$ e inferência

Independentemente do método utilizado, teremos um estimador consistente sob as hipóteses do capítulo 3.2.2, ou seja, podemos agregar os $ATTs$ como a média ponderada que necessitarmos para a análise:

$$\Gamma = \sum_g \sum_{t \geq g} w(g, t) ATT(g, t)$$

Entretanto, diferentemente de Sun and Abraham (2021), não podemos garantir que o método Delta seja aplicável.¹² Callaway and Sant'Anna (2021) propõem realizar a inferência sobre o parâmetro Γ por meio de bootstrap. Um bootstrap totalmente não paramétrico seria adequado; no entanto, os autores sugerem um bootstrap semi-paramétrico, que apresenta menos custo computacional e tem boas propriedades de convergência.¹³ O pacote *did*, do R, implementa integralmente o bootstrap proposto pelos autores.

3.2.2 Hipóteses

Hipóteses

1. (Irreversibilidade do Tratamento) Para todo i , $d_{i,1} = 0$ e para todo i, t :

$$\text{se } d_{i,t-1} = 1 \text{ então, } d_{i,t} = 1.$$

2. (Amostragem aleatória): Todos os indivíduos i foram amostrados de forma aleatória da população.

3. (Antecipação finita) Existe um $0 \leq \delta < \infty$:

$$E[y_t(g)|X, d = g] = E[y_t(0)|X, d = g] \text{ para todo } t < g - \delta$$

4. (Tendências paralelas condicionais - *never treated*): **Para cada** $g \in \mathbf{G}$ e $t \in \{2, \dots, \mathbf{T}\}$ tal que $t \geq g - \delta$, onde δ temos:

$$E[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = E[Y_t(0) - Y_{t-1}(0)|X, C = 1]$$

5. (Tendências paralelas condicionais - *not yet treated*): **Para cada** $g \in \mathbf{G}$ e **cada** $(s, t) \in \{2, \dots, T\} \times \{2, \dots, T\}$ tal que $t \geq g - \delta$ e $t + \delta \leq s < \bar{g}$, onde δ , temos:

$$E[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = E[Y_t(0) - Y_{t-1}(0)|X, D_s = 0, G_g = 0]$$

(Callaway and Sant'Anna, 2021)

¹²Um dos problemas é que as observações de cada um dos estimadores não são independentes. Isso porque, em todos os estimadores os mesmos indivíduos são usados como controle.

¹³Para mais detalhes, veja a seção 4 do paper de Callaway and Sant'Anna (2021).

As hipóteses 1, 2 e 3 são exatamente as mesmas dos outros trabalhos da literatura. Já as hipóteses 4 e 5 correspondem às suposições de tendência paralela, no caso de se usar como grupo de controle os nunca tratados (*never treated*) ou os ainda não tratados (*not yet treated*). A diferença em relação aos outros é que não precisamos impor nenhuma hipótese de homogeneidade para as tendências paralelas condicionais. No entanto, o *trade-off* é que só é possível recuperar o $ATT(g, t)$ caso se especifique corretamente a forma funcional do resultado potencial ou a forma funcional do *propensity score*.

Na literatura aplicada

Graves and Trond (2024) avaliam os efeitos das Leis de Não Discriminação no Emprego (LNDs) na redução de desigualdades sofridas por gays e lésbicas no mercado de trabalho brasileiro. Esse tipo de lei proíbe explicitamente a discriminação em relação à orientação sexual no mercado de trabalho, e foi implementado a nível estadual de maneira escalonada. As primeiras LNDs foram instituídas em 1989 nos estados de Alagoas, Mato Grosso e Sergipe, e, até 2006, por mais outros onze estados. Dessa forma, os grupos de tratamento são compostos por homens gays e mulheres lésbicas residentes em estados que implementaram as LNDs, enquanto grupo de controle inclui homens e mulheres heterossexuais (*never treated*).

Os dados utilizados provêm da Pesquisa Nacional por Amostra de Domicílios (PNAD), abrangendo o período de 1996 a 2011, o que permite uma análise comparativa entre os períodos anterior e posterior à implementação das LNDs em diferentes estados. Para identificar a orientação sexual, os autores construíram uma *proxy* baseada na coabitação com parceiros do mesmo sexo, conforme indicado pela composição dos domicílios. As leis são analisadas em relação a variáveis de resultado como taxa de emprego, formalidade no trabalho, horas semanais trabalhadas, renda mensal e ganhos por hora. As variáveis de controle incluem idade, escolaridade, presença de filhos, raça e efeitos fixos de estado e ano. Os resultados são apresentados de forma desagregada para homens e mulheres, o que permite examinar a heterogeneidade dos efeitos da política entre os públicos feminino e masculino.

Antes da introdução das LNDs, mulheres lésbicas apresentavam um prêmio salarial significativo em relação às mulheres heterossexuais, enquanto homens gays enfrentavam uma penalidade salarial em relação aos homens heterossexuais. Após a implementação das leis, observou-se um aumento significativo na renda dos homens gays, embora não tenha havido redução na informalidade. Para mulheres lésbicas, os resultados foram mistos: houve uma redução na taxa de emprego, mas aquelas que permaneceram no mercado de trabalho experimentaram aumentos salariais, o que evidencia a heterogeneidade de gênero nos efeitos da política de não discriminação.

3.2.3 Código em R com exemplo

Novamente vamos carregar a mesma base de dados utilizada no capítulo anterior, que representa o programa fictício de auxílio para aquisição de fontes renováveis de energia nos municípios de São Paulo.

Código R

```
rm(list=ls())
library(dplyr)

df <- read.csv("base_stagger_did.csv")

head(df)

##   id_estado id_municipio   setor  ano renda
##           1           1001 ServiC'os 2011  55.9
```

```
##          1          1001 ServiC'os 2012  60.0
##          1          1001 ServiC'os 2013  65.0
##          1          1001 ServiC'os 2014  81.0
##          1          1001 ServiC'os 2015  78.0
##          1          1001 ServiC'os 2016  77.0
## escolaridade populacao      y d G Tt d_x_t
##          8.9      117002  9.804006 0 0 0      0
##          8.8      117510 11.145862 0 0 1      0
##          8.8      117754  5.578302 0 0 1      0
##          8.9      118655  4.967904 0 0 1      0
##          8.9      119788  4.748321 0 0 1      0
##          9.0      120792  5.564256 0 0 1      0
```

Agora vamos carregar o pacote *did*, que contém todos os estimadores de Callaway and Sant'Anna (2021), e vamos calcular os $ATT(g, t)$ usando o estimador *double robust* como na equação (17).

Código R

```
library(did)
library(stargazer)

# Rodando Callaway and Sant'ana pelo método 'dr'
ATT_gt = att_gt(yname = 'y',
                 tname = 'ano',
                 idname = 'id_municipio',
                 gname = 'G',
                 xformula = ~ 1 + renda + populacao
                       + escolaridade,
                 data = df,
                 control_group = 'nevertreated',
                 est_method = 'dr'
                 )

ATT_gt

##
## Call:
## att_gt(yname = "y", tname = "ano", idname = "id_municipio",
## gname = "G", data = df, control_group = "nevertreated",
## est_method = "dr")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.
## "Difference-in-Differences with Multiple Time Periods."
## Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021.
## <https://doi.org/10.1016/j.jeconom.2020.12.001>,
## <https://arxiv.org/abs/1803.09015>
##
## Group-Time Average Treatment Effects:
##   Group Time ATT(g,t) Std. Error [95% Simult. Conf. Band]
##   2012 2012  -0.5410      0.1350    -0.9757    -0.1063 *
##   2012 2013  -0.4153      0.1256    -0.8196    -0.0109 *
##   2012 2014  -0.7564      0.1552    -1.2561    -0.2567 *
##   2012 2015  -1.1034      0.1574    -1.6101    -0.5966 *
##   2012 2016  -1.1000      0.1488    -1.5792    -0.6208 *
##   2012 2017  -1.2274      0.1727    -1.7835    -0.6714 *
##   2012 2018  -1.3978      0.1816    -1.9826    -0.8129 *
##   2012 2019  -1.7642      0.1812    -2.3476    -1.1808 *
```

```

## 2012 2020 -1.8876 0.1784 -2.4620 -1.3131 *
## 2014 2012 -0.1807 0.1826 -0.7686 0.4073
## 2014 2013 -0.2479 0.1897 -0.8586 0.3628
## 2014 2014 0.1233 0.1945 -0.5029 0.7494
## 2014 2015 -0.0385 0.1854 -0.6355 0.5585
## 2014 2016 -0.1646 0.1903 -0.7773 0.4480
## 2014 2017 -0.7121 0.2060 -1.3754 -0.0488 *
## 2014 2018 -0.7098 0.2279 -1.4436 0.0241
## 2014 2019 -0.9478 0.2089 -1.6205 -0.2750 *
## 2014 2020 -0.9891 0.2186 -1.6929 -0.2853 *
## 2016 2012 0.0823 0.1435 -0.3798 0.5444
## 2016 2013 -0.1395 0.1471 -0.6130 0.3340
## 2016 2014 0.1614 0.1510 -0.3248 0.6476
## 2016 2015 0.0442 0.1592 -0.4683 0.5568
## 2016 2016 -0.1838 0.1801 -0.7637 0.3961
## 2016 2017 -0.4719 0.1913 -1.0878 0.1441
## 2016 2018 -0.7152 0.2057 -1.3774 -0.0531 *
## 2016 2019 -0.9098 0.1901 -1.5218 -0.2978 *
## 2016 2020 -1.0448 0.1955 -1.6744 -0.4153 *
## 2018 2012 -0.0262 0.1326 -0.4533 0.4009
## 2018 2013 0.0157 0.1456 -0.4531 0.4845
## 2018 2014 0.0516 0.1639 -0.4760 0.5793
## 2018 2015 0.0563 0.1936 -0.5669 0.6796
## 2018 2016 -0.1946 0.1767 -0.7635 0.3742
## 2018 2017 0.1394 0.1587 -0.3714 0.6502
## 2018 2018 -0.2810 0.1895 -0.8910 0.3290
## 2018 2019 -0.6055 0.2158 -1.3003 0.0893
## 2018 2020 -0.8546 0.2178 -1.5560 -0.1532 *
## ---
## Signif. codes: '*' confidence band does not cover 0
##
## P-value for pre-test of parallel trends assumption: 0.69134
## Control Group: Never Treated, Anticipation Periods: 0
## Estimation Method: Doubly Robust

```

Por padrão, o pacote *did* nos retorna o intervalo de confiança de 95% calculado via *bootstrap* semiparamétrico. Caso precisemos calcular via *cluster*, basta adicionar a variável de *cluster* na função *att_gt*, como mostrado anteriormente. É importante lembrar que, como estamos estimando vários parâmetros com possivelmente poucas observações em cada coorte, nem sempre conseguiremos rejeitar a hipótese nula devido à falta de poder estatístico. Além disso, como estamos realizando vários testes de hipótese, é importante tomar cuidado com o problema de *multiple hypothesis test*.

De forma análoga ao que foi feito na seção 3.1, podemos calcular médias ponderadas para responder a perguntas potencialmente interessantes. Isto é:

$$\Gamma = \sum_g \sum_{t \geq g} w(g, t) ATT(g, t)$$

Qual a média dos ATTs entre todos os períodos e grupos?

Por padrão, a função *aggte* do pacote *did* calculará uma média ponderada dos $ATT(g, t)$ pelo tamanho relativo de cada grupo/período na amostra. Como estamos considerando um painel balanceado, vamos ponderar cada grupo por $\frac{N_g}{N \times N_{gt}}$, onde N é o número total de indivíduos tratados, N_g é o número total de indivíduos no grupo g , e N_{gt} é o número total de grupo-períodos, ou seja, o total de ATTs estimados.

Código R

```
# Média dos ATTs

ATT_simples = aggte(ATT_gt, type = 'simple')

ATT_simples

##
## Call:
## aggte(MP = ATT_gt, type = "simple")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.
## "Difference-in-Differences with Multiple Time Periods."
## Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021.
## <https://doi.org/10.1016/j.jeconom.2020.12.001>,
## <https://arxiv.org/abs/1803.09015>
##
##      ATT      Std. Error      [ 95%  Conf. Int.]
##   -0.8493         0.0722    -0.9909      -0.7078 *
##
## ---
## Signif. codes: '*' confidence band does not cover 0
##
## Control Group:  Never Treated,  Anticipation Periods:  0
## Estimation Method:  Doubly Robust
```

Ou seja, a média dos ATTs ponderada pelo tamanho relativo dos grupos é de diminuição de 0.8493 p.p..

Qual o efeito médio do tratamento em cada ano da amostra?

Em muitos casos, é importante analisar o efeito do tratamento a cada ano. Para isso, atribuiremos pesos iguais a cada grupo tratado naquele ano específico. Além disso, por padrão, o *package* did também pondera os efeitos pelo peso relativo de cada grupo naquele ano, considerando o tamanho de cada grupo na construção do estimador.

Código R

```
#Efeito médio por período

att_t = aggte(ATT_gt, type = 'calendar')

summary(att_t)

##
## Call:
## aggte(MP = ATT_gt, type = "calendar")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.
## "Difference-in-Differences with Multiple Time Periods."
## Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021.
## <https://doi.org/10.1016/j.jeconom.2020.12.001>,
## <https://arxiv.org/abs/1803.09015>
##
##
##
## Overall summary of ATT's based on calendar time aggregation:
##      ATT      Std. Error      [ 95%  Conf. Int.]
##   -0.7603         0.0691    -0.8958      -0.6248 *
```

```
##
##
## Time Effects:
## Time Estimate Std. Error [95% Simult. Conf. Band]
## 2012 -0.5410 0.1416 -0.9300 -0.1520 *
## 2013 -0.4153 0.1359 -0.7888 -0.0418 *
## 2014 -0.4632 0.1443 -0.8598 -0.0665 *
## 2015 -0.7484 0.1430 -1.1412 -0.3556 *
## 2016 -0.5868 0.1194 -0.9149 -0.2586 *
## 2017 -0.8611 0.1243 -1.2026 -0.5195 *
## 2018 -0.8333 0.1242 -1.1745 -0.4921 *
## 2019 -1.1249 0.1227 -1.4621 -0.7877 *
## 2020 -1.2689 0.1245 -1.6111 -0.9267 *
## ---
## Signif. codes: '*' confidence band does not cover 0
##
## Control Group: Never Treated, Anticipation Periods: 0
## Estimation Method: Doubly Robust
```

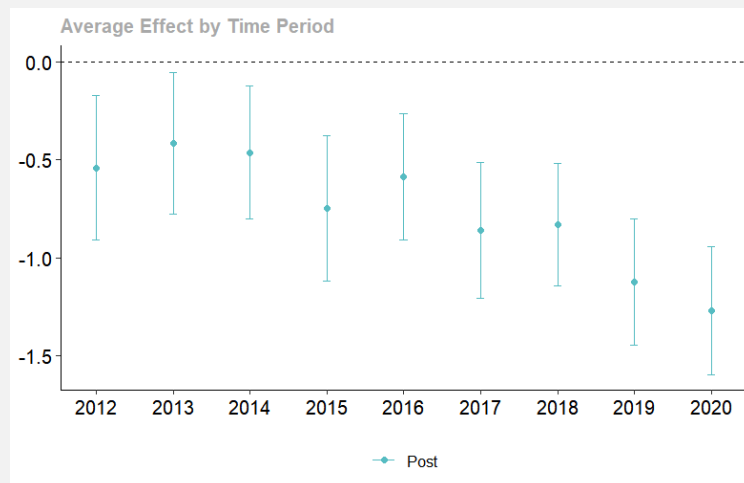
Ou seja, o efeito médio no ano de 2012 entre todos os tratados foi de -0.5410 p.p., enquanto em 2020 foi de -1.2689 . ambos estatisticamente significantes a 95%.

Outro parâmetro estimado que é muito útil em diversas aplicações é o ATT geral, que representa a média simples de cada $ATT(t)$. Nesse caso, ao longo dos 9 anos de tratamento, tivemos um efeito médio anual de -0.7603 .

Também podemos, de forma simples, plotar o gráfico com os efeitos médios estimados para cada ano:

Código R

```
#Plotango gráfico efeito médio por tempo
ggdid(att_t)
```



Qual é o efeito médio do tratamento para cada ano adicional de tratamento?

Na análise anterior, agrupamos todos os tratados no mesmo ano, incluindo efeitos de diferentes exposições ao tratamento. Por exemplo, no ano de 2018, há grupos que foram tratados há 1 período, outros há 3 períodos, e ainda outros há 5 períodos. A média por ano, portanto, combina grupos com diferentes durações de exposição ao programa. No entanto, uma questão frequentemente crucial em análises de políticas públicas é avaliar qual o impacto de cada período adicional de tratamento — ou seja, como o efeito se acumula ao longo do tempo.

Assim, muitas vezes, desejamos estimar qual é o efeito médio do tratamento ao

longo de diferentes períodos de tempo, como o de ser tratado por 1 período, por 2 períodos, e assim por diante. Para isso, construímos uma variável que representa a quantidade de períodos desde o tratamento, ou seja, $d = t - g$. Por exemplo, se estivermos no ano de 2016 e o grupo foi tratado pela primeira vez em 2012 ($g = 2012$), estamos no quarto ano de tratamento, $d = 2016 - 2012 = 4$. Para calcular o efeito médio de ser tratado por 1 período, atribuiremos peso igual para todos os $ATT(g, t)$, onde $d = t - g = 1$. Em outras palavras, consideraremos os $ATT(2012, 2013)$, $ATT(2013, 2014)$, e assim por diante, enquanto atribuímos peso 0 para todos os outros $ATTs$.

Para isso, basta agregarmos como 'dynamic' no pacote *did*.

Código R

```
#Teste Efeito longo prazo x curto prazo

att_d = aggte(ATT_gt, type = 'dynamic')

summary(att_d)
```

```
##
## Call:
## aggte(MP = ATT_gt, type = "dynamic")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.
## "Difference-in-Differences with Multiple Time Periods."
## Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021.
## <https://doi.org/10.1016/j.jeconom.2020.12.001>,
## <https://arxiv.org/abs/1803.09015>
##
##
## Overall summary of ATT's based on event-study/dynamic
## aggregation:
##      ATT      Std. Error      [ 95% Conf. Int.]
## -1.0396      0.079      -1.1945      -0.8847 *
##
##
## Dynamic Effects:
## Event time Estimate Std. Error [95% Simult. Conf. Band]
##      -6    -0.0262    0.1397    -0.4418    0.3895
##      -5     0.0157    0.1544    -0.4439    0.4753
##      -4     0.0670    0.1048    -0.2451    0.3790
##      -3    -0.0416    0.1201    -0.3990    0.3159
##      -2    -0.0576    0.1043    -0.3682    0.2529
##      -1     0.0069    0.1053    -0.3064    0.3202
##       0    -0.2760    0.0866    -0.5337   -0.0183 *
##       1    -0.4142    0.0838    -0.6635   -0.1649 *
##       2    -0.6720    0.0971    -0.9609   -0.3831 *
##       3    -0.9519    0.0976    -1.2424   -0.6614 *
##       4    -0.9949    0.1171    -1.3434   -0.6464 *
##       5    -1.1342    0.1377    -1.5441   -0.7243 *
##       6    -1.2616    0.1423    -1.6851   -0.8381 *
##       7    -1.7642    0.1925    -2.3371   -1.1914 *
##       8    -1.8876    0.1847    -2.4374   -1.3378 *
## ---
## Signif. codes: '*' confidence band does not cover 0
##
## Control Group:  Never Treated,  Anticipation Periods:  0
## Estimation Method:  Doubly Robust
```

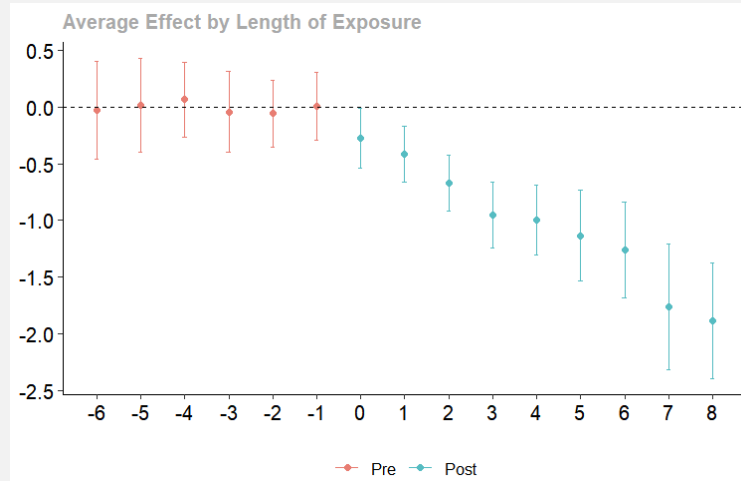
Portanto, no ano da intervenção, em média, observamos um efeito de -0.2760

ponto percentual (p.p.). Um ano após o início da intervenção, o efeito foi de -0.4142 p.p.; após 2 anos, de -0.6720 p.p.; e assim por diante. Além disso, temos a média geral (simples) dos ATTs dinâmicos. No caso do nosso exemplo, -1 p.p.

Podemos plotar um gráfico diretamente por meio do pacote *did*:

Código R

```
# Plotando gráfico efeito dinamico/event study
ggdid(att_d)
```



Este gráfico apresenta um método visual eficaz para avaliar tendências paralelas antes do tratamento (onde $d = -1, -2, \dots, -6$). Por essa razão, é amplamente utilizado em muitos trabalhos científicos.

3.3 De Chaisemartin and d'Haultfoeuille (2020)

3.3.1 Estimador

A abordagem de De Chaisemartin and d'Haultfoeuille (2020) difere consideravelmente dos demais trabalhos discutidos anteriormente, embora, essencialmente, também estime um $ATT(g, t)$ específico. O problema do TWFE é que ele pondera os ATTs de forma inadequada ou, sob outra perspectiva, equivale a aplicar sucessivos modelos DID 2x2 para todos os grupos, o que pode levar ao uso de grupos tratados como controle para outros grupos também tratados, o que distorce a interpretação do efeito médio do tratamento.

Dessa forma, o estimador proposto por De Chaisemartin and d'Haultfoeuille (2020) calcula modelos de diferença-em-diferenças 2 a 2 separadamente e, em seguida, repondera os resultados de forma correta. A principal diferença é que, para cada período, considera-se como grupo tratado apenas os grupos que mudam de status tratamento naquele período, ou seja, aqueles que passam de tratado para controle ou de controle para tratado.

Considere um exemplo simples com três períodos pós-tratamento e três grupos de tratamento, em que o tratamento é escalonado. Ou seja, $t \in \{0, 1, 2, 3\}$ e $G = \{1, 2, 3\}$ e se $g \geq t$, todos do grupo g são tratados em t . Na tabela abaixo, os grupos de controle em cada período são indicados em verde, os tratados em vermelho e "-" significa que aquele(s) grupo(s) não será(ão) usado(s). Além disso, uma outra mudança em relação a Callaway and Sant'Anna (2021) e Sun and Abraham (2021) é que as comparações dois a dois são feitas sempre entre t e $t - 1$.

Diferentemente da literatura padrão de diferença-em-diferenças escalonado, De Chaisemartin and d'Haultfoeuille (2020) possibilitam entradas e saídas do tratamento em qualquer período e em qualquer quantidade de vezes. Entretanto, isso fará com que precisemos ter algumas hipóteses adicionais que serão melhor estudadas na Seção 3.3.2. Uma delas é que, em todos os períodos em que um grupo entrar no tratamento (ou seja, $d_{g,t-1} = 0$ e $d_{g,t} = 1$), deve haver pelo menos um grupo que

t\g	∞	1	2	3
0				
1	•	•	•	•
2	•	-	•	•
3	•	-	-	•

não é tratado naquele período e no período anterior (ou seja, $d_{g',t-1} = d_{g',t} = 0$). E, em todos os períodos em que um grupo deixar de ser tratado (ou seja, $d_{g,t-1} = 1$ e $d_{g,t} = 0$), deve haver pelo menos um grupo que é tratado naquele período e no período passado (ou seja, $d_{g',t} = d_{g',t-1} = 1$).

Essa abordagem é necessária porque, no caso de um grupo que passou a ser tratado, precisamos de pelo menos um grupo que não foi tratado nos dois períodos para servir como controle. Enquanto isso, no caso de um grupo que deixou de ser tratado, precisaremos de um grupo que é tratado nos dois períodos para ser seu controle. O sinal do efeito no segundo caso será invertido, pois estamos comparando "sair" do tratamento contra "ficar". É por isso que De Chaisemartin and d'Haultfoeuille (2020) dividem o estimador em dois componentes: $DID_{+,t}$ e $DID_{-,t}$, que seriam a diferença-em-diferenças padrão e a diferença-em-diferenças "invertida", isso é, multiplicada por -1 .

Por fim, agregamos com pesos corretos cada DID, inclusive com peso negativo para o $DID_{-,t}$. Para construirmos o estimador, definamos:

$$N_{d,d',t} = \sum_{g:d_{g,t}=d, d_{g,t-1}=d'} N_{g,t}$$

$N_{d,d',t}$ é a quantidade de observações no grupo com $d \in \{0,1\}$ no período t e $d' \in \{0,1\}$ no período $t-1$. Ou seja, $N_{0,0,t}$ corresponde à quantidade de observações dos grupos que não eram tratadas em $t-1$ e permaneceram não tratadas no período t . $N_{1,0,t}$ indica a quantidade de observações em todos os grupos que passaram a ser tratadas em t , e assim por diante.

Podemos definir, portanto:

$$DID_{+,t} = \sum_{g:d_{g,t}=1, d_{g,t-1}=0} \frac{N_{g,t}}{N_{1,0,t}} (Y_{g,t} - Y_{g,t-1}) - \sum_{g:d_{g,t}=d_{g,t-1}=0} \frac{N_{g,t}}{N_{0,0,t}} (Y_{g,t} - Y_{g,t-1})$$

$$DID_{-,t} = \sum_{g:d_{g,t}=d_{g,t-1}=1} \frac{N_{g,t}}{N_{1,1,t}} (Y_{g,t} - Y_{g,t-1}) - \sum_{g:d_{g,t}=1, d_{g,t-1}=0} \frac{N_{g,t}}{N_{0,1,t}} (Y_{g,t} - Y_{g,t-1})$$

definamos $N_S = \sum_{(g;t):t \geq 2, D_{g,t} \neq D_{g,t-1}} N_{g,t}$ ¹⁴. Nosso estimador, portanto, será:

$$DID_M = \sum_{t=2}^T \left[\frac{N_{1,0,t}}{N_S} DID_{+,t} + \frac{N_{0,1,t}}{N_S} DID_{-,t} \right]$$

Veja que DID_M é uma média ponderada pelo tamanho relativo de cada grupo em cada período. Isso é possível porque, em cada período, estamos estimando um modelo de diferença-em-diferenças para cada "tipo" de variação no status do tratamento. Ou seja, realizamos modelos DIDs 2x2 separadamente e, em seguida, reponderamos os resultados, em vez de os estimá-los de forma agregada, como no modelo TWFE.

3.3.2 Hipóteses

Entre os artigos da literatura, De Chaisemartin and d'Haultfoeuille (2020) é o que mais flexibiliza o tipo de tratamento ao longo do tempo e entre grupos. Para isso, são necessárias hipóteses um pouco mais fortes. Para mantermos a notação de De Chaisemartin and d'Haultfoeuille (2020), replicaremos a numeração das hipóteses conforme o artigo. Entretanto, algumas hipóteses lá mencionadas são utilizadas

¹⁴ N_S é a quantidade total de unidades que mudam de status de tratamento em t .

apenas para provar o não viés do TWFE. Nesse caso, como não são necessárias para a consistência do DID_M , vamos omiti-las.

Hipóteses

1. **HIPÓTESE 1 (Painel Balanceado de Grupos):** Para todo $(g, t) \in \{1, \dots, G\} \times \{1, \dots, T\}$, $N_{g,t} > 0$.
2. **HIPÓTESE 2 (*Sharp Design*):** Para todo $(g, t) \in \{1, \dots, G\} \times \{1, \dots, T\}$ e $i \in \{1, \dots, N_{g,t}\}$, $D_{i,g,t} = D_{g,t}$.
3. **HIPÓTESE 4 (Exogeneidade Forte):** Para todo $(g, t) \in \{1, \dots, G\} \times \{2, \dots, T\}$,

$$E\left(Y_{g,t}^{(0)} - Y_{g,t-1}^{(0)} | D_{g,1}, \dots, D_{g,T}\right) = E\left(Y_{g,t}^{(0)} - Y_{g,t-1}^{(0)}\right).$$

4. **HIPÓTESE 5 (Tendências paralelas):** Para $t \geq 2$, $E(Y_{g,t}^{(0)} - Y_{g,t-1}^{(0)})$ não varia entre os grupos g .
5. **HIPÓTESE 9 (Exogeneidade Forte para $Y(1)$):** Para todo $(g, t) \in \{1, \dots, G\} \times \{2, \dots, T\}$,

$$E\left(Y_{g,t}^{(1)} - Y_{g,t-1}^{(1)} | D_{g,1}, \dots, D_{g,T}\right) = E\left(Y_{g,t}^{(1)} - Y_{g,t-1}^{(1)}\right).$$

6. **HIPÓTESE 10 (Tendências paralelas para $Y(1)$):** Para $t \geq 2$, $E(Y_{g,t}^{(1)} - Y_{g,t-1}^{(1)})$ não varia entre os grupos g .
7. **HIPÓTESE 11 (Existência de Grupos "Estáveis"):** Para todo $t \in \{2, \dots, T\}$:

- (a) Se existe pelo menos um $g \in \{1, \dots, G\}$ tal que $D_{g,t-1} = 0$ e $D_{g,t} = 1$, então existe pelo menos um $g' \neq g, g' \in \{1, \dots, G\}$ tal que $D_{g',t-1} = D_{g',t} = 0$.
- (b) Se existe pelo menos um $g \in \{1, \dots, G\}$ tal que $D_{g,t-1} = 1$ e $D_{g,t} = 0$, então existe pelo menos um $g' \neq g, g' \in \{1, \dots, G\}$ tal que $D_{g',t-1} = D_{g',t} = 1$.

8. **HIPÓTESE 12 (Independência da Média entre o Resultado de um Grupo e os Tratamentos de Outros Grupos):** Para todos g e t , $E(Y_{g,t}(0)|D) = E(Y_{g,t}(0)|D_g)$ e $E(Y_{g,t}(1)|D) = E(Y_{g,t}(1)|D_g)$.

De Chaisemartin and d'Haultfoeuille (2020)

A hipótese 1 garante que os grupos não "sumam" da base de dados durante o período analisado. Já a hipótese 2 garante que todos os indivíduos em (g, t) terão o mesmo status de tratamento. Entretanto, De Chaisemartin and d'Haultfoeuille (2020) propõem um estimador consistente também quando essa hipótese é flexibilizada. As hipóteses 1 e 2 estão presentes em todos os outros trabalhos da literatura, ainda que de forma não explícita.

As hipóteses 4 e 5 são muito semelhantes às hipóteses de Sun and Abraham (2021). Juntas, elas estabelecem que a tendência paralela deve ser independente do tratamento e igual pra todos os grupos. Isso implica, por exemplo, que o momento em que um indivíduo será tratado não pode ser influenciado por choques passados. Essa é uma das principais diferenças em relação ao trabalho de Callaway and Sant'Anna (2021). Perceba que não temos a restrição adicional de que as tendências paralelas devem ser constantes ao longo do tempo. A hipótese 5 e 10 impõem restrições de que as tendências paralelas devem ser constantes entre os grupos, tanto para o resultado potencial para o caso não tratado (5) quanto para o caso tratado (10). Ou seja, elas podem variar ao longo do tempo, mas não podem ser dependentes, em média, do tratamento ou variar ao longo dos grupos.

A hipótese 9 é, essencialmente, a mesma que a hipótese 4, porém aplicada ao resultado potencial em caso de tratamento. Ou seja, além de exigir que o tratamento seja independente, em média do resultado potencial de não tratamento, como na hipótese 4, ela também deve ser independente para o resultado potencial caso seja tratado. Trata-se, portanto, de uma hipótese mais forte do que aquelas adotadas em Callaway and Sant’Anna (2021) e em Sun and Abraham (2021).

A hipótese 11 exige que, para todo período em que haja ao menos um grupo que passe de não tratado para tratado, haja também pelo menos um grupo que permaneça não tratado nos dois períodos considerados. O mesmo raciocínio vale para os períodos em que algum grupo deixe de ser tratado, ou seja, passe de tratado para não tratado. Neste caso, é necessário haver ao menos um grupo que continue sendo tratado nos dois períodos em questão.

Essa condição é fundamental para garantir que, em todos os períodos, exista um grupo de controle válido para fins de comparação. Por fim, a hipótese 12 estabelece que, condicional ao tratamento, os resultados potenciais dos grupos são independentes entre si.

Na literatura aplicada

Dias and Fontes (2024) investigam os efeitos da Reforma Psiquiátrica Brasileira de 2002, que introduziu os Centros de Atenção Psicossocial (CAPS) como alternativa comunitária ao modelo de internações hospitalares em psiquiatria. Como os CAPS foram implementados a nível municipal e de maneira gradual, os autores puderam usar um modelo de diferença-em-diferenças escalonado usando o estimador de De Chaisemartin and d’Haultfoeuille (2020).

O objetivo foi avaliar os impactos dessa política na prestação de cuidados de saúde mental, hospitalizações psiquiátricas, mortalidade, despesas públicas e indicadores sociais, como taxas de criminalidade. A análise utilizou dados de 5.180 municípios brasileiros de 2002 a 2016, e foram examinados impactos em indicadores de saúde, finanças públicas e taxas de homicídios, com controles para tendências prévias e características socioeconômicas locais.

Os CAPS foram abertos em cada município mediante solicitação ao Governo Federal, a critério do próprio município. Os primeiros CAPS (nos termos da reforma) foram inaugurados em 2002. Até 2016, entre 50 a mais de 200 municípios passaram a contar com seu primeiro CAPS, a cada ano. Dessa forma, cada ano corresponde a um grupo de tratamento (conjunto de municípios que receberam um CAPS naquele ano). A apresentação dos resultados foi feita em formato de *event study* (ATTs para k períodos após o recebimento do tratamento), e ponderada pelo tamanho relativo de cada coorte.

A criação dos CAPS aumentou a oferta de serviços ambulatoriais de saúde mental e reduziu hospitalizações psiquiátricas, especialmente de longa duração e entre pacientes com esquizofrenia. Apesar disso, a economia dos gastos com internações não compensou os custos do programa, gerando um impacto financeiro líquido negativo – ao contrário do que se propunha com a reforma. A política não reduziu significativamente a mortalidade por causas relacionadas à saúde mental e esteve associada a um aumento nas taxas de homicídios. Esse resultado pode estar alinhado à hipótese de Penrose, que associa a redução de internações com maior exposição de pacientes vulneráveis à violência.

3.3.3 Código em R com exemplo

Primeiro, vamos carregar a base de dados.

Código R

```
rm(list=ls())
library(dplyr)

df <- read.csv("base_stagger_did.csv")

head(df)
```

##	id_estado	id_municipio	setor	ano	renda
##	1	1001	ServiC'os	2011	55.9
##	1	1001	ServiC'os	2012	60.0
##	1	1001	ServiC'os	2013	65.0
##	1	1001	ServiC'os	2014	81.0
##	1	1001	ServiC'os	2015	78.0
##	1	1001	ServiC'os	2016	77.0

##	escolaridade	populacao	y	d	G	Tt	d_x_t
##	8.9	117002	9.804006	0	0	0	0
##	8.8	117510	11.145862	0	0	1	0
##	8.8	117754	5.578302	0	0	1	0
##	8.9	118655	4.967904	0	0	1	0
##	8.9	119788	4.748321	0	0	1	0
##	9.0	120792	5.564256	0	0	1	0

Em seguida, carregaremos o pacote *DIDmultiplegt* para usarmos o modelo de De Chaisemartin and d'Haultfoeuille (2020) no R.

Código R

```
library(DIDmultiplegt)
```

Já com o pacote na memória, vamos usar a função *did_multiplegt*, que nos retornará o efeito médio de tratamento. Como notado, tivemos um efeito médio ao longo de toda a mostra de $-0,2499679$. Por padrão, esta função não calcula o desvio-padrão do estimador.

Código R

```
did = did_multiplegt(
  df = df,
  Y = "y",
  G = "G",
  T = "ano",
  D = "d_x_t",
  controls = c("renda", "escolaridade", "populacao"),
  average_effect = T
)

did
```

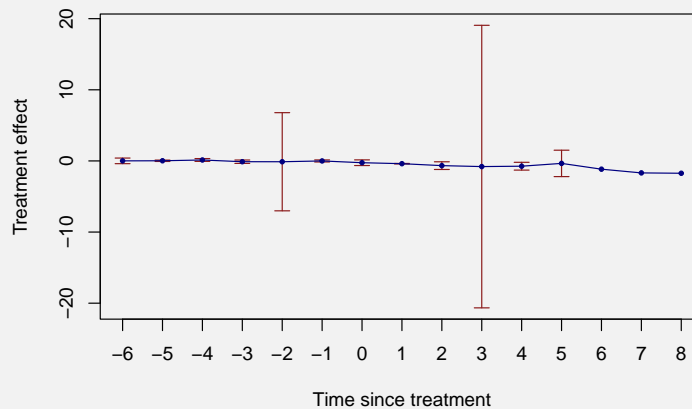
##	\$effect
##	treatment
##	-0.2499679
##	
##	\$N_effect
##	[1] 3660
##	

```
## $N_switchers_effect  
## [1] 720
```

Agora, se quisermos o efeito dinâmico como fizemos com Callaway and Sant’Anna (2021); Sun and Abraham (2021), devemos igualar o parâmetro `dynamic` ao número de períodos pós-tratamento utilizados. No nosso caso, temos 8 períodos pós-tratamento, então `dynamic = 8`. Além disso, para calcularmos os desvios-padrão, precisamos definir `brep`, que é o número de replicações do *bootstrap*. O ideal seria utilizar um número alto de replicações, algo em torno de 1000 ou mais. Para fins de exemplificação, no entanto, usaremos `brep = 2`, pois a implementação do software não funciona muito bem e não é eficiente.

Código R

```
did = did_multiplegt(  
  df = df,  
  Y = "y",  
  G = "G",  
  T = "ano",  
  D = "d_x_t",  
  controls = c("renda", "escolaridade", "populacao"),  
  placebo = 6,  
  dynamic = 8,  
  brep = 2  
)
```



3.4 Borusyak et al. (2024)

3.4.1 Estimador

Novamente, o objetivo será estimar de forma consistente o $ATT(g, t)$ condicional definido por:

$$ATT(g, t) = E[y_{it}(g) - y_{it}(0) | X, d = g]$$

e depois calcular média ponderada dos $ATT(g, t)$ que for mais conveniente para a nossa análise, como fizemos em Callaway and Sant’Anna (2021).

Borusyak et al. (2024) segue pelo mesmo caminho. Entretanto, os autores propõem duas formas alternativas de estimação, sendo que uma requer hipóteses muito mais restritivas, mas é o mais eficiente entre todos os estimadores lineares. Como ambos são, de certa forma, distintos entre si, iremos dividir esta seção em 2 subseções: *Estimação Eficiente*, onde mostraremos o estimador eficiente porém mais

restritivo, e *Estimador de Imputação*, onde mostraremos o principal estimador em Borusyak et al. (2024).

Estimação Eficiente

O grande problema dos modelos de diferença-em-diferenças escalonado usando TWFE é que a heterogeneidade dos efeitos do tratamento pode gerar um estimador que representa uma média ponderada dos $ATT(g, t)$ com pesos pouco informativos ou até negativos.

Para resolver o problema, Borusyak et al. (2024) propõem assumir uma estrutura paramétrica para a heterogeneidade do tratamento. Ou seja, parte-se da premissa de que a forma da heterogeneidade do efeito do tratamento é conhecida, e então estima-se um modelo TWFE impondo essa estrutura paramétrica conhecida.

Caso fôssemos estimar um TWFE padrão, teríamos:

$$y_{it} = \tilde{\alpha}_i + \tilde{\gamma}_t + \tilde{\tau}D_{it} + \varepsilon_{it}$$

onde D_{it} é uma variável indicadora que assume o valor 1 se o indivíduo i é tratado no período t , e zero caso contrário.

Novamente, o problema decorre da heterogeneidade em $\tilde{\tau}$. Portanto, vamos assumir que essa heterogeneidade pode ser descrita como:

$$\tau = \Gamma\theta$$

onde θ é um vetor $(N_1 - M) \times 1$ e Γ uma matriz $N_1 \times (N_1 - M)$ com *rank* completo e conhecida. Nesse caso, θ é desconhecido e contém nossos parâmetros de interesse, que são os diferentes efeitos de tratamento. N_1 ,¹⁵ representa o total de observações tratadas (e, assim, o número máximo de efeitos de tratamento distintos possíveis), e M o número de restrições sobre os parâmetros. De forma que $N_1 - M$ é o número de parâmetros livres a serem estimados. Por exemplo, se assumirmos que o efeito é constante entre todos os grupos e períodos ($ATT(g, t) = ATT(g', t') = \theta_1$ para todo g, g' e para todo $t \geq g$ e $t' \geq g'$), então, $N_1 - M = 1$, $\theta = \theta_1$ e $\Gamma = \underbrace{(1, \dots, 1)'}_{N_1 \text{ vezes}}$.

Apesar de podermos escrever diversos tipos de heterogeneidade com essa versão paramétrica, é preciso assumir que a heterogeneidade é constante em alguma dimensão e, principalmente, que ela é conhecida. Dessa forma, podemos calcular o MQO restrito a essa heterogeneidade:

$$y_{it} = \alpha_i + \gamma_t + D_{it}\Gamma'_{it}\theta + \varepsilon_{it} \quad (18)$$

onde Γ_{it} é conhecido. Isso equivale a saturar o modelo com indicadoras para cada subgrupo onde o tratamento é heterogêneo.

Borusyak et al. (2024) mostram que esse estimador será o mais eficiente entre todos os estimadores lineares para θ . Entretanto, se a especificação escolhida para a heterogeneidade for incorreta, o estimador será inconsistente. Portanto, os autores argumentam que, a menos que exista uma justificativa muito forte ex-ante de que a heterogeneidade terá uma forma exata, é preferível usar o estimador de imputação, mesmo que não seja o mais eficiente.

Estimador de Imputação

Ao estimar o $ATT(g, t)$, o que não se observa é $y_{it}(0)$ para quem os indivíduos que são tratados no período t . O método de imputação consiste em estimar $y_{it}(0)$ para esses indivíduos, e partir dessa estimativa, calcular o $ATT(g, t)$.

Para isso, utilizaremos a hipótese de que $E[y_{it}(0)] = A'_{it}\lambda_i + X'_{it}\gamma$. Podemos escrever:

$$y_{it}(0) = A'_{it}\lambda_i + X'_{it}\gamma + \varepsilon_{it} \quad (19)$$

¹⁵Na notação de Borusyak et al. (2024) $N_1 = |\Omega_1|$, $\Omega_1 = \{it \in \Omega | D_{it} = 1\}$. Isto é, Ω_1 representa o conjunto de observações indivíduo-tempo em que o indivíduo é tratado (e Ω o conjunto de todas as observações), e N_1 é o número de elementos desse conjunto.

onde A_{it} e X_{it} são vetores de variáveis observáveis e conhecidos.

Veja que, se $A_{it} = 1$ para todo i e t , temos o efeito fixo de indivíduo; e se X_{it} contiver uma indicadora para cada período, temos o efeito fixo de tempo. Além disso, A'_{it} pode conter covariáveis interagidas com uma indicadora para cada grupo, de forma que a hipótese 1 da próxima seção torna-se muito próxima da hipótese de tendências paralelas que podem variar entre grupos.

Defina Ω_0 como o conjunto de todas as observações não tratadas e Ω_1 o conjunto de todas as observações tratadas.

1. Estimamos $\hat{\lambda}_i$ e $\hat{\gamma}$ via equação 19 por MQO em todas as observações não tratadas (i.e., $it \in \Omega_0$)
2. Para toda observação tratada ($it \in \Omega_1$), calculamos:

$$\hat{y}_{it}(0) = A'_{it}\hat{\lambda}_i + X'_{it}\hat{\gamma}$$

3. Calculamos o efeito do tratamento para cada observação tratada ($it \in \Omega_1$):

$$\hat{\tau}_{it} = y_{it} - \hat{y}_{it}(0)$$

4. Por fim, agregamos os efeitos com algum peso arbitrário, tal que $\sum_{it} w_{it} = 1$ e $\forall i, t, w_{it} \geq 0$. Ou seja:

$$\hat{\tau} = \sum_{it} w_{it} \hat{\tau}_{it}$$

Esse método calcula, indiretamente, algo muito semelhante ao $ATT(g, t)$. É como se calculássemos cada um dos $ATT(g, t)$ ao reponderar os $\hat{\tau}_{it}$ por $w_{it} / \sum_{i \in g} w_{it}$ para cada observação no grupo g . Se assumirmos que o peso é igual para todos os indivíduos de um mesmo grupo em cada período, estaremos, na prática, estimando exatamente o que propõem Callaway and Sant'Anna (2021): uma média ponderada dos $ATT(g, t)$.

Inferência

Para fins de inferência, Borusyak et al. (2024) mostram que é impossível estimar consistentemente a variância paramétrica de $\hat{\tau}_{it}$ e, por consequência, de $\hat{\tau}$. Entretanto, defina v_{it} os pesos relativos para cada observação (i, t) tal que $\hat{\tau} = \sum_{it} v_{it} y_{it}$,¹⁶ sob as hipóteses da próxima seção, a variância dada por:

$$\hat{\sigma}_w^2 = \sum_i \left(\sum_{t; it \in \Omega_1} v_{it} \tilde{\varepsilon}_{it} \right)^2, \quad \tilde{\varepsilon}_{it} = y_{it} - A'_{it}\hat{\lambda}_i - X'_{it}\hat{\gamma} - D_{it}\hat{\tau}_{it} \quad (20)$$

Trata-se de uma estimativa conservadora da variância assintótica. Isto é, ela converge para uma variância maior ou igual à variância verdadeira. Assim, se rejeitarmos a hipótese nula de efeito zero utilizando essa variância, teremos uma evidência ainda mais forte de que o efeito é, de fato, diferente de zero. Em contrapartida, nosso poder estatístico será menor.

¹⁶Para qualquer vetor de pesos w tal que $\hat{\tau} = \sum_{it} w_{it} \hat{\tau}_{it}$, existe um vetor de pesos v_{it} tal que $\hat{\tau} = \sum_{it} v_{it} y_{it}$. Veja Borusyak et al. (2024), seção 4.2 para mais detalhes.

3.4.2 Hipóteses

Hipóteses

1. **HIPÓTESE 1 (Modelo Geral de $Y(0)$):** Para todo $it \in \Omega_0$,

$$E[Y_{it}^{(0)}] = A'_{it}\lambda_i + X'_{it}\gamma$$

onde λ_i é um vetor de parâmetros não-observados específicos do indivíduo, δ é um vetor de parâmetros não-observados associados a covariáveis comuns, e A_{it} e X_{it} são vetores conhecidos e não-estocásticos.

2. **HIPÓTESE 2 (No-anticipation effects):** $Y_{it} = Y_{it}(0)$ para todo $it \in \Omega_0$
3. **HIPÓTESE 3 (Model of causal effects):** $\tau = \Gamma\theta$ onde θ é um vetor de parâmetros desconhecidos de dimensão $(N_1 - M) \times 1$ e Γ é uma matriz conhecida de dimensão $N_1 \times (N_1 - M)$ de posto completo.
4. **HIPÓTESE 5 (Termos de Erro com Clusters):** Os termos de erro ϵ_{it} são independentes entre as unidades i e têm variância limitada, $Var[\epsilon_{it}] \leq \bar{\sigma}^2$ para $i, t \in \mathbb{N}$ uniformemente.
5. **HIPÓTESE 6 (Condição de Herfindahl):** Ao longo da sequência assintótica, $v_H^2 \equiv \sum_{it \in \mathbb{N}} |v_{it}|^2 \rightarrow 0$, para os pesos v_{it} no estimador linear não viesado $\hat{\tau}_w = \sum_{it \in \mathbb{N}} v_{it} Y_{it}$.

(Borusyak et al., 2024)

A princípio, as hipóteses 1 e 3 podem parecer mais restritivas do que aquelas assumidas em Sun and Abraham (2021), Callaway and Sant'Anna (2021) e De Chaisemartin and d'Haultfoeuille (2020), por exigirem uma forma paramétrica para o resultado potencial em caso de não tratamento. Entretanto, essa forma funcional é bastante geral (abarcando, por exemplo, o modelo TWFE, ou seja, a hipótese de tendências paralelas). Além disso, ao escolher corretamente os vetores A_{it} e X_{it} , é possível especificar uma estrutura de tendências paralelas condicionais que pode ser menos restritiva do que aquelas exigidas por Sun and Abraham (2021) e De Chaisemartin and d'Haultfoeuille (2020). Não é necessário assumir homogeneidade dos efeitos em relação às covariadas. Ou seja, não é necessário que $E[y_{it}(\infty) - y_{i0}(\infty)|\mathbf{G} = \mathbf{g}, \mathbf{X}] = E[y_{it}(\infty) - y_{i0}(\infty)|\mathbf{G} = \mathbf{g}]$.

A hipótese 2 corresponde à hipótese padrão de não antecipação e pode ser relaxada, desde que haja um número suficientemente grande de observações não tratadas.

As hipóteses 5 e 6 são necessárias para a inferência sobre o estimador τ . Elas garantem que a distribuição assintótica dele será normal e que seja possível construir estimadores conservadores da variância.

A principal restrição imposta por essas hipóteses é que as observações devem ser independentes entre os indivíduos, mesmo dentro do mesmo grupo, e que a norma do vetor de pesos deve convergir para zero. Essa última condição é semelhante à hipótese implícita em Sun and Abraham (2021), segundo a qual o tamanho de todos os grupos deve aumentar com o crescimento de N .

3.4.3 Código em R com exemplo

Vamos construir nosso próprio código para rodar o modelo proposto por Borusyak et al. (2024).¹⁷ Primeiramente, vamos carregar a base de dados do exemplo trabalhado até aqui.

¹⁷Poderíamos utilizar o pacote *didimputation* no R. Entretanto, apesar de abarcar os modelos com escalonamento no tratamento, ele atualmente (julho/2024) funciona com muitos erros nessas ocasiões.

Código R

```
rm(list=ls())
library(dplyr)

df <- read.csv("base_stagger_did.csv")

head(df)
```

	id_estado	id_municipio	setor	ano	renda	y
## 1	1	1001	ServiC'os	2011	55.9	8.9
## 2	1	1001	ServiC'os	2012	60.0	8.8
## 3	1	1001	ServiC'os	2013	65.0	8.8
## 4	1	1001	ServiC'os	2014	81.0	8.9
## 5	1	1001	ServiC'os	2015	78.0	8.9
## 6	1	1001	ServiC'os	2016	77.0	9.0

	escolaridade	populacao	d	G	Tt	d_x_t
## 1	117002	9.804006	0	0	0	0
## 2	117510	11.145862	0	0	1	0
## 3	117754	5.578302	0	0	1	0
## 4	118655	4.967904	0	0	1	0
## 5	119788	4.748321	0	0	1	0
## 6	120792	5.564256	0	0	1	0

Agora, seguiremos cada um dos passos para construir o estimador de imputação.

1. Estimamos $\hat{\lambda}_i$ e $\hat{\gamma}$ via equação 19 por MQO em todas as observações não tratadas (i.e., $it \in \Omega_0$).

No nosso caso, a especificação conterá efeitos fixos de tempo e indivíduo, além das covariáveis escolaridade e população. Assim, estimaremos, com todos os indivíduos não tratados ($i \in \Omega_0$):

$$y_{it}(0) = \alpha_g + \delta_t + \gamma_1 \text{escolaridade}_{it} + \gamma_2 \text{populacao}_{it} + \varepsilon_{it}$$

Vamos filtrar a base para manter apenas os não tratados e chamá-la de $df0$:

Código R

```
df$d_x_t = ifelse(df$ano <= df$G | df$G == 0, 0, 1)

df0 = df[df$d_x_t == 0,]
```

Agora, vamos estimar a equação acima por MQO:

Código R

```
reg0 = lm(y ~ as.factor(ano) + as.factor(G)
          + escolaridade + populacao,
          data = df0)

library(stargazer)
stargazer(reg0, title = "Passo 1")
```

Modelo $y(0)$ para imputation

	Dependent variable:
	y
as.factor(ano)2012	-1.220*** (0.059)
as.factor(ano)2013	-2.552*** (0.064)
as.factor(ano)2014	-3.751*** (0.064)
as.factor(ano)2015	-4.937*** (0.067)
as.factor(ano)2016	-6.289*** (0.067)
as.factor(ano)2017	-7.546*** (0.072)
as.factor(ano)2018	-8.841*** (0.073)
as.factor(ano)2019	-9.944*** (0.081)
as.factor(ano)2020	-11.208*** (0.081)
as.factor(G)2012	6.372*** (0.076)
as.factor(G)2014	6.166*** (0.072)
as.factor(G)2016	6.102*** (0.051)
as.factor(G)2018	5.985*** (0.045)
escolaridade	-0.062*** (0.016)
populacao	-0.00000 (0.00000)
Constant	10.568*** (0.146)
Observations	8,280
R ²	0.933
Adjusted R ²	0.932
Residual Std. Error	1.437 (df = 8264)
F Statistic	7,615.220*** (df = 15; 8264)
Note:	*p<0.1; **p<0.05; ***p<0.01

2. Para toda observação tratada ($it \in \Omega_1$), calculamos:

$$\hat{y}_{it}(0) = A'_{it}\hat{\lambda}_i + X'_{it}\hat{\gamma}$$

Código R

```
#Calculando y0 com o modelo anterior
df$y0_hat = predict(reg0, df)
```

3. Calculamos o efeito do tratamento para cada observação tratada ($it \in \Omega_1$):

$$\hat{\tau}_{it} = y_{it} - \hat{y}_{it}(0)$$

Código R

```
# Calculando Tau via imputation
df$tau = 0
df[df$d_x_t==1, 'tau'] = df[df$d_x_t==1, 'y']
                        - df[df$d_x_t==1, 'y0_hat']
```

Por fim, vamos ao passo 4. Juntamente, iremos calcular a variância proposta por Borusyak et al. (2024). Para saber mais detalhes sobre a implementação, consulte o capítulo 4.3 e as proposições 3 e 4 do apêndice de Borusyak et al. (2024).

4. Por fim, agregaremos os efeitos com algum vetor de pesos arbitrário tal que $\sum_{it} w_{it} = 1$ e $\forall i, t, w_{it} \geq 0$. Ou seja:

$$\hat{\tau} = \sum_{it} w_{it} \hat{\tau}_{it}$$

Código R

```
v0 <- function(Z0, Z1, w){

  if(dim(Z0)[2] != dim(Z1)[2]) {
    print('ERROR: Z0 e Z1 precisam ter o mesmo número
    'de colunas.')
    return('')
  }

  dim(Z0 %% solve(t(Z0) %% Z0))

  v0 = - Z0 %% solve(t(Z0) %% Z0) %% t(Z1) %% t(w)
  return(v0)
```

```

summarize_tau <- function(df, id_name, G_name, D_name,
                          formula, restrict = FALSE){

  library(dplyr)
  names_var = names(df)

  for (var in c('tau', id_name, G_name, D_name)){
    if(!var %in% names_var){
      print(c('ERRO: df precisa conter', var))
      return('')
    }
  }

  df$G = df[,G_name]
  df$id = df[,id_name]
  df$D = df[,D_name]

  df1 = df[df$D==1,]
  df0 = df[df$D==0,]

  # Calculando Tau
  df1$w = rep(1/length(df1$tau), length(df1$tau))
  w = matrix(df1$w, nrow = 1)
  tau_it = matrix(df1$tau, ncol = 1)

  TAU = w %*% tau_it

```

```

#Calculando variância de Tau

if(restrict){
  #Calculating tau_g
  df_sum <- df1 %>% group_by(id, G) %>%
    summarise(w = sum(w),
              wtau = sum(w*tau))

  df_g <- df_sum %>% group_by(G) %>%
    summarise(numerador = sum(w*wtau),
              denominador = sum(w^2))

  df_g$tau_g = df_g$numerador/df_g$denominador

  df_res = merge(df, df_g[,c('G', 'tau_g')], by = 'G',
                 all.x = T)
  df_res[is.na(df_res$tau_g), 'tau_g'] = 0
  df_res$res = df_res$y - df_res$y0_hat -
    df_res$d_x_t * df_res$tau_g
} else {
  df_g = NULL
  df_res = df
  df_res$res = df_res$y - df_res$y0_hat -
    df_res$d_x_t * df_res$tau_g
}

#Calculando os pesos implícitos vit
Z = model.matrix(formula, data = df)
Z0 = Z[df$D == 0,]
Z1 = Z[df$D == 1,]

v = rep(NA, dim(Z)[1])

v[df$D == 1] = w
v[df$D == 0] = v0(Z0, Z1, w)

df_res$v = v

#Calculando a variância bia eq 7 borusyak et al.

#Clusterizando no tempo
df_clustered = df_res %>% group_by(id, G) %>%
  summarise(v_e = sum(v*res))

var = sum(df_clustered$v_e^2)
dp = var^(1/2)

t = sqrt(dim(df_clustered)[1]) * TAU/dp
p_value = pnorm(abs(t), lower.tail = F)

return(c(list(TAU=TAU, Sigma2 = var, Sigma = dp, t = t,
p_value=p_value, df = list(df_input = df, df_res = df_res,
df_clustered = df_clustered))))
}

```

```

RES = summarize_tau(df,
  id_name = 'id_municipio',
  G_name = 'G',
  D_name = "d_x_t",
  formula = y ~ as.factor(ano) + as.factor(G)
    + escolaridade + populacao,
  restrict = TRUE)

library(stringr)
TAU = str_pad(round(RES$TAU, 4), 8, side = 'both')
VAR = str_pad(round(RES$Sigma2, 4), 8, side = 'both')
t = str_pad(round(RES$t, 2), 8, side = 'both')
P_v = str_pad(round(RES$p_value, 4), 8, side = 'both')

N1 = str_pad('TAU', 8, side = 'both')
N2 = str_pad('VAR', 8, side = 'both')
N3 = str_pad('t', 8, side = 'both')
N4 = str_pad('P-Valor', 8, side = 'both')

cat( N1, "|", N2, "|", N3, "|", N4, "\n")

cat(TAU, "|", VAR, "|", t, "|", P_v, "\n")

```

```

##    TAU    |    VAR    |    t    | P-Valor
## -0.7899 |  0.0154 | -220.24 |    0

```

Fonte: Construído pelo próprio autor usando Borusyak et al. (2024).

3.5 Diferença entre os estimadores

Nesta seção, vamos dividir a comparação entre os estimadores em dois pontos principais. Primeiro, discutiremos as hipóteses de cada método. Em seguida, compararemos a facilidade de uso de cada um deles.

3.5.1 Hipóteses

A Tabela 1 compara os modelos quanto às hipóteses de tendências paralelas comuns na literatura. Já a Tabela 4 resume outros casos comuns. O símbolo ✓ indica que as hipóteses do modelo abrangem aquele item específico, enquanto o símbolo ✗ indica que, se aquela situação ocorrer, ao menos uma das hipóteses do modelo não será satisfeita.

Hipóteses de Tendências Paralelas

Nesta seção, discutiremos as diferenças entre as hipóteses de tendências paralelas para cada um dos métodos estudados. Se **não houver covariadas** no modelo, por definição, as tendências paralelas serão iguais para todos os grupos. Para visualizar isso, podemos observar que:

Se $\forall g \in G$,

$$E[y_{it}(\infty) - y_{it-1}(\infty)|d = g] = E[y_{it}(\infty) - y_{it-1}(\infty)|d_i = \infty],$$

então, $\forall g, g' \in G$,

$$E[y_{it}(\infty) - y_{it-1}(\infty)|d = g] = E[y_{it}(\infty) - y_{it-1}(\infty)|d = g'].$$

Além disso, **sem covariadas**, o status de tratamento será independente das tendências paralelas (inclusive anteriores), já que elas serão iguais para todos os grupos. Nestes casos, todos os estimadores serão consistentes. Portanto, nesta

seção, focaremos na comparação entre modelos quando temos um vetor de controles X a que as tendências paralelas (TPs) devem ser satisfeitas condicionalmente. Ou seja:

$$E[y_{it}(\infty) - y_{it-1}(\infty)|d = g, X] = E[y_{it}(\infty) - y_{it-1}(\infty)|d_i = \infty, X]$$

Agora, vamos considerar alguns casos particulares da hipótese de tendências paralelas padrão e compará-los entre os métodos utilizados, verificando se suas hipóteses são ou não satisfeitas em cada caso. Para isso, definiremos três cenários particulares relacionados às tendências:

1. Constante entre Períodos e Grupos:

$$\forall g, g' \in G \text{ e } \forall t, t' \in 1, 2, \dots, T:$$

$$E[y_{it}(\infty) - y_{it-1}(\infty)|D = g, X] = E[y_{it'}(\infty) - y_{it'-1}(\infty)|D = g', X]$$

2. Constante entre Períodos:

$$\forall g \in G, \forall t, t' \in 1, 2, \dots, T:$$

$$E[y_{it}(\infty) - y_{it-1}(\infty)|D = g, X] = E[y_{it'}(\infty) - y_{it'-1}(\infty)|D = g, X]$$

3. Constante entre Grupos:

$$\forall t \text{ e } \forall g, g' \in G:$$

$$E[y_{it}(\infty) - y_{it-1}(\infty)|D = g, X] = E[y_{it}(\infty) - y_{it-1}(\infty)|D = g', X]$$

Além disso, alguns métodos exigem, em determinados casos, a hipótese mais restritiva de independência entre as tendências paralelas e o tratamento, também conhecida como exogeneidade forte. Tal hipótese pode ser escrita como:

$$\forall g \text{ e } \forall t, E[y_{it}(\infty) - y_{it-1}(\infty)|D = g, X] = E[y_{it}(\infty) - y_{it-1}(\infty)|X]$$

e, quando necessária, será representada com um * na tabela de comparação abaixo.

Assim, a Tabela 1 resume quais modelos são ou não aplicáveis para cada um dos três tipos de tendências paralelas definidos.

Tabela 1: Hipóteses de Tendências Paralelas

	Constante entre:			
	Períodos e Grupos (1)	Períodos (2)	Grupos (3)	Nada
S&A	✓	✗	✗	✗
C&S	✓	✓	✓	✓
C&H	✓	✗	✓*	✗
Bor et al.	✓	✓*	✓*	✓*

S&A representa o modelo proposto por Sun and Abraham (2021), C&S o modelo de Callaway and Sant'Anna (2021), C&H o modelo de De Chaisemartin and d'Haultfoeuille (2020) e Bor et al. o modelo de Borusyak et al. (2024). ✓- O estimador é utilizável sob a hipótese; ✓* - O estimador é utilizável sob a hipótese, mas ainda é necessário assumir independência entre a tendência e o tratamento; ✗- O estimador não é utilizável sob a hipótese.

No caso 1, no qual podemos assumir que as tendências paralelas são constantes entre todos os grupos e períodos, todos os modelos tratados funcionam. Veja que, nesse caso, por definição, também é satisfeita a hipótese de exogeneidade forte entre

o tratamento e as tendências paralelas, já que elas são iguais para todos os grupos em todos os períodos de tempo (ou seja, são "constantes").

Para todos os outros casos, apenas Callaway and Sant'Anna (2021) não reque-rem a hipótese de exogeneidade forte. Nesse caso, é utilizado o *propensity score* para recuperar a esperança incondicional a partir das esperanças condicionais a X . Ou seja, o custo em termos de hipóteses é que 1) precisamos da hipótese adicional de *overlap*, isto é, a probabilidade do indivíduo ser tratado dado qualquer uma das variáveis de controle não pode ser 1 ou 0, e precisamos assumir que o *propensity score* consiga recuperar corretamente o ATT incondicional. Borusyak et al. (2024) também relaxa essa hipótese para a consistência. Entretanto, não conseguiremos estimar corretamente o ATT incondicional, como argumentado por Callaway and Sant'Anna (2021) e Marcus and Sant'Anna (2021). A forma funcional de Borusyak et al. (2024) não permite que a tendência varie ao longo do grupo e do tempo simultaneamente quando temos controles, a menos que os controles afetem igualmente todos os grupos.

Exemplo numérico

Para facilitar a compreensão das diferentes tendências paralelas, consideremos um exemplo simples em que temos apenas uma covariada $x \in \mathbb{R}$. Suponha, adicionalmente, que x é imutável ao longo do tempo. Essa característica é particularmente importante, pois, em geral, utilizamos controles com essa propriedade para evitar os chamados *bad controls*.

Para atingir esse objetivo, utilizamos os valores da covariada x observados no período em que nenhum indivíduo foi tratado. Assim, mesmo que x_{it} varie entre períodos e indivíduos, definimos uma nova variável x_{it}^* tal que:

$$\forall i, t, \quad x_{it}^* = x_{i0},$$

o que implica que a covariada será constante ao longo do tempo: $x_{it}^* = x_i^*$. Para simplificar a notação, assumiremos que $x_{it} = x_i$, eliminando a necessidade de construir a variável x_i^* . Dessa forma, podemos tratar x diretamente como constante ao longo do tempo, sem perda de generalidade na análise.

Para cada tipo de hipótese de tendência paralela discutida nesta seção, iremos apresentar exemplos de formas funcionais para os resultados potenciais no cenário em que o grupo g não seja tratado, ou seja, $y_{it}(\infty) \mid g$.

Considere que a tendência paralela condicional à covariada X e ao pertencimento ao grupo g seja dada por:

$$E[y_{it}(\infty) - y_{it-1}(\infty) \mid X, d = g] = \phi(x_i, g, t),$$

onde ϕ é uma função arbitrária. Para simplificar nossa análise, assumiremos que:

$$y_{it}(\infty) \mid g = h(x_i, g, t) + \alpha_i + \gamma_t + \varepsilon_{it},$$

em que α_i representa o efeito fixo do indivíduo, γ_g captura o efeito fixo do grupo, e ε_{it} é um termo de erro com as seguintes propriedades:

$$E[\varepsilon_{it} \mid X] = E[\varepsilon_{it}] = 0.$$

Essa formulação nos permitirá ilustrar as diferentes hipóteses de tendências paralelas de forma concreta e analítica, apenas mudando a função $h(x_i, g, t)$.

$$\begin{aligned} E[y_{it}(\infty) - y_{it-1}(\infty) \mid X, d = g] &= E[h(x_i, g, t) + \alpha_i + \gamma_t + \varepsilon_{it} \\ &\quad - h(x_i, g, t-1) + \alpha_i + \gamma_{t-1} + \varepsilon_{it-1} \mid X, d = g] \\ &= h(x_i, g, t) - h(x_i, g, t-1) + \Delta\gamma_t + \underbrace{E[\Delta\varepsilon_{it} \mid X, d = g]}_{\Delta 0=0} \\ &= h(x_i, g, t) - h(x_i, g, t-1) + \Delta\gamma_t \end{aligned}$$

E temos tendências paralelas se:

$$\begin{aligned} E[y_{it}(\infty) - y_{it-1}(\infty) \mid X, d = g] &= E[y_{it}(\infty) - y_{it-1}(\infty) \mid X, d = \infty] \\ h(x_i, g, t) - h(x_i, g, t-1) + \Delta\gamma_t &= h(x_i, \infty, t) - h(x_i, \infty, t-1) + \Delta\gamma_t \\ h(x_i, g, t) - h(x_i, g, t-1) &= h(x_i, \infty, t) - h(x_i, \infty, t-1) \end{aligned}$$

Constante entre Períodos e Grupos

Definimos o resultado potencial do grupo tratado inicialmente no período g (grupo g), no caso em que esses indivíduos não fossem tratados, $y_{it}(\infty) \mid g$, da seguinte forma:

$$\begin{aligned} h(x_i, g, t) &= tx_i, \\ y_{it}(\infty) \mid g &= h(x_i, g, t) + \alpha_i + \gamma_t + \varepsilon_{it}, \\ &= tx_i + \alpha_i + \gamma_t + \varepsilon_{it}. \end{aligned}$$

Ou seja, para todo g e t :

$$\begin{aligned} E[y_{it}(\infty) - y_{it-1}(\infty) \mid X, d = g] &= h(x_i, g, t) - h(x_i, g, t-1) + \Delta\gamma_t, \\ &= tx_i - (t-1)x_i + \Delta\gamma_t, \\ &= x_i + \Delta\gamma_t. \end{aligned}$$

Note que, neste caso, $h(x_i, g, t) - h(x_i, g, t-1) = x_i$, ou seja, condicional a X , essa diferença é constante tanto entre os grupos g quanto ao longo do tempo t .

Constante entre Períodos

Agora, vamos definir o resultado potencial dependendo também do grupo de tratamento:

$$\begin{aligned} h(x_i, g, t) &= t \cdot g \cdot x_i, \\ y_{it}(\infty) \mid g &= h(x_i, g, t) + \alpha_i + \gamma_t + \varepsilon_{it}, \\ &= t \cdot g \cdot x_i + \alpha_i + \gamma_t + \varepsilon_{it}. \end{aligned}$$

Ou seja, dado um g arbitrário, para todo t :

$$\begin{aligned} E[y_{it}(\infty) - y_{it-1}(\infty) \mid X, d = g] &= h(x_i, g, t) - h(x_i, g, t-1) + \Delta\gamma_t, \\ &= t \cdot g \cdot x_i - (t-1) \cdot g \cdot x_i + \Delta\gamma_t, \\ &= gx_i + \Delta\gamma_t. \end{aligned}$$

Diferentemente do caso anterior, a diferença $h(x_i, g, t) - h(x_i, g, t-1) = gx_i$, que é diferente para cada grupo.

Constante entre Grupos

Neste exemplo, o resultado potencial dependerá de forma não linear do tempo. Isso fará com que a diferença dos resultados potenciais seja diferente ao longo do tempo.

$$\begin{aligned} h(x_i, g, t) &= x_i e^t, \\ y_{it}(\infty) \mid g &= h(x_i, g, t) + \alpha_i + \gamma_t + \varepsilon_{it}, \\ &= x_i e^t + \alpha_i + \gamma_t + \varepsilon_{it}. \end{aligned}$$

Então, defina um t arbitrário, para todo g :

$$\begin{aligned}
E[y_{it}(\infty) - y_{it-1}(\infty) \mid X, d = g] &= h(x_i, g, t) - h(x_i, g, t-1) + \Delta\gamma_t, \\
&= x_i e^t - x_i e^{t-1} + \Delta\gamma_t, \\
&= e^{t-1}(e-1)x_i + \Delta\gamma_t.
\end{aligned}$$

Portanto, a diferença $h(x_i, g, t) - h(x_i, g, t-1) = e^t(e-1)x_i$ depende de t .

Antecipação

Tabela 2: Antecipação

	Antecipação
S&A	✗
C&S	✓
C&H	✓
Bor et al.	✓

S&A representa o modelo proposto por Sun and Abraham (2021), C&S o modelo de Callaway and Sant'Anna (2021), C&H o modelo de De Chaisemartin and d'Haultfoeuille (2020) e Bor et al. o modelo de Borusyak et al. (2024). ✓- O método permite antecipação; ✗- O método não permite antecipação.

Todos os modelos, com exceção de Sun and Abraham (2021), aceitam a antecipação do tratamento por parte dos grupos de tratamento, seja diretamente em seus códigos ou com modificações muito simples. A exigência em todos eles é que precisamos de ao menos um período onde os tratados ainda não foram tratados e não anteciparam que seriam tratados no futuro.

Entretanto, mesmo no caso de Sun and Abraham (2021), podemos considerar que o tratamento começa a partir do momento em que o grupo o antecipa, e conseguiremos, reponderando os ATTs, qualquer resultado que desejarmos. Isso faz com que, na prática, qualquer um dos métodos permita a antecipação do tratamento, desde que tenhamos ao menos um período pré-tratamento onde não houve antecipação por parte de nenhum grupo.

Dependência entre os grupos e os períodos

Em um modelo de Diferença-em-Diferenças escalonado, assumimos duas hipóteses principais de independência em relação ao resultado potencial y_{it} : a primeira diz respeito à independência entre **grupos** — ou seja, o indivíduo i pertencente ao grupo g deve ser independente do indivíduo i' pertencente ao grupo g' . A segunda é entre **períodos** — isto é, para o mesmo i , considera-se a correlação de seu resultado potencial ao longo do tempo. Matematicamente:

1. Sem correlação entre indivíduos e períodos

Para todo $i \in g$, $i' \in g'$, t e t' com $g \neq g'$ e $t \neq t'$,

$$\text{cov}(\varepsilon_{it}, \varepsilon_{i't}) = \text{cov}(\varepsilon_{it}, \varepsilon_{i't'}) = \text{cov}(\varepsilon_{it}, \varepsilon_{i't'}) = 0.$$

2. Correlação apenas dentro do mesmo grupo

- Se $i, i' \in g$, para pelo menos algum t : $\text{cov}(\varepsilon_{it}, \varepsilon_{i't}) \neq 0$.
- Se $i \in g$ e $i' \in g'$, para todo t e t' com $t \neq t'$: $\text{cov}(\varepsilon_{it}, \varepsilon_{i't}) = \text{cov}(\varepsilon_{it}, \varepsilon_{i't'}) = 0$.

3. Autocorrelação do mesmo indivíduo ao longo do tempo

- Se $i \neq i'$, $cov(\varepsilon_{it}, \varepsilon_{i't}) = cov(\varepsilon_{it}, \varepsilon_{i't'}) = 0$.
- Existe ao menos um t e um t' tal que $cov(\varepsilon_{it}, \varepsilon_{i't'}) \neq 0$.

4. Correlação entre grupos

- Se $i \in g$ e $i' \in g'$, $cov(\varepsilon_{it}, \varepsilon_{i't'}) \neq 0$.

Tabela 3: correlação dos erros

	Correlação entre:			
	Nada (1)	Dentro do Mesmo Grupo (2)	Autocorrelação (3)	Correlação entre Grupos (4)
S&A	✓	✓	✗	✗
C&S	✓	✓	?	✗
C&H	✓	?	?	✗
Bor et al.	✓*	✓*	✓*	✗

S&A representa o modelo proposto por Sun and Abraham (2021), C&S o modelo de Callaway and Sant'Anna (2021), C&H o modelo de De Chaisemartin and d'Haultfoeuille (2020) e, Bor et al., o modelo de Borusyak et al. (2024). ✓- O paper possibilita esse tipo de independência; ✗- O paper não possibilita esse tipo de independência. ?significa que não fica claro no paper se essa hipótese é ou não satisfeita. * - Para o caso de Borusyak et al. (2024), a hipótese é satisfeita porém, a variância estimada é um conservadora, ou seja maior que a verdadeira.

Conforme a tabela 3, todos os métodos precisam de, ao menos, independência entre os indivíduos de diferentes grupos. Alguns autores não deixam claro se é necessária ou não a hipótese de independência entre os períodos. Mas, em geral, parecem englobar tal hipótese ao colocarmos *clusters* no nível de indivíduo.

Cabe ressaltar que, apesar de Borusyak et al. (2024) ter as hipóteses menos restritivas, conseguimos apenas estimar uma variância conservadora, ou seja, conseguimos estimar uma variância que convergirá para algo maior ou igual à variância verdadeira. Já em Sun and Abraham (2021), Callaway and Sant'Anna (2021) e De Chaisemartin and d'Haultfoeuille (2020) temos variâncias assintóticas exatas.

3.5.2 Usabilidade

Nesta seção, vamos nos concentrar em três pontos principais. O primeiro é se há ou não *packages* em R ou Stata prontos para o uso. O segundo diz respeito à facilidade de uso em diferentes principais modelos de DID escalonados. Por fim, analisaremos facilidade de adaptação do método para casos não padronizados.

Tabela 4: Usabilidade

	Package	Facilidade de Uso	Adaptabilidade
S&A	✗	✓	✓
C&S	✓	✓	✗
C&H	✗	✗	✓
Bor et al.	✓*	✗	✓

S&A representa o modelo proposto por Sun and Abraham (2021), C&S o modelo de Callaway and Sant'Anna (2021), C&H o modelo de De Chaisemartin and d'Haultfoeuille (2020) e Bor et al. o modelo de Borusyak et al. (2024).

Callaway and Sant'Anna (2021) e Borusyak et al. (2024) disponibilizam pacotes prontos em R e Stata, facilitando a aplicação de seus métodos. Por outro lado, De Chaisemartin and d'Haultfoeuille (2020) e Sun and Abraham (2021) não oferecem pacotes dedicados. No entanto, para modelos escalonados de tratamento, o pacote de Borusyak et al. (2024) não funciona totalmente. Em termos de facilidade

de uso, o método de Sun and Abraham (2021), apesar de não ter um pacote pronto, utiliza regressão linear em sua essência, sendo relativamente fácil de implementar.

Por fim, a **adaptabilidade** refere-se à facilidade de incorporar **variações ao método**, como painéis não balanceados, múltiplos tipos ou intensidades de tratamento, heterogeneidades, entre outros. Neste caso, por usar regressão linear como base, Sun and Abraham (2021) tende a ser o mais fácil de adaptar. De Chaisemartin and d'Haultfoeuille (2020) também é adaptável para diferentes necessidades e possibilita uma maior gama de tratamentos, como entrada e saída do tratamento em qualquer momento no tempo. Borusyak et al. (2024), por sua vez, utiliza um modelo de imputação que facilita pequenas mudanças necessárias, mas sua inferência não é trivial. Por fim, o método de Callaway and Sant'Anna (2021) é o mais complexo e difícil de adaptar para casos fora do padrão, tornando-se o menos recomendado nesses casos.

3.6 Como escolher os pesos

Em todos os métodos, direta ou indiretamente, estamos calculando o $ATT(g, t)$ para cada grupo e período e , em seguida, computando uma média ponderada desses $ATT(g, t)$ que representaria, em um número único, um valor para dimensionar o efeito da intervenção:

$$\Gamma = \sum_{g \in G} \sum_{t=T_0}^T w(g, t) ATT(g, t).$$

A agregação de diversos efeitos numa média ponderada é interessante para permitir a visualização sucinta da dimensão do efeito do tratamento em diversos âmbitos (como por grupo, por período, ao longo do tempo, ou outras métricas de interesse). Entretanto, como escolher os pesos $w(g, t)$? À primeira vista, a pergunta pode parecer simples, mas essa escolha pode impactar significativamente as conclusões. A escolha correta dos pesos depende do contexto. Neste capítulo, introduziremos métodos para escolher os pesos de acordo com os objetivos da análise.

Um ponto crucial é que a definição dos pesos deve ser feita **antes da análise** dos dados, o que será detalhado na seção 3.6.1. A Seção 3.6.2 mostra como definimos os pesos com base na pergunta analisada e a Seção 3.6.3 explica os pesos mais utilizados na literatura.

3.6.1 Caçando resultados

A definição dos pesos, sendo uma escolha arbitrária, permite que se construa uma média ponderada que apresente qualquer valor desejado, independentemente do efeito real do tratamento. Por isso, a seleção dos pesos $w(g, t)$ deve ser feita com cautela, a fim de evitar viés de confirmação ou resultados enganosos.

Primeiro, os pesos devem ser determinados com base no objetivo específico da análise, e não com base nos $ATT(g, t)$ estimados. Ou seja, é importante escolher pesos que reflitam o que queremos investigar — como, por exemplo, dar peso proporcionais à representatividade relativa do grupo em questão em relação a população quando estamos analisando se vale a pena ou não expandir o tratamento.

Em segundo lugar, é essencial que essa definição ocorra de forma *ex ante*, antes da coleta e análise dos dados. Dessa maneira, garantimos que nossa análise não seja enviesada pela busca por um resultado estatisticamente significativo ou então em uma direção específica. Se os pesos fossem ajustados conforme os resultados observados, haveria o risco de escolher, dentre todas as configurações possíveis, aquela que mais favorece um resultado específico, o que introduziria viés e comprometeria a validade causal da análise.

Para visualizarmos como a escolha de pesos pode enviesar nossos resultados, considere o exemplo de um tratamento que não tem efeito para nenhum grupo, ou seja, $ATT(g, t) = 0$ para todo t e g . Temos três grupos tratados, $g=1$, $g=2$ e $g=3$, três períodos pós-tratamento $t \in \{1, 2, 3\}$, além do período pré-tratamento, $t = 0$. Mesmo que o verdadeiro $ATT(g, t)$ seja 0, as estimativas não serão iguais a zero. Considere que os ATTs estimados para cada grupo e período sejam iguais aos da Tabela 5

Tabela 5: $ATT(g, t)$ estimado

	t=1	t=2	t=3
g=1	3.2	-2.2	1.5
g=2	-	-2.3	-0.6
g=3	-	-	0.3

Se dermos pesos iguais para cada ATT , ou seja, $w(g, t) = 1/6$ para todo g e t , obtemos uma média ponderada de -0.1. No entanto, se definirmos peso 0 para todos os $ATT(g, t)$ exceto para $ATT(1, 1)$ e $ATT(1, 3)$ que terão peso $1/2$ cada um, a média ponderada passa a ser 1.7.

É claro que, nesse exemplo, escolhemos dar peso positivo apenas para os maiores ATT s estimados com nossa amostra, o que, na prática, nunca é feito. Porém, se testarmos vários tipos de ponderações com os ATT s que estimamos e escolhermos aquelas que dão algum resultado desejado, estaremos enviesando a análise da mesma forma. Em outras palavras, estaríamos "caçando resultados", como fizemos no exemplo acima, porém de forma menos explícita.

3.6.2 Definindo os Pesos

A definição dos pesos depende do objetivo da análise; diferentes pesos respondem a diferentes perguntas. Por exemplo, se o parâmetro de interesse é o efeito médio do tratamento ao longo de toda a amostra, e cada grupo tem a mesma importância relativa, um peso igual para cada $ATT(g, t)$ é o mais indicado. Entretanto, se o objeto de análise é o efeito na minha população de interesse, é melhor ponderar pelo peso relativo de cada grupo nessa população.

Não há certo ou errado quanto à definição dos pesos — essa escolha deve ser orientada pela pergunta de pesquisa principal, e não influenciada pelos resultados observados. Esse cuidado ajuda a garantir que a análise não esteja sujeita ao viés de confirmação ou à "caça aos resultados", em que as escolhas analíticas são moldadas para encontrar um efeito esperado.

Para entendermos melhor o processo de definição dos pesos, é importante definir, em primeiro lugar, qual parâmetro populacional desejamos estimar. No caso do exemplo anterior, cujo objetivo é avaliar o efeito na população de interesse (ao longo de todos os grupos e períodos, com igual importância relativa), devemos estimar:

$$\begin{aligned}
E[y_{it}(g) - y_{it}(\infty)] &= E_{gt}[E[y_{it}(g) - y_{it}(\infty)|G = g]] \\
&= \sum_{t'=1}^T \sum_{g \in G} P(G = g \cap t = t') E[y_{it'}(g) - y_{it'}(\infty)|G = g] \\
&= \sum_{t'=1}^T \sum_{g \in G} P(G = g \cap t = t') ATT(g, t')
\end{aligned}$$

Vamos definir os pesos como: $w(g, t) = \hat{P}(G = g \cap t = t) = \frac{N_g}{NT}$.

Agora, se quisermos avaliar o efeito médio em algum grupo específico, por exemplo $g = 1$, nosso parâmetro de interesse será:

$$\begin{aligned}
E_t[ATT(g = 1, t)] &= E_t[E[y_{it}(g = 1) - y_{it}(\infty)|G = 1]] \\
&= \sum_{t'=1}^T P(t = t'|G = 1) E[y_{it'}(g = 1) - y_{it'}(\infty)|G = 1] \\
&= \sum_{t=1}^T P(t = t'|G = 1) ATT(g = 1, t').
\end{aligned}$$

Como, em geral, estamos em um painel balanceado, $P(t = t'|G = 1) = \frac{1}{T}$. Além disso, definimos $w(1, t) = 1/T$ se $i \in g = 1$ e 0, caso contrário.

Outra possibilidade de definição é a necessidade de testar alguma suposição. Para facilitar, podemos usar os pesos para testarmos alguma hipótese relevante para a análise. Isso é particularmente interessante porque a maioria dos pacotes calculam a variância correta para qualquer $w(g, t)$ que imputarmos. Ou seja, podemos realizar testes de hipóteses com a variância correta facilmente, apenas redefinindo os pesos.

Por exemplo, imagine que queremos saber se o efeito médio do primeiro período de tratamento é maior que o efeito nos meses subsequentes. Ou seja:

$$\begin{aligned} H_0 : E_{gt}[ATT(g, t)|g - t = 0] &\leq E_{gt}[ATT(g, t)|g - t \geq 1] \\ H_1 : E_{gt}[ATT(g, t)|g - t = 0] &> E_{gt}[ATT(g, t)|g - t \geq 1], \end{aligned}$$

H_0 é equivalente a $E_{gt}[ATT(g, t)|g - t = 0] - E_{gt}[ATT(g, t)|g - t \geq 1] \leq 0$. Sejam

$$S_{g-t=0} = \{(g, t) \in G \times \{1, 2, \dots, T\} : g - t = 0\} \text{ e}$$

$$S_{g-t \geq 1} = \{(g, t) \in G \times \{1, 2, \dots, T\} : g - t \geq 1\}.$$

De forma que $S_{g-t=0}$ contém todos os pares grupo-período em que aquele período é o primeiro em que aquele grupo é tratado, e $S_{g-t \geq 1}$ os pares em que o grupo está exposto ao tratamento há ao menos um período. Podemos definir:

$$\begin{aligned} E_{gt}[ATT(g, t)|g - t = 0] &= \sum_{(g, t) \in S_{g-t=0}} P(g, t) ATT(g, t) \\ E_{gt}[ATT(g, t)|g - t \geq 1] &= \sum_{(g, t) \in S_{g-t \geq 1}} P(g, t) ATT(g, t) \end{aligned}$$

Neste caso, definimos:

$$w(g, t) = \begin{cases} P(g, t) & \text{se } (g, t) \in S_{g-t=0} \\ -P(g, t) & \text{se } (g, t) \in S_{g-t \geq 1} \\ 0 & \text{c.c.} \end{cases}$$

Com isso, podemos construir uma estatística-teste para a formulação alternativa de H_0 ($E_{gt}[ATT(g, t)|g - t = 0] - E_{gt}[ATT(g, t)|g - t \geq 1] = \Gamma \leq 0$).

Teremos:

$$\begin{aligned} \Gamma &= E_{gt}[ATT(g, t)|g - t = 0] - E_{gt}[ATT(g, t)|g - t \geq 1] \\ &= \sum_{(g, t) \in S_{g-t=0}} P(g, t) ATT(g, t) - \sum_{(g, t) \in S_{g-t \geq 1}} P(g, t) ATT(g, t) \\ &= \sum_{(g, t) \in S_{g-t=0}} P(g, t) ATT(g, t) + \sum_{(g, t) \in S_{g-t \geq 1}} -P(g, t) ATT(g, t) \\ &= \sum_{(g, t)} w(g, t) ATT(g, t) \end{aligned}$$

E o teste de hipótese original é equivalente a:

$$\begin{aligned} H_0 : \Gamma &\leq 0 \\ H_1 : \Gamma &> 0 \end{aligned}$$

onde $\Gamma = \sum_{(g, t)} w(g, t) ATT(g, t)$.

3.6.3 Pesos Mais Utilizados na Literatura

Os principais pesos utilizados na literatura são aqueles que geram os efeitos por grupo, por período e o efeito dinâmico. No primeiro, estima-se o efeito médio para cada grupo, independentemente de quando o tratamento foi iniciado, atribuindo pesos iguais para cada período. No segundo caso, temos o efeito médio para todos

os tratados naquele período, podendo ou não ponderar pelo tamanho relativo de cada grupo. Por fim, o mais utilizado é o efeito dinâmico, que calcula o efeito médio h períodos após o início do tratamento.

Para exemplificar cada um dos pesos, utilizaremos o mesmo exemplo usado nas seções anteriores. Para simplificar, utilizaremos apenas o $ATT(g, t)$ estimado via Callaway and Sant'Anna (2021). No entanto, todos os pesos propostos aqui podem ser aplicados a qualquer um dos métodos estudados. O *box* abaixo retoma os valores estimados para cada $ATT(g, t)$ na Seção 3.2.

Código R

```
library(did)
library(stargazer)

# Rodando Callaway and Sant'ana pelo método 'dr'
ATT_gt = att_gt(yname = 'y',
                 tname = 'ano',
                 idname = 'id_municipio',
                 gname = 'G',
                 xformula = ~ 1 + renda + populacao
                      + escolaridade,
                 data = df,
                 control_group = 'nevertreated',
                 est_method = 'dr'
                 )

ATT_gt

##
## Call:
## att_gt(yname = "y", tname = "ano", idname = "id_municipio",
## gname = "G", data = df, control_group = "nevertreated",
## est_method = "dr")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.
## "Difference-in-Differences with Multiple Time Periods."
## Journal of Econometrics, Vol. 225, No. 2, pp. 200-230, 2021.
## <https://doi.org/10.1016/j.jeconom.2020.12.001>,
## <https://arxiv.org/abs/1803.09015>
##
## Group-Time Average Treatment Effects:
## Group Time ATT(g,t) Std. Error [95% Simult. Conf. Band]
## 2012 2012 -0.5410 0.1350 -0.9757 -0.1063 *
## 2012 2013 -0.4153 0.1256 -0.8196 -0.0109 *
## 2012 2014 -0.7564 0.1552 -1.2561 -0.2567 *
## 2012 2015 -1.1034 0.1574 -1.6101 -0.5966 *
## 2012 2016 -1.1000 0.1488 -1.5792 -0.6208 *
## 2012 2017 -1.2274 0.1727 -1.7835 -0.6714 *
## 2012 2018 -1.3978 0.1816 -1.9826 -0.8129 *
## 2012 2019 -1.7642 0.1812 -2.3476 -1.1808 *
## 2012 2020 -1.8876 0.1784 -2.4620 -1.3131 *
## 2014 2012 -0.1807 0.1826 -0.7686 0.4073
## 2014 2013 -0.2479 0.1897 -0.8586 0.3628
## 2014 2014 0.1233 0.1945 -0.5029 0.7494
## 2014 2015 -0.0385 0.1854 -0.6355 0.5585
## 2014 2016 -0.1646 0.1903 -0.7773 0.4480
## 2014 2017 -0.7121 0.2060 -1.3754 -0.0488 *
## 2014 2018 -0.7098 0.2279 -1.4436 0.0241
## 2014 2019 -0.9478 0.2089 -1.6205 -0.2750 *
```

```
## 2014 2020 -0.9891 0.2186 -1.6929 -0.2853 *
## 2016 2012 0.0823 0.1435 -0.3798 0.5444
## 2016 2013 -0.1395 0.1471 -0.6130 0.3340
## 2016 2014 0.1614 0.1510 -0.3248 0.6476
## 2016 2015 0.0442 0.1592 -0.4683 0.5568
## 2016 2016 -0.1838 0.1801 -0.7637 0.3961
## 2016 2017 -0.4719 0.1913 -1.0878 0.1441
## 2016 2018 -0.7152 0.2057 -1.3774 -0.0531 *
## 2016 2019 -0.9098 0.1901 -1.5218 -0.2978 *
## 2016 2020 -1.0448 0.1955 -1.6744 -0.4153 *
## 2018 2012 -0.0262 0.1326 -0.4533 0.4009
## 2018 2013 0.0157 0.1456 -0.4531 0.4845
## 2018 2014 0.0516 0.1639 -0.4760 0.5793
## 2018 2015 0.0563 0.1936 -0.5669 0.6796
## 2018 2016 -0.1946 0.1767 -0.7635 0.3742
## 2018 2017 0.1394 0.1587 -0.3714 0.6502
## 2018 2018 -0.2810 0.1895 -0.8910 0.3290
## 2018 2019 -0.6055 0.2158 -1.3003 0.0893
## 2018 2020 -0.8546 0.2178 -1.5560 -0.1532 *
## ---
## Signif. codes: '*' confidence band does not cover 0
##
## P-value for pre-test of parallel trends assumption: 0.69134
## Control Group: Never Treated, Anticipation Periods: 0
## Estimation Method: Doubly Robust
```

Para facilitar a definição de pesos, vamos gerar 3 vetores: G, T e att_gt, que representam, respectivamente, o grupo correspondente àquela entrada (g), o período correspondente àquela entrada (t) e o $ATT(g, t)$.

#definindo o ATT_gt para os pesos.

```
G = ATT_gt$group
T = ATT_gt$t
att_gt = round(ATT_gt$att,3)
```

```
stargazer(cbind(G,T,att_gt))
```

```
##           G      T att_gt      ##           G      T att_gt
## [1,] 2012 2012 -0.541      ## [19,] 2016 2012 0.082
## [2,] 2012 2013 -0.415      ## [20,] 2016 2013 -0.140
## [3,] 2012 2014 -0.756      ## [21,] 2016 2014 0.161
## [4,] 2012 2015 -1.103      ## [22,] 2016 2015 0.044
## [5,] 2012 2016 -1.100      ## [23,] 2016 2016 -0.184
## [6,] 2012 2017 -1.227      ## [24,] 2016 2017 -0.472
## [7,] 2012 2018 -1.398      ## [25,] 2016 2018 -0.715
## [8,] 2012 2019 -1.764      ## [26,] 2016 2019 -0.910
## [9,] 2012 2020 -1.888      ## [27,] 2016 2020 -1.045
## [10,] 2014 2012 -0.181      ## [28,] 2018 2012 -0.026
## [11,] 2014 2013 -0.248      ## [29,] 2018 2013 0.016
## [12,] 2014 2014 0.123      ## [30,] 2018 2014 0.052
## [13,] 2014 2015 -0.039      ## [31,] 2018 2015 0.056
## [14,] 2014 2016 -0.165      ## [32,] 2018 2016 -0.195
## [15,] 2014 2017 -0.712      ## [33,] 2018 2017 0.139
## [16,] 2014 2018 -0.710      ## [34,] 2018 2018 -0.281
## [17,] 2014 2019 -0.948      ## [35,] 2018 2019 -0.605
## [18,] 2014 2020 -0.989      ## [36,] 2018 2020 -0.855
```

Ou seja, $ATT(2012, 2012) = -0.541$, $ATT(2012, 2013) = -0.415$, $ATT(2012, 2014) = -0.756$ e assim por diante.

Efeito por Grupo

Para cada um dos grupos, calcularemos o efeito médio ao longo de todos os períodos em que esse grupo foi tratado. Esses pesos são usados quando o objetivo é analisar diferenças ou efeitos heterogêneos entre os grupos tratados. Como estamos interessados em estimar o efeito médio de um grupo específico independentemente do período em que ele foi tratado e, em geral, lidamos com painéis balanceados (isto é, o número de indivíduos no grupo g é constante ao longo do tempo), a literatura normalmente opta por atribuir peso igual a cada período. Essa escolha também evita o problema da Seção 3.6.1, ao garantir que todos os períodos tenham o mesmo peso *ex-ante* à estimação.

Ou seja, queremos estimar:

$$\begin{aligned}\gamma_g &= E_t[ATT(g, t)|G = g] = \frac{1}{T_{t \geq g}} \sum_{t \geq g} ATT(g, t) \\ &= \frac{1}{T_{t \geq g}} \sum_{t=g}^T E[y_{it}(g) - y_{it}(\infty)|G = g]\end{aligned}$$

onde $T_{t \geq g}$ representa o total de períodos em que o grupo g está sob tratamento. A interpretação de γ_g é, portanto, o efeito médio do tratamento para o grupo g , considerando todos os períodos em que o grupo foi tratado. Essa métrica fornece uma visão clara do impacto médio do tratamento ao longo do tempo para cada grupo, sem se limitar a um período específico.

Por exemplo, para calcular o efeito médio para o grupo 2014 do exemplo acima: $T_{t \geq 2014} = 7$, i.e, 2014, 2015, 2016, 2017, 2018, 2019 e 2020. Ou seja,

$$\begin{aligned}\gamma_{2014} &= \frac{1}{7} [ATT(2014, 2014) + ATT(2014, 2015) + ATT(2014, 2016) + ATT(2014, 2017) \\ &\quad + ATT(2014, 2018) + ATT(2014, 2019) + ATT(2014, 2020)]\end{aligned}$$

O pacote do R disponibilizado pelos autores de Callaway and Sant'Anna (2021), calcula γ_g automaticamente pelo comando `aggte(MP = ATT_gt, type = "GROUP")`.

Código R

Calculando os Pesos por Grupo

A função a seguir calcula os pesos $w(g, t) = 1/T_{t \geq g}$ e γ_g :


```

# Definição dos pesos por grupo
w_g <- function(g, att_gt, G, T){

  # definir 0 como padrão para todos os ATT_gt
  W = rep(0, length(att_gt))

  #definir peso igual para todos os períodos onde o grupo g
  é tratado (definir 1 depois reponderamos)
  W[G==g & T>=g] = 1

  #reponderando para o peso somar 1
  W = W/sum(W)

  # Gamma
  Gamma = W %*% att_gt

  return(list(w=W, Gamma = Gamma))

}

```

Para calcularmos os pesos w e γ_{2014} como no texto dessa seção, usamos o código:

```

# Ex: grupo 2014
Grupo_2014 = w_g(g=2014, G = G, T = T, att_gt = att_gt)

# Pesos
w = round(Grupo_2014$w, 4)
print(cbind(G,T,w))

```

##		G	T	w	##		G	T	w
##	[1,]	2012	2012	0.0000	##	[19,]	2016	2012	0.0000
##	[2,]	2012	2013	0.0000	##	[20,]	2016	2013	0.0000
##	[3,]	2012	2014	0.0000	##	[21,]	2016	2014	0.0000
##	[4,]	2012	2015	0.0000	##	[22,]	2016	2015	0.0000
##	[5,]	2012	2016	0.0000	##	[23,]	2016	2016	0.0000
##	[6,]	2012	2017	0.0000	##	[24,]	2016	2017	0.0000
##	[7,]	2012	2018	0.0000	##	[25,]	2016	2018	0.0000
##	[8,]	2012	2019	0.0000	##	[26,]	2016	2019	0.0000
##	[9,]	2012	2020	0.0000	##	[27,]	2016	2020	0.0000
##	[10,]	2014	2012	0.0000	##	[28,]	2018	2012	0.0000
##	[11,]	2014	2013	0.0000	##	[29,]	2018	2013	0.0000
##	[12,]	2014	2014	0.1429	##	[30,]	2018	2014	0.0000
##	[13,]	2014	2015	0.1429	##	[31,]	2018	2015	0.0000
##	[14,]	2014	2016	0.1429	##	[32,]	2018	2016	0.0000
##	[15,]	2014	2017	0.1429	##	[33,]	2018	2017	0.0000
##	[16,]	2014	2018	0.1429	##	[34,]	2018	2018	0.0000
##	[17,]	2014	2019	0.1429	##	[35,]	2018	2019	0.0000
##	[18,]	2014	2020	0.1429	##	[36,]	2018	2020	0.0000

```

# Gamma
print(Grupo_2014$Gamma)

```

Note que os pesos são iguais a $1/T_{t \geq g} = 1/7 = 0.1429$ para todos os ATT(g,t), tal que $g = 2014$ e zero para qualquer outro ATT. Por fim, para acessarmos o γ_{2014} :

```
##           [,1]
## [1,] -0.4914286
```

Normalmente, queremos estimar os ATTs para todos os grupos da amostra. Para isso, basta executar um *looping* da função anterior para cada grupo, conforme o exemplo abaixo:

```
# Ex: Calculando para todos os grupos:
ATT = NULL
for(g in unique(G)){

  # Calculando w e Gamma para cada g
  Grupo_g = w_g(g=g, G = G, T = T, att_gt = att_gt)

  # Guardando valor do gamma
  ATT = rbind(ATT, c(g, Grupo_g$Gamma))

}

colnames(ATT) <- c("d", "Gamma")
ATT
```

```
##           d      Gamma
## [1,] 2012 -1.1324444
## [2,] 2014 -0.4914286
## [3,] 2016 -0.6652000
## [4,] 2018 -0.5803333
```

Calculando os Pesos por grupo usando Callaway and Sant’Anna (2021)

```
did::aggte(ATT_gt, type = "group")
```

```
##
## Call:
## did::aggte(MP = ATT_gt, type = "group")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant’Anna.
## "Difference-in-Differences with Multiple Time Periods."
## Journal of Econometrics, Vol. 225, No. 2, pp. 200-230,
## 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>,
## <https://arxiv.org/abs/1803.09015>
##
##
## Overall summary of ATT’s based on group/cohort aggregation:
##      ATT      Std. Error      [ 95% Conf. Int.]
## -0.7708      0.0714      -0.9108      -0.6308 *
##
##
## Group Effects:
## Group Estimate Std. Error [95% Simult. Conf. Band]
## 2012 -1.1326      0.1163      -1.4232      -0.8420 *
## 2014 -0.4912      0.1285      -0.8122      -0.1702 *
## 2016 -0.6651      0.1356      -1.0040      -0.3262 *
## 2018 -0.5804      0.1716      -1.0091      -0.1517 *
## ---
## Signif. codes: ‘*’ confidence band does not cover 0
##
## Control Group: Never Treated, Anticipation Periods: 0
## Estimation Method: Doubly Robust
```

Efeito por Período

O cálculo do efeito de tratamento em cada período pós-tratamento é útil quando queremos entender como o impacto do tratamento varia ao longo do tempo, independentemente do grupo específico ou da duração do tratamento de cada indivíduo. Menos utilizada na literatura, essa ponderação, em geral, pode ser feita de duas formas: a primeira é atribuir o mesmo peso a todos os grupos tratados em cada período, o que facilita a comparação entre os períodos, mantendo o foco no efeito médio do tratamento em um dado tempo. Ou então, pode-se ponderar os grupos pelo seu peso relativo em relação ao total de indivíduos tratados naquele período específico. Essa abordagem é particularmente útil quando alguns grupos são mais representativos ou relevantes do que outros em termos de tamanho ou importância na população de interesse.

No caso de pesos iguais para todos os grupos, para cada t :

$$w(g, t) = \begin{cases} \frac{1}{\#_{g \leq t}} & \text{se } g \leq t \\ 0 & \text{c.c.} \end{cases}$$
$$\gamma_t = \frac{1}{\#_{g \leq t}} \sum_{g \leq t} ATT(g, t).$$

onde $\#_{g \leq t}$ é a quantidade de grupos tratados no período t (ou seja, com $g \leq t$). No caso de pesos relativos ao tamanho de cada grupo:

$$\gamma_t = E_g[ATT(g, t)|g \leq t] = \sum_{g \in G} P(g|g \leq t) ATT(g, t)$$
$$= \sum_{g \leq t} \frac{N_g}{N_{g \leq t}} ATT(g, t)$$

onde $N_{g \leq t}$ é a quantidade de indivíduos em todos os grupos tratados no período t , e N_g é o número de indivíduos no grupo g .

Independentemente da escolha, neste tipo de ponderação estimaremos um efeito para cada período pós-tratamento. Se quisermos, por exemplo, calcular o efeito médio para o período de 2016 no nosso exemplo, teremos:

- Dando pesos iguais, $\#_{g \leq 2016} = 3$ pois em 2016 temos 3 grupos tratados, $g = 2012, g = 2014$ e $g = 2016$. Temos, então:

$$\gamma_{2016} = \frac{1}{3} [ATT(2012, 2016) + ATT(2014, 2016) + ATT(2016, 2016)]$$

- Dando peso relativo ao tamanho do grupo. Sejam N_{2012} , N_{2014} e N_{2016} o número de indivíduos em cada grupo. Temos:

$$\gamma_{2016} = \frac{1}{N_{2012} + N_{2014} + N_{2016}} [N_{2012} ATT(2012, 2016) + N_{2014} ATT(2014, 2016) + N_{2016} ATT(2016, 2016)]$$

No box abaixo, temos o cálculo das médias ponderadas por período usando o pacote disponibilizado por Callaway and Sant'Anna (2021). Por padrão, são usados pesos iguais para cada grupo tratado.

Código R

Calculando os Pesos por Grupo

A função a seguir calcula os pesos $w(g, t) = N_g/N_{g \leq t}$ e γ_t :

```

# Definição dos pesos por período
w_t <- function(t, att_gt, G, T, df){

  # definir 0 como padrão para todos os ATT_gt
  W = rep(0, length(att_gt))

  #definir peso 1 para todos os grupos tratados no periodo t.
  (depois reponderamos pelo tamanho relativo do grupo)
  W[G <= t & T == t] = 1

  #peso ponderado pelo tamanho do grupo
  for(g in unique(G)){
    N_g = length(df[df$ano==t & df$G == g,"id_municipio"])
    W[G == g] = W[G==g]*N_g
  }

  #reponderando para o peso somar 1
  W = W/sum(W)

  # Gamma
  Gamma = W %*% att_gt

  return(list(w=W, Gamma = Gamma))
}

```

E, para calcularmos os pesos w e γ_{2016} como no texto dessa seção, usamos o código:

```

# Ex: Período 2016
Periodo_2016 = w_t(t=2016, G = G, T = T, att_gt = att_gt, df=df)

# Pesos
w = round(Periodo_2016$w, 4)
print(cbind(G,T,w))

```

##		G	T	w	##		G	T	w
##	[1,]	2012	2012	0.0000	##	[19,]	2016	2012	0.0000
##	[2,]	2012	2013	0.0000	##	[20,]	2016	2013	0.0000
##	[3,]	2012	2014	0.0000	##	[21,]	2016	2014	0.0000
##	[4,]	2012	2015	0.0000	##	[22,]	2016	2015	0.0000
##	[5,]	2012	2016	0.4444	##	[23,]	2016	2016	0.3333
##	[6,]	2012	2017	0.0000	##	[24,]	2016	2017	0.0000
##	[7,]	2012	2018	0.0000	##	[25,]	2016	2018	0.0000
##	[8,]	2012	2019	0.0000	##	[26,]	2016	2019	0.0000
##	[9,]	2012	2020	0.0000	##	[27,]	2016	2020	0.0000
##	[10,]	2014	2012	0.0000	##	[28,]	2018	2012	0.0000
##	[11,]	2014	2013	0.0000	##	[29,]	2018	2013	0.0000
##	[12,]	2014	2014	0.0000	##	[30,]	2018	2014	0.0000
##	[13,]	2014	2015	0.0000	##	[31,]	2018	2015	0.0000
##	[14,]	2014	2016	0.2222	##	[32,]	2018	2016	0.0000
##	[15,]	2014	2017	0.0000	##	[33,]	2018	2017	0.0000
##	[16,]	2014	2018	0.0000	##	[34,]	2018	2018	0.0000
##	[17,]	2014	2019	0.0000	##	[35,]	2018	2019	0.0000
##	[18,]	2014	2020	0.0000	##	[36,]	2018	2020	0.0000

Para acessarmos o γ_{2016} basta usarmos o código:

```
# Gamma
print(Periodo_2016$Gamma)
```

```
##           [,1]
## [1,] -0.5868889
```

É comum que queiramos estimar os ATTs para todos os períodos. Para isso, basta executar um looping da função anterior para cada período pós-tratamento na nossa amostra:

```
# Ex: Calculando para todos os períodos:
ATT = NULL
for(tt in unique(T)){

  Perodo_t = w_t(t=tt, G = G, T = T, att_gt = att_gt, df=df)
  ATT = rbind(ATT, c(tt, Perodo_t$Gamma))

}

colnames(ATT) <- c("t", "Gamma")
ATT
```

```
##           t      Gamma
## [1,] 2012 -0.5410000
## [2,] 2013 -0.4150000
## [3,] 2014 -0.4630000
## [4,] 2015 -0.7483333
## [5,] 2016 -0.5868889
## [6,] 2017 -0.8608889
## [7,] 2018 -0.8333333
## [8,] 2019 -1.1247500
## [9,] 2020 -1.2691667
```

Calculando os Pesos por Período usando Callaway and Sant'Anna (2021)

```
did::aggte(ATT_gt, type = 'calendar')
```

```
##
## Call:
## did::aggte(MP = ATT_gt, type = "calendar")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.
## "Difference-in-Differences with Multiple Time Periods."
## Journal of Econometrics, Vol. 225, No. 2, pp. 200-230,
## 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>,
## <https://arxiv.org/abs/1803.09015>
##
##
## Overall summary of ATT's based on calendar time aggregation:
##      ATT      Std. Error      [ 95% Conf. Int.]
## -0.7603      0.0724      -0.9021      -0.6185 *
##
##
## Time Effects:
## Time Estimate Std. Error [95% Simult. Conf. Band]
## 2012 -0.5410      0.1407      -0.9229      -0.1591 *
## 2013 -0.4153      0.1353      -0.7825      -0.0481 *
## 2014 -0.4632      0.1419      -0.8481      -0.0782 *
```

```
## 2015 -0.7484 0.1387 -1.1246 -0.3722 *
## 2016 -0.5868 0.1152 -0.8992 -0.2743 *
## 2017 -0.8611 0.1277 -1.2075 -0.5146 *
## 2018 -0.8333 0.1194 -1.1573 -0.5093 *
## 2019 -1.1249 0.1301 -1.4778 -0.7720 *
## 2020 -1.2689 0.1268 -1.6129 -0.9249 *
## ---
## Signif. codes: '*' confidence band does not cover 0
##
## Control Group: Never Treated, Anticipation Periods: 0
## Estimation Method: Doubly Robust
```

Efeito Dinâmico

Os pesos dinâmicos são amplamente utilizados na literatura por permitirem uma análise detalhada de como o efeito do tratamento se comporta ao longo do tempo após a sua introdução. Essa abordagem possibilita observar o comportamento do impacto do tratamento em função do tempo desde que o grupo começou a ser tratado, em vez de se restringir a um único período ou a um grupo específico.

Ao agrupar os $ATT(g, t)$ pela quantidade de períodos em que o grupo foi tratado, é possível analisar padrões temporais como a persistência, o aumento ou a diminuição do efeito do tratamento ao longo do tempo. Por exemplo, ao considerar todos os grupos que foram tratados por dois períodos, podemos observar o efeito médio do tratamento nesses grupos e compará-lo com o efeito médio em grupos tratados por três, quatro ou mais períodos.

Matematicamente, para o efeito k períodos após o início do tratamento:

$$w_k(g, t) = \begin{cases} \frac{N_g}{N_{t-g=k}} & \text{se } t - g = k \\ 0 & \text{c.c} \end{cases}$$

$$\gamma_k = \sum_{g \in G} \sum_{t=1}^T w_k(g, t) ATT(g, t).$$

Essa abordagem é especialmente útil para estudar a durabilidade e a intensidade dos efeitos do tratamento ao longo do tempo, ajudando a responder perguntas sobre a eficácia contínua ou decrescente do tratamento em diferentes horizontes temporais. Por exemplo, com um gráfico dos γ_k para todos os k s possíveis, podemos ver a heterogeneidade do efeito conforme o número de períodos de exposição ao tratamento aumenta.

Outro benefício desse método de ponderação é que ele também pode ser utilizado para a realização de um teste-placebo. Podemos definir o efeito médio do tratamento h meses antes de o grupo ser tratado. Se a hipótese de tendência paralela for satisfeita, esse efeito será próximo de zero. Abaixo, plotamos o resultado padrão do pacote de Callaway and Sant'Anna (2021) para o exemplo dessa seção.

Código R

Calculando os Pesos dinâmicos (event Study)

A função a seguir calcula os pesos $w(g, t)$ tal que:

$$w_d(g, t) = \begin{cases} \frac{N_g}{N_{t-g=d}} & \text{se } t - g = d \\ 0 & \text{c.c} \end{cases}$$

```

# Definição dos pesos como event study
w_d <- function(d, att_gt, G, T, df){

  # Definindo d (tempo desde o tratamento)
  D = T - G

  # definir 0 como padrão para todos os ATT_gt
  W = rep(0, length(att_gt))

  #definir peso 1 para todos os grupos no periodo onde t-g = d.
  (depois reponderamos pelo tamanho relativo do grupo)
  W[D == d] = 1

  #peso ponderado pelo tamanho do grupo
  for(i in 1:length(D)){
    N_gt = length(df[df$G==G[i] & df$ano==T[i], "id_municipio"])
    W[i] = W[i]*N_gt
  }

  #reponderando para o peso somar 1
  W = W/sum(W)

  # Gamma
  Gamma = W %*% att_gt

  return(list(w=W, Gamma = Gamma))
}

```

Para calcularmos os pesos w e γ_1 , isto é, o efeito médio da intervenção 1 ano após sua implementação:

```

# Ex: efeito 1 ano após a implementação do tratamento: d = 1
d_1 = w_d(d=1, G = G, T = T, att_gt = att_gt, df=df)

# Pesos
w = round(d_1$w, 4)
print(cbind(G,T,w))

```

##		G	T	w	##		G	T	w
##	[1,]	2012	2012	0.0000	##	[19,]	2016	2012	0.0000
##	[2,]	2012	2013	0.3333	##	[20,]	2016	2013	0.0000
##	[3,]	2012	2014	0.0000	##	[21,]	2016	2014	0.0000
##	[4,]	2012	2015	0.0000	##	[22,]	2016	2015	0.0000
##	[5,]	2012	2016	0.0000	##	[23,]	2016	2016	0.0000
##	[6,]	2012	2017	0.0000	##	[24,]	2016	2017	0.2500
##	[7,]	2012	2018	0.0000	##	[25,]	2016	2018	0.0000
##	[8,]	2012	2019	0.0000	##	[26,]	2016	2019	0.0000
##	[9,]	2012	2020	0.0000	##	[27,]	2016	2020	0.0000
##	[10,]	2014	2012	0.0000	##	[28,]	2018	2012	0.0000
##	[11,]	2014	2013	0.0000	##	[29,]	2018	2013	0.0000
##	[12,]	2014	2014	0.0000	##	[30,]	2018	2014	0.0000
##	[13,]	2014	2015	0.1667	##	[31,]	2018	2015	0.0000
##	[14,]	2014	2016	0.0000	##	[32,]	2018	2016	0.0000
##	[15,]	2014	2017	0.0000	##	[33,]	2018	2017	0.0000
##	[16,]	2014	2018	0.0000	##	[34,]	2018	2018	0.0000
##	[17,]	2014	2019	0.0000	##	[35,]	2018	2019	0.2500
##	[18,]	2014	2020	0.0000	##	[36,]	2018	2020	0.0000

Veja que, como estamos avaliando $d = 1$, damos pesos não zero apenas para os períodos 2013 para o grupo tratado pela primeira vez em 2012; período 2015 para o grupo tratado pela primeira vez em 2014; período 2017 para o grupo tratado pela primeira vez em 2016; período 2019 para o grupo tratado pela primeira vez em 2018.

```
# Gamma
print(d_1$Gamma)
```

```
##           [,1]
## [1,] -0.4140833
```

Novamente, o mais comum é estimarmos os ATTs para todos os d possíveis na nossa amostra, inclusive os negativos (isso é, o efeito 1 meses antes de aquele grupo ser tratado, $d = -1$). No caso de d positivo, a interpretação é semelhante ao que vimos nas outras seções, com a diferença que agora estamos calculando o efeito médio da intervenção d períodos após o início da intervenção. No caso dos d negativos, como estamos comparando um grupo que não foi tratado, se a tendência paralela for válida, esperamos ter algo perto de zero. Em geral, usamos essa abordagem para testarmos pre-trends. Abaixo o código com o looping que retorna os ATT para todos os d , positivos e negativos, possíveis.

```
# Ex: Calculando para todos os d:
ATT = NULL
for(d in sort(unique(T-G))){

  Perodo_t = w_d(d=d, G = G, T = T, att_gt = att_gt, df=df)
  ATT = rbind(ATT, c(d, Perodo_t$Gamma))

}

colnames(ATT) <- c("d", "Gamma")
ATT
```

```
##           d      Gamma
## [1,] -6 -0.0260000
## [2,] -5  0.0160000
```



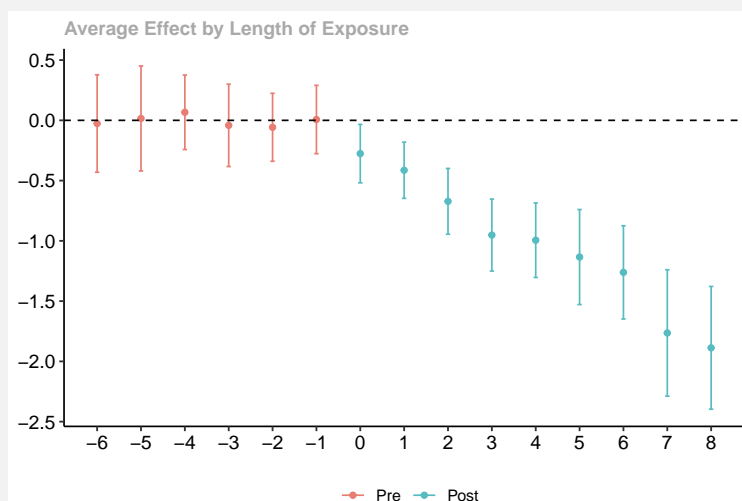
```
## [3,] -4 0.0670000
## [4,] -3 -0.0420000
## [5,] -2 -0.0580000
## [6,] -1 0.0066250
## [7,] 0 -0.2760833
## [8,] 1 -0.4140833
## [9,] 2 -0.6720000
## [10,] 3 -0.9517778
## [11,] 4 -0.9950000
## [12,] 5 -1.1340000
## [13,] 6 -1.2616667
## [14,] 7 -1.7640000
## [15,] 8 -1.8880000
```

Calculando os Pesos dinâmicos (*Event Study*) usando Callaway and Sant'Anna (2021)

```
did::aggte(ATT_gt, type = 'dynamic')
```

```
##
## Call:
## did::aggte(MP = ATT_gt, type = "dynamic")
##
## Reference: Callaway, Brantly and Pedro H.C. Sant'Anna.
## "Difference-in-Differences with Multiple Time Periods."
## Journal of Econometrics, Vol. 225, No. 2, pp. 200-230,
## 2021. <https://doi.org/10.1016/j.jeconom.2020.12.001>,
## <https://arxiv.org/abs/1803.09015>
##
##
## Overall summary of ATT's based on event-study/dynamic agg:
##      ATT      Std. Error    [ 95% Conf. Int.]
## -1.0396      0.0784    -1.1933      -0.886 *
##
##
## Dynamic Effects:
## Event time Estimate Std. Error [95% Simult. Conf. Band]
##      -6    -0.0262    0.1433    -0.4284    0.3760
##      -5     0.0157    0.1571    -0.4250    0.4564
##      -4     0.0670    0.1083    -0.2369    0.3708
##      -3    -0.0416    0.1177    -0.3719    0.2888
##      -2    -0.0576    0.1051    -0.3524    0.2372
##      -1     0.0069    0.1045    -0.2864    0.3002
##       0    -0.2760    0.0828    -0.5083    -0.0437 *
##       1    -0.4142    0.0838    -0.6493    -0.1791 *
##       2    -0.6720    0.0949    -0.9383    -0.4057 *
##       3    -0.9519    0.1015    -1.2368    -0.6670 *
##       4    -0.9949    0.1059    -1.2919    -0.6979 *
##       5    -1.1342    0.1433    -1.5362    -0.7322 *
##       6    -1.2616    0.1380    -1.6488    -0.8743 *
##       7    -1.7642    0.1757    -2.2571    -1.2714 *
##       8    -1.8876    0.1664    -2.3544    -1.4208 *
## ---
## Signif. codes: '*' confidence band does not cover 0
##
## Control Group: Never Treated, Anticipation Periods: 0
## Estimation Method: Doubly Robust
```

```
# Plotango gráfico efeito dinamico/event study  
ggdid(att_d)
```



4 Conclusão

O presente manual técnico sobre diferença-em-diferenças escalonado apresenta uma análise abrangente dos principais conceitos teóricos e práticos discutidos na literatura. Ao longo deste guia, destacamos as nuances do DID escalonado, desde sua formulação básica até as contribuições mais recentes, a saber: Sun and Abraham (2021), Callaway and Sant’Anna (2021), De Chaisemartin and d’Haultfoeuille (2020) e Borusyak et al. (2024). Oferecemos, assim, orientações claras sobre quando e como aplicar cada metodologia, destacando suas diferenças e limitações.

Revisitamos a evolução do DID, comparando suas versões tradicionais com os modelos mais avançados que lidam com múltiplos períodos e tratamentos escalonados. Inicialmente, revisamos o DID padrão com dois períodos e dois grupos — um de tratamento e outro de controle. Em seguida, comparamos esse modelo com abordagens que incorporam múltiplos períodos de pré-tratamento e pós-tratamento, ainda considerando dois grupos: tratado e controle.

Posteriormente, introduzimos o conceito de tratamento escalonado, destacando os desafios metodológicos ao se analisar esse tipo de tratamento com o modelo tradicional de DID. Em resposta, discutimos as principais soluções apresentadas na literatura recente, explorando três pilares fundamentais: o entendimento do modelo proposto, as hipóteses necessárias para garantir uma análise causal confiável e, por fim, um exemplo prático implementado em R. Além disso, na Seção 3.5, detalhamos as diferenças entre os quatro principais trabalhos da literatura.

Outro ponto crucial abordado é a escolha dos pesos para agregar os efeitos do tratamento de diferentes grupos e períodos. Essa escolha deve ser guiada pela pergunta de pesquisa e realizada de forma criteriosa, evitando uma “procura por resultados” que comprometa a validade das conclusões. Na seção 3.6.2, detalhamos o processo de escolha dos pesos.

Embora as metodologias apresentadas sejam robustas, é fundamental destacar que sua aplicação exige atenção especial às suposições subjacentes. Cada estudo possui características específicas que podem demandar adaptações metodológicas ou a inclusão de hipóteses adicionais.

Outro aspecto importante a ser considerado é que os exemplos fornecidos em R são simplificados e utilizam dados já tratados. Em muitas aplicações práticas, será necessário realizar o pré-tratamento de dados brutos, uma etapa essencial que não foi abordada neste manual.

Além disso, a literatura sobre diferença-em-diferenças escalonado está em constante evolução, tornando essencial manter-se atento às novas metodologias que pos-

sam surgir. Vale destacar que há pesquisas em andamento, e até mesmo concluídas, que abordam outras estruturas de tratamento em dados em painel, não contempladas pela literatura discutida neste manual. Exemplos incluem diferença-em-diferenças escalonado com tratamento contínuo e o controle sintético. Essas metodologias podem ser úteis, dependendo da estrutura do problema analisado, e, em alguns casos, até mesmo mais adequadas.

Esperamos que este manual sirva como um guia prático e confiável para economistas e cientistas de dados interessados em compreender e aplicar o DID escalonado. Mais do que uma simples revisão da literatura, nosso objetivo foi construir uma ponte entre a teoria e a prática, capacitando os leitores a realizar análises causais sólidas para a avaliação de políticas públicas baseadas em evidências.

Referências

- Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*, page rdae007.
- Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.
- Carrillo, B. and Feres, J. (2019). Provider supply, utilization, and infant health: Evidence from a physician distribution policy. *American Economic Journal: Economic Policy*, 11(3):156–196.
- De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.
- Dias, M. and Fontes, L. F. (2024). The effects of a large-scale mental health reform: Evidence from brazil. *American Economic Journal: Economic Policy*, 16(3):257–289.
- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.
- Graves, J. and Trond, C. (2024). Employment discrimination and labor market protections for sexual minorities in brazil. *Labour Economics*, 90:102548.
- MaCurdy, T., Chen, X., and Hong, H. (2011). Flexible estimation of treatment effect parameters. *American Economic Review*, 101(3):544–551.
- Marcus, M. and Sant'Anna, P. H. (2021). The role of parallel trends in event study settings: An application to environmental economics. *Journal of the Association of Environmental and Resource Economists*, 8(2):235–275.
- Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.
- Ver Hoef, J. M. (2012). Who invented the delta method? *The American Statistician*, 66(2):124–127.