Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment

Brantly Callaway **Pedro H. C. Sant'Anna**

Working Paper (2020)[1]

Fábio Nishida - Outubro/2020

[1] Esta é a 3ª versão, sendo que a 1ª e a 2ª são de 2018

Estrutura do Artigo

1. Introdução
2. Identificação
   - Estrutura
   - Identificação Não-Paramétrica de ATT(g, t)
3. Sumarização de Efeitos de Tratamento Médio Grupo-Temporal
4. Estimação e Inferência
5. O Efeito da Política de Salário Mínimo no Emprego de Adolescentes
6. Conclusão

Introdução

Diferenças-em-Diferenças na Forma Canônica (1/2)

Diferenças-em-Diferenças (DiD) se tornou um dos designs mais populares para avaliar efeitos causais em intervenções políticas.

Em sua forma tradicional, temos:

2 períodos e 2 grupos:

- no 1º período: nenhum dos grupos é tratado
- no 2º período: um grupo recebe o tratamento (grupo de tratamento) e o outro não recebe (grupo de controle)

Diferenças-em-Diferenças na Forma Canônica (2/2)

No entanto, é necessário supor que, na ausência de tratamento, os grupos de tratamento e de controle teriam trajetórias paralelas ao longo do tempo. Porém, muitas aplicações empíricas de DiD desviam da forma padrão, possuindo:

(i) múltiplos períodos de tempo, e

(ii) variação do efeito no tempo de tratamento.

Exemplo do Problema Fundamental (2018)

Figure 1: Example of Selective Treatment Timing and Dynamic Treatment Effects

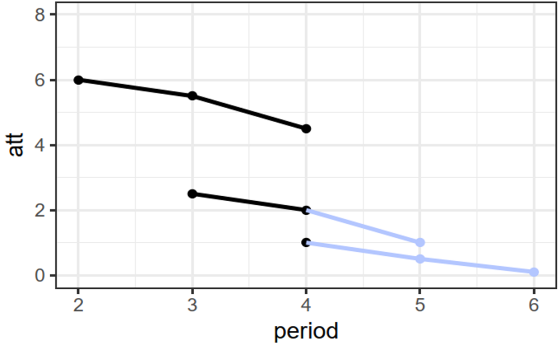


Figure: Todos os grupos G2, G3 e G4 apresentam efeitos de tratamento dinâmicos declinantes e também possuem tempo de tratamento seletivo - grupos tratados no início obtém maiores efeitos do tratamento (heterogeneidade).

Abordagem Proposta

- O modelo proposto envolver analisar DiD em três passos separados:
  (i) identificação dos parâmetros causais desagregados relevantes;
  (ii) agregação destes parâmetros para formar medidas de resumo dos efeitos causais;
  (iii) estimação e inferência sobre estes diferentes parâmetros desejados.
- Usa-se o parâmetro causal desagregado chamado de efeito de tratamento médio grupo-temporal, $ATT(g, t)$, pois não restringe:
  - a heterogeneidade em relação às variáveis observadas,
  - os período em que os grupos foram inicialmente tratados, e
  - a evolução dos efeitos de tratamento ao longo do tempo.
- Logo, é mais geral e flexível em relação ao modelo com efeitos fixos bidirecionais padrão (*standard two-way fixed effects* - TWFE), que não permite esses efeitos.

Medidas de Efeito Causal de Resumo

- Ao considerar g grupos e t período, podemos ter um grande número de efeitos de tratamento médio grupo temporal, $ATT(g,t)$'s.
- Podemos agregá-los em parâmetros de resumo, em especial, em que os efeitos de tratamento médio variam:
  (a) com o tamanho de exposição do tratamento - **efeitos de estudo de evento**;
  (b) entre os grupos - **efeitos específicos de grupo**; e
  (c) dado um período no tempo - **efeitos de tempo de calendário**.
- Para validar assintoticamente a inferência, usa-se o procedimento *bootstrap* do tipo multiplicador. (não será abordado na apresentação).
- Para ilustrar, será feita análise do efeito do salário mínimo no emprego de adolescentes.

Identificação

Notações

- Considere um caso com $\mathcal{T}$ períodos e denote um período de tempo particular como t, tal que $t = 1, ..., \mathcal{T}$. Note que, na configuração padrão de DiD, $\mathcal{T} = 2$ e nenhum grupo é tratado no período 1.
- Seja D_t a variável binária igual a 1 se o indivíduo é tratado no período t.
- Seja G_g a variável binária igual a 1 se o indivíduo teve o 1º tratamento no período g.
- Seja C a variável binária que é igual a 1 se o indivíduo nunca é tratado.
- O *propensity score* generalizado é dado por

$$p_{g,s}(X) = P(G_g = 1 \mid X, G_g + (1 - D_s)(1 - G_g) = 1).$$

ou seja, é a probabilidade de um indivíduo ser tratado no tempo g, condicional às covariadas de pré-tratamento X, e ao fato de ser membro do grupo g, $G_g = 1$, ou membro do grupo ainda não-tratado no período s, $(1 - D_s)(1 - G_g)$.

Efeito de Tratamento Médio Grupo-Temporal - *ATT(g, t)*

- O efeito de tratamento médio nos tratados no caso canônico ($\mathcal{T} = 2$ com 2 grupos) é

$$ATT = E[Y_2(1) - Y_2(0) \mid G_2 = 1].$$

Note que, neste caso, o grupo $g = 2$ é o único grupo tratado.

- Quando temos mais de 2 períodos e com variação no tempo de tratamento, usaremos:

$$ATT(g, t) = E[Y_t(1) - Y_t(0) \mid G_g = 1],$$

- Este é parâmetro causal de **efeito de tratamento médio grupo-temporal**.
- Em particular, note que $ATT(2, 2)$ colapsa para o ATT padrão.

Hipóteses (1/4)

- **Hip. 1 (Irreversibilidade do Tratamento).** Para $t = 2, ..., \mathcal{T}$,

$$D_{t-1} = 1 \implies D_t = 1$$

Ninguém é tratado em $t = 1$ e, uma vez que o indivíduo é tratado, ele será tratado no próximo período. Esse efeito é chamado de adoção de tratamento escalonado (*staggered treatment adoption*) na literatura. Pode ser interpretado como: os indivíduos não se "esquecem" sobre a experiência do tratamento.

- **Hip. 2 (Amostragem Aleatória).** $\{Y_{i1}, Y_{i2}, ..., Y_{i\mathcal{T}}; X_i; D_{i1}, D_{i2}, ..., D_{i\mathcal{T}}\}_{i=1}^{n}$ é independente e identicamente distribuído (iid).
Permitirá o uso de dados em painel, pois permite entendermos todos resultados potenciais como aleatórios.

Hipóteses (2/4)

- **Hip. 3 (Antecipação de Tratamento Limitada).** Existe um $\delta \geq 0$ tal que

$$E[Y_t(g) \mid X, G_g = 1] = E[Y_t(0) \mid X, G_g = 1] \;\; a.s.$$
$$\forall g \in \mathcal{G}, \;\; t \in \{1, ..., \mathcal{T}\}, \;\text{tal que}\; t < g - \delta.$$

Restringe antecipação do tratamento para todos grupos "eventualmente tratados":

- quando $\delta = 0$: indivíduos não antecipam o tratamento.
- quando $\delta = 1$: os indivíduos antecipam o tratamento em 1 período, ou seja, o tratamento pode afetar o comportamento no período anterior.

Se indivíduos antecipam o tratamento, precisamos considerar isso nas escolhas dos grupos que servirão como contrafactual dos grupos tratados.

Hipóteses (3/4)

- **Hip. 4 (Tendências Paralelas Condicionais com o Grupo "Nunca-Tratado").** Defina δ como na Hipótese 3. Para cada $t \in \{2, ..., \mathcal{T}\}$, $g \in \mathcal{G}$ tal que $t \geq g - \delta$,

$$E[Y_t(0) - Y_{t-1}(0) \mid X, G_g = 1] = E[Y_t(0) - Y_{t-1}(0) \mid X, C = 1] \; a.s..$$

- **Hip. 5 (Tendências Paralelas Condicionais com Grupos "Ainda-Não-Tratados").** Defina δ como na Hipótese 3. Para cada $(s, t) \in \{2, ..., \mathcal{T}\} \times \{2, ..., \mathcal{T}\}$, $g \in \mathcal{G}$ tal que $t \geq g - \delta$ e $s \geq g + \delta$,

$$E[Y_t(0) - Y_{t-1}(0) \mid X, G_g = 1] = E[Y_t(0) - Y_{t-1}(0) \mid X, D_s = 0, G_g = 0] \; a.s..$$

Hipóteses 4 e 5 generalizam a hipótese de tendências paralelas em apenas 2 períodos para o caso de múltiplos períodos de tempo e diversos grupos de tratamento.

Hipóteses (4/4)

- **(Continuação) Hipóteses 4 e 5**:
  - Hipótese 4 postula que, condicionada às covariadas X, os resultados médios para o grupo que inicia o tratamento no período g e para o grupo de "nunca-tratados" deveriam seguir trajetórias paralelas na ausência de tratamento.
  - Hipótese 5 impõe tendências paralelas condicionais entre o grupo g e grupo de "ainda não tratados" no período $t + \delta$. Observe que, para um $\delta = 1$, o grupo que será tratado no período seguinte poderia não seguir tendências paralelas ao grupo tratado no período t, dada a antecipação de tratamento.

- **Hipótese 6 (Sobreposição/Overlap)**. Para todo $t = 2, ..., \mathcal{T}$, existe algum $\varepsilon > 0$, tal que $P(G_g = 1) > \varepsilon$ e $p_{g,t}(X) < 1 - \varepsilon$ a.s..
  Para proporção positiva da população que começou a ser tratada no período g e que, para todo g e t, o *propensity score* generalizado é uniformemente limitado e longe de 1. Ou seja, exclui possibilidade de "identificação irregular".

Identificação Não-Paramétrica de $ATT(g,t)$ (1/3)

Para a identificação não-paramétrica de $ATT(g,t)$, podemos utilizar as abordagens

- de regressão de resultados (OR) de Heckman et al. (1997, 1998),
- de ponderação probabilidade inversa (IPW) de Abadie (2005), e
- duplamente robusto (DR) de Sant'Anna e Zhao (2020).

Sejam,

$$m_{g,t,\delta}^{nev}(X) = E[Y_t - Y_{g-\delta-1} \mid X, C = 1]$$
$$m_{g,t,\delta}^{ny}(X) = E[Y_t - Y_{g-\delta-1} \mid X, D_{t+\delta} = 0, G_g = 0]$$

as regressões de resultados populacionais para o grupo nunca-tratado (*never-treated* - *nev*) e para o ainda-não-tratado (*not-yet-treated* - *ny*) pelo grupo do tempo $t + \delta$.

Identificação Não-Paramétrica de $ATT(g,t)$ (2/3)

Sejam, para grupo nunca-tratado,

$$ATT_{ipw}^{nev}(g,t;\delta) = E\left[\left(\frac{G_g}{E[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{E\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)(Y_t - Y_{g-\delta-1})\right], \tag{2.2}$$

$$ATT_{or}^{nev}(g,t;\delta) = E\left[\frac{G_g}{E[G_g]}\left(Y_t - Y_{g-\delta-1} - m_{g,t,\delta}^{nev}(X)\right)\right], \tag{2.3}$$

$$ATT_{dr}^{nev}(g,t;\delta) = E\left[\left(\frac{G_g}{E[G_g]} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{E\left[\frac{p_g(X)C}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-\delta-1} - m_{g,t,\delta}^{nev}(X)\right)\right]. \tag{2.4}$$

Identificação Não-Paramétrica de $ATT(g,t)$ (3/3)

Analogamente, sejam, para grupos ainda-não-tratados,

$$ATT^{ny}_{ipw}(g,t;\delta) = E\left[\left(\frac{G_g}{E[G_g]} - \frac{\frac{p_g(X)(1-D_{t+\delta})(1-G_g)}{1-p_g(X)}}{E\left[\frac{p_g(X)(1-D_{t+\delta})(1-G_g)}{1-p_g(X)}\right]}\right)(Y_t - Y_{g-\delta-1})\right], \tag{2.5}$$

$$ATT^{ny}_{or}(g,t;\delta) = E\left[\frac{G_g}{E[G_g]}\left(Y_t - Y_{g-\delta-1} - m^{ny}_{g,t,\delta}(X)\right)\right], \tag{2.6}$$

$$ATT^{ny}_{dr}(g,t;\delta) = E\left[\left(\frac{G_g}{E[G_g]} - \frac{\frac{p_g(X)(1-D_{t+\delta})(1-G_g)}{1-p_g(X)}}{E\left[\frac{p_g(X)(1-D_{t+\delta})(1-G_g)}{1-p_g(X)}\right]}\right)\left(Y_t - Y_{g-\delta-1} - m^{ny}_{g,t,\delta}(X)\right)\right]. \tag{2.7}$$

Teorema 1 (1/2)

Teorema 1. Sob hipóteses 1, 2, 3 e 6,

(i) Se Hipótese 4 é válida, então, para todo $g \in \mathcal{G}_\delta$, $t \in \{2, ..., \mathcal{T} - \delta\}$ e $t \geq g - \delta$,

$$ATT(g,t) = ATT_{ipw}^{nev}(g,t;\delta) = ATT_{or}^{nev}(g,t;\delta) = ATT_{dr}^{nev}(g,t;\delta)$$

(ii) Se Hipótese 5 é válida, então, para todo $g \in \mathcal{G}_\delta$, $t \in \{2, ..., \mathcal{T} - \delta\}$ e $t \geq g - \delta$,

$$ATT(g,t) = ATT_{ipw}^{ny}(g,t;\delta) = ATT_{or}^{ny}(g,t;\delta) = ATT_{dr}^{ny}(g,t;\delta)$$

Ou seja, podemos recuperar $ATT(g,t)$ explorando diferentes partes do *dgp*:

- Regressão de resultados (OR) que depende na modelagem da expectativa condicional da evolução de resultados para os grupos de comparação;
- Ponderação Probabilidade Inversa (IPW) que depende da modelagem da probabilidade condicional em ser do grupo g; e
- Duplamente Robusto (DR) que utiliza ambos OR e IPW

Teorema 1 (2/2)

Para aplicar os resultados de Heckman et al. (1997, 1998), Abadie (2005), e Sant'Anna e Zhao (2020) para múltiplos períodos e grupos, precisamos lidar com:

- Período de tempo referência apropriado: pelo Teorema 1 podemos usar o período de tempo $t = g - \delta - 1$ como referência. É o período mais recente quando resultados potenciais não-tratados pode ser observados para o grupo g. Quanto maior δ, mais temos que "voltar no tempo".
- Grupos de comparação apropriado:
  - Sob Hipótese 4, podemos tratar unidades nunca-tratadas como grupo de controle fixo para todas unidades "eventualmente tratadas";
  - sob Hipótese 5, podemos usar unidades ainda não-tratadas no período $t + \delta$ como grupo de controle válido para aqueles que são tratados inicialmente no período g.

Sumarização de Efeitos de Tratamento Médio Grupo-Temporal

Sumarização de $ATT(g,t)$

- Quando os números de grupos e de períodos de tempo são grandes, pode ser difícil interpretar muito efeitos de tratamento médio grupo-temporal, $ATT(g,t)$.
- Veremos algumas agregações que podem ser feitas:
  - Único parâmetro de efeito de tratamento global;
  - Agregações relacionadas ao entendimento de efeitos dinâmicos;
  - Agregações por grupo; e
  - Agregações por tempo de calendário.
- Esquemas de agregação têm a seguinte forma geral:

$$\theta = \sum_{g \in \mathcal{G}} \sum_{t=2}^{\mathcal{T}} w(g,t).ATT(g,t), \tag{3.1}$$

em que $w(g,t)$ são funções de ponderação especificados pelos pesquisadores tal que θ possa ser utilizado.

Problema de Ponderação Negativa (1/2)

- As medidas mais utilizadas para sumarizar ATT em DiD com múltiplos períodos são as especificações de efeitos fixos bidirecionais (TWFE) estático e dinâmico:

$$Y_{i,t} = \alpha_t + \alpha_g + \beta D_{i,t} + \varepsilon_{i,t}, \tag{3.2}$$

$$Y_{i,t} = \alpha_t + \alpha_g + \sum_{e=-K}^{-2} \delta_e^{antecip}.D_{i,t}^e + \sum_{e=0}^{L} \beta_e.D_{i,t}^e + v_{i,t}, \tag{3.3}$$

respectivamente, em que:

  - α_t é um efeito fixo no tempo;
  - α_g é um efeito fixo de grupo;
  - $\varepsilon_{i,t}$ e $v_{i,t}$ são termos de erro;
  - $D_{i,t}^e = 1\{t - G_i = e\}$ é um indicador para a unidade i estando e períodos depois do tratamento inicial no tempo t; e
  - K e L são constantes positivas.

Problema de Ponderação Negativa (2/2)

- Normalmente, o parâmetro de interesse na especificação TWFE estático é o β, interpretado como efeito global de participação no tratamento entre grupos e períodos.
- Na especificação TWFE dinâmico, o foco é normalmente no β_e, com $e = t - g \geq 0$, e esses parâmetros são tipicamente interpretados como o efeito de participar do tratamento em diferentes durações do tratamento, e.

- **Problema**: β recupera uma média ponderada de alguns parâmetros de efeito de tratamento, mas alguns pesos podem ser negativos.
- O "problema do peso negativo" ocorre quando efeitos de tratamento evoluem ao longo do tempo. O mesmo pode ocorrer na especificação dinâmica ao β_e.

- Por facilidade, considera-se $\delta = 0$ nesta seção, ou seja, pessoas não antecipam tratamento.

Como ATT's variam com a duração de exposição ao tratamento? (1/3)

- Seja $e = t - g$ o tempo decorrido desde o tratamento. Relembre que G denota o período de tempo em que a unidade foi tratada inicialmente.
- Para destacar a heterogeneidade do efeito de tratamento em relação à duração e (**estudo de evento** - es), agregamos $ATT(g,t)$'s em

$$\theta_{es} = \sum_{g \in \mathcal{G}} \mathbf{1}\{g + e \leq \mathcal{T}\} P(G = g | G + e \leq \mathcal{T}) ATT(g, g + e). \tag{3.4}$$

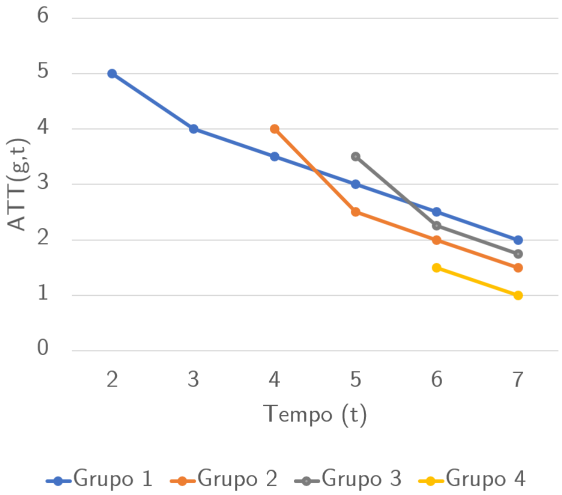
- Esse é efeito médio de participar no tratamento por e períodos após o início dos tratamentos, entre todos os grupos que participaram de exatamente por e períodos.
- Limitação: a diferença entre dois $\theta_{es}(e)$'s não é exatamente o efeito dinâmico de participar do tratamento. Isso só é válido quando $ATT(g, t + e)$ são homogêneos entre os g, ou seja, efeitos dinâmicos são comuns entre grupos.

Como ATT's variam com a duração de exposição ao tratamento? (2/3)

- Uma alternativa é "balancear" os grupos por agregação de $ATT(g,t)$'s para um conjunto de grupos que são expostos ao tratamento por um nº particular de períodos.
- Em participar, para algum tempo de evento e', com $0 \leq e \leq e' \leq \mathcal{T} - 2$, seja

$$\theta_{es}^{bal}(e; e') = \sum_{g \in \mathcal{G}} \mathbf{1}\{g + e' \leq \mathcal{T}\} ATT(g, g+e) P(G = g \mid G + e' \leq \mathcal{T}). \tag{3.6}$$

- $\theta_{es}^{bal}(e; e')$ considera apenas grupos que participaram do tratamento por pelo menos e' períodos, e calcula o efeito para para as unidades cujo tempo de evento é igual a e.
- A limitação é que haverá menos grupos para computar o estimando do tipo estudo-evento $\rightarrow$ *Trade-off*: Robustez $\times$ Eficiência.



- Considere o exemplo em que queremos calcular $\theta_{es}^{bal}(e; e')$ a partir dos $ATT(g,t)$ mostrados no gráfico ao lado.
- Seja $e' = 3 \leq 4 = \mathcal{T} - 2$.
- Logo, estamos excluindo ambos grupos 3 e 4, já que a duração máxima de ambos é $e_3 = 7 - 5 = 2$ e $e_4 = 7 - 6 = 1$, respectivamente.
- Podemos computar as medidas de resumo para todo $e \leq e' = 3$:

$$\theta_{es}^{bal}(0;3); \quad \theta_{es}^{bal}(1;3);$$
$$\theta_{es}^{bal}(2;3); \quad \theta_{es}^{bal}(3;3);$$

Como ATT's variam entre grupos?

- Estes parâmetros são menos comuns, mas são úteis para entender se o efeito em participar no tratamento foi maior em grupos tratados mais cedo em relação a grupos que foram tratados depois (**seleção** - *sel*).
- Para considerar efeitos heterogêneos entre grupos, considere o seguinte:

$$\theta_{sel}(\tilde{g}) = \frac{1}{\mathcal{T} - \tilde{g} + 1} \sum_{t=\tilde{g}}^{\mathcal{T}} ATT(\tilde{g}, t). \tag{3.7}$$

$\theta_{sel}(\tilde{g})$ é o efeito médio de participar no tratamento entre unidades no grupo g, em todos os seus períodos de pós-tratamento.

O que é ATT cumulativo da política em todos os grupos até o tempo $\tilde{t}$?

- Pode ser interessante destacar a heterogeneidade do efeito de tratamento em relação ao **tempo de calendário** - c. O efeito médio de participar do tratamento no período t, entre grupos que adotaram o tratamento até o período, é dado por

$$\theta_c(t) = \sum_{g \in \mathcal{G}} \mathbf{1}\{\tilde{t} \geq g\} P(G = g \mid G \leq \tilde{t}) ATT(g, t) \tag{3.8}$$

- Uma possível extensão é ver o **efeito acumulado** de participar no tratamento até algum período particular $\tilde{t}$. Considere:

$$\theta_c^{cumu}(\tilde{t}) = \sum_{t=2}^{\tilde{t}} \theta_c(t). \tag{3.9}$$

$\theta_c^{cumu}(\tilde{t})$ pode ser interpretado como efeito de tratamento médio acumulado entre as unidades tratadas até tempo $\tilde{t}$.

Agregações em Parâmetros de Efeito de Tratamento Global (1/4)

- Uma ideia simples é apenas tomar a média de todos efeitos de tratamento médio grupo-temporal identificados:

$$\theta_W^O = \frac{1}{\mathcal{K}} \sum_{g \in \mathcal{G}} \sum_{t=2}^{\mathcal{T}} \mathbf{1}\{t \geq g\} ATT(g,t) P(G = g \mid G \leq \mathcal{T}) \qquad (3.10)$$

em que $\mathcal{K} = \sum_{g \in \mathcal{G}} \sum_{t=2}^{\mathcal{T}} \mathbf{1}\{t \geq g\} P(G = g \mid G \leq \mathcal{T})$.

- θ_W^O é uma média ponderada de cada $ATT(g,t)$, colocando mais peso nos grupos com maiores tamanhos.
- Desvantagem: θ_W^O coloca sistematicamente mais peso aos grupos que participam no tratamento por maior tempo.

- Os autores sugerem o uso do seguinte parâmetro para sumarizar o efeito médio em participar do tratamento:

$$\theta_{sel}^{O} = \sum_{g \in \mathcal{G}} \theta_{sel}(g) P(G = g | G \leq \mathcal{T}) \tag{3.11}$$

em que $\theta_{sel}(g)$ é o efeito médio de participar do tratamento para unidades no grupo g como definido na equação (3.7).

- θ_{sel}^{O} computa o efeito médio para cada grupo (entre todos períodos) via θ_{sel}, e então toma a média desses efeitos para sumarizar o efeito médio global.
- Logo, θ_{sel}^{O} é o efeito médio de participar do tratamento de todas as unidades que participaram do tratamento. A interpretação é a mesma do DiD canônico com 2 períodos e 2 grupos.

Agregações em Parâmetros de Efeito de Tratamento Global (3/4)

- Seguindo o mesmo raciocínio, alguém poderia definir os parâmetros de efeito de tratamento global tomando a média de θ_{es} entre todos os tempos de evento (estudo de evento - e), ou de $\theta_c(t)$ entre todos os períodos de tempo (calendário - c), i.e.,

$$\theta_{es}^{O} = \frac{1}{\mathcal{T}-1}\sum_{e=0}^{\mathcal{T}-2}\theta_{es}(e) \qquad \theta_{c}^{O} = \frac{1}{\mathcal{T}-1}\sum_{t=2}^{\mathcal{T}}\theta_{c}(t) \tag{3.12}$$

- Problema: θ_{es}^{O} tem a dificuldade imposta pela composição dos grupos entre diferentes valores de e. Similarmente, isso ocorre para θ_{c}^{O} também.

Agregações em Parâmetros de Efeito de Tratamento Global (4/4)

- Deveríamos balancear a amostra em relação ao tempo de evento, logo um parâmetro de sumarização único (local) é dado por:

$$\theta_{es}^{O,bal}(e') = \frac{1}{e'+1} \sum_{e=0}^{e'} \theta_{es}^{bal}(e, e') \tag{3.13}$$

- Esse é o efeito médio de participar do tratamento sobre os primeiros e' períodos de exposição ao tratamento.

- Observação: Nenhum destes parâmetros de efeito global são iguais aos demais, exceto no caso em que $ATT(g,t)$ é o mesmo para todos os grupos g e em todos os períodos t. Neste caso, todos parâmetros agregados, incluindo o β da regressão TWFE, são iguais.

Estimação e Inferência

Procedimentos de Estimação e Inferência

Note que os resultados de identificação no Teorema 1 sugerem uma estratégia de estimação em 2 etapas para estimar os $ATT(g,t)$'s:

1. Estima-se as funções para cada grupo g e período t:
   - $p_g(x)$ e/ou $m^{nev}_{g,t,\delta}(X)$ (se depende da hipótese 4); e
   - $p_{g,t+\delta}(x)$ e/ou $m^{ny}_{g,t,\delta}(X)$ (se depende da hipótese 5).
2. Aplica os valores ajustados dessas funções estimadas na amostra análoga do estimando de $ATT(g,t)$, que está sendo considerado para obter estimativas de efeito de tratamento médio grupo-temporal.

Qual abordagem utilizar?

- Uma questão natural é qual abordagem utilizar na prática: a regressão de resultado (OR), ponderação de probabilidade inversa (IPW), ou a duplo-robusta (DR)?
- Embora sejam equivalentes pela perspectiva de **identificação**, não são equivalentes na perspectiva de **estimação/inferência**:
  - Abordagem OR: requer a modelagem correta da evolução do resultado do grupo de controle para estimar os $ATT(g,t)$'s.
  - Abordagem IPW: evita a modelagem da evolução do resultado do grupo de controle, mas requer a modelagem correta do probabilidade condicional da unidade i ser do grupo g dadas covariadas X.
  - Abordagem DR: combina ambos OR e IPW, pois modela ambos evolução do resultado para o grupo de controle, e *propensity score*, porém requer que apenas um dos dois seja corretamente especificado (Sant'Anna e Zhao, 2020). Portanto, tem uma robustez adicional contra erros de especificação e também permite o uso de outros métodos de estimação, como penalização e tipos de modelo de seleção.

O Efeito da Política de Salário Mínimo no Emprego de Adolescentes

Objetivo

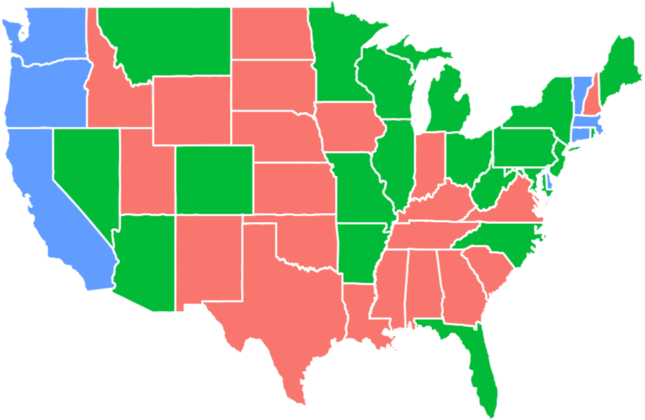
- Neste seção, esses métodos propostos serão aplicado para estudar o efeito do aumento do salário mínimo no emprego de adolescentes.
- Serão comparados os resultados usando a especificação TWFE (mais utilizado em aplicações) aos resultados do método proposto, como forma de entendermos as limitações do TWFE.

Dados

- São usados dados de emprego de adolescentes a nível de condado obtidos na *Quarterly Workforce Indicators* (QWI).
- Características de pré-tratamento foram obtidas no *County Data Book* 2000:
  - População do condado em 2000, Fração da população que é branca, Características educacionais de 1990, Renda mediana em 1997, e Fração da população abaixo do nível da pobreza em 1997.
- Em 1997, o salário mínimo federal aumentou de $4,75 por hora para $5,15 por hora, e permaneceu igual até Julho de 2007.
- Foram usados dados de 2001 a 2007 para 29 dos 40 Estados que possuíam as informações necessárias. Em todos estes, o salário mínimo era inicialmente igual ao salário mínimo Federal.
- Parte não elevou o salário em 2007 (não-tratados) e houve grupos que elevaram o salário mínimo em 2004, 2006 e 2007 (Tratados).

Distribuição Espacial dos Estados por Política de Salário Mínimo (2018)

Figure 2: The Spatial Distribution of States by Minimum Wage Policy



- Estados em verde aumentaram seu salário mínimo (inicialmente igual ao Federal) entre 200 e 2007.
- Estados em azul possuíam o salários mínimos em patamares superiores ao Federal já no início de 2000.
- Estados em vermelho não aumentaram o salário entre 2000 e 2007, logo serão controles.
- Alguns Estados foram omitidos por falta de dados relevantes ou por estarem na região Norte dos EUA, em que nenhum estado aumentou o salário mínimo entre 2000 e 2007.

Resumo das Estatísticas Principais

Table 2: Summary Statistics for Main Dataset

	Treated Counties	Untreated Counties	Diff.	P-val on Diff.
Midwest	$\boxed{0.59}$	0.34	0.25	0.00
South	0.27	$\boxed{0.59}$	-0.32	0.00
West	0.14	0.07	0.07	0.00
Population (1000s)	$\boxed{94.32}$	53.43	40.89	0.00
White	0.89	0.83	0.06	0.00
HS Graduates	0.59	0.55	0.04	0.00
Poverty Rate	0.13	0.16	-0.03	0.00
Median Inc. (1000s)	33.91	31.89	2.02	0.00

Notes: Summary statistics for counties located in states that raised their minimum wage between Q2 of 2003 and Q1 of 2007 (treated) and states whose minimum wage was effectively set at the federal minimum wage for the entire period (untreated). The sample consists of 2284 counties. *Sources:* Quarterly Workforce Indicators and 2000 County Data Book

Figure: Maior parte dos condados tratados eram do Midwest. Eles também têm uma população maior (em média 94.000 em comparação com 53.000 nos condados não-tratados). A proporção de brancos e a mediana de renda são um pouco superiores nos condados tratados. E a diferença é pequena em formados no ensino médio e taxa de pobreza.

Resultados (1/9)

- Compara-se os casos em que assume-se tendências paralelas não-condicionadas, e quando ocorre após controle de características X.
- Considera-se o caso em que os condados nunca-tratados são o grupo de controle e que não permitimos efeitos de antecipação (i.e., $\delta = 0$). No Apêndice Suplementar, são mostrados resultados parecidos para $\delta = 1$.
- Todos procedimentos de inferência usem erros padrão *clustered bootstrapped* a nível de condado, e leva em conta a autocorrelação dos dados.
- As estimativas de efeito de tratamento médio grupo-temporal sustenta a visão de que o crescimento do salário mínimo leva à redução do emprego de adolescentes.

Figure 1: Minimum Wage Group-Time Average Treatment Effects

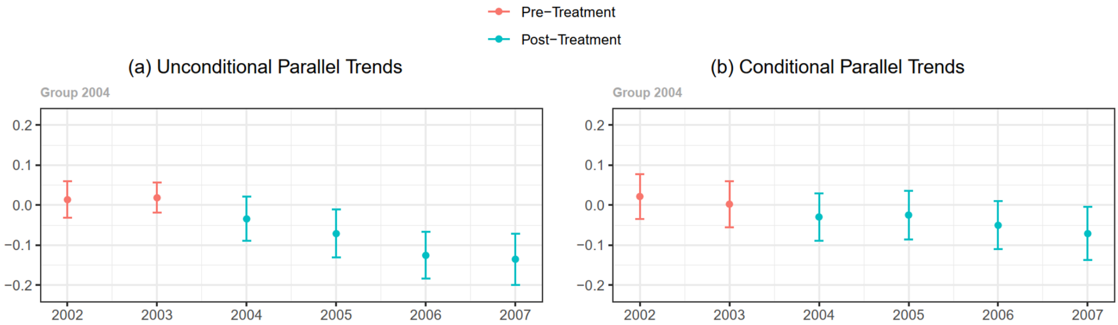


Figure: Os $ATT(g,t)$'s "individuais" diminuíram entre 2,3% a 13,6% o emprego de adolescentes (entre todos os grupos e todos os períodos).

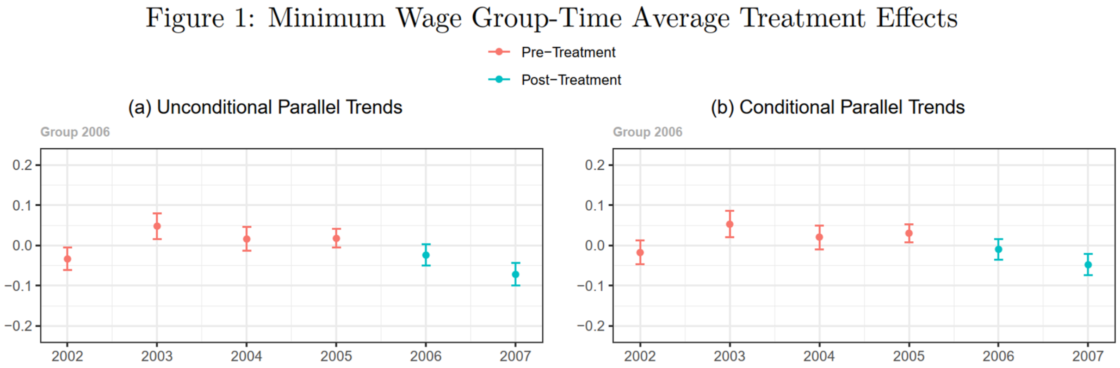


Figure: Os $ATT(g,t)$'s "individuais" diminuíram entre 2,3% a 13,6% o emprego de adolescentes (entre todos os grupos e todos os períodos).

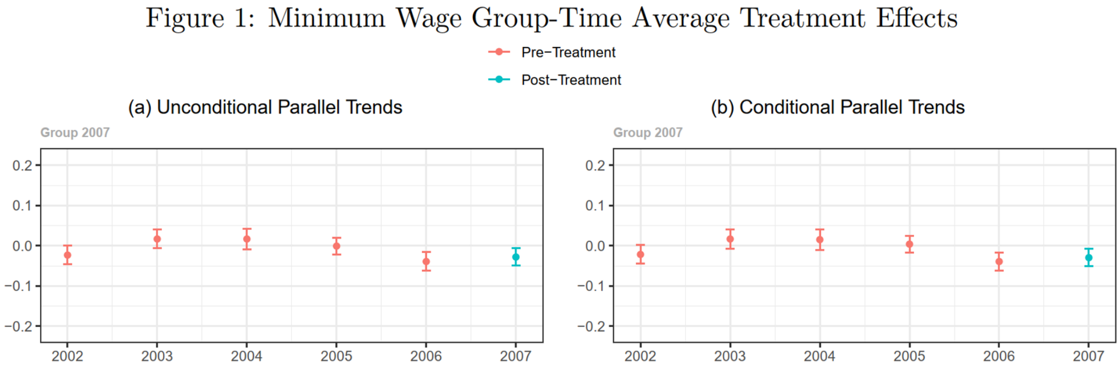


Figure: Os $ATT(g,t)$'s "individuais" diminuíram entre 2,3% a 13,6% o emprego de adolescentes (entre todos os grupos e todos os períodos).

Table 3: Minimum Wage Aggregated Treatment Effect Estimates

(a) Unconditional Parallel Trends

	Partially Aggregated				Single Parameters
TWFE					-0.037 (0.006)
Simple Weighted Average					-0.052 (0.006)
Group-Specific Effects	g=2004	g=2006	g=2007		
	-0.091 (0.019)	-0.047 (0.008)	-0.028 (0.007)		-0.039 (0.007)
Event Study	e=0	e=1	e=2	e=3	
	-0.027 (0.006)	-0.071 (0.009)	-0.125 (0.021)	-0.136 (0.023)	-0.090 (0.013)
Calendar Time Effects	t=2004	t=2005	t=2006	t=2007	
	-0.034 (0.019)	-0.071 (0.02)	-0.055 (0.009)	-0.050 (0.006)	-0.052 (0.013)
Event Study w/ Balanced Groups	e=0	e=1			
	-0.027 (0.009)	-0.071 (0.009)			-0.049 (0.008)

Resultados (6/9)

- Uma média simples (ponderada apenas pelo tamanho do grupo) resultado em uma diminuição de 5,2% do emprego de adolescentes.
- O efeito médio do aumento do salário mínimo entre todos os grupos (corresponde ao θ^{O}_{sel}) é a diminuição de 3,9% do emprego do adolescente.
- O modelo de efeitos fixos bidirecionais (TWFE) também resultou em um resultado similar, diminuição de 3,7%.
- No estudo de evento, vemos que o efeito de aumento do salário mínimo aparenta ser negativo e crescente em magnitude ao longo do tempo.
- Quando restringimos a amostra para incluir apenas grupos que tiveram aumento de salário mínimo por pelo menos 1 ano inteiro (grupos 2004 e 2006), as estimativas, neste caso, são iguais ao do Estudo de Eventos

Table 3: Minimum Wage Aggregated Treatment Effect Estimates

(b) Conditional Parallel Trends

	Partially Aggregated				Single Parameters
TWFE					-0.008 (0.006)
Simple Weighted Average					-0.033 (0.007)
Group-Specific Effects	g=2004 -0.044 (0.020)	g=2006 -0.029 (0.008)	g=2007 -0.029 (0.008)		-0.031 (0.007)
Event Study	e=0 -0.024 (0.006)	e=1 -0.041 (0.009)	e=2 -0.050 (0.022)	e=3 -0.071 (0.026)	-0.046 (0.013)
Calendar Time Effects	t=2004 -0.030 (0.022)	t=2005 -0.025 (0.021)	t=2006 -0.030 (0.009)	t=2007 -0.049 (0.007)	-0.033 (0.012)
Event Study w/ Balanced Groups	e=0 -0.016 (0.010)	e=1 -0.041 (0.009)			-0.028 (0.008)

Resultados (8/9)

- Agora assumiu-se hipótese de tendências paralelas condicionais, ou sejam supomos que apenas condados com as mesmas características iriam seguir a mesma tendência de emprego de adolescentes na ausência de tratamento.
- Foram usadas as seguintes características mostradas na Tabela 2.
- Foi usado o procedimento duplamente-robusto de Sant'Anna e Zhao, 2020)
- Então, na 1ª etapa, foi feita a estimação do *propensity score* generalizado e a regressão de resultado.
  - Cada *propensity score* generalizado foi estimado um modelo logit que inclui as características de cada condado, junto de termos quadráticos para população e para a renda mediana.
  - Para regressão de resultado, foi usada a mesma especificação para as covariadas.

Resultados (9/9)

- Controlando pelas covariadas, obtemos resultados diferentes da abordagem de efeitos fixos bidirecionais (TWFE), sendo que esta não é estatisticamente significante.
- De maneira geral, a estimação pelos métodos propostos mantiveram os efeito negativos e significativos.
- A principal conclusão é que, mantendo hipóteses de identificação constantes, em uma aplicação econômica que possui muitas funcionalidades (heterogeneidade de efeito de tratamento, efeitos dinâmicos, e adoção de tratamento escalonado), a escolha do método de estimação pode potencialmente levar a diferentes conclusões.

Conclusão

Conclusão

- Considerou DiD em que há mais de 2 períodos de tempo e mais de 2 unidades tratadas.
- Propôs efeitos de tratamento médio grupo-temporal, $ATT(g,t)$.
- Quando obtidos diversos $ATT(g,t)$, foram mostradas possíveis agregações.
- Essa abordagem é adequada quando:
  - (i) a hipótese de tendências paralelas é válida apenas após condicionar nas covariadas;
  - (ii) há uso de diferentes grupos de controle, como nunca-tratados ou ainda-não-tratados;
  - (iii) as unidades podem antecipar o tratamento e ajustar seus comportamentos antes do tratamento ser implementado.
- Provou a consistência e a normalidade assintótica para os estimadores propostos.
- Custos computacionais são baixos e pode ser implementado pelo pacote de R *did*.
- Pelo exemplo notou-se diferenças entre o TWFE e a abordagem proposta, o que sugere que esta seja considerada por pesquisadores.