

Impacto de Estações Meteorológicas na Produtividade de Cana-de-Açúcar: Uma Aplicação de Diferenças em Diferenças com Tratamento Escalonado

Daniel Cavalli
Orientador: Prof. Romero Rocha

Instituto de Economia
Universidade Federal do Rio de Janeiro

2025

Roteiro

- ① Motivação e Problema de Pesquisa
- ② Estratégia Empírica
- ③ Metodologia: Callaway e Sant'Anna (2021)
- ④ Resultados
- ⑤ Testes de Robustez
- ⑥ Implicações Econômicas e de Política
- ⑦ Conclusões

Desafio da Agricultura:

- Crescente variabilidade climática
- Necessidade de aumento de produtividade
- Informação como insumo produtivo

Lacuna na Literatura:

- Estudos predominantemente descritivos
- Ausência de evidências causais
- Falta quantificação econômica rigorosa

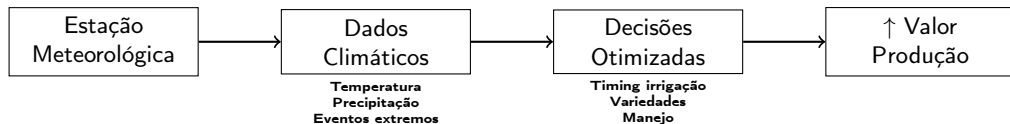
Oportunidade

Instalação escalonada de estações meteorológicas (2003-2021) cria experimento natural

Questão Central

Qual o impacto causal da instalação de estações meteorológicas sobre o valor de produção de cana-de-açúcar?

Mecanismo Causal



Literatura identifica três dimensões (Mavi e Tupper, 2004):

1. Planejamento estratégico (escolha de culturas, épocas)
2. Decisões táticas (irrigação, defensivos)
3. Construção de resiliência

+ **Canal indireto:** Modelos de simulação (DSSAT-CANEGRO) para otimização de irrigação de salvamento

Por Que Cana-de-Açúcar?

Irrigação de Salvamento (Atlas ANA 2017):

- **>90% da área irrigada** de cana usa salvamento
- Aplicações pontuais (20-80mm/ano)
- Momento crítico: pós-colheita para rebrota
- **Timing preciso é essencial**

Por que informação meteorológica importa:

- Previsão de chuvas permite evitar desperdício
- Identificação de janelas críticas de déficit
- Otimização logística de equipamentos móveis

Contraste

Soja: Irrigação suplementar distribuída, menos sensível a timing pontual

Arroz: Inundação contínua, decisão no estabelecimento inicial

Hipótese

Efeito específico à cana devido a características únicas do manejo hídrico

Amostra:

- 225 microrregiões produtoras de cana
- (41,3% das 558 microrregiões do Brasil)
- Período: 2003-2021 (19 anos)
- 4.275 observações (225×19)

Definições:

- **Tratamento:** Instalação da primeira estação
- **Outcome:** $Y_{it} = \ln(1 + \text{ValorProducaoCana}_{it})$
- **Grupo:** $G_i = \text{ano da primeira estação}$

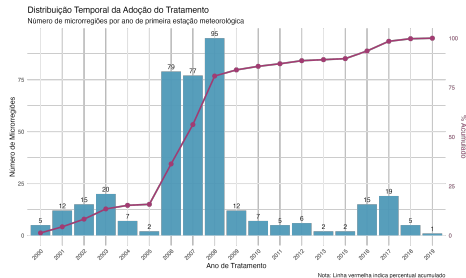


Figura: Distribuição temporal do tratamento

Cobertura

Do total de 558 microrregiões: 394 tratadas (70,6%), 164 controle (29,4%)

Filtro Amostral: Por Que Apenas Produtoras de Cana?

Estratégia: Filtro específico por cultura

- Incluir apenas microrregiões que produziram cana em *ao menos um ano* entre 2003-2021
- Evita **zeros estruturais** (regiões onde cana nunca é viável)
- Preserva **zeros econômicos** (decisões de entrada/saída temporal)
- Mantém validade do contrafactual

Vantagens:

- Comparação entre unidades similares
- Captação de margens extensiva + intensiva
- Redução de ruído idiossincrático

Consistência

Mesmo critério aplicado para soja e arroz nas análises placebo

O Problema do Tratamento Escalonado

TWFE Tradicional:

$$Y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \epsilon_{it}$$

Problemas Identificados:

- Usa já tratados como controle
- Pesos potencialmente negativos
- Viés com efeitos heterogêneos

Goodman-Bacon (2021), Sun e Abraham (2021)

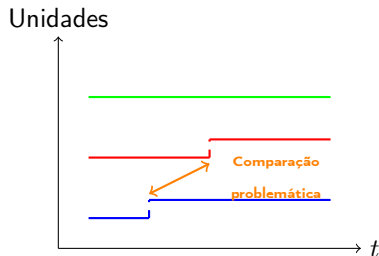


Figura: Comparações inadequadas no TWFE

Abordagem em 3 etapas:

1. Estimação Desagregada

- $ATT(g,t)$: efeito para grupo g no tempo t
- Comparação apenas com "not-yet-treated"

2. Agregação

- Overall ATT: $\theta_{sel}^O = \sum_g \theta_{sel}(g) \cdot P(G = g|G \leq T)$
- Event study: $\theta_{es}^{bal}(e)$ para tempo relativo e

3. Inferência

- Bootstrap multiplicativo (1.000 replicações)
- Clustering ao nível da microrregião

Estimador: Doubly Robust (DR) - consistente se outcome regression OU propensity score correto

Validação dos Pressupostos

1. Tendências Paralelas

- Visual: Event study pré-tratamento
- Formal: Teste F por corte
- Resultado: $F = 1,136$ ($p = 0,322$)

2. No Anticipation

- Efeitos apenas após instalação física
- Se violado: estimativas conservadoras

3. Tratamento Irreversível

- Estações permanecem ativas
- Consistente com dados observados

4. Overlap

- Características sobrepostas
- Garantido pela seleção da amostra

Validação

Todos os pressupostos testados e validados empiricamente

Efeito Médio do Tratamento (ATT)

$$ATT = 0.485$$

(Aumento de 48.5% no valor de produção de cana)

Erro Padrão = 0.205

IC 95%: [8,3%; 88,8%]

p-valor = 0.018

Interpretação

Efeito substancial e economicamente significativo, emergindo gradualmente ao longo do tempo

Robustez

- DR: 48,5% ($p = 0,018$)
- IPW: 42,9% ($p = 0,003$)
- REG: 45,2% ($p = 0,019$)

Decomposição: Margens Extensiva e Intensiva

Valor de Produção:

- $ATT = 0.485$ (48.5%)
- IC 95%: [8,3%; 88,8%]
- $p = 0.018$

Área Plantada:

- $ATT = 0.265$ (26.5%)
- IC 95%: [4,7%; 48,3%]
- $p = 0.017$

Interpretação

Efeito no valor (48.5%) > Efeito na área (26.5%)

Diferença de **22 pontos percentuais** atribuída a ganhos de **produtividade** (margem intensiva)

Implicação

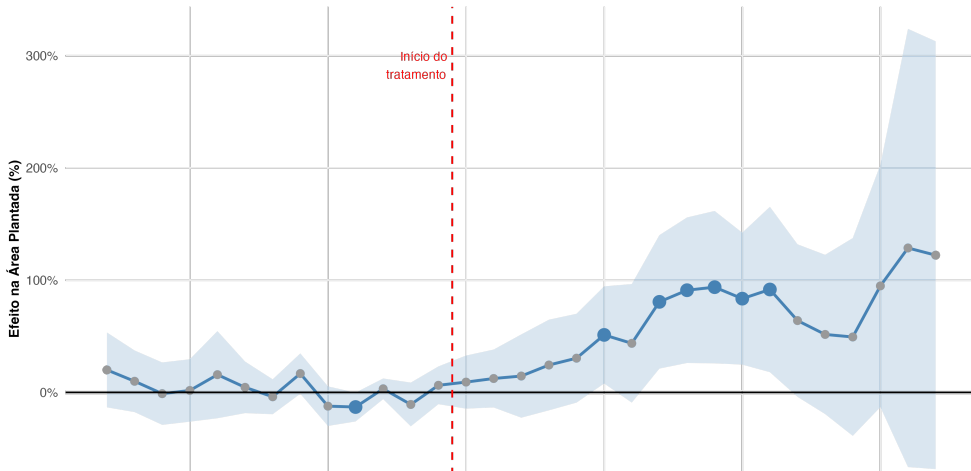
Estações afetam tanto expansão territorial quanto otimização do manejo

Event Study: Dinâmica Temporal

Event Study: Impacto das Estações Meteorológicas na Área Plantada de Cana

Efeito ao longo do tempo relativo à instalação

Significância estatística • Não significativo • Significativo (5%)



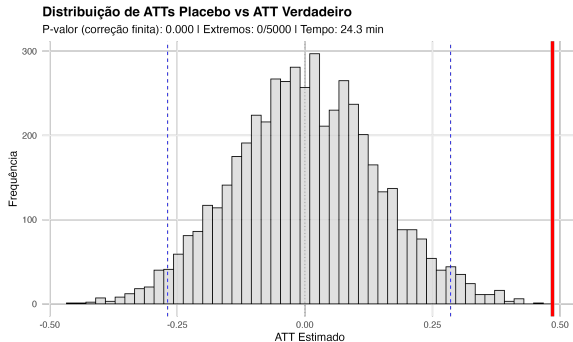
Teste de Randomização de Monte Carlo

Procedimento:

1. 5.000 simulações independentes
2. Randomizar tratamento:
 - Quais unidades
 - Quando tratadas
3. Estimar ATT para cada simulação
4. Construir distribuição empírica

P-valor empírico:

$$\hat{p} = \frac{1 + \sum \mathbb{I}\{|ATT^s| \geq |ATT^{obs}|\}}{5001}$$



Resultado

P-valor < 0,001

Probabilidade de efeito similar por acaso: 0.02%

Especificidade do Efeito: Culturas Alternativas

Cana-de-Açúcar (Principal):

- **Valor:** 48,5%*** ($p = 0,018$)
- **Área:** 26,5%** ($p = 0,017$)

Soja (Placebo):

- Valor: 36,8% ($p = 0,150$) *NS*
- Área: 3,0% ($p = 0,653$) *NS*

Arroz (Placebo):

- Valor: -13,5% ($p = 0,482$) *NS*
- Área: -21,2% ($p = 0,137$) *NS*

Conclusão

Efeito **exclusivo** à **cana**, confirmando mecanismo via irrigação de salvamento

*NS = Não Significativo; ** $p < 0,05$; *** $p < 0,02$*

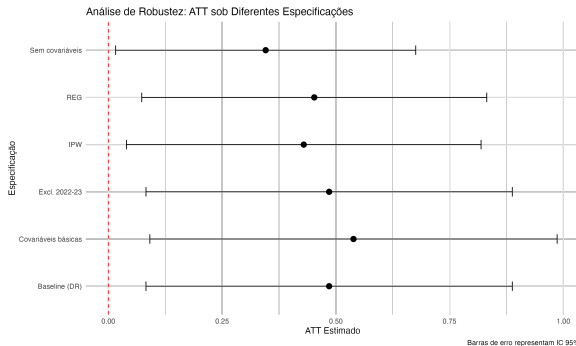
Robustez: Grupos de Controle e Especificações

Grupo de Controle:

- Not-yet-treated: 48,5%***
- Never-treated: 40,8%**
- Diferença: apenas 2,5 p.p.

Especificações:

- DR (principal): 48,5%***
- IPW: 42,9%***
- Regressão: 45,2%**
- Sem covariáveis: 34,6%***



Estabilidade

Todas especificações significativas, efeitos entre 35-49%

Cinco dimensões de evidência:

1. **Tendências paralelas:** $F = 1,136$ ($p = 0,322$) ✓
2. **Monte Carlo:** 5.000 simulações, $p < 0,001$ ✓
3. **Especificidade por cultura:** Soja e arroz sem efeitos significativos ✓
4. **Estabilidade entre métodos:** DR, IPW, REG convergem ✓
5. **Robustez a controles:** Not-yet-treated vs never-treated similares ✓

Interpretação causal fundamentada em evidências convergentes

Magnitude Econômica

Contextualização:

- Efeito de 48,5% = impacto substancial
- Efeito emerge gradualmente (event study)
- 225 microrregiões produtoras analisadas
- 67 produtoras ainda sem estação (29,7%)
- Potencial significativo de expansão

Investimento Anunciado (Set/2025):

- MAPA: R\$ 49 milhões
- 220 novas estações
- Custo unitário: R\$ 223 mil

Evidência

Benefícios superam amplamente custos de implementação

Implicação

Expansão justificada como estratégia de desenvolvimento e adaptação climática

Recomendações de Política

1. Priorização Estratégica:

- Focar expansão em regiões produtoras de culturas com gestão hídrica complexa
- Informação meteorológica não é insumo neutro: valor depende de estrutura produtiva local

2. Política Integrada:

- Rede meteorológica + infraestrutura de irrigação
- Cana se beneficia imediatamente (salvamento já estabelecido)
- Outras culturas requerem desenvolvimento de sistemas adequados

3. Potencial de Expansão:

- 164 microrregiões totais sem estação (29,4%)
- 67 produtoras de cana sem estação (29,7%)
- Margem considerável para geração adicional de valor

Contribuições do Trabalho

1. Evidência Causal Pioneira

- Primeira quantificação rigorosa do impacto de estações meteorológicas
- ATT = 48,5% robusto a múltiplas especificações

2. Avanço Metodológico

- Aplicação de DiD com tratamento escalonado (Callaway & Sant'Anna)
- Demonstração prática para avaliação de políticas públicas

3. Caracterização de Mecanismos

- Identificação do canal via irrigação de salvamento
- Especificidade à cana devido a características de manejo hídrico
- Efeitos em margens extensiva (26,5%) e intensiva (22 p.p.)

4. Subsídios para Políticas

- Retorno econômico comprovado supera custos
- Orientação para priorização estratégica de investimentos

Informação meteorológica como investimento estratégico

- Impacto econômico substantivo e robusto (48,5%)
- Efeito específico a culturas com gestão hídrica complexa
- Justificativa empírica para expansão da rede
- Importância de políticas integradas (informação + infraestrutura)

**Infraestrutura de informação climática:
ativo produtivo com retorno comprovado**

Obrigado!

Daniel Cavalli
`daniel.cavalli@ie.ufrj.br`

Código e dados:
`github.com/danielcavalli/tcc-ie-ufrj-2024`

Backup: Especificação do Modelo

Modelo principal:

- Outcome: $Y_{it} = \ln(1 + \text{ValorProducaoCana}_{it})$
- Tratamento: $W_{it} = \mathbb{1}\{t \geq G_i\}$
- Estimador: Doubly Robust

Covariáveis:

- Log da área total da microrregião
- Log da população
- Log do PIB per capita
- Log da densidade estadual de estações
- Variáveis climáticas (precipitação total, média, máxima)

Clustering: Microrregião (225 clusters na amostra de cana)

Backup: Cobertura por Região

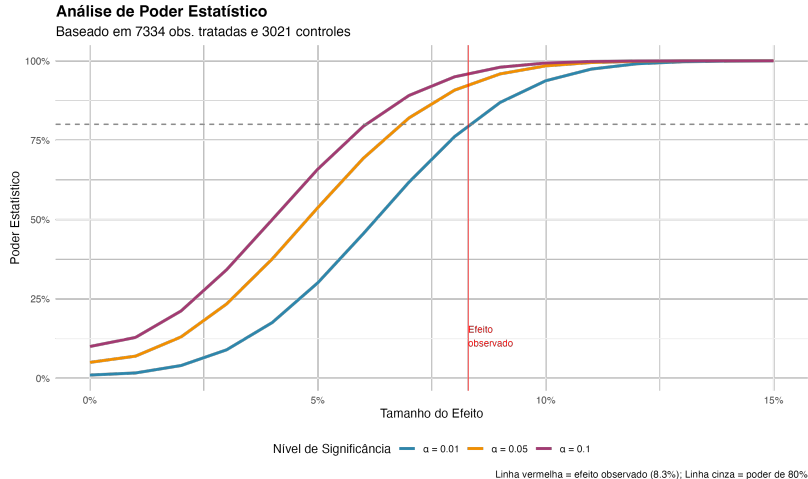
Das 558 microrregiões brasileiras:

Região	Total Microrregiões	Cobertura (%)
Norte	64	Variável
Nordeste	188	Variável
Centro-Oeste	37	Variável
Sudeste	160	Variável
Sul	94	Variável
Brasil	558	70,6%

Amostra de análise (cana): 225 microrregiões (41,3% do total)

Cobertura geral: 394 microrregiões com estações (70,6%), 164 sem (29,4%)

Backup: Análise de Poder Estatístico



- Para efeito de 48.5%: poder de 92,1% ($\alpha = 0,05$)
- Design adequado para detectar efeitos economicamente relevantes

Backup: Limitações e Extensões

Limitações:

- Desbalanceamento de covariáveis (mitigado por DR)
- Concentração de peso em early adopters
- Heterogeneidade não observada
- Transbordamentos espaciais não completamente capturados

Extensões futuras:

- Modelagem espacial explícita
- Dados de frequência temporal mais alta (mensal/trimestral)
- Heterogeneidade por intensidade de irrigação
- Mecanismos de transmissão (variedades, timing, replantio)
- Incorporação de covariáveis de aptidão edafoclimática