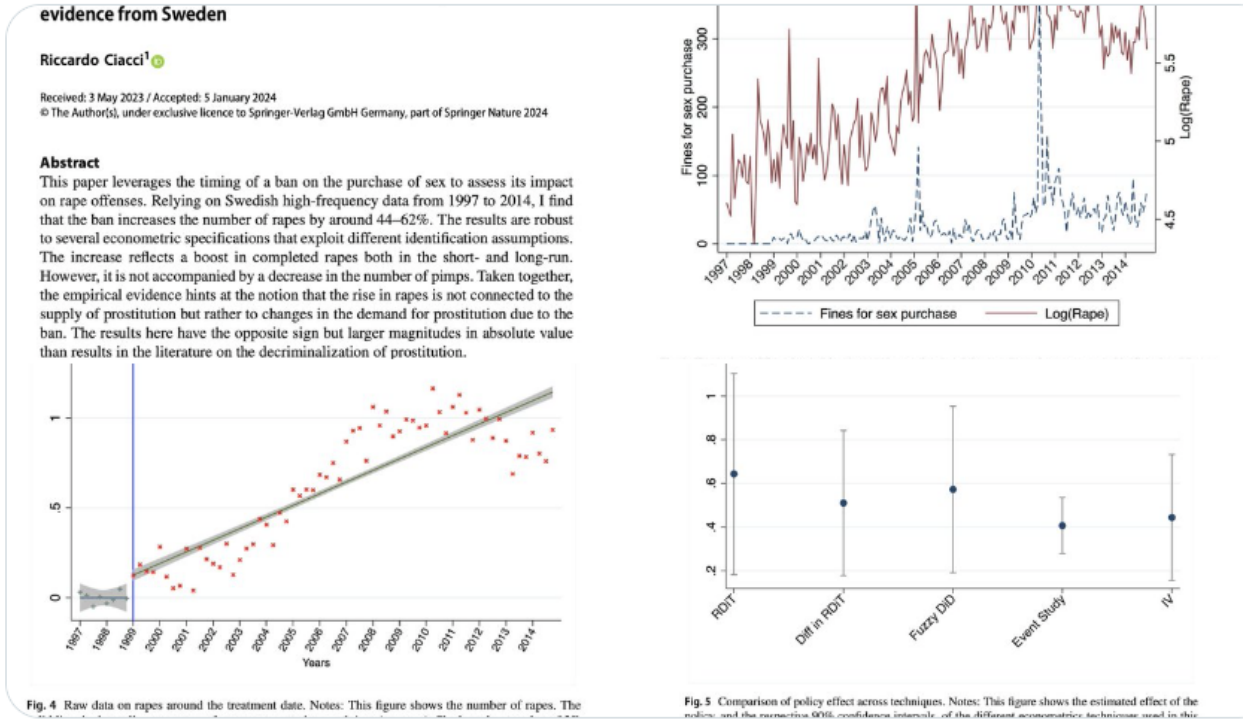
Does banning the purchase of sex increase cases of rape? Evidence from Sweden

Joop Adema (University of Innsbruck), Olle Folke (Stockholm School of Economics), & Johanna Rickne (Stockholm University),

John B. Holbein
@JohnHolbein1

Banning the purchase of sex increases cases of rape.

link.springer.com/article/10.100...



evidence from Sweden

Riccardo Ciacci[1]

Received: 3 May 2023 / Accepted: 5 January 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

This paper leverages the timing of a ban on the purchase of sex to assess its impact on rape offenses. Relying on Swedish high-frequency data from 1997 to 2014, I find that the ban increases the number of rapes by around 44–62%. The results are robust to several econometric specifications that exploit different identification assumptions. The increase reflects a boost in completed rapes both in the short- and long-run. However, it is not accompanied by a decrease in the number of pimps. Taken together, the empirical evidence hints at the notion that the rise in rapes is not connected to the supply of prostitution but rather to changes in the demand for prostitution due to the ban. The results here have the opposite sign but larger magnitudes in absolute value than results in the literature on the decriminalization of prostitution.

Fig. 4 Raw data on rapes around the treatment date. Notes: This figure shows the number of rapes. The

Fig. 5 Comparison of policy effect across techniques. Notes: This figure shows the estimated effect of the

Effect of the policy on rape in %

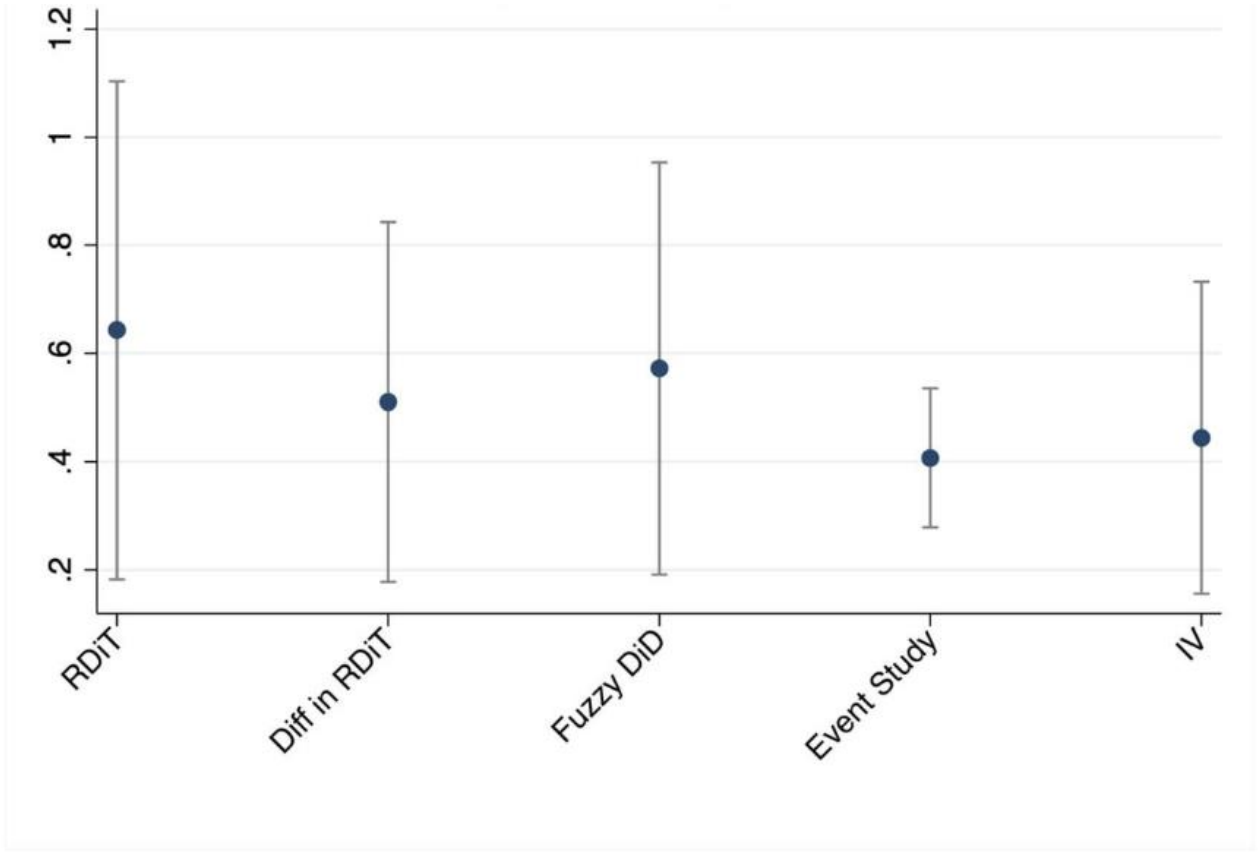
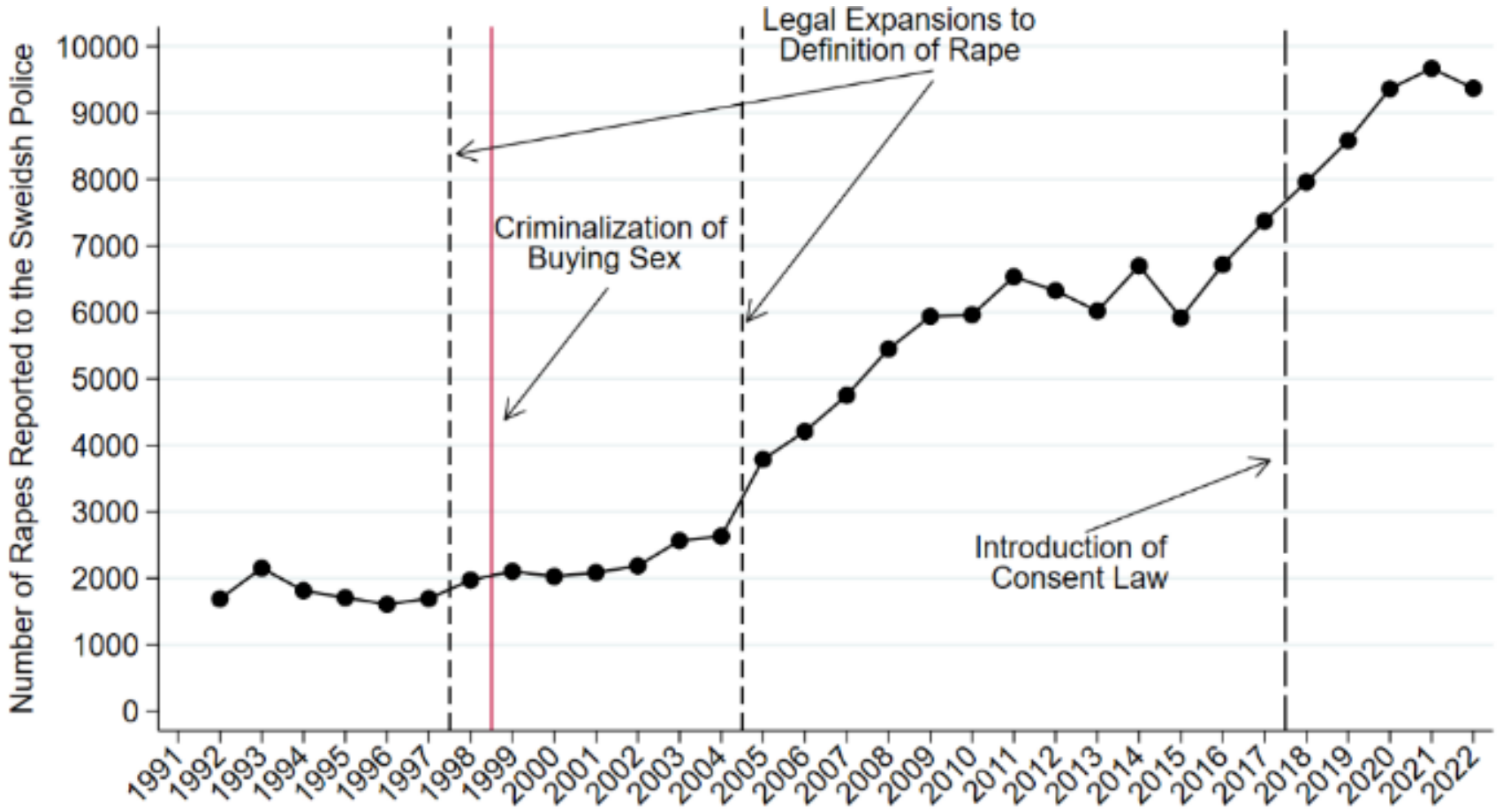


Fig. 5 Comparison of policy effect across techniques. Notes: This figure shows the estimated effect of the policy, and the respective 90% confidence intervals, of the different econometrics techniques used in this paper. Confidence intervals overlap across specifications (i.e., estimates are statically equal)

Source: Ciacci (2024)

National trend in reported rapes



Source: Swedish National Council for Crime Prevention

This talk

- Replication timeline.
- Statistical approaches to evaluating national policies.
- Outline problematic research practices in Ciacci (2024, 2025).
- What can we learn about the reform's impact on rape?
- What can we learn about *questionable research practices* and *research misconduct* in studies using observational data?

Replication timeline

- March 2024: publication of Ciacci (2024)
  - No replication material from author or journal.
  - Code for the main identification strategy is shared.
  - We inform the journal that a coding error produces these results.
- March 2025: publication of two responses
  - Zimmerman (2025): The error exists, but no misconduct
  - Ciacci (2025): Results in the original paper are “not robust”
  - Replication package posted by the journal
- May 2025: Comprehensive reanalysis desk-rejected, complaints to Springer Nature and COPE.
- June 2025: Retraction of Ciacci (2024).

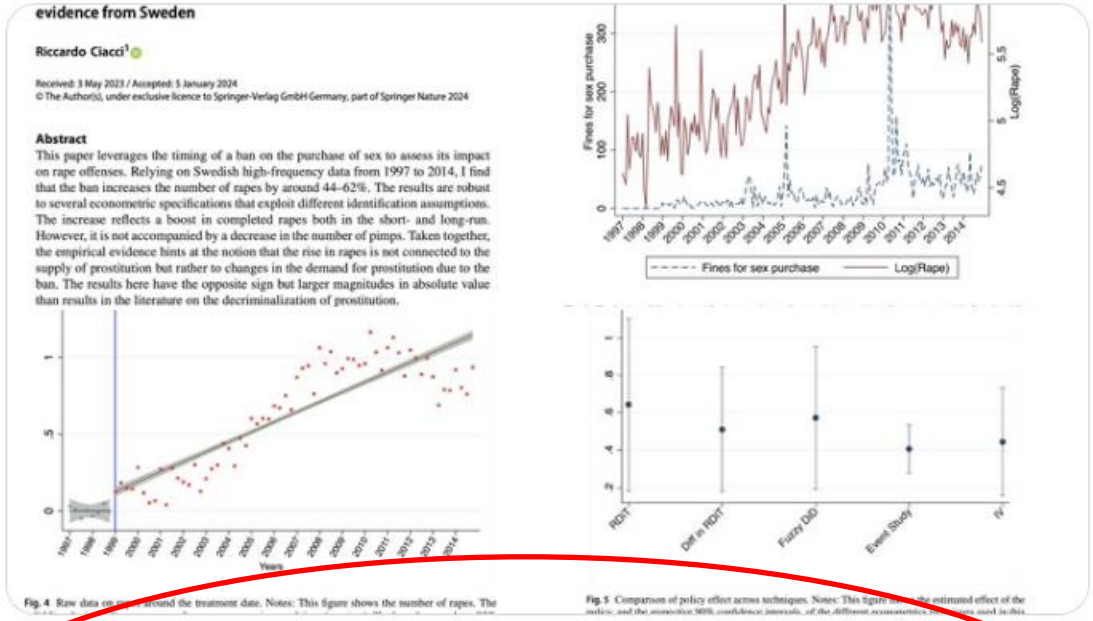
Retraction note

The Editor-in-Chief has retracted this article. Following publication, concerns were raised regarding the analysis presented in this article. Post publication review (Zimmermann 2025) concluded that the original results and conclusions are incorrect and are not supported by the data, as confirmed by a re-analysis of the data by the author (Ciacci 2025). The author disagrees with this retraction.

John B. Holbein
@JohnHolbein1

Banning the purchase of sex increases cases of rape.

link.springer.com/article/10.100...



evidence from Sweden

Riccardo Ciacci[1]

Received: 3 May 2023 / Accepted: 5 January 2024

Abstract

This paper leverages the timing of a ban on the purchase of sex to assess its impact on rape offenses. Relying on Swedish high-frequency data from 1997 to 2014, I find that the ban increases the number of rapes by around 44–62%. The results are robust to several econometric specifications that exploit different identification assumptions. The increase reflects a boost in completed rapes both in the short- and long-run. However, it is not accompanied by a decrease in the number of pimps. Taken together, the empirical evidence hints at the notion that the rise in rapes is not connected to the supply of prostitution but rather to changes in the demand for prostitution due to the ban. The results here have the opposite sign but larger magnitudes in absolute value than results in the literature on the decriminalization of prostitution.

Fig. 4 Raw data on rapes around the treatment date. Notes: This figure shows the number of rapes. The

Fig. 5 Comparison of policy effect across techniques. Notes: This figure shows the estimated effect of the

Readers added context they thought people might want to know

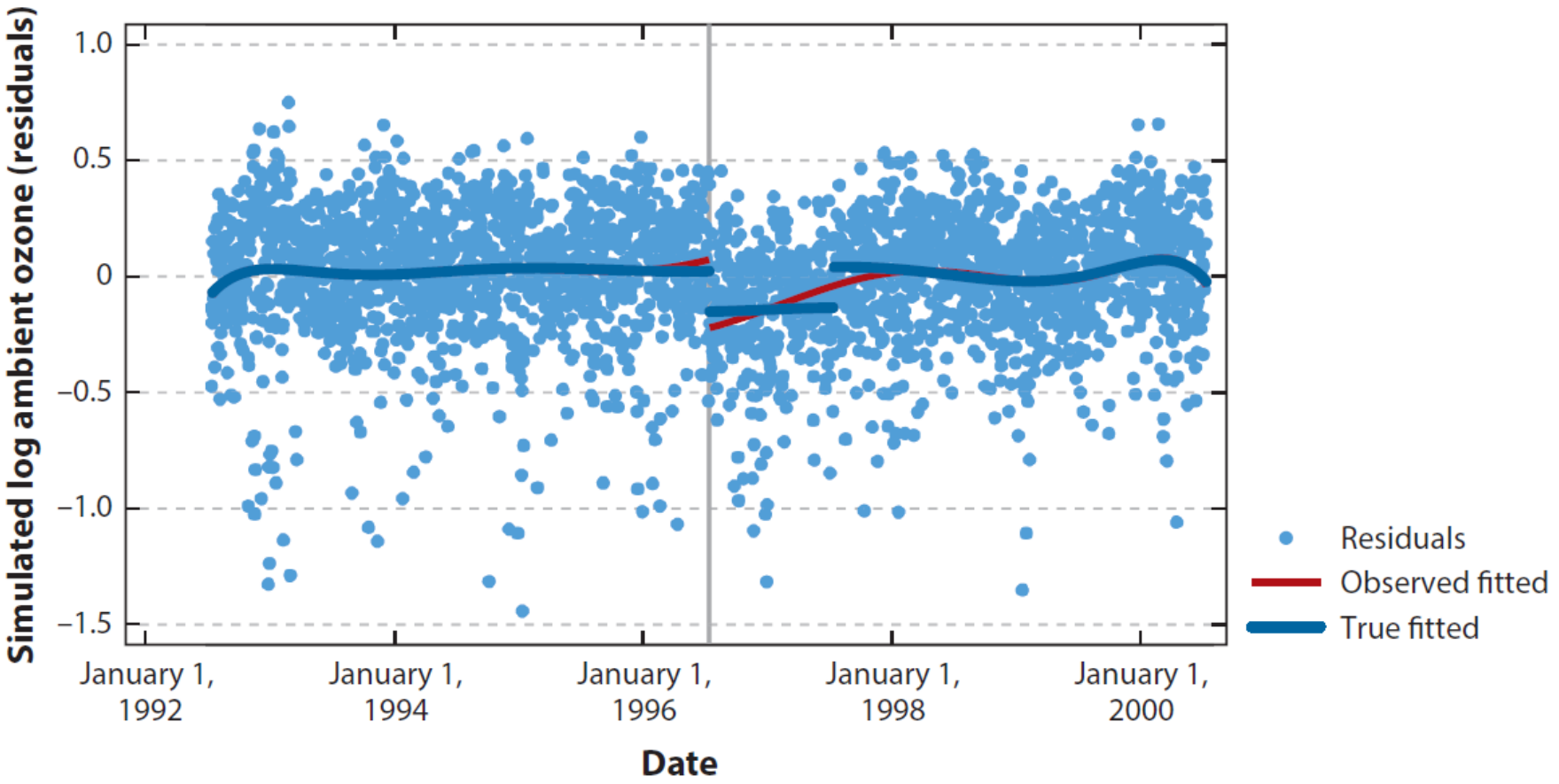
The analysis in this paper is possibly wrong due to a programming error, a corrected analysis did not find the claimed effect.

twitter.com/johannarickne/...

Evaluating national policies

- The single reform date precludes standard differences-in-differences.
- Regression discontinuity in time using only the time-series variation
  - Expanding N adds spatially correlated data
  - Forcing variable imbalanced by definition → sensitivity to omitted variables (Hausman and Rapson 2018)
- "Even with covariates included, bias is possible—for instance, a global polynomial control may overfit" (Hausman and Rapson 2018).
  - Always a key concern in RDD (Cattaneo et al 2020) and larger for discrete forcing variables (Cattaneo et al. 2024)
- Solutions? Do not use RDD (Cattaneo et al. 2024); Plot the raw data and show different control functions (Hausman and Rapson 2018).

Example of overfitting



Source: Hausman and Rapson (2018)

Policy evaluation of national policies

- Cross-sectional variation can sometimes be found.
- Ciacci (2024) theorizes that police-issued fines for buying sex signals relative price increase and triggers men to substitute prostitution for rape.
- Challenges with this approach
  - No public data on fines at the region-month level
  - Very few fines (10 issued in 1999; expected cost ↑ 0.001 USD)
  - Nearly exclusively in the three largest cities (SOU 2010:49)

How can the paper find these results?

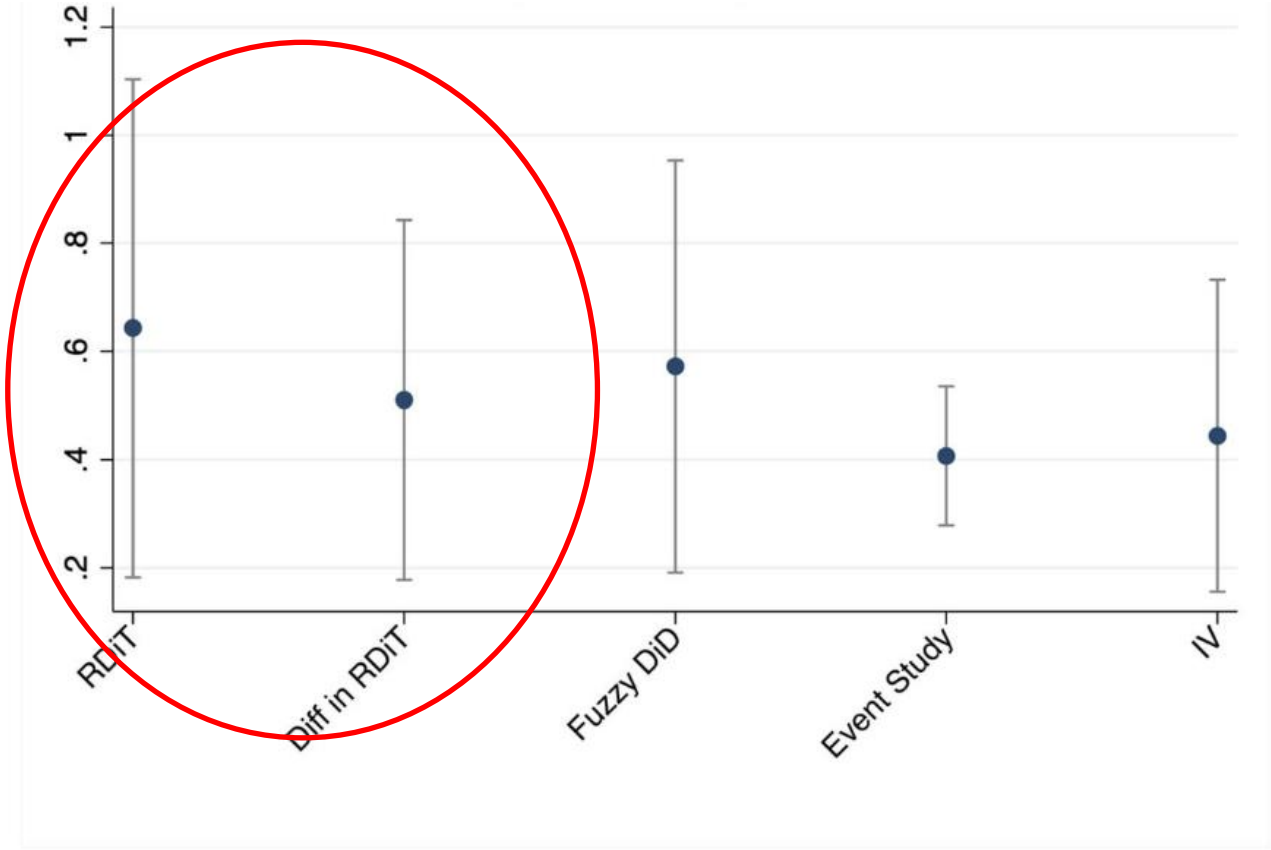


Fig. 5 Comparison of policy effect across techniques. Notes: This figure shows the estimated effect of the policy, and the respective 90% confidence intervals, of the different econometrics techniques used in this paper. Confidence intervals overlap across specifications (i.e., estimates are statically equal)

Source: Ciacci (2024)

Problematic research practices

1. Describing the results as showing an RDiT effect while implementing a regression command that estimates a different parameter.
2. Claiming to use an optimal bandwidth from a specific command but selecting a different value in the implemented analysis.

Regression equation

In monthly data for regions (N=21), Ciacci (2024) estimates

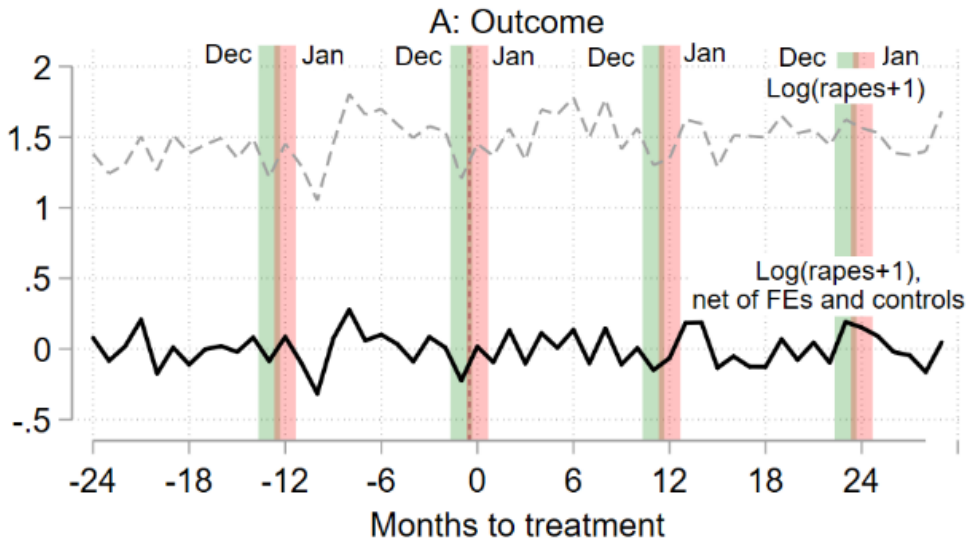
$$log(rape_{rmy}) = \beta_1 \mathbb{I}\{y \geq Jan99\} + \beta_2 F\{y \geq Jan99\} + \gamma officers_{ry} + \alpha_r + \alpha_m + \alpha_y + \epsilon_{rmy}$$

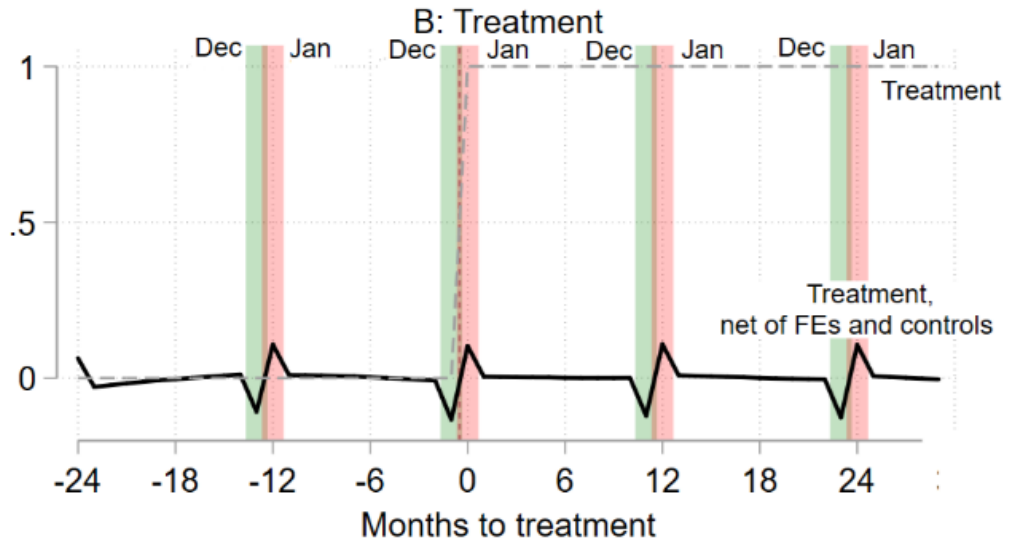
- Treatment = 1 from January 1999 and later, 0 before
- Region, month, and year dummies
- Perfect collinearity precludes estimation
- One longer (“whole”) and one shorter (“restricted”) time sample

How does estimation happen?

- Use the reg command with i.s in front of the categorical variables, but omitting the xi: prefix
- Stata prioritize obtaining point estimates based on variables' order of appearance in the regression equation
- Stata drops variables that appear later in the equation
  - In all samples: the dummy for December (in all samples)
  - In the short sample: the last year dummy (1999 or 2001)
  - In the short with 2nd order polynomials: the dummy for November

Remaining variation in outcome and treatment





The regression in the whole sample estimates a January-December seasonality

Bandwidth choice

- Not using the rdrobust command avoids automatic selection
- Manually run rdbwselect, save optimal bandwidth number, insert in regression command
- Rdbwselect provides optimal bandwidths for “the RD treatment effect estimator”, but also for “the bias of the RD treatment effect estimator”
- Use the latter instead of the former

Table 1. Re-analysis of RDiT results in Ciacci (2024).

	Restricted sample			Whole sample		
	Original estimate (Table 3, column 1, row 4; Figure 5)	Re-analysis of column (1) without coding error	Corrected analysis	Original estimate (Table 3, column 1, row 2)	Re-analysis of columns (4) without coding error	Corrected analysis
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.643**	N/A	0.156	0.548*	N/A	0.066
	(0.280)	N/A	(0.121)	(0.306)	N/A	(0.072)
p-value	*0.022*	*N/A*	*0.188*	*0.075*	*N/A*	*0.260*
Observations	399	N/A	373	1,113	N/A	1,003
Year FE	x	x		x	x	
Region FE	x	x	x	x	x	x
Month FE	x	x	x	x	x	x

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses and p-values are in italics. Columns 1 and 4 show the original reform effects from estimating Equation (1) in the restricted and whole data samples. Columns 2 and 5 show that running those regressions without Ciacci's Stata coding error causes the program to drop the treatment dummy and return an error message. Columns 3 and 6 report estimates from a corrected RDiT analysis. We correct the original estimation by (a) dropping the year fixed effects, (b) removing monthly seasonality by residualization, (c) applying the correct optimal bandwidth, and (d) using robust bias-corrected p-values.

Table 2. Re-analysis of Difference in RDiT results in Ciacci (2024).

	Restricted sample			Whole sample		
	Original estimate (Table C.1, column 1, row 4; and Figure 5)	Re-analysis of column (1) without coding error	Corrected analysis	Original estimate (Table C1, column 1, row 2)	Re-analysis of columns (4) without coding error	Corrected analysis
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.510**	N/A	0.138	0.572**	N/A	0.065
	(0.202)	N/A	(0.159)	(0.275)	N/A	(0.082)
p-value	*0.012*	*N/A*	*0.315*	*0.038*	*N/A*	*0.321*
Observations	483	N/A	373	1,113	N/A	1,003
Year FE	X	x		x	x	

Note: *** p<0.01, ** p<0.05, * p<0.1 Standard errors in parenthesis and p-values in italics. Columns 1 and 3 show the original reform effect from estimating Equation 1 in the "restricted" and "whole" data samples (quotations in original). Columns 2 and 4 show that running those regressions without the Stata coding error makes the program drop the treatment dummy and return an error message instead of a regression output. Columns 3 and 6 report estimates from a re-analysis under the incorrect assumption of more than one treatment cluster. We correct the original estimation by (a) dropping the year fixed effect, (b) correcting for monthly seasonality by residualization, (c) using the optimal bandwidth rather than a different number, and (d) using robust bias-corrected p-values.

Additional evidence in Ciacci (2025)

- Remove year, region, and month fixed effects
- Exclude data after 2005
- Test sensitivity to population weights
- Test sensitivity to number of rapes instead of log(N) as DV
- Controls for season FE and the number of police officers

Treatment effects remain in reasonable tests?

- Population weights → no results remain at 5% level
- N of rapes as DV → no results remain

Table 2 Robustness: no weights

	(1)	(2)	(3)	(4)
RD_Estimate	0.241	0.042	0.558	0.037
Observations	2268	2268	2268	2268
BW above cutoff	7.854	84	7.169	84
BW below cutoff	7.854	24	7.169	24
Effective *N* above cutoff	168	1764	168	1764
Effective *N* below cutoff	147	504	147	504
p-value	0.538	0.792	0.0174	0.699

No population weights, log(rapes) as the outcome, and controls for season FE and # of police officers

Source: Ciacci (2025)

Problematic research practices

1. Describe the results as showing an RDiT effect, but coding a control so that the estimate captures a different parameter
   - Code season dummies so that perfect collinearity is created with the treatment dummy
2. Present the treatment effect as valid, despite deriving it from overfitted control functions.

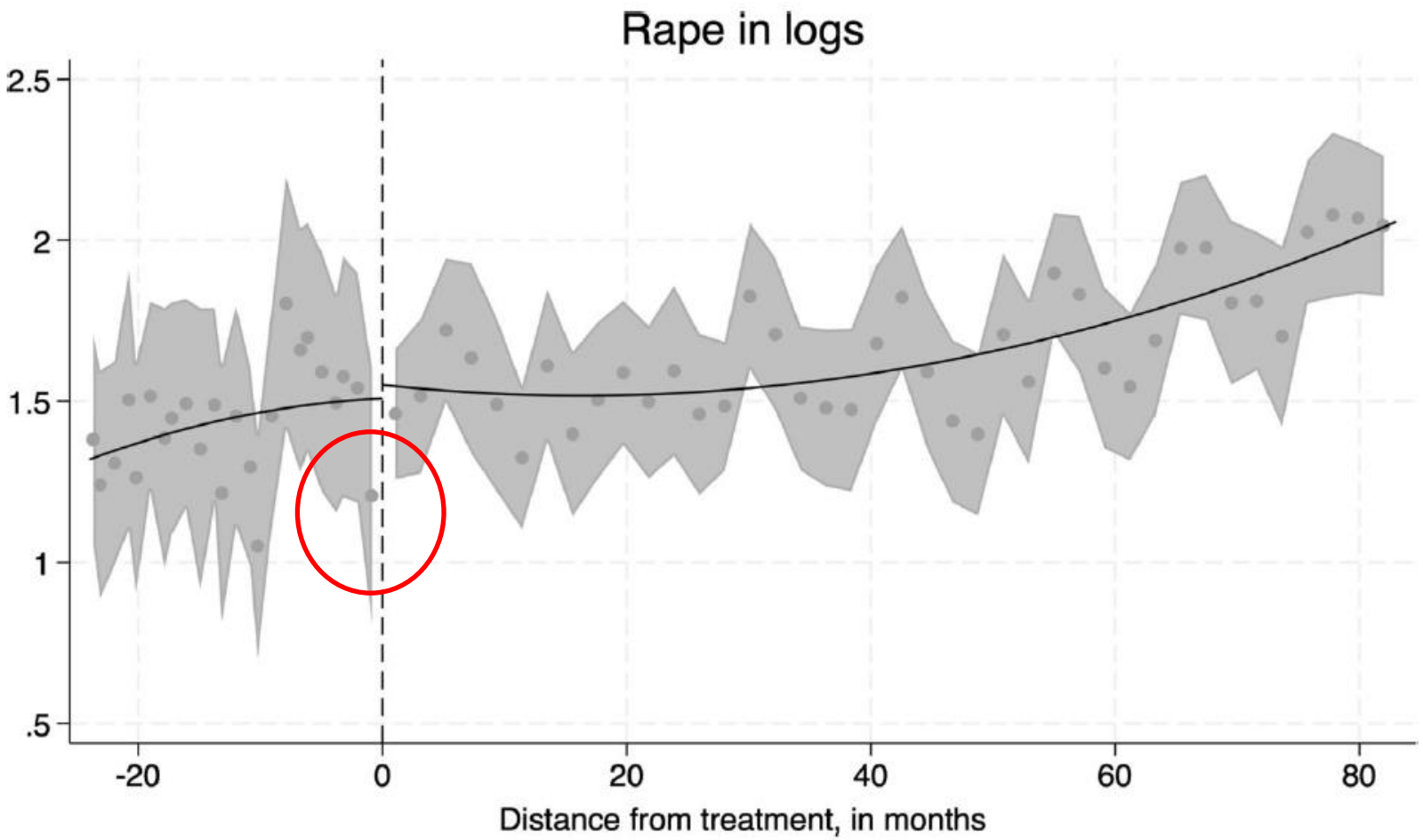
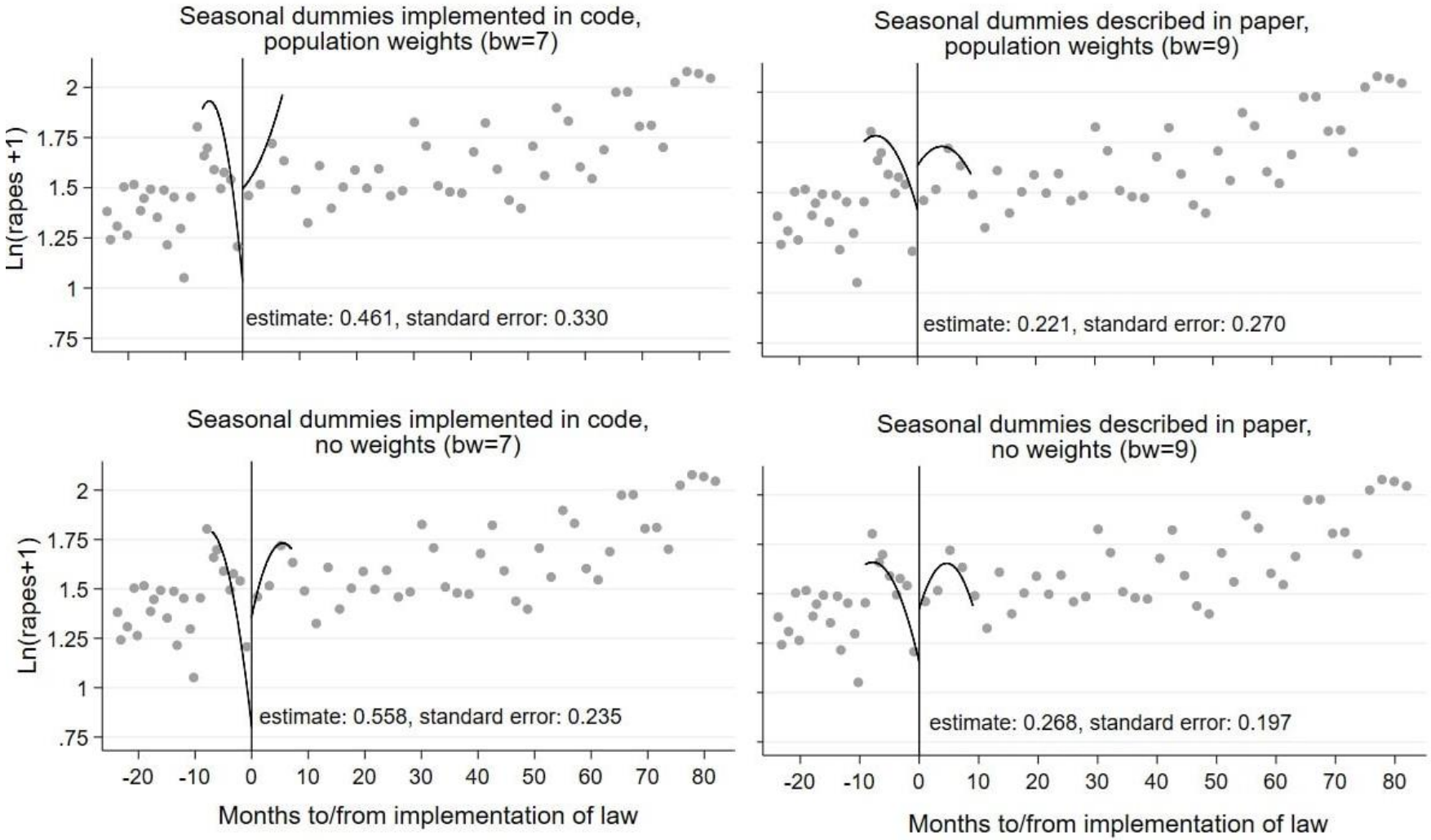
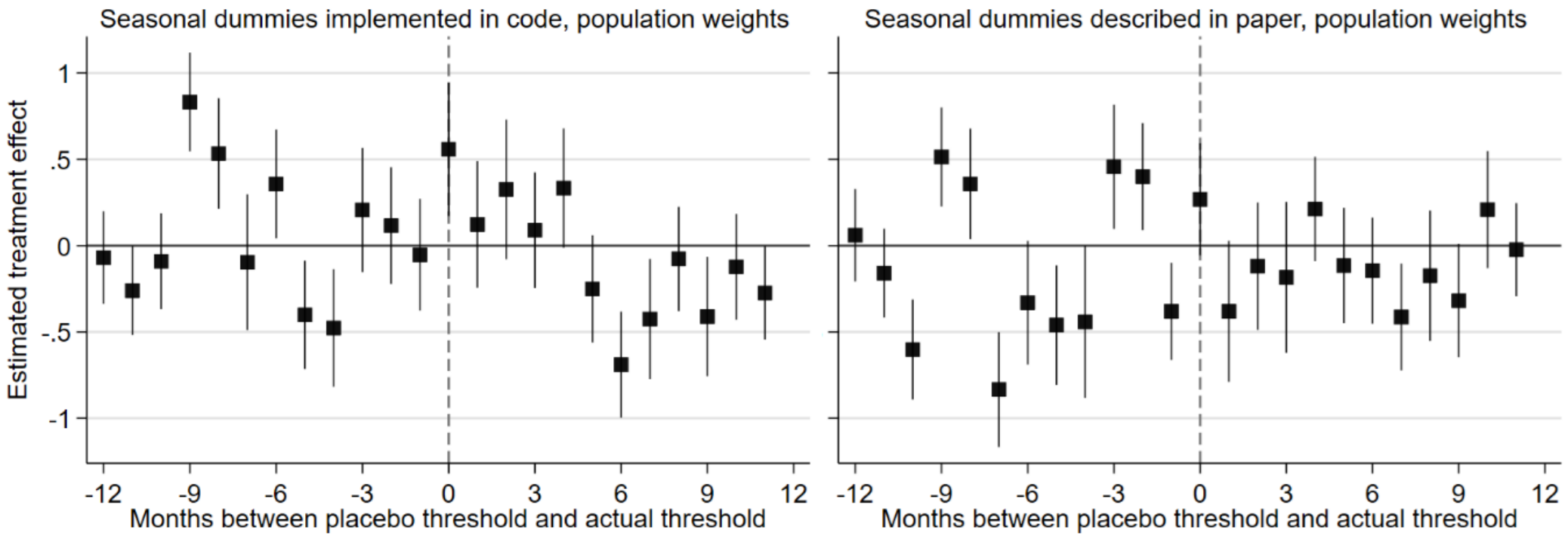


Fig. 1 RDiT plot. Notes: This shows the regression discontinuity plot of Eq. 1. The confidence interval is for the binned data

Source: Ciacci (2025)


Seasonal dummies implemented in code, population weights (bw=7)
Ln(rapes +1)
2
1.75
1.5
1.25
1
.75
estimate: 0.461, standard error: 0.330
Seasonal dummies described in paper, population weights (bw=9)
estimate: 0.221, standard error: 0.270
Seasonal dummies implemented in code, no weights (bw=7)
Ln(rapes+1)
estimate: 0.558, standard error: 0.235
-20
-10
0
10
20
30
40
50
60
70
80
Months to/from implementation of law
Seasonal dummies described in paper, no weights (bw=9)
estimate: 0.268, standard error: 0.197
Months to/from implementation of law



The specification providing the large and significant treatment effect in January 1999 also does so for many other placebo thresholds.

Research designs using fines in Ciacci (2024)

- Fuzzy DiD
- Event study
- IV

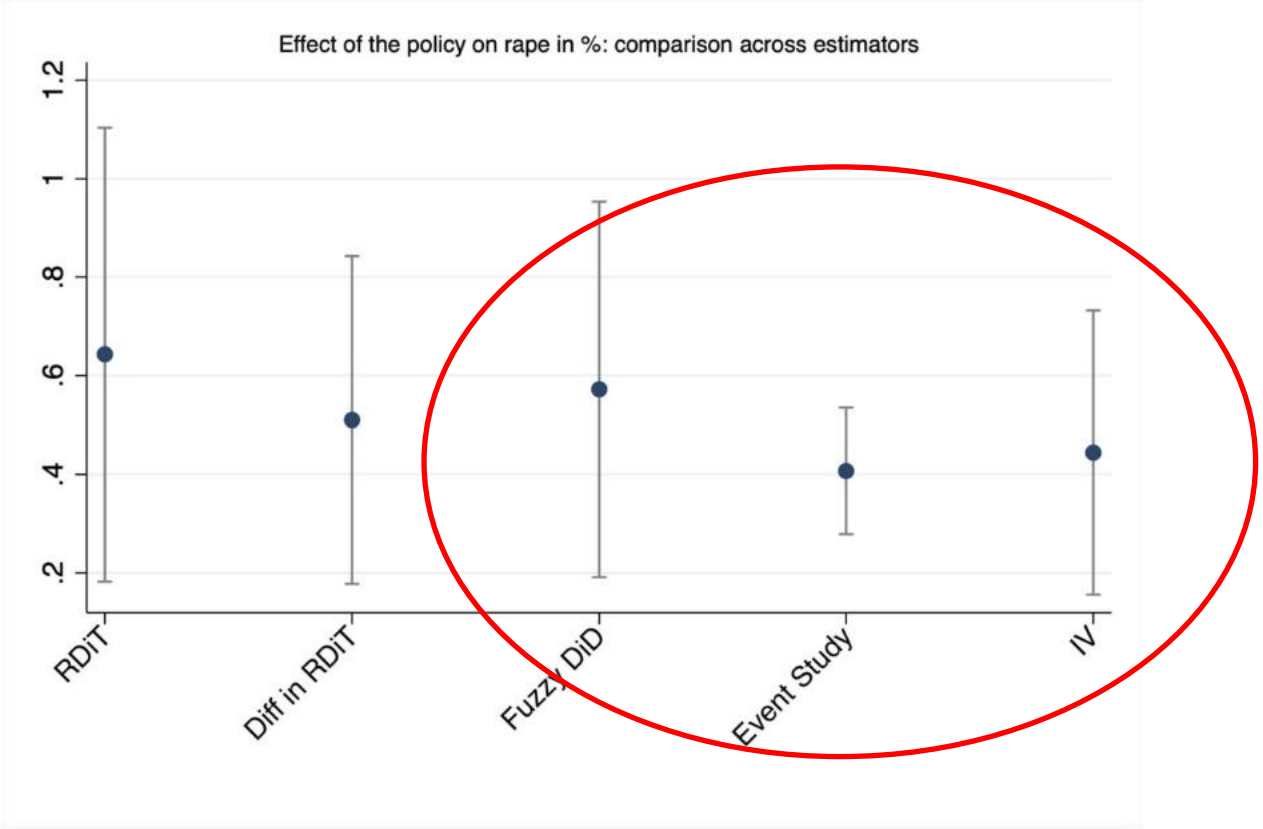


Fig. 5 Comparison of policy effect across techniques. Notes: This figure shows the estimated effect of the policy, and the respective 90% confidence intervals, of the different econometrics techniques used in this paper. Confidence intervals overlap across specifications (i.e., estimates are statically equal)

Problematic research practices

Replace data on fines with data on police reports for purchasing sex

- In 1999, 10 fines and 94 police reports
- Fine # are 5—50% of report # in each year
- Time lag makes police reports an imprecise proxy for fines
  - Problematic as the identification in each method relies on the exact timing of fines

IGNORING THIS PROBLEM, each design has problematic practices similar to the RDiT case

Fuzzy DiD: Problematic research practices

The method in brief: Exploit combination of timing of treatment (when fines are issued) with intensity of treatment (sum of previous fines issued) to estimate the treatment effect.

- Code the treatment and time variables differently than called for by method described in the paper
  - Define the treatment variable based on monthly time variation but use yearly instead of monthly variation for the event time indicators in the code.
- Describe a method in the paper without presenting its results.

Stockholms universitet

Table 3. Re-analysis of Fuzzy DiD results in Ciacci (2024).

	Wald-DID		Wald-DID TC	
	Original estimate, (Table S.5, column 1, row 1)	Estimation with the correct time variable	Estimation with the incorrect time variable	Estimation with the correct time variable
	(1)	(2)	(3)	(4)
Fine events	0.044**	0.017	0.055	0.138
	(0.018)	(0.016)	(0.106)	(0.189)
p-value	*0.014*	*0.310*	*0.893*	*0.467*
Observations	4,536	4,536	4,536	4,536

Note: *** p<0.01, ** p<0.05, * p<0.1 Standard errors in parenthesis and p-values in italics. Results from estimating the Wald-DiD Fuzzy-DiD estimator. Column 1 estimates the same model as in the do-file in the replication package, i.e., using calendar year as the time variable and including region, year-, and month-specific time trends. Column 2 uses the correct year-month variable as a time variable, columns 3 and 4 repeat the two specifications with the Wald-DID TC proposed by de Chaisemartin and D'Haultfoeuille (2018).

Event study: problematic research practices

1. Claiming to use a specific clustering method in the paper but implementing a different clustering method in the analysis
   - Replace clustering at the region–time level with a user-written three-way clustering method
2. Non-standard specification of variables
   - Include all out-of-window observations in the reference category, t=0
   - No reference category for consecutive events.

Table 4. Re-analysis of event study results in Ciacci (2024).

Clustering method	Three-way Original estimate (Table S.10, column 1	Standard one-way				
Clustering level	Region, year, and month	Region	Region-year	Region-month	Region-year-month	No clustering
	(1)	(2)	(3)	(4)	(5)	(6)
Dummy for t = -2	-0.006	-0.006	-0.006	-0.006	-0.006	-0.006
	(0.022)	(0.028)	(0.026)	(0.028)	(0.027)	(0.029)
p-value	*0.779*	*0.825*	*0.809*	*0.822*	*0.816*	*0.829*
Dummy for t = 0	0.031***	0.031	0.031	0.031	0.031	0.031
	(0.006)	(0.023)	(0.024)	(0.025)	(0.023)	(0.026)
p-value	*0.000*	*0.193*	*0.185*	*0.203*	*0.181*	*0.225*
Dummy for t = +1	-0.013	-0.013	-0.013	-0.013	-0.013	-0.013
	(0.048)	(0.043)	(0.039)	(0.040)	(0.038)	(0.037)
p-value	*0.787*	*0.757*	*0.731*	*0.737*	*0.727*	*0.717*
Dummy for t = +2	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.043)	(0.036)	(0.038)	(0.039)	(0.038)	(0.040)
p-value	*0.989*	*0.986*	*0.987*	*0.987*	*0.987*	*0.988*
Observations	4,536	4,536	4,536	4,536	4,536	4,536

Note: *** p<0.01, ** p<0.05, * p<0.1 Standard errors in parenthesis, p-values in italics. Column 1 estimates the same model as in the do-file in the replication package, which uses three-way clustering to calculate the standard

Table 5. Re-analysis of event study results in Ciacci (2024).

	Original estimate (Table S.10, column 1)	Original estimate after adding an out-of-window dummy	Dropping observations outside any event window
	(1)	(2)	(3)
Dummy for t = -2	-0.006	-0.010	-0.017
	(0.022)	(0.041)	(0.041)
p-value	*0.779*	*0.807*	*0.688*
Dummy for t = 0	0.031***	0.027	0.021
	(0.006)	(0.044)	(0.044)
p-value	*0.000*	*0.541*	*0.637*
Dummy for t = +1	-0.013	-0.017	-0.013
	(0.048)	(0.067)	(0.070)
p-value	*0.787*	*0.801*	*0.852*
Dummy for t = +2	-0.001	-0.005	0.004
	(0.043)	(0.065)	(0.069)
p-value	*0.989*	*0.944*	*0.960*
Observations	4,536	4,536	1,773

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parenthesis, p-values in italics. Column 1 uses the specification from the replication package that produces the estimates in Column 1 in Table S.10. Column 2 uses the same specification but includes a dummy for observations that are outside of the event window, while column 3 drops all observations outside of any event window.

IV: Additional problematic research practices

The method in brief: Instrument for fines with a complex variable based on data for (i) region-airport distances and (ii) flight numbers and types.

1. Describe the instrument as capturing a specific data feature but coding the variable to exclude most of this variation.
   - Code distance to the region's closest airport in a way that excludes five of Sweden's six largest airports.
2. State and evaluate the wrong exclusion restriction; true exclusion restriction clearly contradicted by the paper's theory.

What can we learn about whether the Swedish reform increased rape?

Summary of evidence

- Flat time trend strongly suggests “no”
- No declines in other Nordic countries in the relevant time frame.
- Lessons learned about other margins (reviewed in SOU 2010:49)
  - Strong reduction in street prostitution
  - Slowed growth of internet-based prostitution
  - Prevented the establishment of international criminal networks
  - Security for women in prostitution did not decrease due to prostitution “going underground”

What can we learn about questionable research practices and misconduct?

Summary evidence for problematic practices

- Many across the two papers.
- Some increase the size of the treatment effect, other make it more precisely estimated.
- Not a single result “survives” correction.
- The problematic practices clearly produce the evidence of large and significant treatment effects of the Swedish reform on rape.

Definitions: Questionable research practices

Minor infractions or research practices, including avoidable errors, which fall short of the definition of intentional research misconduct. They may arise due to a lack of knowledge or attention to detail, negligence, or deliberate action, and may occur where there is no evident intention to deceive.

Source: UK Concordat to support research integrity; COPE

Definitions: Research misconduct

Research misconduct means fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results

a) Fabrication is making up data or results and recording or reporting them.

b) Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

c) Plagiarism […]

d) Research misconduct does not include honest error or differences of opinion.

Source: US Department of Health and Human Services; COPE

Definitions: Research misconduct

Research misconduct means fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results

a) Fabrication is making up data or results and recording or reporting them.

b) Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.

c) Plagiarism […]

d) Research misconduct does not include honest error or differences of opinion.

Source: US Department of Health and Human Services; COPE

Studies with observational data

- Well-known cases of research misconduct in the social sciences nearly always regard fabrication of data.
- These tend to be lab or field experiments.
- Detection risk deters fabrication when using observational data?
- Falsification is more relevant?

Possible falsification with observational data

Am I inaccurately representing my research record or results if I...

- …say I estimate one quantity but actually estimate another?
- …say I use one variable but actually use another?
- …say I use one clustering method but actually use another?
- …say my variable measures one quantity but code it to exclude nearly all that variation?

And what can journals do about these practices besides posting replication packages?

Stockholms universitet

Thank you for listening!