Revisiting the Impacts of Anti-Discrimination Employment Protections on American Businesses

Jack Fitzgerald*

September 23, 2024

Abstract

Greene & Shenoy (2022) – henceforth GS22 – find that the staggered adoption of U.S. state-level protections against racial discrimination in employment decreased both the profitability and leverage of affected businesses. However, these results arise from two-way fixed effects (TWFE) difference-in-differences models. Such models are now known to return inaccurate estimates of average treatment effects on the treated (ATTs) when treatment assignment is staggered, as some firm-year ATTs can enter the TWFE estimator with negative weight. I find that 21-36% of firm-year ATTs in GS22's sample enter the TWFE estimator with negative weight. I then replicate GS22's results using recently-developed difference-in-differences estimators that return valid ATT estimates under staggered adoption. None of these new ATT estimates are statistically significantly different from zero.

Keywords: Replication, difference-in-differences, two-way fixed effects, staggered adoption, racial discrimination

*Email: j.f.fitzgerald@vu.nl. This project has approval from the Ethical Review Board at Vrije Universiteit Amsterdam's School of Business and Economics. The code necessary to reproduce these findings can be found at https://osf.io/6q4k5/, and the necessary replication data can be found at https://services.informs.org/dataset/download.php?doi=mnsc.2021.4057.

1 Introduction

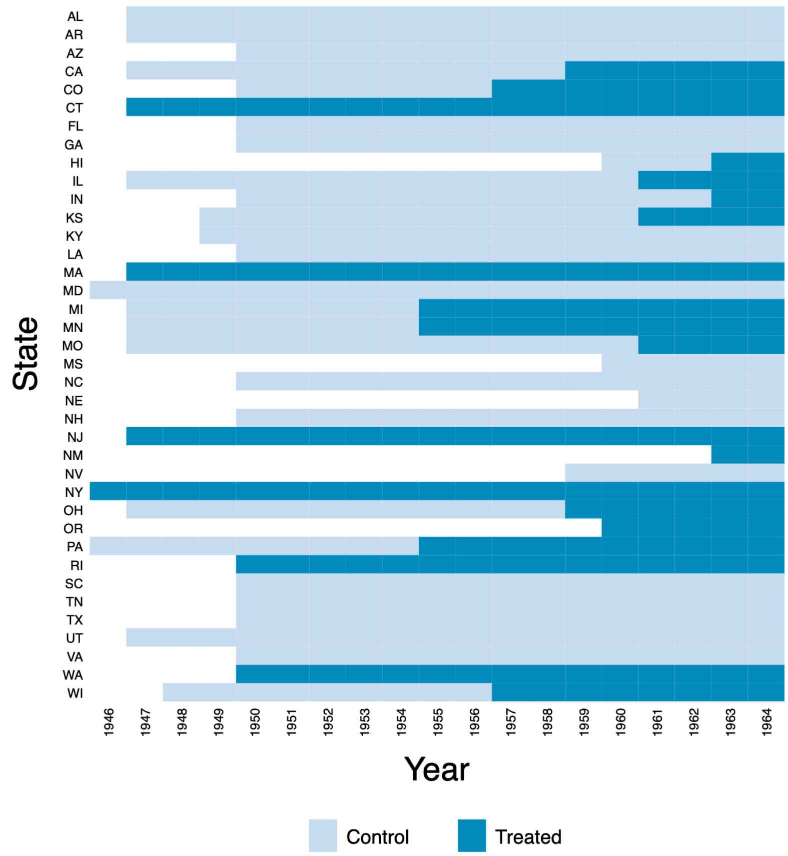
Greene & Shenoy (2022) – henceforth GS22 – investigate the firm-level impacts of state-level laws against racial employment discrimination in the United States. Prior to the passage of the Civil Rights Act of 1964, 22 states passed such anti-discrimination laws at different points in time. Leveraging this staggered adoption, GS22 use difference-in-differences (DID) models to estimate the impacts of these laws on firms' operating profitability, return on assets, and employment growth, as well as several measures of firms' leverage. GS22 find that prior to the passage of the Civil Rights Act, the adoption of anti-discrimination laws reduced treated firms' operating profitability, and firms responded to anti-discrimination laws by reducing the risk profiles of their capital structure, decreasing leverage.

This paper re-examines the robustness of GS22's findings in light of recent developments in the econometrics literature, which highlight problems with DID models that estimate the impacts of staggered treatments. GS22's models are estimated using two-way fixed effects (TWFE) models. A recent literature makes clear that such models can produce highly inaccurate estimates of treatment effects of interest when treatment adoption is staggered and treatment effects are heterogeneous (see de Chaisemartin & D'Haultfœuille 2023; Roth et al. 2023). Under these conditions, the average treatment effects on the treated (ATTs) for some firm-years can enter the computation of the TWFE estimator with negative weight. This can cause the TWFE estimator to misidentify an intervention's ATT so severely that all ATTs in the sample may be positive, but the TWFE estimate is still negative (or vice versa; see de Chaisemartin & D'Haultfœuille 2020).

I show empirically that these issues with the TWFE estimator render GS22's main findings non-robust. I first document that 21-36% of firm-year ATTs enter the baseline TWFE estimator with negative weight. I then replicate GS22's estimates using two DID estimators that produce accurate ATT estimates under staggered adoption. The estimates from these staggered adoption-robust DID estimators tend to be considerably attenuated compared to GS22's TWFE estimates. None of these staggered adoption-robust estimates are statistically significantly different from zero.

2 Staggered Adoption and Difference-in-Differences Weighting

GS22's research setting is characterized by staggered treatment adoption. Figure R-I displays how the timing of states' adoption of anti-discrimination laws shows up in GS22's data, based on GS22's coding and data availability. Over the 19 years in GS22's time horizon, there are eight different



Note: Dark shading indicates that the state has an active employment anti-discrimination law in that year and that at least one firm's profitability in that state is observed in that year. White shading indicates that no firm's profitability in that state is observed in that year. The graph is constructed using the `panelview` command in Stata (Mou, Liu, & Xu 2023).

Figure R-I: Staggered Adoption of State Anti-Discrimination Employment Laws

timepoints at which firms in GS22's data are newly exposed to anti-discrimination laws.[1] State s can be observed adopting an anti-discrimination law in year t either because state s genuinely first adopts an anti-discrimination law in year t, or because the first firm from state s is observed in GS22's data in year t (even if state s first adopts an anti-discrimination law prior to year t). Figure R-I largely replicates GS22's Table 1.[2]

Per Equation (1) in GS22, treatment effects are estimated in two-way fixed effects (TWFE)

[1] This count excludes Massachusetts, who adopts an anti-discrimination law in 1946 and is always treated throughout the duration of GS22's time horizon.

[2] Differences emerge because some states adopt anti-discrimination laws before the first year in which a firm from that state appears in GS22's data.

difference-in-differences (DID) models of the form

$$Y_{i,s,t} = \beta_1 \text{ADLaw}_{s,t} + Z_{i,s,t}\gamma + \alpha_i + v_t + \epsilon_{i,s,t}. \tag{E1}$$

Here $Y_{i,s,t}$ is the dependent variable of interest (which changes across models), $\text{ADLaw}_{s,t}$ indicates whether state s has an active anti-discrimination law in year t, $Z_{i,s,t}$ is a matrix of control variables, and α_i and v_t represent firm and year fixed effects (respectively). β_1 is the coefficient of interest, which GS22 interpret as the effect of anti-discrimination laws.

However, TWFE models do not generally return accurate treatment effect estimates in research settings such as GS22's, which is characterized by staggered adoption and treatment effects that are likely heterogeneous. de Chaisemartin & D'Haultfœuille (2020) show that under a parallel trends assumption,[3] which is also implicitly invoked by GS22, the TWFE estimator β_1 from Equation E1 can be decomposed as

$$\beta_1 = \mathbb{E}\left[\sum_{(i,t):\ \text{ADLaw}_{i,t}=1} W_{i,t}\Delta_{i,t}\right]. \tag{E2}$$

Here $\Delta_{i,t}$ is the average treatment effect on treated firm i in year t, and $W_{i,t}$ is the weight with which $\Delta_{i,t}$ enters the TWFE computation. The weights $W_{i,t}$ must sum to one, but need not be positive, and negative $W_{i,t}$ can emerge in staggered adoption settings. The negative weights can make the TWFE coefficient β_1 an extremely inaccurate treatment effect estimator. de Chaisemartin & D'Haultfœuille (2020) show that with sufficient negative weighting and heterogeneous treatment effects, all average treatment effects $\Delta_{i,t}$ of a policy can be positive, but the TWFE estimator can be negative (and vice versa). Later works confirm and extend this result (see Goodman-Bacon 2021; Roth et al. 2023; de Chaisemartin & D'Haultfœuille 2023).

GS22's staggered setting results in many firm-years' average treatment effects on the treated (ATTs) to enter the TWFE estimator with negative weight. Table R-I shows the extent of negative ATT weighting in GS22's baseline TWFE specifications (i.e., those without any control variables) across the six outcome variables examined in Table 4 and Table 6. The extent of negative weighting is computed under two different assumptions about the data-generating process. Panel A computes ATT weights after simply imposing de Chaisemartin & D'Haultfœuille's (2020) parallel trends

[3] Applied to GS22's setting, the parallel trends assumption invoked by de Chaisemartin & D'Haultfœuille (2020) posits that the evolution in untreated potential outcomes for $Y_{i,s,t}$ between consecutive periods is identical for all firms i and all years $t \geq 1947$; see Assumption 5 in de Chaisemartin & D'Haultfœuille (2020).

	Operating Profitability	Return on Assets	Employment Growth	Book Leverage	Total Leverage	Market Leverage
Panel A: Parallel Trends Assumed						
# Positively-Weighted ATTs	544	583	397	604	611	572
# Negatively-Weighted ATTs	295	332	206	311	304	302
Proportion of ATTs Negatively-Weighted	35.2%	36.3%	34.2%	34%	33.2%	34.6%
Sum of Negative ATT Weights	-0.8575	-0.8111	-0.8701	-0.7763	-0.8881	-0.6525
Panel B: Parallel Trends and Temporally Constant ATTs Assumed						
# Positively-Weighted ATTs	57	55	51	51	54	48
# Negatively-Weighted ATTs	16	16	21	20	19	21
Proportion of ATTs Negatively-Weighted	21.9%	22.5%	29.2%	28.2%	26%	30.4%
Sum of Negative ATT Weights	-0.0057	-0.0013	-0.0019	-0.0006	-0.0012	-0.0005

Note: Results indicate the counts and proportions of ATTs that enter the TWFE estimator's computation with positive/negative weight in the baseline specifications underlying GS22's Tables 4 and 6 (i.e., those regressions without controls), along with the sum of weights for all negatively-weighted ATTs. Columns indicate the outcome variable used. Panels indicate which assumptions are made about the data-generating process. Results arise from the `twowayfeweights` command in Stata (de Chaisemartin & D'Haultfœuille 2020).

Table R-I: Extent of Negative ATT Weighting in GS22's TWFE Regressions

assumption, whereas Panel B computes those weights after additionally imposing the assumption that ATTs do not vary over time. Under just a parallel trends assumption, 33-36% of firm-year ATTs $\Delta_{i,t}$ enter the TWFE estimator with negative weight. Assuming both parallel trends and that ATTs are constant over time, 21-30% of ATTs enter the TWFE estimator with negative weight.

These negative ATT weights can substantially decrease the accuracy of TWFE estimates under sufficient treatment effect heterogeneity, and GS22 themselves posit and test several ways in which anti-discrimination laws may heterogeneously impact firms. For example, GS22 note that some state commissions responsible for securing anti-discrimination law compliance initially lacked enforcement power, and explicitly test for heterogeneity in treatment effects on this dimension. Thus even firms in different states which respond in identical ways to anti-discrimination law enforcement will often exhibit heterogeneous ATTs because they are exposed to different treatment intensities.

There is also reason to believe that variation in treatment effects over time may adversely impact the accuracy of GS22's TWFE estimates. Specifically, the impacts of anti-discrimination laws likely grow over time, as it takes time for states to develop effective strategies for enforcing anti-discrimination laws. Additionally, a key channel by which anti-discrimination laws impact firms is through changing norms around hiring practices, which also takes time. These theoretically intuitive mechanisms are consistent with prior evidence that anti-discrimination laws have stronger impacts in states that instated such laws earlier (e.g., see Collins 2003). Given that Table R-I shows that temporal variation in ATTs yields much of the potential negative ATT weighting on GS22's

TWFE estimates, this temporal heterogeneity in the effects of anti-discrimination laws likely has adverse impacts on the accuracy of the TWFE estimator.

3 Staggered Difference-in-Differences Estimates

GS22's data and model specifications limit the staggered adoption-robust DID estimators that can be used to replicate their results. Figure R-I shows that GS22's panel data is unbalanced, which makes computing many popular staggered-adoption DID estimators computationally infeasible (e.g., Callaway & Sant'Anna 2020; Sun & Abraham 2021; Borusyak, Jaravel, & Spiess 2024). Further, GS22 control for state-specific time trends, which are not feasible to specify in estimators that incorporate covariates via demeaning (e.g., de Chaisemartin & D'Haultfœuille 2024).

Using the replication data available at the online version of GS22, I replicate GS22's results in Tables 4 and 6 using two DID estimators that are robust to staggered adoption and can accommodate GS22's data and specifications. The first estimator that I consider is the DID_{M} estimator developed by de Chaisemartin & D'Haultfœuille (2020). In GS22's context, the DID_{M} estimator computes the average treatment effect of anti-discrimination laws on firms whose exposure to anti-discrimination laws changes at some point over the time horizon, given a series of parallel trends and strong exogeneity assumptions (see de Chaisemartin & D'Haultfœuille 2020). The second estimator that I employ is the FEct estimator developed by Liu, Wang, & Xu (2024). This estimator directly imputes counterfactual values of outcome $Y_{i,s,t}$ for treated firms using a weighted combination of $Y_{i,s,t}$ for untreated firms, using weighting constraints that eliminate the negative weighting induced by TWFE models. The FEct estimator identifies the average treatment effect of anti-discrimination laws on treated firms under the assumptions that error terms are strictly exogenous from anti-discrimination laws, that error terms can be estimated using linear fixed effects, and that some regularity conditions hold (see Liu, Wang, & Xu 2024). For the DID_{M} estimator, I use the `did_multiplegt_old` command in Stata, and incorporate all of GS22's controls except for state-level time trends in the `controls()` option, which adjusts covariates via demeaning. I adjust for state-level time trends by passing state indicators through the `trends_lin()` option. For the FEct estimator, I use the `fect` command in Stata, and incorporate all controls, including linear state-level time trends, in the command's `cov()` option.

Table R-II displays my reproductions of GS22's main outcomes of interest – operating profitability and book leverage – using both their original TWFE models, the DID_{M} estimator, and the

	Operating Profitability	Operating Profitability	Operating Profitability	Book Leverage	Book Leverage	Book Leverage
Original TWFE Estimate	-0.008	-0.007	-0.011	-0.013	-0.015	-0.014
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)
	[-2.174]	[-2.191]	[-3.214]	[-2.539]	[-3.146]	[-2.86]
DID_M Estimate	-0.002	-0.004	-0.003	0.001	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
	[-0.7]	[-1.43]	[-1.071]	[0.389]	[0.345]	[0.278]
FEct Estimate	-0.008	-0.007	0.028	-0.007	-0.001	0.061
	(0.005)	(0.005)	(0.161)	(0.008)	(0.008)	(0.156)
	[-1.636]	[-1.522]	[0.176]	[-0.891]	[-0.118]	[0.389]
GS22 Controls		Y	Y		Y	Y
State-Level Time Trends			Y			Y

Note: Estimated ATTs of anti-discrimination law introduction on firms' total leverage are displayed along with standard errors in parentheses and t-statistics in brackets. Original TWFE estimates are reproduced from GS22's code. DID_M estimates are computed using `did_multiplegt_old` from de Chaisemartin & D'Haultfœuille (2020), whereas FEct estimates are computed using `fect` from Liu, Wang, & Xu (2024). Standard errors are obtained from 1000 bootstrap replications. GS22's controls for both outcome variables include logarithmic total assets, fixed assets, a dividend payer dummy, and state income growth; book leverage specifications additionally control for return on assets.

Table R-II: Main DID Estimates

FEct estimator. Though the point estimates of GS22's TWFE estimates perfectly match those in GS22, the standard errors are wildly different. This is because GS22 do not in fact report standard errors in parentheses for Tables 4 and 6, instead reporting t-statistics. I thus report t-statistics in brackets beneath standard errors in all tables. The t-statistics in my reproductions nearly exactly replicate those in GS22, differing only down to rounding error at three decimal places.

The DID_M estimates are all considerably closer to zero than the original TWFE estimates. Compared to the TWFE estimates, the DID_M estimates on operating profitability all attenuate by at least 40%, and none are statistically significantly different from zero. The DID_M estimates for book leverage are all positive, which would imply that being exposed to anti-discrimination laws actually increases leverage if these estimates are taken at face value, though none of these estimates are statistically significantly different from zero. The DID_M estimates on book leverage are at least 90% smaller than the TWFE estimates on book leverage.

None of the FEct estimates are statistically significantly different from zero. The FEct estimates on operating profitability actually do not attenuate compared to the respective TWFE estimates; the point estimates without state-level time trends are the same down to three decimal points. Controlling for state-level time trends causes the operating profitability FEct estimate to more than double in magnitude compared to the respective TWFE estimate and flip signs, with a standard error that explodes in magnitude. Looking exclusively at the operating profitability results might lead one to conclude that the FEct estimator produces the same results as the original TWFE estimates, that the FEct estimator's standard error is simply less efficient than that for the TWFE estimator, and that the FEct estimator appears to handle state-level time trends idiosyncratically. Though the latter two conclusions appear to hold true throughout my reproductions (see also Appendix Tables A-I and A-II), the first observation is a sheer coincidence. The FEct estimates on book leverage without state-level time trends attenuate considerably compared to the respective TWFE estimates, and the FEct estimate with state-level time trends again flips signs compared to the respective TWFE estimate.

Appendix Tables A-I and A-II repeat this exercise for the other outcome variables in Tables 4 and 6 (respectively). The reproductions in Appendix Table A-I do not change statistical significance conclusions for return on assets or for employment growth. As for the TWFE estimates for return on assets and employment growth in Table 4, no staggered adoption estimate in Appendix Table A-I is statistically significantly different from zero. Appendix Table A-II shows that the differences between TWFE estimates and staggered adoption DID estimates for total leverage and market leverage broadly resemble those same estimator differences for book leverage. The only major difference is for the FEct model on market leverage which controls for state-level time trends, whose magnitude explodes similar to other FEct estimates in this paper, but becomes considerably more *negative*, rather than exhibiting a sign flip compared to the respective TWFE estimate. However, this estimate is particularly noisy, and like all other staggered adoption DID estimates in this paper, it is not statistically significantly different from zero.

4 Conclusion

I revisit GS22's evaluation of how American protections against racial employment discrimination impacted firm outcomes. I document that the staggered rollout of anti-discrimination laws causes TWFE estimates of the impact of those laws to suffer from issues related to negative ATT weighting. I then show that GS22's estimates of the impacts of anti-discrimination laws are not robust to this issue. Replicating GS22's main estimates with multiple staggered DID estimators that are not impacted by negative ATT weighting yields no estimate that is statistically significantly different from zero.

The robustness issues that I point out in this paper are not issues that GS22 should be expected to have known at the time this study was first developed. GS22 first submitted this manuscript in July 2019, and the published literature that exposed the issues with the TWFE estimator in staggered DID settings largely did not emerge until after this point (see Roth et al. 2023; de Chaisemartin & D'Haultfœuille 2023). However, these issues are known now, and as this exploding DID literature progresses, more research findings arising from staggered DID designs will need to be re-evaluated for robustness. GS22's findings on the business impacts of anti-discrimination laws are among these many findings that are not robust when ATTs are computed using appropriate staggered DID estimators.

References

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2024). "Revisiting event-study designs: Robust and efficient estimation". *Review of Economic Studies*. DOI: 10.1093/restud/rdae007.

Callaway, Brantly and Pedro H.C. Sant'Anna (2021). "Difference-in-differences with multiple time periods". *Journal of Econometrics* 225.2, pp. 200–230. DOI: 10.1016/j.jeconom.2020.12.001.

Chaisemartin, Clément de and Xavier D'Haultfœuille (2020). "Two-way fixed effects estimators with heterogeneous treatment effects". *American Economic Review* 110.9, pp. 2964–2996. DOI: 10.1257/aer.20181169.

— (2023). "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey". *The Econometrics Journal* 26.3, pp. C1–C30. DOI: 10.1093/ectj/utac017.

— (2024). "Difference-in-differences estimators of intertemporal treatment effects". *Review of Economics and Statistics*. DOI: 10.1162/rest_a_01414.

Collins, William J. (2003). "The labor market impact of state-level anti-discrimination laws, 1940–1960". *ILR Review* 56.2, pp. 244–272. DOI: 10.1177/001979390305600203.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". *Journal of Econometrics* 225.2, pp. 254–277. DOI: 10.1016/j.jeconom.2021.03.014.

Greene, Daniel and Jaideep Shenoy (2022). "How do anti-discrimination laws affect firm performance and financial policies? Evidence from the post-World War II period". *Management Science* 68.5, pp. 3813–3833. DOI: 10.1287/mnsc.2021.4057.

Mou, Hongyu, Licheng Liu, and Yiqing Xu (2023). "Panel data visualization in R (panelView) and Stata (panelview)". *Journal of Statistical Software* 107.7. DOI: 10.18637/jss.v107.i07.

Roth, Jonathan et al. (2023). “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature”. *Journal of Econometrics* 235.2, pp. 2218–2244. DOI: 10.1016/j.jeconom.2023.03.008.

Sun, Liyang and Sarah Abraham (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. *Journal of Econometrics* 225.2, pp. 175–199. DOI: 10.1016/j.jeconom.2020.09.006.

Appendix

	Return on Assets	Return on Assets	Return on Assets	Employment Growth	Employment Growth	Employment Growth
Original TWFE Estimate	-0.001	-0.001	-0.003	-0.011	-0.012	-0.008
	(0.002)	(0.002)	(0.002)	(0.007)	(0.007)	(0.008)
	[-0.75]	[-0.872]	[-1.78]	[-1.584]	[-1.756]	[-0.987]
DID_M Estimate	0.001	0.001	0.001	0.002	-0.002	-0.004
	(0.002)	(0.002)	(0.002)	(0.016)	(0.017)	(0.018)
	[0.724]	[0.344]	[0.534]	[0.121]	[-0.138]	[-0.199]
FEct Estimate	-0.001	0	0.016	-0.003	-0.003	0.187
	(0.002)	(0.002)	(0.076)	(0.011)	(0.011)	(0.292)
	[-0.222]	[-0.202]	[0.21]	[-0.313]	[-0.233]	[0.639]
GS22 Controls		Y	Y		Y	Y
State-Level Time Trends			Y			Y

Note: Estimated ATTs of anti-discrimination law introduction on firms' return on assets are displayed along with standard errors in parentheses and t-statistics in brackets. Original TWFE estimates are reproduced from GS22's code. DID_M estimates are computed using `did_multiplegt_old` from de Chaisemartin & D'Haultfœuille (2020), whereas FEct estimates are computed using `fect` from Liu, Wang, & Xu (2024). Standard errors are obtained from 1000 bootstrap replications. GS22's controls for all specifications include logarithmic total assets, fixed assets, a dividend payer dummy, and state income growth.

Table A-I: DID Estimates for Return on Assets and Employment Growth

	Total Leverage	Total Leverage	Total Leverage	Market Leverage	Market Leverage	Market Leverage
Original TWFE Estimate	-0.019	-0.021	-0.016	-0.02	-0.02	-0.02
	(0.006)	(0.006)	(0.005)	(0.009)	(0.008)	(0.008)
	[-3.203]	[-3.618]	[-3.009]	[-2.115]	[-2.516]	[-2.592]
DID_{M} Estimate	0.004	0.001	0.002	0.001	0.001	0.001
	(0.004)	(0.003)	(0.004)	(0.006)	(0.005)	(0.006)
	[0.967]	[0.193]	[0.572]	[0.179]	[0.278]	[0.149]
FEct Estimate	-0.022	-0.017	0.097	-0.012	-0.004	-0.133
	(0.009)	(0.009)	(0.161)	(0.015)	(0.011)	(0.376)
	[-2.379]	[-1.932]	[0.6]	[-0.836]	[-0.338]	[-0.355]
GS22 Controls		Y	Y		Y	Y
State-Level Time Trends			Y			Y

Note: Estimated ATTs of anti-discrimination law introduction on firms' total leverage are displayed along with standard errors in parentheses and t-statistics in brackets. Original TWFE estimates are reproduced from GS22's code. DID_{M} estimates are computed using `did_multiplegt_old` from de Chaisemartin & D'Haultfœuille (2020), whereas FEct estimates are computed using `fect` from Liu, Wang, & Xu (2024). Standard errors are obtained from 1000 bootstrap replications. GS22's controls for this specification include logarithmic total assets, return on assets, fixed assets, a dividend payer dummy, and state income growth.

Table A-II: DID Estimates for Leverage