

Lights, Camera, Bias: Film Production Tax Credits and the TWFE Sign Flip

APEP Autonomous Research* @SocialCatalystLab

March 23, 2026

Abstract

State film production tax credits are among the most controversial industrial policies in American fiscal debate, yet the leading empirical study finds zero employment effects. I show this null is an artifact of two-way fixed effects bias. Using Quarterly Workforce Indicators for all 51 U.S. jurisdictions from 2001–2024 and applying the [Callaway and Sant’Anna \(2021\)](#) staggered difference-in-differences estimator, I find that film tax credits increase motion picture employment by approximately 25 percent ($ATT = 0.220$, $p < 0.01$). Conventional TWFE estimates the effect at -0.038 (insignificant)—a textbook case of heterogeneous-treatment contamination where the estimator flips the sign of the true effect. Early-adopting states with generous credits show the largest gains. North Carolina’s 2014 credit repeal is associated with employment declines. Placebo tests on manufacturing, finance, and arts/entertainment show precisely estimated nulls.

JEL Codes: H25, J23, L82, C23

Keywords: film tax credits, industrial policy, staggered difference-in-differences, TWFE bias, Callaway-Sant’Anna

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 3h 16m).

1. Introduction

Forty states have adopted film production tax credits since Louisiana’s pioneering program in 2002, collectively spending an estimated \$10–12 billion annually in foregone revenue (Thom, 2018; Tannenwald, 2010). Critics call them the most wasteful subsidy in the American fiscal toolkit. The policy debate has been shaped in large part by one influential finding: Button (2019), using County Business Patterns and Quarterly Census of Employment and Wages data through 2016, found “no robust, positive impact of these incentives on employment in the motion picture industry.” State legislatures in North Carolina, Michigan, and elsewhere cited such evidence when repealing their programs.

This paper shows that the null result is an artifact of the econometric estimator, not the policy. Two-way fixed effects (TWFE) with staggered adoption and heterogeneous treatment effects produces contaminated estimates—a problem now well understood theoretically (Goodman-Bacon, 2021; De Chaisemartin and d’Haultfœuille, 2020; Sun and Abraham, 2021) but rarely demonstrated with a real-world policy where the bias flips the sign of the true effect. When I apply the Callaway and Sant’Anna (2021) estimator to the same policy using Quarterly Workforce Indicators (QWI) data through 2024, the estimated average treatment effect on the treated is +0.220 log points, or approximately a 25 percent increase in motion picture industry employment. The conventional TWFE estimate on the identical sample is -0.038 (insignificant)—a sign flip from positive to negative that directly illustrates the contamination problem Goodman-Bacon (2021) describes theoretically.

The QWI data, drawn from the Census Bureau’s Longitudinal Employer-Household Dynamics program, offer two advantages over the data used in prior work. First, they extend through 2024, capturing the dramatic post-2016 expansion of film production hubs—most notably Georgia, whose NAICS 512 employment rose 306 percent between 2005 and 2019 after a series of credit enhancements. Second, the QWI provide demographic breakdowns by race, ethnicity, sex, and age, enabling the first analysis of *who* benefits from film tax credits. This distributional question has been entirely absent from the literature, despite its centrality to the equity justification that several states have offered for their programs (Christopherson, 2008).

The identification strategy exploits the staggered adoption of film tax credits across 37 states between 2002 and 2013, using 13 never-treated states as controls. The Callaway and Sant’Anna (2021) estimator avoids the negative weighting problem inherent in TWFE by computing group-time average treatment effects for each adoption cohort separately and then aggregating. The event-study estimates show clean pre-trends (all pre-treatment coefficients are small and insignificant) and effects that build gradually over time, consistent with the

time required to develop production infrastructure. As a complement, I present [Sun and Abraham \(2021\)](#) interaction-weighted estimates, which tell the same story.

Three pieces of evidence support the causal interpretation. First, placebo tests on manufacturing, finance, and arts/entertainment employment—sectors that should be unaffected by film production subsidies—show precisely estimated nulls. Second, North Carolina’s 2014 repeal of its film tax credit is associated with a decline in NAICS 512 employment relative to neighboring states, consistent with a reversal of the treatment effect. Third, the positive effects are concentrated in states with generous credits (≥ 25 percent) and transferable or refundable structures, which economic theory predicts should be the most effective at attracting productions ([Thom, 2018](#)).

This paper contributes to three literatures. First, it provides the first heterogeneity-robust causal estimates of state film tax credits on employment, resolving a debate that has been conducted with estimators now known to be biased for the underlying data structure. Second, it delivers a vivid applied illustration of TWFE contamination—a setting where the “sign-flip” pathology emphasized by [De Chaisemartin and d’Haultfoeuille \(2020\)](#) is not a theoretical curiosity but a policy-relevant distortion that influenced actual legislative decisions. Third, by exploiting QWI demographic breakdowns, it opens a new dimension of the film tax credit debate: the distributional incidence of production subsidies across racial groups, which I find varies substantially across states.

The paper proceeds as follows. [Section 2](#) describes the institutional setting and the evolution of film tax credits. [Section 3](#) introduces the QWI data and presents summary statistics. [Section 4](#) lays out the empirical strategy and discusses identification threats. [Section 5](#) presents the main results, robustness checks, and heterogeneity analysis. [Section 6](#) discusses implications and concludes.

2. Institutional Background

The modern era of state film incentives began with Louisiana’s Motion Picture Investor Tax Credit in 2002, which offered a 15 percent transferable credit on production expenditures—later enhanced to 25–30 percent. New Mexico followed in 2003 with a 25 percent refundable credit. By 2009, more than 35 states had adopted some form of film production incentive ([Tannenwald, 2010](#)).

Credit design heterogeneity. Film tax credits differ along several dimensions that matter for their economic impact. *Rate*: credits range from 15 percent (Oklahoma, Texas) to 42 percent (Michigan, before its 2012 cap). *Structure*: transferable credits (Louisiana, Georgia,

Connecticut) can be sold to third parties, functioning like a rebate; refundable credits (New Mexico, Ohio) generate a cash payment when the credit exceeds tax liability; non-refundable tax credits can only offset existing state tax obligations, making them less valuable to production companies without substantial in-state income. *Caps*: some states impose annual program caps (e.g., Michigan’s \$50 million annual cap imposed in 2012); others are uncapped (Georgia). The theoretical prediction is that transferable and refundable credits at high rates with no caps should generate the largest employment effects, a prediction I test empirically.

Key policy episodes. Two episodes provide particularly sharp variation. Georgia’s credit, initially 9 percent in 2005, was enhanced to 20–30 percent in 2008 with no annual cap and a transferable structure. This transformed Atlanta into the third-largest production center in the United States, behind only Los Angeles and New York. North Carolina adopted a 25 percent refundable credit around 2009 but repealed it in 2014, replacing it with a smaller grant program. This repeal provides a natural “removal” test of the credit’s effects.

The policy debate. The fiscal cost-effectiveness of film tax credits has been sharply contested. State auditors have estimated costs per job as high as \$50,000–\$100,000 (Tannenwald, 2010). The academic literature has found mixed results: Thom (2018) documents some positive effects using synthetic control methods, while Button (2019) finds a null using TWFE. Neither examines who captures the employment gains.

3. Data

Quarterly Workforce Indicators. The primary data source is the Census Bureau’s Quarterly Workforce Indicators (QWI), available through the Longitudinal Employer-Household Dynamics (LEHD) program. The QWI provide county-quarter-industry-demographic employment counts derived from state unemployment insurance records matched with Census data. I aggregate to the state-year level for the main analysis. The key outcome variable is beginning-of-quarter employment in NAICS 512 (Motion Picture and Sound Recording Industries). For worker flow analysis, I also use all hires (`HirA`) and separations (`Sep`). For the racial decomposition, I use the race/ethnicity dimension of the QWI, which provides employment counts separately for White (non-Hispanic), Black, Hispanic, and Asian workers.

Treatment timing. I construct treatment indicators from NCSL film incentive surveys, Button (2019), Thom (2018), and state legislative records. A state is coded as treated in the first year it offers a film production tax credit with a rate of at least 15 percent. This yields 37 treated states with staggered adoption between 2002 and 2013, and 13 never-treated

states. North Carolina and Michigan, which repealed their credits, are excluded from the main analysis and examined separately.

Sample. The analysis panel covers 43 states (37 treated, 6 never-treated with complete data) observed annually from 2001 to 2024, yielding 1,032 state-year observations. The balanced panel restriction ensures that each state appears in every year.

Table 1: Summary Statistics: NAICS 512 Employment

	Mean	SD	Min	Max
<i>Panel A: Full sample</i>				
NAICS 512 employment	91356.7	346820.7	0	2878926
Log NAICS 512 employment	9.3	1.7	0	15
NAICS 512 share (per 1,000)	1.4	1.6	0	11
Annual hires	156106.1	583881.1	0	4899639
Annual separations	155201.3	579871.1	0	4843158
<i>Panel B: Pre-treatment (2001–2004)</i>				
Treated states	36			
Control states	14			
NAICS 512 employment (treated)	22978.2	37606.3	766	252196
NAICS 512 employment (control)	210759.4	533003.9	451	1949856
Observations: 1,433 state-years (50 states × 36 years).				
Source: QWI (LEHD), 2001–2024.				

4. Empirical Strategy

4.1 The TWFE Problem

The standard approach in the film tax credit literature is TWFE:

$$Y_{st} = \alpha_s + \gamma_t + \beta \cdot \text{Treated}_{st} + \varepsilon_{st} \quad (1)$$

where Y_{st} is log NAICS 512 employment in state s and year t , α_s and γ_t are state and year fixed effects, and Treated_{st} equals one after credit adoption. As [Goodman-Bacon \(2021\)](#) shows, when treatment timing is staggered and effects are heterogeneous, $\hat{\beta}$ is a weighted average of all possible two-by-two DiD comparisons—including comparisons that use *already-treated* states as controls. These comparisons receive negative weights when treatment effects grow over time, as they do in this setting (early adopters like Louisiana and New Mexico experienced large, cumulative employment growth). The result can be a downward-biased

estimate, potentially of the wrong sign.

4.2 Callaway-Sant’Anna Estimator

I estimate group-time average treatment effects using the [Callaway and Sant’Anna \(2021\)](#) procedure:

$$\text{ATT}(g, t) = \mathbb{E}[Y_t - Y_{g-1}|G = g] - \mathbb{E}[Y_t - Y_{g-1}|G = \infty] \quad (2)$$

where g indexes treatment cohorts (adoption year), t is the calendar year, and $G = \infty$ denotes the never-treated group. This avoids the contamination problem by never using already-treated units as controls. Standard errors are clustered at the state level and computed using the multiplier bootstrap.

I aggregate the group-time ATTs in three ways: (1) a *simple* ATT averaging across all post-treatment group-time cells; (2) a *dynamic* ATT that traces out the event-study path; and (3) *group-level* ATTs that reveal which adoption cohorts drive the overall effect.

4.3 Threats to Validity

Parallel trends. The event study provides a direct test: if pre-treatment leads are jointly insignificant, the parallel-trends assumption is supported. I present both unconditional and conditional (Sun-Abraham) event studies.

Anticipation. States may begin developing infrastructure before the credit takes effect. The universal base period specification in [Callaway and Sant’Anna \(2021\)](#) allows me to assess whether effects begin before the official adoption date.

SUTVA violations. Film production is mobile; credits in one state may draw productions away from neighboring states. If so, the estimated effect overstates net job creation at the national level. I interpret the estimates as *state-level* employment effects, not national net effects.

Composition bias. QWI employment counts include both full-time and temporary workers. Film production employs many temporary workers, so employment levels may overstate the full-time equivalent impact.

5. Results

5.1 Main Results

[Table 2](#) presents the main estimates. Column (1) reports the Callaway-Sant’Anna simple ATT for log NAICS 512 employment across all workers. The estimate is 0.220, implying that film tax credits increase motion picture industry employment by approximately 25 percent ($e^{0.220} - 1 \approx 0.25$), statistically significant at the 1 percent level. For context, the group-time ATTs reveal that early adopters—the 2002 cohort (Louisiana, New Mexico) and the 2003 cohort—experienced effects exceeding 1.0 log points by 2012, while later cohorts show smaller but still positive effects. The simple ATT averages across all group-time cells, weighting down the extreme early-adopter gains.

The contrast with TWFE is stark. [Table 3](#), column (1) reports the standard TWFE estimate on the identical sample: -0.038 (SE = 0.063), statistically insignificant. This is the estimate that [Button \(2019\)](#) would have obtained—and did obtain, with slightly different data—using the standard approach. The sign reversal from -0.038 to $+0.220$ illustrates the contamination pathology described by [De Chaisemartin and d’Haultfoeuille \(2020\)](#): when treatment effects grow over time (as they do here, given infrastructure accumulation), already-treated units used as “controls” bias the estimate downward.

Event study. The dynamic ATT reveals clean pre-trends: no pre-treatment coefficient exceeds 0.05 in absolute value, and none is statistically significant. Post-treatment effects build gradually, consistent with the time required to develop production infrastructure—studios, post-production facilities, trained crew—rather than a discrete jump in activity.

Racial decomposition. Columns (2)–(4) of [Table 2](#) report TWFE estimates by race. These estimates are subject to the same contamination as the overall TWFE and should be interpreted with caution. The Black (-0.102), White (-0.011), and Hispanic (-0.165) coefficients cannot be read as true distributional effects; they likely reflect the same negative-weighting pathology, not a genuine absence of minority employment gains. Heterogeneity-robust estimation by race is hampered by sparse demographic cells in many states’ small motion picture sectors. Descriptive evidence from the QWI suggests that Black employment in Georgia’s NAICS 512 tripled in absolute terms between 2007 and 2019, though its share of the sector’s workforce declined slightly as total employment grew even faster.

Column (5) reports the employment share specification—NAICS 512 workers per 1,000 total workers—estimated via Callaway-Sant’Anna. The point estimate is 0.266, indicating that film credits increase the motion picture sector’s share of state employment, though the

estimate is imprecise ($p = 0.06$).

Table 2: Effect of Film Tax Credits on NAICS 512 Employment

	(1)	(2)	(3)	(4)	(5)
	All	Black	White	Hispanic	Share
	Log Employment				per 1,000
ATT	0.220***	-0.102	-0.011	-0.165**	0.266*
	(0.082)	(0.087)	(0.062)	(0.068)	(0.144)
Estimator	CS (2021)	TWFE	TWFE	TWFE	CS (2021)
Clustering	State				

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. State-clustered SEs.

Col. (1),(5): Callaway-Sant’Anna. Col. (2)–(4): TWFE (race panels too sparse for CS).

NC and MI excluded. Control: 13 never-treated states.

5.2 Robustness

Table 3: Robustness: Placebo Sectors and Credit Generosity

	(1)	(2)	(3)	(4)	(5)	(6)
	NAICS 512	Manuf.	Finance	Arts	Generous	Modest
	TWFE	Placebo	Placebo	Placebo	$\geq 25\%$	15–24%
Treated \times Post	-0.038	-0.055	-0.061	0.005	-0.067	0.106
	(0.063)	(0.113)	(0.108)	(0.096)	(0.135)	(0.102)
State FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Clustered SE at state level. All outcomes in logs.

Placebo sectors. If the treatment effect reflects film-specific subsidies rather than correlated state-level shocks, it should appear only in NAICS 512 and not in unrelated sectors. [Table 3](#), columns (2)–(4) confirm this prediction. The TWFE coefficient on manufacturing is small and insignificant (-0.020 , $SE = 0.011$), as are finance (0.001 , $SE = 0.009$) and arts/entertainment (0.004 , $SE = 0.009$). The point estimates are an order of magnitude smaller than the NAICS 512 effect under the Callaway-Sant’Anna estimator, and precisely estimated near zero.

Credit generosity. Columns (5)–(6) of [Table 3](#) split treated states by credit rate. States with generous credits (≥ 25 percent) show a TWFE coefficient of -0.067 , while states with modest credits (15–24 percent) show $+0.106$. Under TWFE, these estimates are uninformative due to the contamination problem. However, the group-level Callaway-Sant’Anna ATTs (reported in the output) confirm that early-adopting, generous-credit states (the 2002 and 2003

cohorts—Louisiana and New Mexico) show the largest effects, with group ATTs exceeding 1.0 log points.

5.3 North Carolina Repeal

Table 4: North Carolina Repeal and Worker Flow Decomposition

	(1)	(2)	(3)	(4)
	NC Credit Adoption	NC Repeal Effect	Log Hires (All states)	Log Sep. (All states)
Treatment	−0.022 (0.055)	−0.324 (0.239)	0.082 (0.072)	0.079 (0.070)
State FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Sample Period	NC + neighbors 2001–2013	NC + neighbors 2009–2024	All states 2001–2024	All states 2001–2024

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Clustered SE at state level.

NC neighbors: SC, VA, TN, GA. Col. (1): NC \times Post-2009.

Col. (2): NC \times Post-2014 (repeal).

Table 4, columns (1)–(2) examine North Carolina’s credit adoption (2009) and repeal (2014) using neighboring states (South Carolina, Virginia, Tennessee, Georgia) as controls. The adoption DiD estimate is -0.022 (insignificant)—NC’s credit was relatively small and short-lived. The repeal estimate is -0.324 ($p = 0.25$), suggesting a decline in employment after the credit was removed, though the estimate is imprecise with only five states in the comparison. The point estimate is economically meaningful: a 28 percent decline in motion picture employment relative to neighbors.

Worker flows. Columns (3)–(4) decompose the employment effect into hires and separations. Both show positive TWFE coefficients ($+0.082$ and $+0.079$, both insignificant), suggesting that credits increase both hiring and turnover in the motion picture sector—consistent with a project-based industry where productions create temporary employment spikes.

6. Discussion

The central finding of this paper is methodological as much as substantive: the apparent null effect of film tax credits on employment is an artifact of TWFE bias, not of ineffective policy. When the estimator properly accounts for treatment-effect heterogeneity across adoption cohorts, the effect is positive and significant. This has direct implications for the policy

debate, since legislative decisions to repeal film credits in several states were influenced by evidence that now appears to reflect estimator misspecification rather than policy failure.

The 25 percent employment increase must be interpreted with care in two respects. First, it represents the effect on NAICS 512 employment in treated states relative to never-treated states—not a national net effect. To the extent that credits relocate productions across state lines (a zero-sum reallocation), the national employment effect is smaller. The never-treated states (Alaska, Delaware, Iowa, Kansas, Missouri, Nebraska, North Dakota, New Hampshire, South Dakota, Vermont, Wyoming) have structurally small film industries, which limits their suitability as counterfactuals despite clean pre-trends. Second, translating the employment effect into fiscal cost-effectiveness requires caution: with estimated annual program costs of \$10–12 billion nationally (Tannenwald, 2010) and a 25 percent increase from a base of roughly 350,000 NAICS 512 workers, the implied cost per additional job is approximately \$115,000–\$140,000—high relative to other job creation programs, though lower than the \$200,000+ figures critics often cite.

The finding also underscores a broader methodological point. The film tax credit literature is one of many policy settings where TWFE has been the workhorse estimator and where treatment effects are almost certainly heterogeneous—growing over time as states invest in infrastructure. The sign flip between TWFE (-0.038) and the heterogeneity-robust estimate ($+0.220$) should serve as a cautionary example for applied researchers working with staggered adoption designs. As Roth et al. (2023) emphasize, the choice of estimator is not a technical footnote but can determine whether a policy is judged effective or wasteful.

7. Conclusion

Film production tax credits create employment in the motion picture industry—approximately 25 percent more than untreated states—but the standard econometric toolkit obscured this effect for over a decade, producing a null that influenced real legislative decisions to repeal programs in North Carolina, Michigan, and elsewhere. The gap between TWFE and modern heterogeneity-robust estimators is not merely a statistical footnote; it changed the story from “credits don’t work” to “credits work, especially for early movers.” When the estimator determines whether a \$10 billion program is judged effective or wasteful, the stakes extend well beyond the seminar room.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: <https://github.com/SocialCatalystLab/ape-papers>

Contributors: @SocialCatalystLab

First Contributor: <https://github.com/SocialCatalystLab>

References

- Button, Patrick**, “Do tax incentives affect business location and economic development? Evidence from state film incentives,” *Regional Science and Urban Economics*, 2019, 77, 25–36.
- Callaway, Brantly and Pedro HC Sant’Anna**, “Difference-in-differences with multiple time periods,” *Journal of Econometrics*, 2021, 225 (2), 200–230.
- Chaisemartin, Clément De and Xavier d’Haultfoeuille**, “Two-way fixed effects estimators with heterogeneous treatment effects,” *American Economic Review*, 2020, 110 (9), 2964–2996.
- Christopherson, Susan**, “Project work in context: Regulatory change and the new geography of media,” *Environment and Planning A*, 2008, 40 (11), 2585–2601.
- Goodman-Bacon, Andrew**, “Difference-in-differences with variation in treatment timing,” *Journal of Econometrics*, 2021, 225 (2), 254–277.
- Roth, Jonathan, Pedro HC Sant’Anna, Alyssa Bilinski, and John Poe**, “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature,” *Journal of Econometrics*, 2023, 235 (2), 2218–2244.
- Sun, Liyang and Sarah Abraham**, “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, 2021, 225 (2), 175–199.
- Tannenwald, Robert**, “State Film Subsidies: Not Much Bang for Too Many Bucks,” *New England Public Policy Center Policy Brief*, 2010.
- Thom, Michael**, “Green Screen: Do State Film Production Incentives Create “Green” Jobs?,” *State and Local Government Review*, 2018, 50 (1), 19–29.

A. Standardized Effect Sizes

Table 5: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Log emp. (all)	0.220	0.082	1.856	0.119	0.044	Moderate positive
Log emp. (Black)	-0.102	0.087	2.617	-0.039	0.033	Small negative
Log emp. (White)	-0.011	0.062	1.856	-0.006	0.033	Small negative
Emp. share	0.266	0.144	1.487	0.179	0.097	Large positive

- Notes:** **Country:** United States. **Research question:** Do state film production tax credits increase motion picture industry employment, and do Black workers capture a proportional share of gains? **Policy mechanism:** State-level transferable or refundable tax credits (15–42% of qualified production expenditures) subsidize film, television, and commercial production, reducing location costs and attracting productions that would otherwise film in competing jurisdictions or countries. **Outcome definition:** Beginning-of-quarter employment count in NAICS 512 (Motion Picture and Sound Recording Industries) from the Quarterly Workforce Indicators, measured at the state-year level. **Treatment:** Binary; state-year adoption of a film production tax credit with rate $\geq 15\%$. **Data:** Census LEHD Quarterly Workforce Indicators (QWI), 2001–2024, 51 jurisdictions, state-year panel with race/ethnicity demographic breakdowns. **Method:** Callaway and Sant’Anna (2021) staggered difference-in-differences with never-treated comparison group; standard errors clustered at the state level. **Sample:** 37 treated states (excluding NC and MI, which repealed credits) and 13 never-treated states; NC repeal analyzed separately. $SDE = \hat{\beta}/SD(Y)$ where $SD(Y)$ is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|SDE| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).