From Your Pocket to Your Boss's: Worker Tax Credits & Wage Theft Reporting

Taylor Watson *

April 16, 2024

Abstract

Wage theft is a pervasive issue, yet it often goes unreported due to the threat of retaliatory firing. To better understand the tradeoff between remaining silent (and employed) and reporting the crime (and risking unemployment), this paper considers the role of tax credits — an important resource for the often low-income victims. Incorporating tax credits into the standard model of workplace crime, I show that when credits are contingent upon employment they decrease reporting propensities, as they increase the relative value of remaining silent (and employed). Using variation in state Earned Income Tax Credits over time in a Poisson regression, I show that an increase in credits of 10 percentage points is associated with a decrease in reporting counts of 4-11%. I argue this relationship is not driven by decreases in actual wage theft by showing that there is no negative effect of credits on several measures of wage theft intensity. Event study analysis shows that the effect takes roughly a year to occur but is strongly consistent thereafter.

*Boston University, watson@bu.edu

I Introduction

Wage theft is a pervasive issue where employers avoid paying due wages. It takes many forms, such as failing to meet the minimum wage or withholding overtime pay, and is present at firms as large as Amazon or as small as two-person shops (Sainato 2023). While it is widespread (costing workers an estimated $50 billion annually) it is not well-studied due to underreporting — only around $1 billion is recovered each year, as victims usually do not report the crime (Meixell and Eisenbrey 2014).

One of the key reasons that workers do not report wage theft is the threat of retaliatory firing. A similar mechanism has been well-documented in other workplace crimes such as sexual assault or safety violations: workers must choose between remaining silent (and employed) or speaking up (and risking unemployment), leading many to prefer the former.[1]

Unfortunately, due to underreporting wage theft is still poorly understood, in particular as it interfaces with the tax and transfer system.[2] The crime of wage theft falls heavily on workers in low-wage industries who are also likely to rely on transfer programs, making the interaction of the two important (Meixell and Eisenbrey 2014). This paper begins to fill this gap in the literature by exploring how worker tax credits affect the reporting behavior of wage theft victims. This contributes not only to the understanding of worker responses to wage theft, but also to the broader understanding of the retaliatory firing mechanism in workplace crime.

When studying the tradeoffs in reporting workplace crime, previous studies have primarily focused on the unemployment risk. For example, in Boone et al. (2011), the authors show how worsening a unemployment outlook discourages reporting of safety violations. In this paper, I instead consider variation in the value of remaining employed, specifically by looking at tax credits which are work-contingent. This amounts to a different kind of variation in

[1] Two recent examples: low-income workers with fewer outside options are less comfortable discussing workplace issues (Hertel-Fernandez 2020), and reductions in unemployment benefits exacerbate underreporting of sexual harassment in the workplace (Dahl and Knepper 2021).

[2] The lack of research in this area was highlighted by a recent review article, Marinescu and Rosenfeld (2022). The literature on employer behavior is much more robust, including what factors make firms more likely to engage in theft (Ji and Weil 2010; Clemens and Strain 2022) and how the government can best deter this behavior (Weil 2010, 2018; Stansbury 2021).

the tradeoff underlying the same retaliatory firing mechanism of reporting deterrence.

To study the effects of tax credits, I enhance the workplace reporting model of Boone and Van Ours (2006) to incorporate work-contingent credits. Simple comparative statics suggest that an increase in these credits should decrease reporting propensities, as the value of employment rises higher above the potential downside of reporting and becoming unemployed.

Taking this model to the data on wage theft (documented by the Federal Wage & Hour Division) this prediction is borne out. As a measure of work-contingent credits, I use differences in state Earned Income Tax Credit (EITC) matching generosity. States match the federal EITC at varying rates (from zero to as high as 85%), and many have adjusted those rates over time, generating plausibly exogenous variation. Analyzing this relationship at the ZIP-code level via Poisson regression, I find that an increase in state EITC of 10 percentage points is associated with a decrease in wage theft reports of approximately 4.2%. Scaling that up to the state level, this implies roughly 26 fewer reports per year (from an average of 600) and and additional $286,000 in unreported stolen wages.

While I emphasize the effect of the state EITC matching rate, it is more accurate to consider the joint effect of the matching rate and the local EITC participation rate. My theoretical model focuses on a representative worker and eschews the participation rate, but since my empirical work is at the ZIP code level, the participation rate can be incorporated. The participation rate modulates the mechanism of effect through which the matching rate operates — if the participation rate was zero, changes in the matching rate should not affect wage theft reports at all. I incorporate this interaction with similar results: the interaction between matching and participation rates has a significant negative effect on wage theft reporting. Additionally, the matching rate alone has a nearly null effect while the participation rate alone has a significant negative effect. This likely reflects the role of the federal EITC operating via a similar mechanism as the state matching rate.

While I find a negative relationship between tax credits and wage theft reporting, it is not clear *a priori* that this finding reflects my proposed mechanism. Reported wage theft is a product of not just the propensity of workers to report, but also the propensities of employers to engage in the underlying wage theft (Marinescu et al. 2020). The importance of this distinction was made clear by companion papers Boone and Van Ours (2006) and Boone et al. (2011) in the context of workplace safety. While Boone and Van Ours (2006) show that empirically it appears workplace safety violations fall during recessions, Boone et al. (2011) clarified that the effect is largely driven by changes in reporting behavior, rather than actual workplace safety changes.

Changes in EITC rates could be affecting this second factor, the frequency of wage theft itself. This would result in a lower reporting frequency mechanically, as there is less real theft for workers to report. In other words, the negative relationship between credits and observed reports could be driven by a mechanism of worker behavior (lowering their reporting propensities) or alternatively through a mechanism of employer behavior (lowering their wage theft propensities).

To understand this alternative mechanism, one can imagine an employer who sees the increased worker tax credits as insulating their employee against the costs of being unemployed, thus diminishing the potency of the employer's retaliatory firing threat. Such an employer would likely decrease their wage theft behavior in response to rising tax credits, in order to forestall reports from their newly flush workers, which could be costly to the employer.

In order to differentiate between these two mechanisms, I analyze the effect of changes in tax credits on secondary outcomes which proxy for employer behavior. These measures tell us something about the nature of the wage theft being reported. If the observed decrease in reporting is being driven by the alternative mechanism of employer behavior, one would expect that these secondary measures would also show some adjustment. If this were not the case, then employers would be adjusting their frequency of wage theft without changing their behavior within those wage theft episodes, which seems unlikely to be the case. Specifically, I use three metrics of employer behavior: the number of "empty" reports (where no theft or violations are found), the number of legal violations found during investigations of reports, and the amount of civil monetary penalties assessed per report. Across these three outcomes, I find no evidence of reduced wage theft intensity or changes in behavior on the part of employers consistent with reduced frequency.

To better understand the dynamics of the relationship between credits and wage theft, I extend my estimation strategy to include an event study design. While the estimated effect is not significant in the first year of state EITC availability, it becomes negative in the second year, and persists for at least five years thereafter. This finding of a negative relationship is consistent with my static results, and the timing of the effect may reflect a need for workers to learn about the availability of tax credits before they begin claiming it and incorporating the additional work-contingent income into their reporting decision-making.

These findings help explain why victims of wage theft so infrequently report their crime to the authorities, allowing many employers to get away without punishment. The key role of retaliatory firing must be addressed by future policy if progress is to be made in boosting the reporting rates of these abuses and building worker agency in the workplace.

II Theory

To understand why wage theft is often unreported this section walks through a simple reporting model, well-established in the literature, and expands it to incorporate the role of work-contingent tax credits. This model is based off that of Boone and Van Ours (2006) for the similar reporting issue of workplace safety, as demonstrated in static form and applied by Dahl and Knepper (2021) to sexual assault reporting (my version is similarly static). Comparative statics predict that increased credit generosity will decrease reporting probabilities — the intuition being that credits increase the value of employment relative to unemployment such that the threat of being fired becomes costlier.

Consider an environment with a representative worker earning wage $w > 0$ and facing exogenous wage theft, $t \in [0, w]$. The worker must choose whether or not to report that theft to the authorities. I abstract from the employer side of this relationship for now, focusing on how workers respond to a fixed level of theft.

If the worker chooses not to report the theft, they will receive their wage net of theft, $(w - t)$, and further net of taxes τ, represented as $(w - t)(1 + \tau)$. Note that in this setup $\tau \in [0, 1]$ will be assumed positive, similar to the real-world Earned Income Tax Credit.

If the worker chooses to report the theft, their outcome is less certain. A report will trigger an investigation into the situation by the government, which will succeed with a exogenous probability p. This uncertainty reflects the administrative challenges in proving wage theft, as well as that employers may be able to evade detection. A successful report results in repayment of backwages, equal to the amount taken t, plus potentially some multiplicative additional damages, represented as $\delta > 1$.[3] An unsuccessful report returns nothing for the worker.

Simultaneously, reporting creates the risk of a retaliatory firing by the firm, occurring with probability θ. A fired worker will either find a replacement job immediately (at new wage level w'), or they will remain unemployed and earn welfare benefits b. These events

[3] This income would also be subject to taxes.

occur with probability q and $1-q$, respectively. Accordingly, the value for the reporting worker of not being fired, V_{NF} is their status quo net income, $(w-t)(1+\tau)$. The value for the worker of being fired, V_F is the probability-weighted sum of re-employment and unemployment, or $qw'(1+\tau)+(1-q)b$.

Formalizing this setup, we can calculate the expected value for the worker of making a report, and consider when it may outweigh the welfare value of not making a report:

$$p(t\delta)(1+\tau)+\theta V_F+(1-\theta)V_{NF} \geq (w-t)(1+\tau) \tag{1}$$

The first term represents the potential payout of the report, the next two terms the uncertainty of potential retaliatory firing, and the right-hand side the status quo. This can be rewritten to compare the expected gains from reporting on one side to the expected losses on the other:

$$p(t\delta)(1+\tau) \geq \theta[(w-t)(1+\tau)-qw'(1+\tau)-(1-q)b] \tag{2}$$

With this simplification, note that the gains on the left-hand side are increasing with respect to the amount of theft t, which is to say that the potential gains from choosing to report are increasing in theft. In contrast, the right-hand side is decreasing in t, as greater wage theft reduces the take home pay of the worker, and thus their potential losses from remaining employed and silent. Combining these two properties, it is clear that there must also therefore exist a level of theft, t^* such that this condition holds with equality. Thus, there exists a tipping point degree of wage theft which can induce the worker to report or not report accordingly. Formally, there exists some t^* such that:

$$p(t^*\delta)(1+\tau) = \theta[(w-t^*)(1+\tau)-qw'(1+\tau)-(1-q)b] \tag{3}$$

which can therefore be solved for the expression of that level t^*:

$$t^* = \frac{\theta[(w - qw')(1+\tau) - (1-q)b]}{(p\delta + \theta)(1+\tau)} \tag{4}$$

This value is the indifference point for the worker: at a given level of wage theft, t, the decision to report boils down to whether that theft is above or below this indifference value. If $t > t^*$, the worker should report, and if $t < t^*$, they should not.

To understand how changes in the tax structure may affect reporting, consider now a mass of identical workers at identical wages facing a distribution of theft levels. If t^* rises, then some fraction of those workers will be pushed from a position where they prefer reporting into a position where they prefer not reporting, as their respective t falls below the threshold. In other words, a rise in t^* predicts a fall in reporting.

To assess whether taxes, τ, in this model will increase t^*, the derivative of the previous expression can be taken with respect to τ:

$$\frac{\partial t^*}{\partial \tau} = \frac{\theta(1-q)b}{[(p\delta + \theta)(1+\tau)]^2} \tag{5}$$

This expression must always be positive given the parameters involved, thus predicting that an increase in the tax transfer will increase the critical value t^*. Accordingly, this suggests that an increase in transfers should decrease the propensity of workers to report. This is the key prediction of the model which will be tested in this paper: an increase in cash transfers via the EITC should be associated with a decrease in reporting of wage theft.

III Data

Taking this model to the data requires three components: (1) wage theft reports, (2) state EITC rates, and (3) relevant control variables.

III.A Wage Theft Data

The most-used archive of wage theft reports comes from the US Wage & Hour Division, who maintain thorough documentation of each investigation. Their data extends back to 2005, and includes several variables required for this analysis: the address of each complaint, the amounts claimed and awarded, the number of employees involved, etc. While this data is quite rich, there are two challenges with using this data.

Firstly, this dataset contains two kinds of cases: cases where an employee complained to the government for assistance (worker-initiated case), and cases where the federal government proactively investigated a firm (agency-initiated case). Only the first case type is useful for this analysis, as an instance of a worker reporting wage theft. Fortunately, worker-initiated cases are a strong majority of those recorded (roughly 75%), but the two case types are not distinguished in the data. From the perspective of my analysis, this creates an overreporting bias in the count data, a form of nonclassical measurement error.

I will take this mixed data as given, treating all cases as worker-initiated reports. I offer a robustness test in Appendix A, which attempts to overcome the issue through a restricted sample. Using what is known about the Department of Labor's targeting strategy for agency-initiated cases, I limit my sample to the cases which are most likely to be worker-initiated. The DoL consistently targets their investigations to locations where there is known to be high wage theft, to industries with high prevalence of wage theft, and to areas where they believe their investigation will be able to create strong local deterrence effects (Weil 2010). Accordingly, I drop from my sample the top 0.25% of ZIP codes by wage theft report count.[4] These ZIP codes account for roughly 17% of all wage theft reports in my sample — for comparison, the share of all cases in the data which are agency-initiated is roughly 25% per year, meaning that this data restriction likely drops primarily agency-initiated cases. Repeating my analysis on this restricted sample returns similar results, suggesting that my findings are not being driven by agency-initiated cases.

[4] Specifically, I drop all ZIP codes which experience greater than 10 wage theft cases in any year. These are all ZIPs which regularly experience high levels of cases, and do not appear to be "one-off" instances which otherwise experience low wage theft.

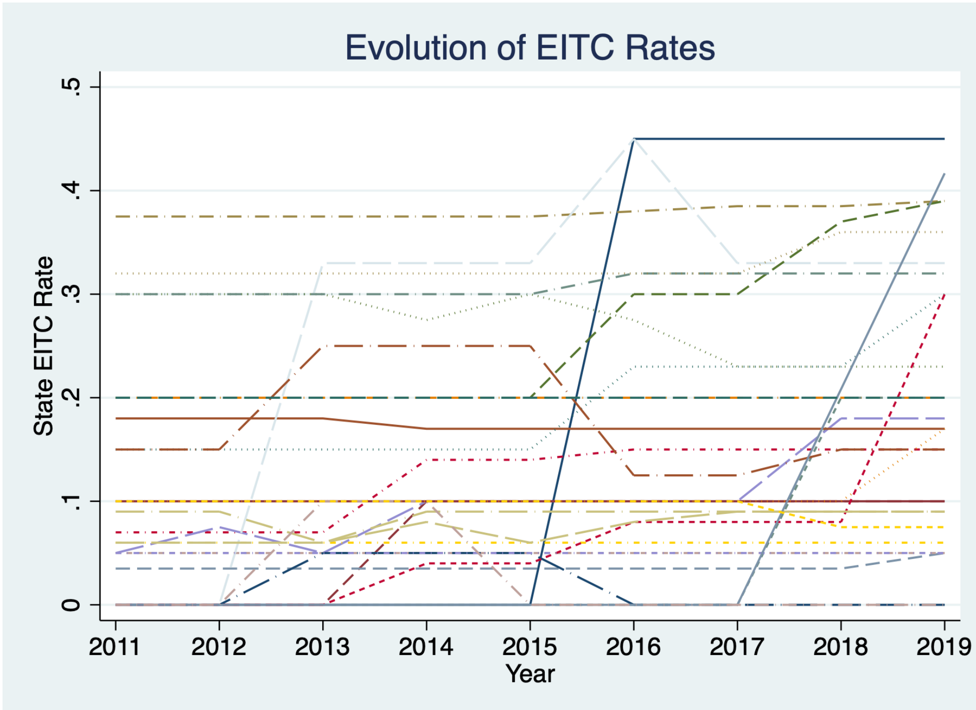
The second issue with the Wage & Hour dataset is that it is only a subset of all wage theft reports. Workers in the US have two avenues by which to seek redress of wage grievances: the federal DoL and their local state agency. Thus, not only does the federal file measure just a subset of all reports, it is likely a selected sample: individuals in states with effective local complaint management may be more likely to report there than in states with less effective local management, who may prefer the federal option. Unfortunately, collection of individual state wage theft archives is impossible due to the differences in data management across states.

Variation in state agency engagement would present a potential source of bias if there was strong correlation between agency engagement and state EITC rates. For example, if states with generous EITC matching also had strong state agency enforcement, this could create a negative correlation between state EITC and federal wage theft reports. This is unlikely to be the case. Firstly, state EITC rates vary across years within states frequently, creating variation within the relatively unchanging state agency enforcement context, as state wage and hour laws are infrequently changed. Secondly, a study of state-level policies specifically found that many policy changes taken by states in recent years have had an insignificant effect on wage theft reporting (Galvin 2016). To address any remaining potential bias from this source, I include state fixed effects in my analysis and control for the political party of the state governor in each year (as they have the greatest control over the state agencies enforcement activity and intensity). My results are therefore unlikely to be significantly altered by bias stemming from this data construction.

III.B State EITC Data

The explanatory variable of interest, state EITC matching rates, were sourced from the Tax Policy Center at the Urban Institute & Brookings Institution (Tax Policy Center 2023), with gaps filled in using the NBER TAXSIM calculator (Shapiro 2019). In most states, a single matching rate is provided, meaning that the state matches that percentage of the federal EITC; however, in some states there are different rates depending on factors like number of children. These nuances necessitate some interpretive choices, which I explain in more detail in Appendix B. The evolution of rates under my interpretation are summarized in Figure 1.

Figure 1: Evolution of Rates



III.C Control Data

With the outcome and explanatory variables in hand, I then incorporate relevant control variables

Tax Return Statistics

Detailed information on the tax returns filed in each ZIP code are available at the IRS Statistics of Income archive (Internal Revenue Service 2022). In particular, I pulled the total number of tax returns and the share of returns which claimed an EITC benefit for inclusion in my analysis. This is what I will use as a measure of the local EITC participation rate, particularly in the interaction analysis in Section V.

Business Patterns

The Census County Business Patterns archive includes detailed data at the ZIP code level (U.S. Census 2023b). This includes the number of establishments, the employment share of firms with 50 employees or less, and the share of establishments in several key wage-theft-prone industries: agriculture, construction, and food services.

Demographics

Demographic control variables were taken from the American Community Survey Five-Year Data, including variables for race, education, nativity, poverty, and public assistance (U.S. Census 2023a). There are two variables which were not available from the ACS, namely the population density and unemployment rates at the ZIP level. These are available at five-year frequency from the National Neighborhood Data Archive, and I use them as follows: the estimate of 2008-2012 applied over those years, for 2013-2015 applied to those years, and for 2016-2020 applied to those years (Melendez et al. 2023).

State-Level Variables

Three key variables were not available at the ZIP code level, so I used the available state-year level data. State minimum wages in 2020 dollar equivalents were sourced from the Wage & Hour Division (US Department of Labor Wage & Hour Division 2023).[5] Unionization rates were sourced from Unionstats (Macpherson and Hirsch 2023). The party of state governors was sourced from the ICPSR (Kaplan 2023) and validated via the National Governor's Association, and this is encoded as a binary variable taking on a value of 0 if the governor is a Republican and a value of 1 otherwise.

III.D Sample Construction

The various datasets were merged on the ZIP-code year basis, then the sample was trimmed in various ways to avoid outliers driving the results. Firstly, I restricted to those ZIP codes which correspond to the fifty US states, and exclude other ZIP codes for territories, Washington DC, or other administrative units. Second, ZIP codes which at any point reported fewer than 250 tax returns were dropped across all years, as tax information was not reported for

[5]My thanks to Joe Lisle who makes these data more easily available through a compilation available here.

any ZIP below this threshold, and is essential to the analysis.[6] Third, I winsorize the main outcomes at the 1% and 99% levels, to reduce the influence of outliers.

Temporally, I restrict to 2011-2019 to ensure control coverage and avoid the COVID-19 pandemic. These years display radically different wage theft behaviors, and are subject to a host of other concerns that they may otherwise be different in unaccountable ways.

The final sample contains 25,339 ZIP codes, almost all of which are observed in all nine years 2011-2019 (for a total of 218,947 ZIP-year observations). This sample and the variables used in my analysis are summarized in Table 1.

[6]These dropped zip codes account for roughly 2% of all wage theft cases in the data, making their removal unlikely to drive the results.

Table 1: Summary Statistics

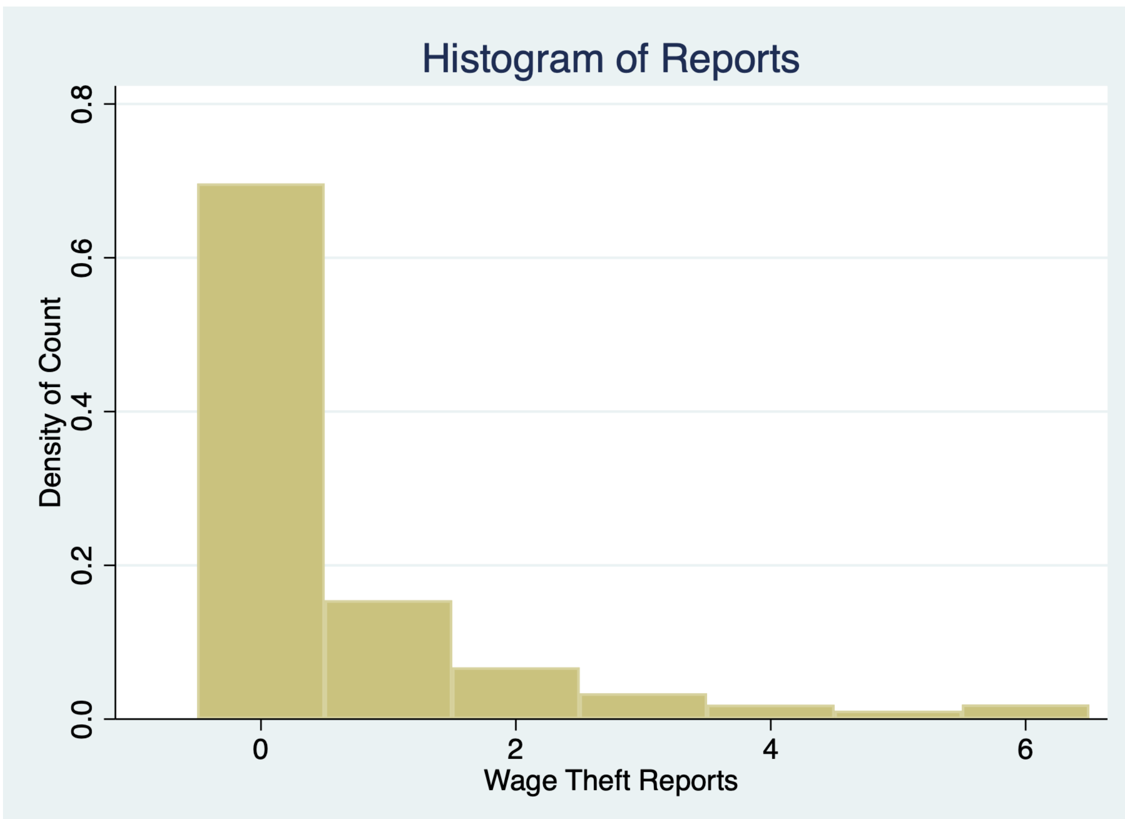
	Mean	Std. Dev.	Min	Max
Wage Theft Reports	0.63	1.25	0.00	6.00
State EITC Rate	0.10	0.16	0.00	0.85
State Unionization Rate	0.12	0.05	0.03	0.26
State Minimum Wage	8.37	1.48	2.02	13.66
Tax Returns	5,862	7,066	250	59,000
EITC Participation Rate	0.33	0.12	0.00	0.84
Share Below Poverty	0.10	0.08	0.00	1.00
Share Black	0.09	0.16	0.00	1.00
Share <HS Education	0.13	0.09	0.00	0.92
Share Noncitizen	0.04	0.06	0.00	0.61
Share Receiving Public Assistance	0.25	0.17	0.00	1.00
Share Establishments in Construction	0.13	0.10	0.00	1.00
Share Establishments in Agriculture	0.01	0.03	0.00	1.00
Share Establishments in Food Services	0.08	0.06	0.00	1.00
Share Employment in <50 Worker Establishments	0.97	0.03	0.00	1.00
Population Density	1,545	5,478	0.05	159,897
Unemployment Rate	0.05	0.03	0.00	0.74

IV Reporting Analysis

IV.A Reporting - Empirical Approach

Given the outcome of interest (the number of wage theft reports in a ZIP code-year) is a count variable, I will accordingly use a Poisson regression approach. The distribution of report counts across ZIP-years (seen below in Figure 2) is positively skewed with a large proportion of zeroes.[7]

Figure 2: Histogram of Reports



Even with the correct regression specification, the relationship between EITC rates and report frequencies may be corrupted by endogeneity, primarily by omitted variables and reverse causality. Addressing these in turn, my covariates adequately control for the relevant variables. The relative sizes of ZIP codes are accounted for through the number of

[7]While just over 71% of observations have no reports, less than a quarter of ZIP codes never experience a report in any year.

establishments and the number of tax returns filed. The labor market conditions are measured through the ZIP unemployment rate, poverty level, public assistance engagement, and the state unionization rate. Furthermore, the industrial makeup of the ZIP is controlled for through the establishment shares of key wage theft industries (agriculture, construction, food service) and the share of employment at smaller (<50 worker) firms. On top of these detailed controls, I account for unobserved differences in state environments through state fixed effects, as well as for national shocks through year effects. This comprehensive set of controls is sufficient to ensure that the estimated effect of state EITC rates is free of contamination from omitted variables — such a variable would need to comove with rates within states over time, where the among of variation (both positive and negative, varying in magnitude) makes it highly unlikely.

With regards to reverse causality, it seems highly unlikely that the setting of state EITC generosity is done in consideration of the level of wage theft. These rates are almost always considered in the context of broader tax reforms, which are subject to numerous higher-order concerns such as fairness, budgetary needs, and political dynamics. That the occurrence of wage theft at a particular level would be a contributing factor to these decisions seems implausible.

To estimate the effect of state EITC rates via maximum likelihood, I assume the following conditional fixed effects Poisson model:

$$\mathrm{E}[R_{zy}|\delta_s, \gamma_y, \mathrm{SEITC}_{s(z)y}, X_{zy}] = \exp(\delta_s + \gamma_y + \beta \cdot \mathrm{SEITC}_{s(z)y} + \Theta X_{zy}) \tag{6}$$

where R_{zy} is the number of wage theft reports in ZIP code z in year y, δ_s are state fixed effects, γ_y are year effects, X_{yz} are the various control variables mentioned, and β is the coefficient of interest on the variable $\mathrm{SEITC}_{s(z)y}$, the state EITC rate for the state s of ZIP code z in year y. Estimation of Equation (6) by quasi-maximum likelihood with fixed effects leads to consistent estimates (Wooldridge 1999). While I have clustered my standard errors at the level of the state-year to allow for correlation of the errors within those units, this in turn creates an issue of heterogeneously sized clusters. I address concerns related to this issue in Appendix C.

IV.B Reporting - Results

Modelling of wage theft reporting is summarized in Table 2. Beginning with a naive regression using only the state EITC rate, Column 1 estimates a statistically significant negative effect, with a coefficient of -1.311. The sign and significance of this estimate is robust to the inclusion of two-way fixed effects for state and year in Column 2 as well as a battery of controls in Column 3 (though the magnitude falls to -0.402). To further assess the strength of this result, Column 4 uses ZIP code fixed effects in place of state fixed effects, controlling for greater unobserved heterogeneity, producing highly similar results. Finally, Column 5 estimates the effect in a binary probit model, where the outcome is whether any wage theft reports occurred in a given ZIP-year. Here as well I find a significant negative effect.

While not the focus of this analysis, it is worth noting the coefficient on the ZIP code unemployment rate, for comparison with previous work like Dahl and Knepper (2021) and Boone et al. (2011). They use unemployment conditions as the main variation of interest in explaining changes in reporting behavior, finding that worsening unemployment conditions have a suppressive effect on reporting of workplace crime. In line with those papers, I find that an increase in the unemployment rate has a negative correlation with reporting, with a sizeable effect estimate in Columns 3 and 5.

Table 2: Wage Theft Report Results

	(1) Naive Poisson	(2) TWFE Poisson	(3) Complete Poisson	(4) ZIP FE Poisson	(5) Probit
State EITC Rate	-1.311***	-0.367***	-0.402***	-0.361***	-0.203**
	(0.283)	(0.120)	(0.136)	(0.133)	(0.102)
Unemployment Rate			-1.413***	-0.017	-0.809***
			(0.348)	(0.282)	(0.217)
State & Year FE		✓	✓	ZIP & Year	✓
Controls			✓	✓	✓
R^2	0.00	0.06	0.32	0.31	0.31
N	218,930	218,930	218,930	145,975	218,930

Note: Standard errors are clustered at the state-year level. * = p < 0.1, ** = p < 0.05, *** = p < 0.01. Complete estimation results are available in Appendix Section D.

V Interaction Analysis

The foregoing analysis abstracts away the interaction between EITC matching rates and local EITC participation rates. There is well-known heterogeneity in participation across ZIP codes, which could in turn modulate the effect of a state's EITC matching rate (Chetty et al. 2013). To capture this, I recreate my results while regressing on not just the state EITC matching rate and the ZIP code EITC participation rate (as before), but also on the interaction between the two.

Theoretically, the effect of the matching rate by itself should be null in a regression which also includes its interaction. The interpretation of this coefficient is how reports vary with the matching rate when participation rate is zero, which should forestall any effect of changing rates. In contrast, the coefficient on the participation rate by itself should be negative, as there the federal EITC is available to workers in all states. More individuals participating in the EITC locally follows a similar logic to the model in Section II: more workers with work-contingent income will value employment more relative to unemployment, and be discouraged from reporting. Note that in Table 2, the coefficient on participation rate was negative and highly statistically significant in both the Poisson and probit specifications. Finally, the coefficient on the interaction term of matching and participation rates should be negative for the same reasons argued for the rate alone in the foregoing analysis without the interaction.

The results of this extended analysis are presented in Table 9. Taking the coefficients in turn, the coefficient on the EITC participation rate is negative and highly statistically significant, aligning to the theoretical value. When more individuals in a locality participate in the EITC, the value of work relative to unemployment increases, and discourages reporting workplace crime such as wage theft. This result suggests that a state with no matching of the federal EITC would still see a decline in wage theft reports if workers increased their participation. The coefficient estimate of -0.622 implies that an increase in participation rate of 10 percentage points would decrease wage theft reports by about 6.2%.

Next, the coefficient on the interaction term is large, negative, and highly statistically significant, confirming the theoretical prediction and validating the previous analysis in a more nuanced specification. Interpreting this value at the mean of state EITC participation rate (33%), the coefficient of -3.279 suggests that increasing the matching rate by 10 percentage points would decrease wage theft reporting by roughly 11% (ignoring for now the contribution of the coefficient on state EITC rate itself, which would tamper this effect slightly). While only an approximation, this would correspond to a state-level decrease of 66 fewer wage theft reports (off an average of 600 annually) and about $726,000 in additional unreported stolen wages.

Finally, the state EITC matching rate alone displays a positive effect, significant at the 10% level. This is puzzling given the theoretical prediction of a null result, and may reflect unobserved heterogeneity between those ZIP codes with no EITC filers and those with some EITC filers — for example, they are likely to be higher-income and homogeneous compared to the average ZIP. There are only 750 ZIP-year observations in my dataset with no EITC filers, with an average of 0.05 reports across them. Ideally this value would be zero, but as my estimate is just a linear approximation, some deviation is tolerable.

Table 3: Interaction Results

	(1) Complete Poisson	(2) Probit
State EITC Rate	0.680*	0.314
	(0.386)	(0.248)
EITC Participation Rate	-0.622***	-0.172*
	(0.125)	(0.089)
State EITC Rate × EITC Participation Rate	-3.279***	-1.596**
	(1.097)	(0.686)
State & Year FE	✓	✓
Controls	✓	✓
R^2	0.34	0.33
N	218,947	218,947

Note: Standard errors are clustered at the state-year level. * = p < 0.1, ** = p < 0.05, *** = p < 0.01.

VI Secondary Outcomes Analysis

VI.A Secondary Outcomes - Empirical Approach

The results of the above estimations alone are insufficient to claim that there is a negative effect of transfers on reporting — it could be the case that transfers reduce theft frequency itself, mechanically generating a decline in reporting as there is less theft to report. To argue against this alternative mechanism (which is characterized by a change in employer behavior, rather than worker behavior) I move beyond the outcome of report frequency and assess the characteristics of the reports themselves. If this alternative mechanism is driving the observed change in reporting, one would expect also to see changes in these measures as proxies of the intensity of theft and employer behavior. I present each measure in turn:

Total Violations

In a wage theft investigation, the authorities record each individual violation of the law discovered. This is a proxy for the intensity of the wage theft taking place, as employers stealing from their workers in multiple ways (violating multiple laws) can steal more from their workers. If employers are responding to changes in tax credits by reducing their wage theft (out of greater fear of being reported) it seems highly likely they would seek to reduce the number of laws they are breaking in the process, or otherwise reduce the different kinds of wage theft in which they engage. In other words, if employers are reducing their extensive margin of wage theft, it seems likely they are also reducing their intensive margin. I use a similar Poisson model for estimating the effect of tax credits on violation counts.

Civil Monetary Penalties

In addition to ordering repayment of backwages, wage theft investigations can lead to civil penalties for the employer. By similar thinking to total violations, this variable proxies for the intensive margin of wage theft, and should be decreasing if employers change their behavior as predicted under the alternative mechanism. An increased probability of being reported increases the expected cost of wage theft, and should therefore decrease the intensity of wage theft engagement.

Unfortunately, this variable is not a count, but a continuous variable (specifically I use the log of the amount of penalties per employee involved in the case) so a Poisson model is inappropriate for estimation. Instead, I use a Heckman selection model to estimate the effect of credits on civil penalties. The sample of reports wherein penalties are found are certainly a selected sample among wage theft reports, thus estimating a simple linear model would produce a biased estimate. For identification, I omit the following variables from the penalties estimation but include them in the binary step: Governor's Party, State Minimum Wage, and share of the population earning over 400% of the poverty line. Each of these variables likely has an effect on whether or not penalties are assessed on a report, but do not influence in turn the amount of those penalties. I estimate this model in one step MLE.

Empty Reports

When a wage theft report is documented with no findings (i.e., no backwages owed, no legal violations found, and no civil monetary penalties assessed) I refer to that as an "empty" report.[8] If employers were responding to transfers via an alternative mechanism and engaging in less wage theft, I would predict a positive effect of transfers on the count of empty reports, as there would be less wage theft to find, and thus, more empty reports. In contrast, if the effect of transfers on reporting is driven by the worker-side mechanism, then we would expect no change. I estimate the effect via similar model to Equation (6), using the count of empty reports.

VI.B Secondary Outcomes - Results

I test my battery of secondary outcomes in Table 10. Interpreting these results (beginning in Column 1), the effect of EITC rates on the number of "empty" reports, where no wage theft or other workplace crime is found, is null. These empty reports are difficult to characterize — some may be agency-initiated investigations which fail to find evidence. This difficulty notwithstanding, if it was true that employers were decreasing their wage theft behavior in

[8]It is worth noting that these are more common among agency-initiated reports relative to worker-initiated ones, and that my previous results on reporting are robust to measuring only non-empty reports as the outcome.

response to rising EITCs, this would suggest a positive coefficient estimate, rather than the observed negative (though insignificant) estimate in Column 1.

Another null result is found for total violations in Column 2; however, not for the effect on civil monetary penalties in Column 3. This result is less straightforward to interpret, yet it does not support the mechanism of a reduction in employer intensity. The estimated effect is positive and statistically significant (even when bootstrapping in Appendix C). This suggests that more penalties are assessed on wage thefts reported at higher levels of state EITC, whereas a decrease in wage theft intensity should result in fewer penalties.

Returning to the model in Section II, this result agrees with the intuition. At higher EITC levels, workers are willing to "put up" with more intense wage theft, as the value of work rises such that the risk of retaliatory firing is not worth running. Accordingly, those workers who do still choose to report are likely experiencing more severe wage theft, as for them the benefits of reporting still outweigh the potential costs. The results in Column 3 suggest that the wage theft reported at higher EITC levels is more severe or intense (more worthy of being penalized) than at lower EITC rates. A similar finding is reported in the context of sexual assault in Dahl and Knepper (2021), that the kinds of reports filed become more severe as outside options become less attractive relative to remaining employed.

Table 4: Intensive Margin Results

	(1) Empty Reports Poisson	(2) Total Vltns. Poisson	(3) Civil Penalties Heckman
State EITC Rate	-0.163	0.131	-0.357
	(0.252)	(0.228)	(0.475)
EITC Participation Rate	-0.374**	-0.061	-0.415
	(0.184)	(0.137)	(0.318)
State EITC Rate × EITC Participation Rate	0.293	-0.829	4.030***
	(0.624)	(0.566)	(1.143)
State & Year FE	✓	✓	✓
Controls	✓	✓	✓
N	66,421	66,421	66,421

Note: Standard errors are clustered at the state-year level. * = $p < 0.1$, ** = $p < 0.05$, *** = $p < 0.01$.

The results presented here rebuke the argument that the decreased reporting found in Section IV and Section V is being driven by changes in employer behavior.

VII Event Study Analysis

VII.A Event Study - Empirical Approach

To better understand the nature of the relationship between tax credits and wage theft reporting, I extend my estimation strategy to include an event study design. This allows some understanding of the dynamics of the effect, compared to the simple aggregates presented earlier.

Given the staggered timing of treatment and the nonlinear outcome of interest, the optimal estimator in this context is the Mundlak approach developed by Wooldridge (2021), using fixed effects Poisson quasi-maximum likelihood estimation.[9] Under a modified version of the parallel trends assumption, this estimator has been shown to identify average treatment effects on the treated (hereafter ATTs) for each treatment cohort and time unit.

Specifically, I estimate the following exponential mean function:

$$\begin{aligned} \mathrm{E}[R_{zy}|\delta_z, \mathrm{SEITC}_{s(z)y}, X_z] = \\ \delta_z \exp\Bigg[\sum_{s=2}^{Y} \theta_s f_{s_y} + \sum_{s=2}^{Y} \pi_s (f_{s_y} \times X_z) + \sum_{r=q}^{Y}\sum_{s=r}^{Y} \tau_{rs} (\mathrm{SEITC}_{s(z)y} \times d_{zr} \times f_{s_y}) \\ + \sum_{r=q}^{Y}\sum_{s=r}^{Y} \rho_{rs} (\mathrm{SEITC}_{s(z)y} \times d_{zr} \times f_{s_y} \times \dot{X}_{zr})\Bigg] \end{aligned} \tag{7}$$

where R_{zy} is the number of wage theft reports in ZIP code z in year y, δ_s are ZIP-code fixed effects, $\mathrm{SEITC}_{s(z)y}$ is equal to 0 if the year y is before the start of the state s of ZIP code z's EITC and is equal to the average EITC rate since introduction thereafter, X_z are the control variables of ZIP code z averaged over the panel, f_{s_y} are binary variables equal to unity if $s = y$, d_{zr} is a binary variable equal to unity if ZIP code z of treatment cohort r is ever treated, and $\dot{X}_{zr}$ are within-treatment-cohort demeaned covariates. Estimation

[9] The implementation of this approach in Stata was developed by Rios-Avila (2023), to whom I am quite grateful for its development and broader Stata expertise.

of Equation (7) by quasi-maximum likelihood leads to consistent estimates of the ATTs (Wooldridge 2021).

Unlike many event studies, the treatment in this context is a continuous variable, rather than a binary one. Deviating from the previous results, I employ two alternative definitions of treatment which are better suited to this analysis. In both cases, I define the treatment such that it is fixed for treated units from the start of treatment through the end of the sample. In the first case, I use the average of the ZIP code EITC participation rate since the state began offering a supplemental EITC, and interact it with the average rate of that state's EITC since introduction. In the second case, I use the average of the state's EITC participation rate since introduction, again interacted with the average of the state matching rate. The results are largely similar across these two specifications.[10]

The key identifying assumptions for estimation of this model are those of no anticipation and parallel trends. In the first case, as argued previously it seems highly unlikely that individual workers would adjust their behavior in anticipation of shifts in tax policy at the state level, which are relatively complex. The immediate consequences of the reporting decision relative to the delayed-benefit of state EITC matching make for an unlikely anticipatory change in outcomes ahead of treatment. For the second assumption, I will follow Wooldridge (2022) in using a joint cluster-robust Wald test of significance on leading event study coefficient estimates to test whether the parallel trends assumption is clearly violated.

VII.B Event Study - Results

The results of the event study estimation are summarized in Table 5. Traditional event study coefficients of leads and lags were aggregated from the regression output using the same Stata techniques mentioned previously. Beginning in Column 1, the ATTs for treatment defined at the ZIP code level are provided from five years prior to treatment through five years after treatment. While the coefficients are generally not statistically significant, it is clear that

[10] Some states in my sample undergo treatment reversal, which would violate the assumptions required for identification of this estimator (i.e., some states introduce a state EITC and then remove it). I drop these states (North Carolina and Washington) from this part of the analysis.

the coefficients after treatment are negative. This aligns with the previous results in this paper, that exposure to additional work-contingent income reduces the propensities of wage theft reports.

The effect does not appear immediately, as in the year of first treatment (coefficient row of Treatment × Lag 0), the effect is insignificant and positive in sign. This may suggest that workers require some time to learn about the availability and benefits of their state EITCs. The coefficient estimates are presented graphically in Figure 3 below.

Another feature of the estimated ATTs is their relative consistent magnitude post-treatment. This suggests that the effect, once treated, is relatively constant, rather than waxing or waning. This is in turn gives confidence that units which are always treated in the sample period (i.e., treated on or before 2011) should display relatively similar treatment effects to those which are treated in the sample period. This is presented more formally in Column 2, where I add to the estimation those states which are always treated in the sample period, assigning them a treatment year according to when their EITC began, such that they can contribute to ATT estimation where possible relative to the available data. These results offer more precise estimates of the treatment effect on the lags, with slightly larger magnitude coefficient estimates, and increased statistical significance.

Assessing the parallel trends assumption, the Wald test for joint significance of the lead coefficients fails to reject the null hypothesis. This provides confidence that the estimation approach is valid in this context.

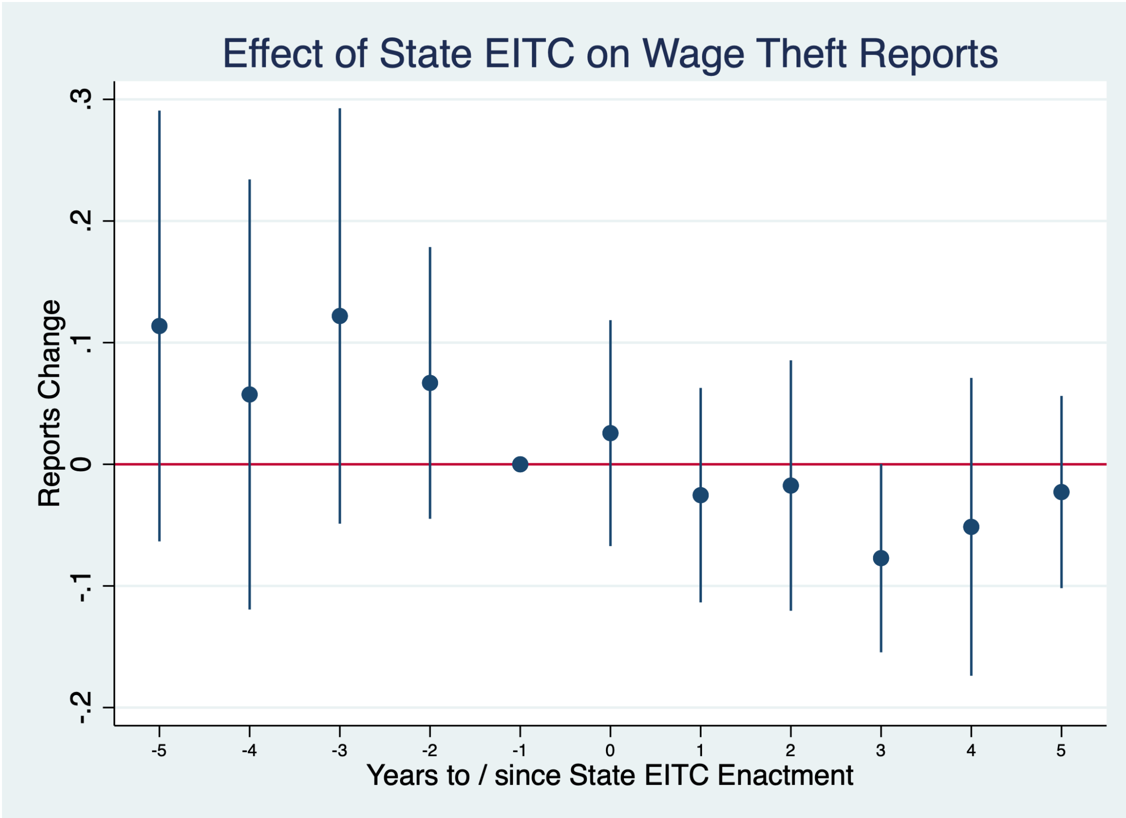
Finally, Columns 3 and 4 repeat the estimation of Columns 1 and 2, but use the treatment defined at the state level rather than that at the ZIP code, and produce relatively consistent results. The graphical results for all specifications are presented in Appendix E.

Table 5: Event Study Results

	ZIP-Level Treatment		State-Level Treatment	
	(1)	(2)	(3)	(4)
Treatment × Lead 5	0.114	0.097	0.074	0.059
	(0.090)	(0.091)	(0.086)	(0.085)
Treatment × Lead 4	0.057	0.032	0.022	-0.001
	(0.090)	(0.084)	(0.087)	(0.080)
Treatment × Lead 3	0.122	0.122	0.090	0.095
	(0.087)	(0.087)	(0.084)	(0.083)
Treatment × Lead 2	0.066	0.066	0.046	0.048
	(0.057)	(0.056)	(0.060)	(0.058)
Treatment × Lag 0	0.026	-0.024	-0.006	-0.011
	(0.047)	(0.037)	(0.062)	(0.037)
Treatment × Lag 1	-0.025	-0.069**	-0.052	-0.059*
	(0.045)	(0.035)	(0.055)	(0.035)
Treatment × Lag 2	-0.017	-0.058*	-0.038	-0.048
	(0.053)	(0.031)	(0.057)	(0.031)
Treatment × Lag 3	-0.077*	-0.047**	-0.096**	-0.040*
	(0.040)	(0.022)	(0.044)	(0.023)
Treatment × Lag 4	-0.051	-0.063*	-0.067	-0.050
	(0.062)	(0.032)	(0.071)	(0.034)
Treatment × Lag 5	-0.023	-0.037	-0.031	-0.024
	(0.040)	(0.030)	(0.045)	(0.030)
Wald Test χ^2 Stat	3.44	3.36	1.79	2.11
Stat P-Value	(0.48)	(0.50)	(0.77)	(0.72)
Include Always-Treated		✓		✓
Controls	✓	✓	✓	✓
N	109,143	200,007	109,143	200,007

Note: Standard errors are clustered at the state-year level. * = p < 0.1, ** = p < 0.05, *** = p < 0.01. Coefficients correspond to years relative to treatment, such that Lead 2 refers to the effect two years prior to treatment, and Lag 0 refers to the effect in the first year of treatment.

Figure 3: Event Study ATT Estimates - ZIP Treatment



VIII Conclusion

Wage theft is an understudied workplace crime, even though it poses a significant cost to American workers. Despite its prevalence, few workers choose to report to the government when they are the victims of wage theft, frustrating a deeper understanding of its dynamics. I explore this underreporting through the established literature mechanism of retaliatory firing deterrence.

Workers faced with potential retaliatory firing for reporting workplace crime must weigh the value of employment against the costs of potential unemployment. As shown in recent works such as Dahl and Knepper (2021) and Johnson et al. (2023), when unemployment conditions worsen, workers are less likely to report, as the cost of reporting has effectively risen. I make two contributions to this literature. Firstly, I extend the mechanism of retaliatory firing into the specific context of wage theft, finding consistency with the results in other crimes.

Secondly, I augment the mechanism by altering not the cost of unemployment, but rather the value of employment itself, through work-contingent government transfers. Building this variable into the standard reporting model, I show how increased transfers should lower reporting propensities. Taking this model to the data, I confirm this result, showing that state EITC rates are negatively associated with wage theft reports. This finding is robust to a variety of empirical specifications.

I further defend this finding from the critique that a decrease in reports could reflect decreased wage theft activity, rather than a decrease in reporting propensities of workers. I show through several proxy measures that the intensity of wage theft does not have a negative relationship with state EITC rates, making it unlikely that employers are reducing their wage theft behavior in response to state matching rates. This makes the mechanism of reduced worker reporting the likely explanation for the observed phenomena. This finding improves the understanding of wage theft, and may inform policy designed to increase reporting propensities.

References

Boone, J. and Van Ours, J. (2006). Are recessions good for workplace safety? *Journal of Health Economics*, 25:1069–93.

Boone, J., van Ours, J. C., Wuellrich, J.-P., and Zweimüller, J. (2011). Recessions are bad for workplace safety. *Journal of Health Economics*, 30(4):764–773.

Chetty, R., Friedman, J. N., and Saez, E. (2013). Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings. *American Economic Review*, 103(7):2683–2721.

Clemens, J. and Strain, M. R. (2022). Understanding"wage theft": Evasion and avoidance responses to minimum wage increases. *Labour Economics*, 79:102285.

Dahl, G. B. and Knepper, M. M. (2021). Why is workplace sexual harassment underreported? the value of outside options amid the threat of retaliation. Working Paper 29248, National Bureau of Economic Research.

Galvin, D. J. (2016). Deterring wage theft: Alt-labor, state politics, and the policy determinants of minimum wage compliance. *Perspectives on Politics*, 14(2):324–350.

Hertel-Fernandez, A. (2020). American workers' experiences with power, information, and rights on the job: A roadmap for reform. Report, Roosevelt Institute.

Internal Revenue Service (2022). SOI Tax Stats - Individual Income Tax Statistics - ZIP Code Data (SOI).

Ji, M. and Weil, D. (2010). Does ownership structure influence regulatory behavior? the impact of franchising on labor standards compliance. *Industrial and Labor Relations Review*, 68(5):977–1006.

Johnson, M., Schwab, D., and Koval, P. (2023). Legal protection against retaliatory firing improves workplace safety. *Review of Economics and Statistics*.

Kaplan, J. (2023). United States Governors 1775-2020. Inter-university Consortium for Political and Social Research.

Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, 1(1):23–41.

MacKinnon, J. G., Ørregaard Nielsen, M., and Webb, M. D. (2023). Cluster-robust inference: A guide to empirical practice. *Journal of Econometrics*, 232(2):272–299.

Macpherson, D. and Hirsch, B. (2023). Five decades of cps wages, methods, and union-nonunion wage gaps at unionstats. com. *Industrial Relations: A Journal of Economy and Society*.

Marinescu, I. and Rosenfeld, J. (2022). Worker power and economic mobility: A landscape report. Report, Urban Institute WorkRise report.

Marinescu, I. E., Qiu, Y., and Sojourner, A. (2020). Wage inequality and labor rights violations. SSRN Scholarly Paper ID 3673495, Social Science Research Network, Rochester, NY.

Meixell, B. and Eisenbrey, R. (2014). An epidemic of wage theft is costing workers hundreds of millions of dollars a year. Issue Brief 385, Economic Policy Institute.

Melendez, R., Clarke, P., Khan, A., Gomez-Lopez, I., Li, M., and Chenoweth, M. (2023). National Neighborhood Data Archive (NaNDA): Socioeconomic Status and Demographic Characteristics of Census Tracts, United States, 2008–2020. Inter-university Consortium for Political and Social Research.

Rios-Avila, F. (2023). Jwdid: Stata module to estimate difference-in-difference models using mundlak approach. Statistical Software Components S459114, Boston College Department of Economics.

Roodman, D., Ørregaard Nielsen, M., MacKinnon, J. G., and Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal*, 19(1):4–60.

Sainato, M. (2023). ‘I have not seen one cent’: billions stolen in wage theft from US workers”. *The Guardian.*

Shapiro, I. (2019). State EITC provisions 1977-2018.

Stansbury, A. (2021). Do US firms have an incentive to comply with the FLSA and the NLRA? Working Paper 21-9, Peterson Institute for International Economics.

Tax Policy Center (2023). State EITC as Percentage of the Federal EITC.

U.S. Census (2023a). American Community Survey 5-Year Data (2009-2021). Database, U.S. Department of Commerce.

U.S. Census (2023b). County Business Patterns Database. Database, U.S. Department of Commerce.

US Department of Labor Wage & Hour Division (2023). Changes in basic minimum wages in non-farm employment under state law: Selected years 1968 to 2022.

Weil, D. (2010). Improving workplace conditions through strategic enforcement. *Boston U. School of Management Research Paper*, (2010-20).

Weil, D. (2018). Creating a strategic enforcement approach to address wage theft: One academic’s journey in organizational change. *Journal of Industrial Relations*, 60(3):437–460.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1):77–97.

Wooldridge, J. M. (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators. Working paper.

Wooldridge, J. M. (2022). Simple Approaches to Nonlinear Difference-in-Differences with Panel Data. Working paper.

Appendix

A Robustness - Dropping Top ZIPs by Wage Theft

One concern using the Wage & Hour Division data is that I cannot distinguish between worker-initiated and agency-initiated cases. The latter is not relevant to my analyses. To partially overcome this issue, I attempt to remove some of the cases which are most likely to be agency-initiated, specifically all those ZIP codes which experience at least 10 wage theft reports in any year. This should target those ZIP codes with the highest amount of wage theft, which are precisely those which are most often targeted for agency-initiated cases.

The Wage & Hour Division targets their investigations in areas with high density of wage theft and probable wage theft establishments, in order to maximize the deterrent effect of their investigations. That makes this drop likely to remove those ZIP codes where agency-initiated reports dominate the count of reports. In total, I drop about 0.10% of observations, accounting for about 17% of total wage theft cases in my sample. Given that the Wage & Hour Division reports roughly 25% per year, this seems likely to drop a high share of the agency-initiated cases, without significantly altering my dataset from the winsorized version.

Results from recreating the main analyses of this paper are presented below in Table 6. The key coefficients maintain similar size, magnitude, and significance levels, supporting the belief that my analysis would likely be robust to a sample of only worker-initiated reports. The exception is the result in Column 3, for the secondary outcome measure of empty reports count — the results for matching and participation rates are significant and negative, while the result for their interaction is significant and positive.

The combination of positive and negative effects here is puzzling — it is hard to conceive of a mechanism by which increased EITC matching and participation rates leads to less empty reports, while their interaction would have a positive association. This could mean that the outcome of empty report counts is simply not a reliable measure of employer wage theft intensity, and is generating spurious results. Even if not, it is difficult to rationalize this result with a decrease in employer wage theft intensity, given the contradictory results.

Table 6: Robustness - Dropping Top ZIPs

	Reports		Secondary Outcomes		
	(1) Complete Poisson	(2) Probit	(3) Empty Reports Poisson	(4) Total Violations Poisson	(5) Civil Penalties Heckman
State EITC Rate	0.680*	0.302	-0.481**	0.008	-0.622
	(0.374)	(0.236)	(0.239)	(0.204)	(0.466)
EITC Participation Rate	-0.486***	-0.141	-0.315**	0.083	-0.067
	(0.115)	(0.086)	(0.160)	(0.137)	(0.336)
State EITC Rate × EITC Participation Rate	-3.380***	-1.556**	1.182**	-0.500	4.591***
	(1.032)	(0.650)	(0.503)	(0.493)	(1.174)
State Unionization Rate	-0.820	-0.795	-1.498	0.114	-0.057
	(0.911)	(0.662)	(1.453)	(0.906)	(2.172)
State Minimum Wage	0.003	-0.013	0.001	0.007	
	(0.016)	(0.010)	(0.020)	(0.014)	
Governor Party	0.008	-0.007	-0.002	-0.049*	
	(0.026)	(0.020)	(0.039)	(0.026)	
Log Tax Returns	0.874***	0.631***	0.217***	0.257***	-0.250***
	(0.009)	(0.006)	(0.014)	(0.010)	(0.032)
Share Below Poverty	1.407***	0.884***	1.037***	0.306*	-0.396
	(0.112)	(0.090)	(0.198)	(0.168)	(0.402)
Share Black	-0.110**	0.119**	-0.186***	-0.085	0.205
	(0.055)	(0.047)	(0.068)	(0.057)	(0.142)
Share <HS Education	-0.680***	-0.372***	-0.883***	-0.494***	-0.102
	(0.110)	(0.091)	(0.171)	(0.134)	(0.482)
Share Noncitizen	0.970***	1.549***	0.774***	1.256***	-0.173
	(0.138)	(0.128)	(0.201)	(0.145)	(0.364)
Share on Pub Assist.	0.533***	0.126***	0.112	0.087	-0.471***
	(0.064)	(0.044)	(0.111)	(0.086)	(0.175)
Share Est in Cons	-2.065***	-0.872***	0.025	-1.031***	0.264
	(0.114)	(0.065)	(0.169)	(0.140)	(0.274)
Share Est in Ag	-1.151***	-0.056	0.117	0.121	-1.149
	(0.405)	(0.170)	(0.798)	(0.570)	(1.116)
Share Est in Food Svcs.	0.760***	0.906***	-0.601**	0.000	-0.037
	(0.163)	(0.098)	(0.254)	(0.201)	(0.334)
Share Emp <50	-7.544***	-5.372***	-4.923***	-4.715***	2.784***
	(0.525)	(0.204)	(0.265)	(0.240)	(0.583)
Log Population Density	-0.011*	-0.002	-0.013	-0.003	0.042
	(0.006)	(0.005)	(0.009)	(0.006)	(0.027)
Unemployment Rate	-1.367***	-0.777***	-0.288	-0.581*	0.302
	(0.331)	(0.215)	(0.473)	(0.341)	(0.653)
State & Year FE	✓	✓	✓	✓	✓
N	215,761	215,761	63,408	63,408	63,408

Note: Standard errors are clustered at the state-year level. * = $p < 0.1$, ** = $p < 0.05$, *** = $p < 0.01$.

B EITC Rate Definitions

The values used for state EITC rates are discussed in Section III but bear further explanation. While many states use a simple matching rate, several use alternative structures. Here I review those alternatives and how I encoded values from the description:

California

California's credit has a smaller eligible income range than the federal credit, with a maximum which rose over the active years in my sample of analysis. The Tax Policy Center has summarized this divergence in recent years by listing the rate as a 45% match (rather than the statutory 85% figure). I follow this example and use the 45% rate in my set.

Maryland

Maryland employs two separate matching rates, one for refundable credits (used in the baseline) and one for non-refundable at 50%. To approximate the effective rate of this combination, I use the average of the varying statutory rate and 50%.

Minnesota

Minnesota's credit is different from other states in that it is not structured as a percentage of the federal credit. The Tax Policy Center reports an average rate of 33%, which I employ in the baseline. I do not adjust this rate for my preferred set, but simply report this usage for completeness.

New Jersey

New Jersey limits their match to a maximum income of $20,000. This is roughly half the total federal EITC range, though it is in the "plateau" region of the maximum credit. Accordingly, as an alternative to the statutory rate, I use three-quarters of that rate to reflect the diminished significance.

New York

New York City has its own supplemental 5% EITC. Given that roughly 40% of the state lives in the city, I adjust the statutory rate of 30% up to 32% in my preferred set.

Ohio

Ohio's credit is limited to 50% of liability for state taxable income above $20,000. I halve Ohio's rates to attempt and reflect this diminished significance.

Oregon

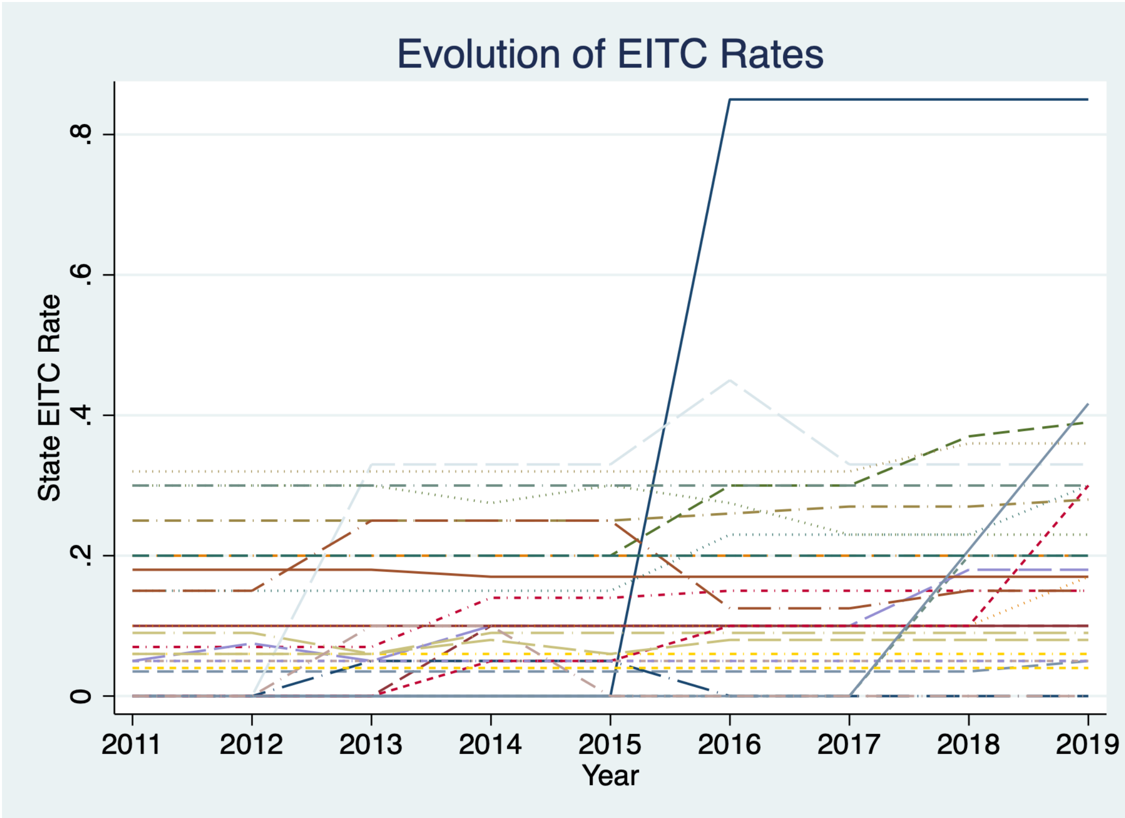
Oregon switched from a single rate for all filers to having a separate rate for filers with children under the age of three in 2017. Accordingly, I adjust the statutory rate of 8% up to 9% as my preferred rate, to reflect that a share of filers are receiving the new separate higher rate of 11%.

Wisconsin

Wisconsin's credit is only available to workers with qualified children, and uses separate rates by number of children, as one (4%), two (11%), and three (34%). To try and match this in an aggregate, I use the 4% figure in the statutory and adjust to 10% in the preferred rate.

Finally, the evolution of the statutory rates can be visualized below in Figure 4. While labelling every state is impractical, in anticipation of interest I note that the state which transitioned from no rate to 85% match in 2016 is California.

Figure 4: Evolution of Statutory Rates



C Cluster Size Heterogeneity

While clustering at the state-year level ensures that I have a relatively large number of clusters and allows for the appropriate level of error correlation, there is significant size heterogeneity among those clusters. This is due to the variation in number of underlying ZIP codes per state, which vary by an order of magnitude at the extremes (despite most being fairly similar in the several hundreds count). This kind of heterogeneity can lead to over-rejection in hypothesis testing, as the larger clusters will tend to have much higher leverage, meaning that asymptotic theory may not carry through reliably. To overcome this issue, I employ the wild score cluster bootstrap approach, as developed by Kline and Santos (2012) and recommended in this setting by MacKinnon et al. (2023).

Implementing this via the Stata command boottest (Roodman et al. 2019), I recreate the key portions of Tables 9 and 10 below with the significance levels calculated accordingly. The results are mostly unchanged, giving confidence that the cluster size heterogeneity is not driving inference misleadingly. The effect on reports is significant and negative, and the results for the first two secondary outcomes remain null. The one exception is in Column 5, for the Heckman of civil monetary penalties. The significance of the interaction term remains high, and as-argued this is in line with the model and counter to a mechanism of decreased employer intensity; however, the significance of the two individual terms is now also high. It is difficult to interpret this result, but it seems given the coefficient sizes that even if the significance is taken as legitimate, the overall effect of increased EITC rates would be positive, and thus would not align with a mechanism of decreased employer intensity.

Table 7: Main Results - Bootstrapped Significance

	Reports		Secondary Outcomes		
	(1)	(2)	(3)	(4)	(5)
	Complete Poisson	Probit	Empty Reports Poisson	Total Violations Poisson	Civil Penalties Heckman
State EITC Rate	0.680*	0.314	-0.163	0.131	-0.357***
Bootstrap P-value	(0.056)	(0.176)	(0.804)	(0.558)	(0.002)
EITC Participation Rate	-0.622***	-0.172**	-0.374	-0.061	-0.415**
Bootstrap P-value	(0.000)	(0.048)	(0.257)	(0.653)	(0.016)
State EITC Rate × EITC Participation Rate	-3.279***	-1.596**	0.293	-0.829	4.030***
Bootstrap P-value	(0.002)	(0.012)	(0.967)	(0.136)	(0.000)
N	218,947	218,947	66,421	66,421	66,421

Note: P-values are presented in parentheses, and were calculated using the wild score cluster bootstrap. * = $p < 0.1$, ** = $p < 0.05$, *** = $p < 0.01$.

D Complete Estimation Results

Table 8: Wage Theft Report - Full Results

	(1) Naive Poisson	(2) TWFE Poisson	(3) Complete Poisson	(4) ZIP FE Poisson	(5) Probit
State EITC Rate	-1.311***	-0.367***	-0.402***	-0.361***	-0.203**
	(0.283)	(0.120)	(0.136)	(0.133)	(0.102)
State Unionization Rate			-0.680	-0.577	-0.828
			(0.856)	(0.833)	(0.660)
State Minimum Wage			0.004	-0.000	-0.011
			(0.015)	(0.014)	(0.010)
Governor Party			0.011	0.010	-0.005
			(0.026)	(0.025)	(0.020)
Log Tax Returns			0.861***	0.194***	0.635***
			(0.009)	(0.070)	(0.005)
EITC Participation Rate			-0.908***	0.280	-0.337***
			(0.156)	(0.209)	(0.096)
Share Below Poverty			1.523***	0.306**	0.961***
			(0.105)	(0.136)	(0.088)
Share Black			-0.117*	0.390**	0.114**
			(0.068)	(0.184)	(0.050)
Share <HS Education			-0.431***	-0.167	-0.360***
			(0.108)	(0.188)	(0.092)
Share Noncitizen			0.866***	-0.078	1.618***
			(0.152)	(0.245)	(0.130)
Share on Pub Assist.			0.537***	0.037	0.125***
			(0.064)	(0.069)	(0.044)
Share Est in Cons			-2.348***	0.478*	-0.917***
			(0.123)	(0.289)	(0.066)
Share Est in Ag			-1.438***	0.297	-0.055
			(0.404)	(0.758)	(0.168)
Share Est in Food Svc.			0.592***	0.580**	0.900***
			(0.177)	(0.294)	(0.101)
Share Emp <50			-7.550***	-1.358***	-5.556***
			(0.559)	(0.355)	(0.208)
Log Population Density			-0.012*	0.040	-0.002
			(0.007)	(0.031)	(0.005)
Unemployment Rate			-1.413***	-0.017	-0.809***
			(0.348)	(0.282)	(0.217)
State & Year FE		✓	✓	ZIP & Year	✓
R^2	0.00	0.06	0.32	0.31	0.31
N	218,930	218,930	218,930	145,975	218,930

Note: Standard errors are clustered at the state-year level. * = p < 0.1, ** = p < 0.05, *** = p < 0.01.

Table 9: Interaction - Full Results

	(1) Complete Poisson	(2) Probit
State EITC Rate	0.680*	0.314
	(0.386)	(0.248)
EITC Participation Rate	-0.622***	-0.172*
	(0.125)	(0.089)
State EITC Rate × EITC Participation Rate	-3.279***	-1.596**
	(1.097)	(0.686)
State Unionization Rate	-0.636	-0.795
	(0.856)	(0.659)
State Minimum Wage	0.001	-0.012
	(0.014)	(0.010)
Governor Party	0.008	-0.007
	(0.025)	(0.020)
Log Tax Returns	0.862***	0.636***
	(0.009)	(0.005)
Share Below Poverty	1.481***	0.941***
	(0.108)	(0.089)
Share Black	-0.135**	0.105**
	(0.065)	(0.050)
Share <HS Education	-0.438***	-0.358***
	(0.114)	(0.092)
Share Noncitizen	0.882***	1.644***
	(0.153)	(0.128)
Share on Pub. Assist.	0.549***	0.130***
	(0.066)	(0.044)
Share Est in Cons	-2.336***	-0.908***
	(0.125)	(0.066)
Share Est in Ag	-1.413***	-0.071
	(0.400)	(0.170)
Share Est in Food Svcs.	0.630***	0.917***
	(0.182)	(0.102)
Share Emp <50	-7.510***	-5.534***
	(0.551)	(0.207)
Log Population Density	-0.009	-0.001
	(0.007)	(0.005)
Unemployment Rate	-1.368***	-0.797***
	(0.345)	(0.217)
State & Year FE	✓	✓
R^2	0.34	0.33
N	218,947	218,947

Note: Standard errors are clustered at the state-year level. * = p < 0.1, ** = p < 0.05, *** = p < 0.01.

Table 10: Intensive Margin - Full Results

	(1) Empty Reports Poisson	(2) Total Vltns. Poisson	(3) Civil Penalties Heckman
State EITC Rate	-0.163	0.131	-0.357
	(0.252)	(0.228)	(0.475)
EITC Participation Rate	-0.374**	-0.061	-0.415
	(0.184)	(0.137)	(0.318)
State EITC Rate × EITC Participation Rate	0.293	-0.829	4.030***
	(0.624)	(0.566)	(1.143)
State Unionization Rate	-1.422	0.299	0.843
	(1.394)	(0.910)	(2.186)
State Minimum Wage	0.005	0.013	
	(0.020)	(0.014)	
Governor Party	-0.002	-0.053**	
	(0.036)	(0.026)	
Log Tax Returns	0.255***	0.285***	-0.211***
	(0.013)	(0.011)	(0.030)
Share Below Poverty	1.279***	0.687***	-0.682*
	(0.193)	(0.169)	(0.372)
Share Black	-0.294***	-0.154**	0.364***
	(0.084)	(0.067)	(0.140)
Share <HS Education	-0.676***	-0.465***	0.356
	(0.175)	(0.145)	(0.370)
Share Noncitizen	0.966***	1.509***	0.403
	(0.170)	(0.137)	(0.361)
Share on Pub Assist.	0.192*	0.109	-0.374**
	(0.108)	(0.091)	(0.173)
Share Est in Cons	-0.399**	-1.428***	0.502*
	(0.190)	(0.160)	(0.287)
Share Est in Ag	-0.202	0.032	0.525
	(0.863)	(0.601)	(1.040)
Share Est in Food Svcs.	-0.772***	-0.166	0.168
	(0.274)	(0.229)	(0.334)
Share Emp <50	-5.511***	-5.288***	2.862***
	(0.277)	(0.245)	(0.613)
Log Population Density	-0.023**	-0.010	-0.008
	(0.010)	(0.006)	(0.023)
Unemployment Rate	-0.493	-0.765**	0.313
	(0.480)	(0.370)	(0.651)
State & Year FE	✓	✓	✓
N	66,421	66,421	66,421

Note: Standard errors are clustered at the state-year level. * = $p < 0.1$, ** = $p < 0.05$, *** = $p < 0.01$.

E Event Study Figures

The ATT estimates from additional specifications of the event study analysis are presented below in figures.

Figure 5: Event Study ATT Estimates - State Treatment

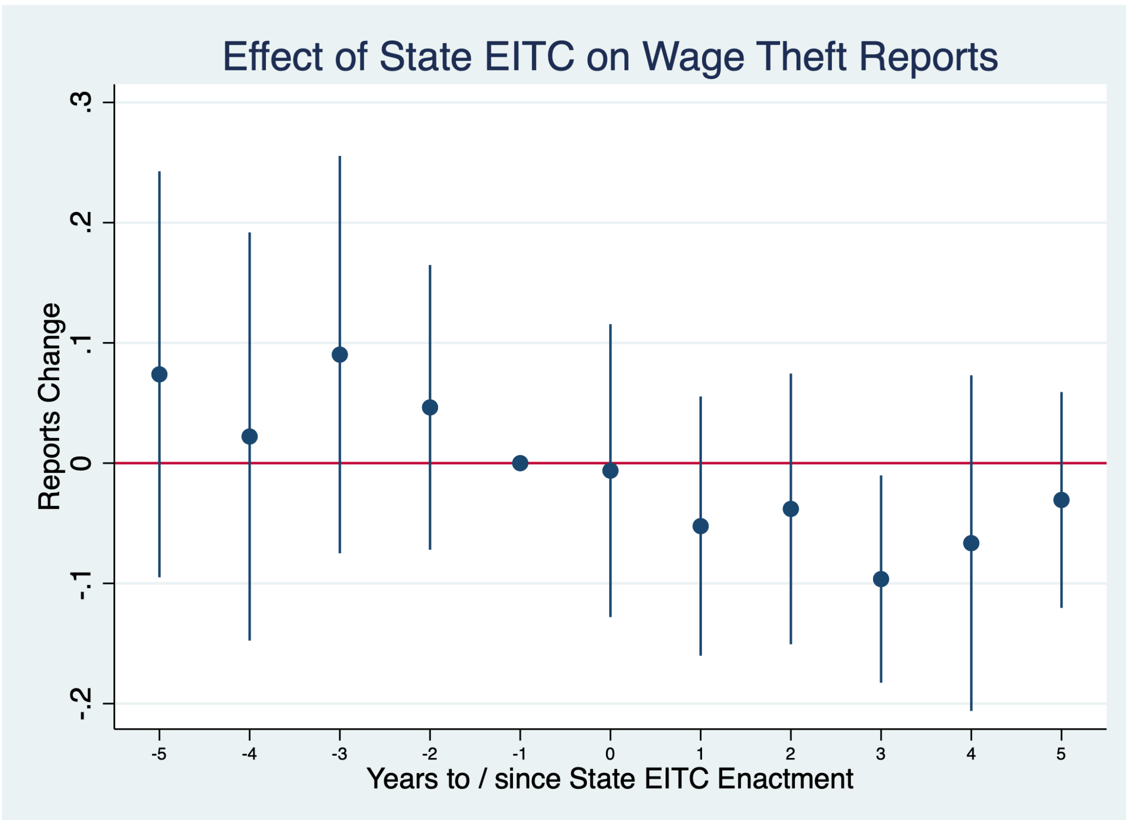


Figure 6: Event Study ATT Estimates - ZIP Treatment, Always-Treated Included

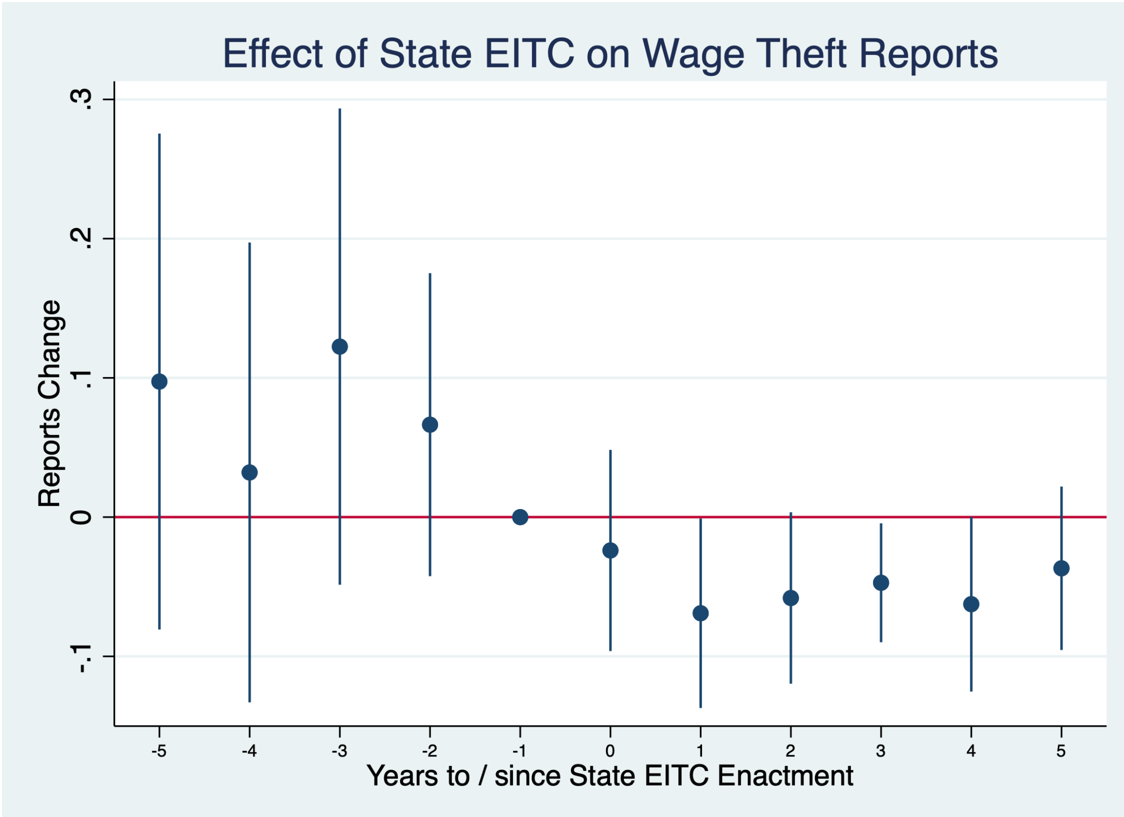


Figure 7: Event Study ATT Estimates - State Treatment, Always-Treated Included