

When Does Tax Monitoring Matter? Evidence from Brazil's Large-Taxpayer Program

Vagner Marques

vagner.marques@ufes.br

Universidade Federal do Espírito Santo

Research Article

Keywords: Tax Monitoring, Corporate Taxation, Large Taxpayers, Staggered Difference-in-Differences, Deterrence Theory, Brazil

Posted Date: October 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10942730/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

When Does Tax Monitoring Matter? Evidence from Brazil's Large-Taxpayer Program

Vagner Antônio Marques^{1*}

^{1*}Department of Accounting, Universidade Federal do Espírito Santo (UFES), Vitória, ES, Brazil.

Corresponding author(s). E-mail(s): vagner.marques@ufes.br;

Abstract

Does escalating enforcement intensity change firms' realized corporate tax? I provide causal evidence from Brazil's Receita Federal large-taxpayer monitoring program, which separates a broad differentiated classification tier from a higher-intensity special tier. Entry into differentiated monitoring produces no detectable average change across four tax outcomes, a bounded null consistent with administrative classification functioning as screening rather than enforcement. Transitioning from differentiated to special monitoring raises the total tax burden by 10.76 percentage points and the benchmark-adjusted effective tax rate (ETR) by 17.26 percentage points above the industry-year median. The effects that exceed by an order of magnitude the two-percentage-point response to marginal audit probability increases documented in the U.S. context and span the full tax base, including income, consumption, and payroll taxes. Using 311 publicly listed firms from 2010 to 2019 and Staggered Difference-in-Differences models following [Callaway and Sant'Anna \(2021\)](#), the paper provides the first causal evidence separating administrative classification from enforcement escalation in an emerging-market large-taxpayer program; documents that escalation to intensive monitoring generates economically large compliance effects while classification alone does not, and shows that heterogeneous responses concentrate among firms with higher pre-entry fiscal pressure, consistent with the information-traceability mechanism. The findings have direct implications for enforcement resource allocation, tax-risk assessment, and after-tax earnings forecasts.

Keywords: Tax Monitoring, Corporate Taxation, Large Taxpayers, Staggered Difference-in-Differences, Deterrence Theory, Brazil

JEL Classification: H26 , H25 , H32 , H83

1 Introduction

The corporate income tax backstops the personal income tax by preventing high-income individuals from reclassifying labor income as retained profits (Slemrod, 2004), taxes economic rents that would otherwise escape the personal tax base (Mintz, 1995), and contributes to the perceived fairness of tax systems where corporate equity ownership is concentrated (Fuest & Neumeier, 2023). Protecting the corporate tax base is therefore a first-order policy priority, particularly as globalization increases tax competition and complicates profit attribution across jurisdictions.

Firms have structural incentives to minimize their tax liabilities, and the resulting compliance gap imposes significant fiscal and social costs (Carnahan, 2015; Okunogbe & Tourek, 2024). Enforcement closes that gap through credible monitoring, third-party information systems, and strategic audit targeting (Plumley, 2002; Pomeranz, 2015). Because the government holds a de facto equity claim on corporate profits, tax authority monitoring also disciplines managerial behavior beyond the tax return. Hanlon, Hoopes, and Shroff (2014) find that higher IRS enforcement improves financial reporting quality, and Mason and Williams (2022) show it reduces the likelihood that managers initiate accounting fraud.

Resource-constrained tax authorities respond by concentrating enforcement on the largest taxpayers, a strategy endorsed by the IMF and OECD (Benon, Baer, & Toro R., 2002). A small number of large firms account for most corporate tax collections, making their compliance behavior disproportionately important to aggregate revenue. Almunia and Lopez-Rodriguez (2018) demonstrate that firms in Spain bunch below the Large Taxpayers Unit eligibility threshold to avoid stricter monitoring, with compliance responses strongest where transactions leave a clearer paper trail. In Brazil, the Receita Federal do Brasil (RFB) has operated a Differentiated monitoring program since 2010, with regulatory ordinances setting selection thresholds based on gross revenue, tax liabilities, payroll, and contribution to aggregate federal tax collection (Martini, 2022; Oliveira, 2022). Section 2 documents the evolution of these thresholds across the ordinances that governed the sample period, and formally defines the two treatment variables used in the empirical design.

The RFB program distinguishes two tiers: a broad Differentiated screen that identifies fiscally relevant firms and a higher-intensity Special tier reserved for the largest and most complex taxpayers. That design creates an observable institutional sequence. Firms enter Differentiated monitoring when they cross lower fiscal-relevance thresholds; a smaller group later transitions to Special monitoring after meeting substantially stricter criteria. Despite the program’s institutional significance, rigorous microeconomic evidence on its causal effect on firms’ realized tax burden remains absent. This paper fills that gap.

Monitoring effects on firms’ realized tax positions operate through complementary channels, each yielding distinct predictions for the Brazilian two-tier design. Deterrence theory predicts that monitoring raises compliance when firms perceive a higher probability of detection and a higher expected cost of aggressive tax positions (Devos,

2007; Hoopes, Mescall, & Pittman, 2012; Scholz, 1997). Information-based enforcement predicts stronger effects where transactions and book-tax positions leave observable signals for the tax authority (Almunia & Lopez-Rodriguez, 2018; Mills, 1998; Pomeranz, 2015). The government-as-monitor mechanism predicts that tax authority scrutiny constrains managerial reporting choices beyond the tax return (Hanlon et al., 2014; Mason & Williams, 2022). The empirical design separates two institutionally distinct margins: (H_1) entry into Differentiated monitoring; and (H_2) transition from Differentiated to Special monitoring within the post-Differentiated risk set.

Prior evidence rarely separates broad administrative classification from subsequent escalation to a higher-intensity tier. Studies of IRS audit probabilities document that enforcement reduces tax avoidance, shapes disclosure behavior, and disciplines financial reporting (Ehinger, Lee, Stomberg, & Towery, 2025; Hanlon et al., 2014; Hoopes et al., 2012; Mason & Williams, 2022). Studies of large-taxpayer units show that size-based monitoring changes reporting incentives around thresholds (Almunia & Lopez-Rodriguez, 2018). The Brazilian two-tier design permits the separation those studies cannot make.

Using 311 firms from 2010 to 2019, I estimate group-time average treatment effects via the doubly robust estimator of Callaway and Sant’Anna (2021), which accommodates heterogeneous treatment effects across cohorts and avoids the two-way fixed-effects (TWFE) biases documented in staggered-adoption settings (Baker, Callaway, Cunningham, Goodman-Bacon, & Sant’Anna, 2026; Borusyak, Jaravel, & Spiess, 2024). The paper contributes along three fronts. Causally, it provides the first evidence separating administrative classification from enforcement escalation in an emerging-market large-taxpayer program, extending the evidence base of Hoopes et al. (2012) and Almunia and Lopez-Rodriguez (2018) to a new institutional setting. Empirically, it documents heterogeneous effects across pre-treatment fiscal pressure and potential tax exposure, with all splits defined before monitoring entry to avoid post-treatment conditioning. Methodologically, it supplements the main DiD estimates with pre-treatment life-cycle controls (Dickinson, 2011) and a regression discontinuity design (RDD) around the statutory revenue threshold, providing local corroboration and a direct test for bunching.

Section 2 describes the Brazilian large-taxpayer monitoring program and formally defines the two treatment events. Section 3 develops the theoretical framework and Section 4 describes the data, sample, and empirical strategy. Section 5 presents the results. Section 6 concludes.

2 Institutional Setting

2.1 Brazil’s Large-Taxpayer Monitoring Program

The RFB’s structured large-taxpayer monitoring program, introduced in 2010, is governed by a series of regulatory ordinances (*portarias*) that set explicit revenue and fiscal-liability thresholds for program entry and escalation. The program identifies two monitoring tiers: a broad *Differentiated* tier for firms with fiscal relevance above a lower revenue threshold, and a *Special* tier reserved for the largest and most fiscally complex taxpayers.

Each ordinance specifies the thresholds that govern the *following* calendar year, using data from the second year prior to the monitoring year as the reference base.¹ Formally, the RFB evaluates a firm’s Q4 annual revenue from year $t - 2$ against the thresholds in force for monitoring year t .

Once classified, firms remain in the program regardless of subsequent revenue changes, because the ordinances establish that firms remain in their current monitoring condition until a superseding act establishes new criteria.² Treatment is therefore absorbing.

2.2 Threshold Schedule

Table 1 reports the annual threshold schedule for the Differentiated and Special tiers from 2010 to 2023, documenting three distinct regulatory eras. During 2010–2014, governed by Port. n° 2.356/2010, thresholds were set with modest adjustments. Port. n° 641/2015 raised the Special threshold to R\$1 billion. From 2019 onward, Port. n° 2.176/2018 and its successors maintained the Special threshold at R\$1 billion until Port. n° 252/2022 raised it to R\$2 billion for the 2023 monitoring year.

¹The two-year lag is codified across multiple ordinances: Port. RFB n° 5.018/2020, Art. 2° §3°; Port. RFB n° 252/2022, Art. 4°; Port. RFB n° 390/2023, Art. 4°.

²Port. RFB n° 2.176/2018, Art. 5°: “*As pessoas jurídicas submetidas ao monitoramento diferenciado ou especial durante o ano de 2019 permanecerão nessa condição durante os anos subsequentes até que ato normativo superveniente estabeleça novos critérios de indicação.*” Identical absorbing-status language appears in Port. RFB n° 2.135/2019, Art. 1° §2°.

Table 1: Evolution of RFB Large-Taxpayer Monitoring Criteria (2010–2023)

Criterion	Ord. 2.36/2010	Ord. 641/2015	Ord. 4.89/2020	Tier
<i>Panel A — Differentiated Monitoring</i>				
Gross revenue (R\$ million/year)	≥ 90	≥ 90	≥ 75	Differentiated
Federal tax debt (R\$ million)	≥ 10	≥ 10	≥ 10	Differentiated
Payroll (R\$ million/year)	≥ 13	≥ 13	≥ 13	Differentiated
Share of federal tax collection (%)	≥ 0.01	≥ 0.01	≥ 0.01	Differentiated
<i>Panel B — Special Monitoring</i>				
Gross revenue (R\$ billion/year)	≥ 1.50	≥ 1.50	≥ 1.50	Special
Federal tax debt (R\$ billion)	≥ 1.00	≥ 1.00	≥ 1.00	Special
Foreign-trade value (R\$ billion/year)	–	–	≥ 1.00	Special

Notes: Thresholds apply to the prior calendar year. Firms may qualify under any single criterion. Ordinance n° 4.89/2020 (effective January 2021) reduced the Differentiated revenue threshold from R\$90m to R\$75m, added foreign-trade exposure as a Special criterion, and extended monitoring to tax-substitution participants. The sample period (2010–2019) predates this structural revision. Sources: [Martini \(2022\)](#); [Oliveira \(2022\)](#).

2.3 Two Institutionally Distinct Treatment Events

The two-tier architecture generates two economically and institutionally distinct causal margins, which define the treatment variables in the empirical design.

T1 — Differentiated entry (G_i^D). The first period in which firm i becomes eligible for Differentiated monitoring. Eligibility is determined mechanically: the firm’s Q4 annual revenue from year $t-2$ exceeds the Differentiated lower bound $\underline{c}_t$ in force for monitoring year t . This margin captures the transition from unmonitored to classified (the extensive margin of entry), dated at Q1 of the first monitoring year for which the condition is met:

$$G_i^D = \min\{t : \text{Rev}_{i,Q4(t-2)} > \underline{c}_t\}. \quad (1)$$

T2 — Transition to Special monitoring (G_i^{SD}). The first period in which firm i , already under Differentiated monitoring, becomes eligible for Special monitoring. The condition is met in Q4 of year s when Q4 revenue from year $s-2$ exceeds

the Special threshold $\bar{c}_{s+2}$ for monitoring year $s + 2$; the transition event is dated at Q1 of year $s + 2$:

$$G_i^{S|D} = \min\{t : G_i^D < t, \text{Rev}_{i,Q4(t-2)} > \bar{c}_t\}, \quad (2)$$

with $G_i^{S|D} = \infty$ (coded as zero in the panel) for firms that never cross the Special threshold after Differentiated entry. This margin captures the intensive escalation from administrative classification to full cross-base enforcement scrutiny.

These two margins are not interchangeable. T1 identifies the effect of entering the monitoring information set; T2 identifies the incremental effect of escalation to intensive scrutiny, conditional on prior Differentiated status. The sample contains 14 firms that make the transition (T2), compared with 123 firms classified into Differentiated monitoring (T1). Firms that enter the Special tier directly without a prior Differentiated spell (9 firms) are excluded from the T2 estimand and reported as a supplementary benchmark.

Table 2 summarises cohort sizes, pre-treatment windows, and the event-time support available for each treatment margin.

Table 2: Staggered DiD cohort support for G_i^D and $G_i^{S|D}$

Cohort	Firms i	Pre- t support	Post- t support	Estimable	Min T_i	Status / reason if not estimated
<i>Panel A: H1 – Differentiated entry</i>						
Never/not-yet ($g = 0$)	188	0	0	0	1	Comparison cohort; not a treated ATT cohort.
$g = 17$	46	46	46	46	4	Estimable.
$g = 21$	6	6	6	6	5	Estimable.
$g = 25$	26	26	26	26	4	Estimable.
$g = 29$	7	7	7	7	3	Estimable.
$g = 33$	9	9	9	9	9	Estimable.
$g = 37$	23	23	23	23	8	Estimable.
$g = 41$	2	2	2	2	18	Estimable.
$g = 45$	4	4	4	4	5	Estimable.
<i>Panel B: H2 – Transition to Special</i>						
Never/not-yet ($g = 0$)	109	0	0	0	1	Comparison cohort; not a treated ATT cohort.
$g = 21$	2	2	2	2	11	Estimable.
$g = 29$	2	2	2	2	7	Estimable.
$g = 37$	1	1	1	1	14	Estimable.
$g = 41$	8	8	8	8	9	Estimable.
$g = 45$	1	1	1	1	23	Estimable.

Notes: Cohort support for Callaway–Sant’Anna (2021) $ATT(g, t)$ estimation. Firms are indexed by i and quarters by t ; sector j is used only in sector-stratified analyses. G_i^D denotes first entry into Differentiated monitoring; $G_i^{S|D}$ denotes first Special-monitoring quarter after Differentiated entry. Pre- t and post- t support count firms with at least one observed quarter before or on/after the cohort treatment period. The H2 panel is computed within the Differentiated risk set.

3 Literature Review and Hypotheses

3.1 Corporate Taxation and the Role of Enforcement

Three overlapping rationales motivate taxing corporations: the corporate tax serves as a benefit tax, backstops the personal income tax by preventing income reclassification, and captures economic rents that would otherwise escape the personal tax base (Mintz, 1995). Cross-country evidence supports the backstop rationale: countries with higher personal income tax rates maintain higher corporate rates, consistent with the two taxes functioning as a system (Slemrod, 2004). A fourth rationale centers on progressivity, given that corporate equity ownership is concentrated among high-income households (Fuest & Neumeier, 2023).

Enforcement amplifies these rationales by shrinking the compliance gap. Plumley (2002) estimates that IRS activities generate measurable indirect effects on voluntary compliance across the full taxpayer distribution, not only among audited firms. Nessa, Schwab, Stomberg, and Towery (2020) show that IRS resource constraints reduce both audit incidence and deficiency magnitude, with the resulting revenue losses exceeding the budget savings from cutting enforcement. Enforcement is therefore not merely a cost of tax administration but a net revenue-generating input (Johns & Slemrod, 2010).

Realized tax burden is a more informative outcome than statutory rates in enforcement research because it integrates audit intensity, settlement risk, and the interaction between financial and tax reporting. Jacob (2022) argues that causal identification in corporate tax research requires settings where enforcement variation can be linked to firm-level behavior. The Brazilian large-taxpayer program provides that variation through observable threshold-crossing events before and after monitoring escalation.

Analyzing multiple tax outcomes matters because enforcement affects several reporting margins simultaneously: income taxes, value-added taxes, payroll taxes, and other bases each respond to monitoring intensity (Almunia & Lopez-Rodriguez, 2018). The outcome architecture in this study reflects that breadth. ETR captures the income-tax channel; TB captures the total fiscal burden borne through Brazil's multi-tax system; AbnETR and AbnTB add industry-year benchmarks to isolate firm-specific changes from sector-level shocks.

3.2 Deterrence Theory and Monitoring

The Brazilian two-tier design generates distinct predictions across the deterrence, governance, information-traceability, and continuous-monitoring channels.

Deterrence. The baseline mechanism holds that compliance responds to the expected cost of noncompliance, defined by the probability of detection and the magnitude of sanctions (Allingham & Sandmo, 1972; Benon et al., 2002). Scholz (1997) extends this baseline by identifying bounded rationality as an additional channel: inspections are effective not primarily through their monetary cost but through the managerial attention they force firms to pay to overlooked risks, an extension directly relevant when firms manage large and complex tax positions. Devos (2007) adds that compliance responds to perceptions of detection risk rather than objective audit

probabilities, and [Ariel \(2012\)](#) cautions that communication-based deterrence alone is insufficient without substantive monitoring effort. [Hoopes et al. \(2012\)](#) provide the empirical benchmark: raising audit probability from the 25th to the 75th percentile increases cash effective tax rates by nearly two percentage points.

Continuous monitoring versus one-off audits. The RFB regime differs structurally from the occasional audits studied by [Hoopes et al. \(2012\)](#) and [DeBacker, Heim, Tran, and Yuskavage \(2015\)](#). A one-off audit updates beliefs about future audit risk and produces a U-shaped dynamic response ([DeBacker et al., 2015](#)). Continuous monitoring instead changes the firm’s recurring information environment, so managers know the authority can compare filings over time and across tax bases, eliminating the re-adjustment phase that generates the U-shape. Unlike the transient response to one-off audits ([DeBacker et al., 2015](#)), sustained monitoring permanently shifts the firm’s equilibrium tax position because managers cannot readily reverse compliance adjustments made under continuous cross-base scrutiny ([Allingham & Sandmo, 1972](#)). $ATT(g, t)$ estimates are therefore expected to remain stable or deepen from the first post-entry period.

Governance spillovers. The government’s equity claim on corporate profits gives the tax authority an incentive to monitor managerial behavior independently of shareholders. [Hanlon et al. \(2014\)](#) exploit that logic to show that higher IRS enforcement improves financial reporting quality, with the effect strongest where other governance mechanisms are weakest. [Mason and Williams \(2022\)](#) extend this to accounting fraud: firms facing greater IRS monitoring initiate fraud at lower rates. For the RFB’s Special tier, which increases authority access to firm records and enables cross-matching across federal, state, and municipal tax obligations, the governance channel constrains not only tax-planning opportunities but also the scope for managerial rent extraction that would otherwise reduce taxable income.

Information traceability. Monitoring effort and information traceability are complements rather than substitutes ([Pomeranz, 2015](#)): randomized experiments with over 400,000 Chilean firms show that announcing additional monitoring has less compliance impact on transactions already covered by a paper trail. [Mills \(1998\)](#) provides direct support in the audit context: larger book-tax differences predict larger IRS audit adjustments, consistent with authorities using observable signals to target positions with greater reporting risk. The RFB’s multi-base cross-matching makes this channel particularly salient for Special-tier firms.

These four mechanisms generate a tier-specific prediction and guide the heterogeneity design. Fiscal pressure, measured from pre-treatment ETR, identifies where the deterrence effect should be largest. Potential Tax Exposure indexes pre-treatment signals available to the RFB. Life-cycle classification follows [Dickinson \(2011\)](#) using pre-treatment cash-flow signs, avoiding any conditioning on post-treatment outcomes.

3.3 Hypotheses

The deterrence, governance, and information-traceability mechanisms each predict that monitoring effects should increase with tier intensity. The Differentiated tier identifies fiscally relevant firms and creates administrative visibility, but its institutional

function is primarily screening and classification. Under the bounded-rationality extension (Scholz, 1997), that classification may not yet impose the managerial attention necessary to change tax positions. The first hypothesis therefore allows for a meaningful effect or a bounded null, with the null predicting that entry-tier classification functions as screening without generating a discrete enforcement shock:

H₁: Entry into Differentiated monitoring changes firms’ realized effective tax burden.

The Special tier assigns firms to intensive cross-base monitoring, with greater authority access to records and higher expected cost of aggressive positions. Deterrence theory predicts a positive compliance response (Hoopes et al., 2012); the governance channel predicts a constraint on managerial rent extraction (Hanlon et al., 2014); and the traceability channel predicts amplified effects where firms carry larger observable tax bases (Mills, 1998; Pomeranz, 2015). All three mechanisms converge on a positive prediction for the escalation margin:

H₂: Transition from Differentiated to Special monitoring increases firms’ realized effective tax burden above the level attained under Differentiated monitoring.

4 Methods

4.1 Data and Sample

The analysis draws on quarterly financial data from CVM Open Data (Dados Abertos da Comissão de Valores Mobiliários, Brazil’s securities and exchange regulator) covering Brazilian publicly listed firms from 2010 Q1 to 2019 Q4. This period covers the first decade of the RFB large-taxpayer monitoring program and provides sufficient pre-treatment observations to assess the parallel trends assumption. The starting panel contains the universe of available listed-firm ticker-quarter observations with the variables needed to construct the treatment, outcomes and controls. The sample is then reduced for reasons that follow directly from the research design. Ticker-quarter records that duplicate the same firm through consolidations, corporate restructuring or repeated vendor identifiers are removed, since retaining them would artificially inflate the treated group. Firm-quarters that lack the accounting inputs for a given outcome enter only those models for which the inputs exist. ETR requires positive EBIT as a meaningful denominator, so negative-EBIT observations are excluded from ETR regressions while remaining available for TB and benchmark-adjusted outcomes. The H_2 transition design additionally restricts to firm-periods observed after first Differentiated entry, because only firms already under monitoring can escalate to the Special tier. These restrictions convert the gross potential panel of 12,440 firm-quarters into a final deduplicated panel of 6,044 firm-quarters. Table 3 details the attrition.

Financial firms and utilities are not discarded mechanically. They are kept when the required variables are observed, but their distinct regulatory environments are examined in supplementary sector analyses. This choice avoids losing economically important monitored firms while still acknowledging that regulated sectors may have different tax bases, accounting rules and enforcement exposure.

None of these exclusions reflects discretionary sample trimming. Each follows from measurement requirements or from the institutional structure of the monitoring design, consistent with established criteria for panel data in corporate finance research (Coles, Daniel, & Naveen, 2008; Dickinson, 2011; Linck, Netter, & Yang, 2008). Large-taxpayer programs target economically important firms, including regulated entities, and excluding them mechanically would narrow the external relevance of the setting.

Table 3: Sample construction for the two-stage monitoring design

Sample step	Firm-quarters	Firms	Periods
Final deduplicated listed-firm panel	6,044	311	40
Firms ever entering Differentiated monitoring	3,062	123	38
Firm-periods in the Differentiated-to-Special risk set	2,104	123	32
Firms transitioning from Differentiated to Special	242	14	32
Firms remaining Differentiated without observed Special transition	1,862	109	32

Notes: The transition risk set contains firm-periods observed after first entry into Differentiated monitoring. Firms transitioning to Special define the H2 treated group; firms not yet Special or never observed as Special provide the comparison observations within the risk set.

Table 4 defines all variables used in the analysis. The sample spans 311 firms and 40 quarterly periods (2010 Q1 to 2019 Q4), yielding 6,044 firm-quarter observations. For T1 (Differentiated entry), 123 firms are estimable treated units across eight quarterly cohorts. For T2 (Transition to Special), 14 firms are estimable treated units across five estimable cohorts after applying the minimum group-size filter described in Section 2. The distribution of treatment timing is highly staggered: the largest H1 cohort ($g = 17$, 2014-Q1) contains 46 firms, while the remaining cohorts range from 2 to 26 firms. This staggering is institutionally driven because the RFB applies threshold criteria annually against prior-year revenue data, so firms enter the monitoring program as their revenue crosses the applicable threshold, not through any centralized assignment.

Table 4: Variable Definitions and Data Sources

Variable	Symbol	Definition	Source
<i>Outcome variables</i>			
Effective tax rate	ETR	Income tax charges (IRPJ + CSLL) divided by EBIT. Defined only for EBIT > 0.	CVM Open Data
Tax burden	TB	Total tax charges across all government levels divided by gross value added (DVA).	CVM Open Data
Abnormal ETR	AbnETR	ETR minus the industry-year median ETR ($\widehat{ETR}_{j,t}$).	Computed
Abnormal TB	AbnTB	TB minus the industry-year median TB ($\widehat{TB}_{j,t}$).	Computed
<i>Treatment variables</i>			
Differentiated entry	G_i^D	First quarter firm i crosses the gross-revenue threshold for Differentiated monitoring.	RFB Ordinances
Special transition	$G_i^{S D}$	First Special quarter after prior Differentiated entry (defined only in H_2 risk set).	RFB Ordinances
<i>Control variables</i>			
Firm size	Size	Natural log of total assets ($\ln AT_{it}$).	CVM Open Data
Leverage	Lev	Total debt (current + non-current liabilities) divided by total assets.	CVM Open Data
Profitability	ROA	Earnings before income tax divided by total assets ($LAIR_{it}/AT_{it}$).	CVM Open Data
Tangibility	Tang	Net PP&E divided by total assets ($IMOB_{it}/AT_{it}$).	CVM Open Data
Big-4 auditor	Big4	Indicator equal to 1 if external auditor is one of the four largest global audit firms.	CVM Open Data
Life-cycle stage	LCS	Pre-treatment modal Dickinson (2011) cash-flow stage (introduction, growth, mature, shake-out, decline).	CVM Open Data
<i>Heterogeneity / moderator variables</i>			
Fiscal pressure	FP	Median pre-treatment ETR; firms split at sample median (predetermined).	Computed
IFRS 15 period	IFRS15	Indicator equal to 1 for fiscal years 2018–2019 (post-mandatory adoption in Brazil).	Calendar
Potential tax exposure	PET	First principal component of z-scored pre-treatment IMOB, RB2 and LAIR (PCA-validated).	Computed

Notes: All continuous controls are winsorised at the 1st and 99th percentiles within each year. Pre-treatment moderators use information available before first monitoring entry. The PET index is validated via Bartlett sphericity test ($p = 0$), KMO = 0.71, and PC1 explaining 80.40% of total variance with all three inputs loading in the same direction.

The treated group (Differentiated or Special monitoring) is disproportionately large and more likely to have Big-4 auditors than the comparison group, consistent with the size-based selection rule. This pre-treatment imbalance is expected rather than accidental because large-taxpayer monitoring is explicitly size based. It motivates the doubly robust CS estimator and the supplementary entropy-balancing and PSM robustness checks reported in the Online Resource (Appendix E). The four outcome variables capture complementary dimensions of the firm’s fiscal position. ETR and TB measure absolute levels. AbnETR and AbnTB measure deviations relative to industry-year benchmarks, isolating idiosyncratic changes from sector-level trends.

Table 5 presents descriptive statistics for the full panel. The mean ETR across all firm-quarters is 0.239, with considerable dispersion ($SD = 0.186$). Mean leverage is 0.584, reflecting the high debt-to-assets ratios typical among Brazilian listed firms. Pre-treatment mean differences confirm the overlap challenge: for H_1 , treated firms have a mean log total assets of 21.14 vs. 22.82 for comparison firms (adjusted difference = -1.58 , $p < .001$), and a Big-4 auditor rate of 79% vs. 88% ($p = .028$); the full set of period-adjusted pre-treatment comparisons for both hypotheses is reported in the supplementary material. These imbalances are expected given the size-based monitoring rule and do not threaten identification provided conditional parallel trends hold after controlling for the observed covariates in X_{it} .

Table 5: Descriptive statistics for the full analysis sample

Variable	Description	N	Mean	SD	P25	P75
<i>Outcome variables</i>						
ETR_{it}	Effective tax rate	6,044	0.24	0.15	0.12	0.34
TB_{it}	Total tax burden	6,044	0.27	0.16	0.16	0.37
$AbnETR_{it}$	Benchmark-adjusted effective tax rate	5,910	0.02	0.18	-0.02	0.02
$AbnTB_{it}$	Benchmark-adjusted total tax burden	5,858	0.01	0.07	-0.01	0.01
<i>Control variables</i>						
$Size_{it}$	Log total assets	6,044	21.80	1.62	20.52	22.91
Lev_{it}	Leverage	6,042	0.67	0.66	0.47	0.73
ROA_{it}	Pre-tax return on assets	6,044	0.04	0.16	-0.01	0.08
$Tang_{it}$	Tangibility	5,676	0.27	0.21	0.07	0.40
$Big4_{it}$	Big-4 auditor indicator	6,044	0.75	0.43	0	1
<i>Treatment and moderator variables</i>						
D_{it}^D	Differentiated-monitoring indicator	6,044	0.35	0.48	0	1
$D_{it}^{S D}$	Differentiated-to-Special transition indicator	6,044	0.02	0.15	0	0
PET_i	Potential tax exposure (pre-treatment PC1 score)	311	0.00	1.60	-0.54	-0.18
FP_i	Fiscal pressure measured before treatment	311	0.24	0.12	0.15	0.31

Note. Statistics are computed from the analysis panel for 2010–2019. Subscript i indexes firms and t indexes quarters. Outcomes, controls, and treatment indicators are summarized at the firm-quarter level; PET_i and FP_i are predetermined firm-level measures and therefore use one observation per firm. Indicator-variable quartiles are reported without decimal places.

4.2 Identification Strategy

The empirical strategy follows a design-based quasi-experimental logic. The two treatment variables are formally defined in Section 2. For H_1 , G_i^D (Equation 1) denotes the first period of Differentiated entry and $D_{it}^D = \mathbf{1}(t \geq G_i^D)$. For H_2 , the sample is restricted to firms already in the Differentiated risk set (i.e., with $G_i^D > 0$). Within this risk set, $G_i^{S|D}$ (Equation 2) denotes the first transition to Special monitoring and $D_{it}^{S|D} = \mathbf{1}(t \geq G_i^{S|D})$. Firms entering Special directly without a prior Differentiated spell (9 firms) are excluded from H_2 and reported as a supplementary benchmark only. Both treatment indicators are absorbing, consistent with the ordinance-level provisions documented in Section 2. Table 6 summarises the estimand architecture.

Table 6: Two-stage institutional estimands and comparison groups

Estimand		Sample	Cohort	Comparison group	Interpretation
H1	Differentiated entry	Full listed-firm panel	G_i^D : first Differentiated-monitoring period	Not-yet-Differentiated firms	Extensive-margin effect of entering the broad large-taxpayer monitoring tier.
H2	Transition to Special	Firm-periods after first Differentiated entry	$G_i^{S D}$: first Special period after Differentiated entry	Differentiated firms not yet Special, plus Differentiated firms never observed as Special	Escalation effect of moving from broad monitoring to the higher-intensity Special tier.
Supplement:	direct Special entry	Full listed-firm panel	G_i^S : first Special-monitoring period	Not-yet-Special firms	Legacy benchmark only; not the main institutional estimand.

Notes: Firms are indexed by i and quarters by t ; sector j is used in sector-stratified analyses. The transition design is restricted to firms already at risk of escalation because they have entered Differentiated monitoring.

The identifying parameter is the group-time average treatment effect on the treated, $ATT(g, t)$. For H_1 , treated firms are compared with firms that have not yet entered Differentiated monitoring at period t . For H_2 , treated firms are compared with firms already in the Differentiated risk set that have not yet transitioned to Special monitoring or are never observed transitioning during the sample. This choice follows the clean-comparison logic in modern staggered DiD (Baker et al., 2026; Callaway & Sant’Anna, 2021). It avoids using already-treated firms as controls for later-treated firms, which can bias TWFE estimates when effects are dynamic or heterogeneous.

The design rests on three assumptions. Parallel trends requires that treated and comparison firms would have followed similar tax trajectories absent monitoring, conditional on observed covariates. No anticipation requires that firms do not alter reported tax positions before the formal monitoring event. The stable unit treatment value assumption (SUTVA) requires that one firm’s monitoring status does not mechanically alter another firm’s outcome. I assess these assumptions through event-study pre-trends, treatment-support diagnostics, placebo checks, sector concentration checks and supplementary matching or weighting diagnostics. These tests do not prove the assumptions, but they make the identifying conditions explicit and auditable.

This implementation follows best practice for staggered adoption settings. The literature warns that TWFE regressions can combine treated units with already-treated controls and produce misleading estimates when treatment effects vary over cohorts or time. The group-time approach instead starts from interpretable $ATT(g, t)$ parameters, uses not-yet-treated or not-yet-escalated comparison firms, and then aggregates only after the cohort-time effects are estimated.

Identification caveat for H_1 and H_2 . Because the revenue proxy captures only one dimension of the multi-criterion RFB screen, the H_1 null admits two interpretations: (i) the broad Differentiated tier genuinely produces no average compliance effect, or (ii) the proxy attenuates the true effect toward zero. Given 123 treated firms and a standard error of 0.0185, the minimum detectable effect at 80% power is approximately 2.6 percentage points. The bounded null should be read alongside the H_2 estimates. Because the official RFB monitoring list is confidential, treatment status is reconstructed using the publicly observable criteria established by the relevant ordinances. The estimates should therefore be interpreted as intent-to-treat effects based on eligibility criteria rather than direct administrative assignment. The same limitation applies to the H_2 transition estimand: the Special-monitoring indicator is likewise reconstructed from public criteria, so the H_2 estimates identify the effect of becoming eligible for escalation, not the effect of a confirmed administrative decision to escalate a specific firm.

4.3 Empirical Model

Let Y_{ijt} denote the realized tax-burden outcome for firm i in industry j and period t :

$$Y_{ijt} = \alpha_{g,t} \cdot \mathbf{1}(G_i = g) \cdot \mathbf{1}(t \geq g) + X'_{it}\beta + \gamma_i + \lambda_j + \delta_t + \varepsilon_{ijt} \quad (3)$$

where $\alpha_{g,t}$ is the group-time ATT, X_{it} is a vector of time-varying firm-level controls, γ_i are firm fixed effects, λ_j captures industry differences, and δ_t are period fixed effects. Estimation uses the doubly robust method (inverse probability weighting,

IPW + outcome regression); for two ETR specifications with a singular propensity-score design matrix, the estimator falls back to IPW only. Inference uses clustered bootstrap SE at the firm level (999 replications):

$$\widehat{ATT}^{agg} = \sum_{g \in \mathcal{G}} w_g \left(\frac{1}{T_g} \sum_{t=g}^T ATT(g, t) \right) \quad (4)$$

The four outcomes are the Effective Tax Rate (Kubick, Lynch, Mayberry, & Omer, 2016): $ETR_{it} = IT_{it}/EBIT_{it}$, where IT_{it} is income tax charges and $EBIT_{it}$ is earnings before interest and taxes; Tax Burden (Cunha, Ribeiro, & Santos, 2005): $TB_{it} = TT_{it}/AVD_{it}$, where TT_{it} is total tax charges across all government levels and AVD_{it} is value added distributed (*valor adicionado distribuído*, the denominator used in the Brazilian value-added statement); and their abnormal (benchmark-adjusted) counterparts: $AbnETR_{ijt} = ETR_{ijt} - \widetilde{ETR}_{jt}$ and $AbnTB_{ijt} = TB_{ijt} - \widetilde{TB}_{jt}$, where tildes denote industry-year medians.

AbnETR is not a heterogeneity variable. It is an outcome that compares firm i with the contemporaneous median ETR of industry j in period t . The notation i, j, t is used consistently throughout the paper to distinguish firms, industries and time periods. This distinction matters because heterogeneity analyses split firms by pre-treatment characteristics, whereas AbnETR asks whether the treated firm moves away from the industry-year benchmark after monitoring.

The control vector X_{it} includes firm size ($\ln AT$, the natural log of total assets, AT), leverage, pre-tax profitability (ROA), tangibility (PP&E/assets), sector code, Big-4 auditor indicator (Klassen, Lisowsky, & Mescall, 2016), and firm life-cycle stage following Dickinson (2011). Governance and auditor variables are reported only at Q4 in the CVM Open Data filings (annual filing cadence). For Q1–Q3 observations, I apply bounded last-observation-carried-forward (LOCF) imputation within each firm’s observed window, following the within-year propagation practice of Larcker and Rusticus (2010) and Coles et al. (2008); no imputation is applied before a firm’s first observed value or after its last, and no LOCF is applied to outcome variables. The modal pre-treatment stage is mature ($\approx 54\%$ of firms), followed by growth (24%), shake-out (12%), decline (7%) and introduction (3%). All continuous controls are winsorised at the 1st and 99th percentiles within each year. The industry-year benchmark ETR and TB used to construct AbnETR and AbnTB are computed from clean-control observations only (firms that have not yet entered any monitoring tier at period t), following Sant’Anna and Zhao (2020) to avoid contaminating the benchmark with treatment-induced changes.

Potential Tax Exposure (PET) is constructed only from pre-treatment information. I z-score standardize tangible assets (PP&E), gross revenue (rb_1), and pre-tax income (LAIR), then extract the first principal component via PCA. All three variables reflect distinct taxable-base dimensions available to the RFB before monitoring entry; using pre-treatment values avoids conditioning on post-treatment outcomes (following the bad-control logic of Borusyak et al., 2024; Callaway & Sant’Anna, 2021). The PCA is used as a dimension-reduction device rather than as an additional causal model. Bartlett’s sphericity test rejects independence among the inputs ($p < 0.001$), the Kaiser–Meyer–Olkin (KMO) statistic is 0.71 (classified as middling adequate by

the Kaiser scale), the first component explains 80.4% of total variance, and all three variables load positively on PC1 (loadings: PP&E 0.58, gross revenue 0.55, LAIR 0.60), consistent with PC1 representing a common taxable-exposure dimension rather than a contrast. These diagnostics support the use of PET as a compact pre-treatment exposure measure. The scree plot and loading diagram are in the Online Resource (Appendix B).

5 Results and Discussion

5.1 Main Staggered DiD Estimates

Table 7 reports the core evidence. The table separates T1 (Differentiated entry, $n = 123$ treated firms) from T2 (Transition to Special, $n = 14$ treated firms across five estimable cohorts), because these are institutionally distinct margins.³ The diagnostic evidence supporting this design is reported in the supplementary material: all estimable cohorts have pre- and post-entry observations, pre-treatment event-study coefficients are jointly consistent with parallel trends, and matching/entropy diagnostics confirm that size and auditor imbalance are addressed through the doubly robust CS estimator rather than ignored.

Table 7: Main two-stage staggered DiD estimates: entry into Differentiated monitoring and transition to Special monitoring

Outcome	H1: Entry into Differentiated monitoring			H2: Transition to Special monitoring		
	ATT	SE	95% CI	ATT	SE	95% CI
ETR_{it}	0.01	(0.02)	[-0.02, 0.04]	-0.03	(0.06)	[-0.15, 0.08]
TB_{it}	0.01	(0.02)	[-0.02, 0.04]	0.11**	(0.04)	[0.02, 0.19]
$AbnETR_{it}$	-0.01	(0.01)	[-0.04, 0.01]	0.17***	(0.04)	[0.09, 0.25]
$AbnTB_{it}$	-0.01	(0.01)	[-0.02, 0.01]	0.03*	(0.02)	[-0.00, 0.06]

Notes: Callaway and Sant'Anna (2021) doubly robust ATT aggregations. Firms are indexed by i and quarters by t ; sector j is used in sector-stratified specifications. H1: first entry into Differentiated monitoring; comparison group: not-yet-treated. H2: risk set restricted to firms already in Differentiated monitoring; treatment = first Special quarter. n.a. indicates that the requested ATT aggregation was not estimable; cohort-level reasons are reported in Table 2. Effects in pp: $100 \times$ ATT. Clustered bootstrap standard errors (999 replications) in parentheses; 95% simultaneous confidence intervals in brackets. *, **, *** significance at 10%, 5%, 1%.

³The combined indicator $g_{grcontr}$ is not used as a main estimand because it conflates entry into Differentiated monitoring with escalation to Special monitoring. The two-stage design separates these institutionally distinct margins, following the RFB ordinance architecture formalised in Section 2.

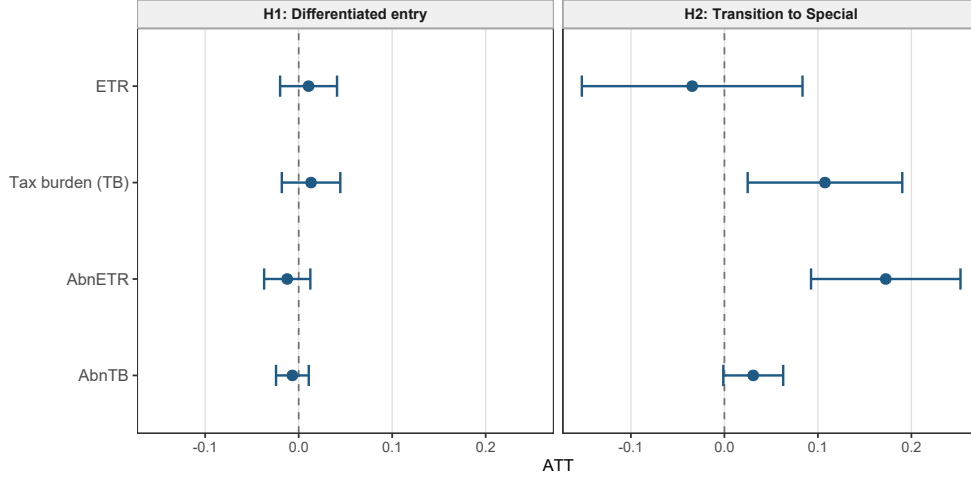


Fig. 1: ATT estimates across the four main outcomes: H1 entry into Differentiated monitoring and H2 transition to Special monitoring. Each dot is the aggregate ATT point estimate with 95% simultaneous confidence interval. The dashed vertical line marks zero.

The evidence is sharply tier-specific. Entry into Differentiated monitoring does not produce a statistically precise change in realized tax burden: the H_1 estimates for ETR, TB, AbnETR, and AbnTB are all statistically indistinguishable from zero. This bounded null is consistent with the institutional interpretation in Section 3: the Differentiated tier records fiscal relevance and assigns firms to routine monitoring, but it does not yet impose the intensive cross-base scrutiny needed to alter tax positions.

The transition to Special monitoring produces the economically relevant response. TB increases by about 11 percentage points, AbnETR by about 17 percentage points, and AbnTB by about 3 percentage points; the raw ETR estimate is negative but imprecise. The H2 sample is small and must be read with that power constraint in mind.⁴ Still, the sign pattern is economically coherent. ETR captures income-tax charges relative to EBIT, whereas TB measures the total fiscal burden across income, consumption, payroll, and value-added taxes. In a Special-tier relationship, the RFB observes the firm’s full tax-filing portfolio; therefore TB and benchmark-adjusted outcomes better capture the cross-base compliance response than raw ETR alone. The divergence is consistent with how each measure is built: ETR responds only to adjustments booked through the income-tax line, which firms can offset with deferrals, credits, and loss carryforwards that leave the reported income-tax charge largely unchanged even when underlying compliance behavior shifts. TB and AbnTB instead aggregate across every base the RFB can cross-check once a firm enters intensive monitoring, so a compliance

⁴Two singleton cohorts ($g = 37, n = 1$; $g = 45, n = 1$) are excluded from the estimation via a minimum group-size filter of 3 firms per cohort (following Callaway & Sant’Anna, 2021). The dynamic event-study window is restricted to $[-4, +8]$ quarters, the interval with support in at least three cohorts (following the declared-window convention of Autor, Dorn, & Hanson, 2013). The ATT simple aggregation uses all estimable cohorts.

response concentrated in consumption or payroll taxes shows up there even when it is muted or offset in the income-tax base alone. The imprecise raw ETR estimate is therefore better read as evidence that escalation does not operate primarily through the income-tax margin, not as evidence against a compliance response overall.

5.2 Dynamic Event-Study Evidence

Figure 2 provides the key dynamic check for the Special-transition margin, with one panel for each outcome. Pre-treatment coefficients are generally centered around zero, supporting the conditional parallel-trends interpretation for the risk set in which H2 is estimated. After transition, the stronger post-treatment signal appears in TB and benchmark-adjusted outcomes, which is the dynamic counterpart of Table 7. Figure 1 summarises why the dynamic figure focuses on H2: the Differentiated-entry margin is close to zero across the four outcomes, whereas escalation to Special monitoring carries the economically relevant response.

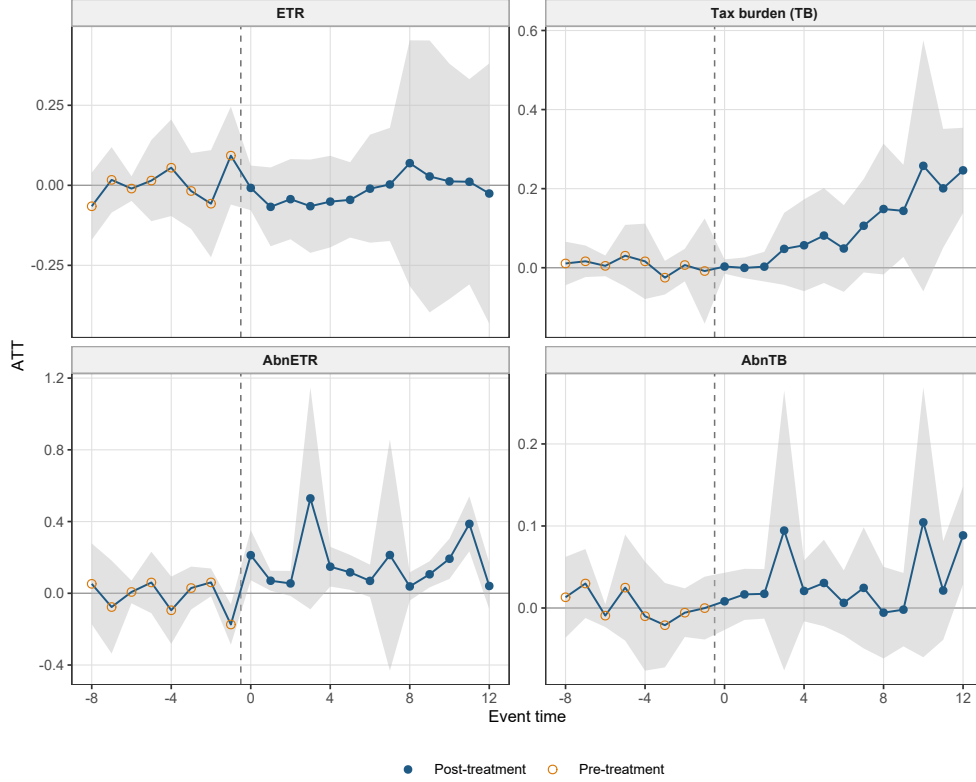


Fig. 2: Dynamic event-study coefficients for H2, the transition from Differentiated to Special monitoring, by outcome. Panels report ETR, Tax burden (TB), AbnETR, and AbnTB. Filled circles denote post-treatment coefficients; open circles denote pre-treatment coefficients. Shaded bands are 95% simultaneous confidence intervals (Callaway and Sant'Anna, 2021). The dashed vertical line marks the treatment quarter.

5.3 Heterogeneity Analysis

Table 8 reports the only heterogeneity table kept in the main text. The splits are measured before treatment or defined by calendar timing, so they describe effect heterogeneity rather than post-treatment selection. The strongest subgroup evidence is concentrated in the escalation design: low fiscal-pressure firms show a negative H2 ETR response, while the post-IFRS 15 and PET splits are less precise. These results should be interpreted as mechanism evidence rather than separate causal designs. Fiscal pressure captures pre-entry tax position; PET captures pre-entry observable exposure; and the IFRS 15 split checks whether the accounting regime shift is driving the estimates. These splits are exploratory: they were not pre-registered, the subgroup comparisons are not adjusted for multiple testing, and the resulting patterns should

guide future hypothesis-driven work rather than be read as confirmatory tests in their own right.

Table 8: Heterogeneity of the monitoring effect on the effective tax rate

Dimension	Split	Yvar	Estimate	Std Error	Conf Interval
Fiscal pressure	Low fiscal pressure	ETR	0.01	(0.0227)	[-0.0362, 0.0527]
Fiscal pressure	High fiscal pressure	ETR	-0.09**	(0.0347)	[-0.1557, -0.0198]
IFRS 15 period	pre	ETR	-0.05**	(0.0223)	[-0.0903, -0.0030]
IFRS 15 period	post	ETR	-0.05	(0.0477)	[-0.1392, 0.0476]
Compliance/rating	Low compliance/rating	ETR	-0.10	(0.0633)	[-0.2259, 0.0222]
Compliance/rating	High compliance/rating	ETR	-0.03	(0.0338)	[-0.0949, 0.0375]
Tax exposure	Low tax exposure	ETR	-0.07***	(0.0248)	[-0.1183, -0.0209]
Tax exposure	High tax exposure	ETR	0.06	(0.1107)	[-0.1551, 0.2789]

Notes: Heterogeneity estimates use pre-treatment or predetermined strata. The outcome is ETR. These are effect-heterogeneity estimates, not mediation decompositions.

5.4 Robustness Checks

The supplementary material reports the full robustness battery. The main patterns are stable in leave-one-cohort-out checks; placebo timing does not produce a comparable effect; Goodman-Bacon and HonestDiD diagnostics do not overturn the interpretation; and the [Oster \(2019\)](#) coefficient-stability diagnostic is not identified for this specification: the TWFE proxy regression exhibits near-zero R^2 movement once firm and period fixed effects are included, so the bound is undefined rather than confirmatory (reported for transparency in the supplementary material). RDD estimates around the Special-monitoring revenue threshold are imprecise, which is consistent with the institutional design: assignment depends on multiple fiscal-relevance criteria rather than a single manipulable cutoff. Together, the robustness evidence supports the same conclusion as the main estimates: the measurable response comes from escalation to Special monitoring, not from broad administrative classification.

6 Conclusion

The evidence shows that enforcement intensity matters, but only at the margin where monitoring becomes substantively stronger. Entry into the Differentiated tier does not generate a detectable average change in firms' realized tax burden. Transition to Special monitoring, by contrast, raises total tax burden by 10.76 percentage points and produces positive benchmark-adjusted responses. The central contribution is therefore the separation between administrative classification and enforcement escalation: classification improves screening, while Special monitoring generates the measurable change in tax position.

This pattern is consistent with deterrence and information-traceability mechanisms. Broad classification may increase the probability of being observed, but Special

monitoring creates a continuing relationship in which cross-base scrutiny constrains tax planning. The result aligns with [Scholz \(1997\)](#), [Hoopes et al. \(2012\)](#), and [Almunia and Lopez-Rodriguez \(2018\)](#): compliance responses are strongest when detection risk is persistent, salient, and supported by detailed information.

The policy implication is direct. Large-taxpayer programs should not be evaluated only by threshold entry or by the size of the monitored population. The relevant design question is whether the administration can identify the firms for which escalation to intensive monitoring is most informative. The heterogeneity evidence suggests that pre-entry fiscal pressure and observable tax exposure can help refine that targeting, although subgroup results should be treated as mechanism evidence rather than standalone causal estimates.

Brazil’s two-tier design is one implementation of a monitoring model used elsewhere. Large-taxpayer units organized around graduated compliance relationships, rather than a single monitoring status, are a recurring feature of tax administrations reviewed by international bodies such as the OECD, and [Martini \(2022\)](#) situates Brazil’s cooperative-compliance approach explicitly within that broader international practice. The escalation result reported here, that a compliance response appears at the more intensive tier rather than at the point of routine classification, is consistent with the graduated-intensity logic behind those programs. Whether the same pattern holds in jurisdictions with different disclosure requirements, enforcement capacity, or firm populations is an empirical question this single-country design cannot answer; the contribution here is evidence on how one such program in one context performs, not a claim of generality across large-taxpayer regimes.

The study is bounded by the H2 sample size, the restriction to publicly listed firms, and the use of observable proxies for monitoring assignment. These limits do not overturn the main finding, but they define its scope. Future work with confidential RFB administrative records could validate the assignment proxy, identify which tax bases drive the total-burden response, and test whether the same dose-response relationship between monitoring intensity and compliance holds outside publicly listed firms.

Supplementary information. Online Resource 1 (ESM_1) reports the full identifying-assumption battery for the staggered DiD design (parallel trends, no anticipation, SUTVA), the PET/PCA validation (Appendix B), matching and entropy-balancing diagnostics (Appendix E), the RDD diagnostics around the Special threshold, the complete robustness and sensitivity battery (leave-one-cohort-out, placebo timing, the Goodman-Bacon decomposition, HonestDiD sensitivity, and the Oster coefficient-stability diagnostic), additional outcome estimates (TB, AbnETR, and AbnTB) for the main results, event study, heterogeneity, and RDD analyses, an annual-frequency replication and sector/regulation robustness samples, additional heterogeneity dimensions (governance and litigation risk), and sample-construction and estimation-diagnostic detail.

Acknowledgments. The author acknowledges the Espírito Santo Research and Innovation Support Foundation (Fundação de Amparo à Pesquisa e Inovação do Espírito Santo, FAPES) and the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e

Tecnológico, CNPq) for the grant resources supporting this research, awarded under the FAPES/CNPq Call (Chamada) n° 23/2025 (Adhesion Agreement SEI n° 01300.009416/2025-38; CNPq process 311211/2026-7).

Declarations

Funding. This research was supported by a Research Productivity Fellowship (Bolsa de Produtividade em Pesquisa, PQ) awarded to the author by the Espírito Santo Research and Innovation Support Foundation (Fundação de Amparo à Pesquisa e Inovação do Espírito Santo, FAPES) and the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq), under the FAPES/CNPq Call (Chamada) n° 23/2025, formalized through the FAPES–CNPq Adhesion Agreement (Acordo de Adesão) SEI n° 01300.009416/2025-38 (CNPq process 311211/2026-7).

Competing interests. The author declares no competing interests.

Ethics approval. Not applicable. This study uses publicly listed firms’ financial-statement and regulatory data; it involves no human participants, human tissue samples, or live vertebrate subjects.

Consent to participate. Not applicable.

Consent for publication. Not applicable.

Availability of data and materials. The raw financial-statement data underlying this study are publicly available from CVM Open Data (Dados Abertos da Comissão de Valores Mobiliários, <https://dados.cvm.gov.br>), Brazil’s securities and exchange regulator. The constructed analysis panel (the cleaned, merged firm-quarter dataset used for estimation) and the RFB monitoring ordinance texts on which the treatment-timing variables are based will be deposited in Harvard Dataverse (or an equivalent public repository) upon acceptance, with the dataset DOI added at that stage.

Code availability. The R code implementing the Callaway–Sant’Anna estimator, the RDD, and the full robustness battery will be deposited alongside the data in the same Harvard Dataverse (or equivalent) repository upon acceptance.

Authors’ contributions. Not applicable (single-authored work).

Generative AI use. The author used AI-based tools to assist with reviewing analysis scripts, checking methodological implementation details, and copyediting English-language grammar and phrasing. All prose, critical analysis, and interpretation of the results are the author’s own. No generative AI tool is credited as an author, and no AI tool was used to generate data, analyses, or substantive content.

References

- Allingham, M.G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3–4), 323–338, [https://doi.org/10.1016/0047-2727\(72\)90010-2](https://doi.org/10.1016/0047-2727(72)90010-2)
- Almunia, M., & Lopez-Rodriguez, D. (2018). Under the Radar: The Effects of Monitoring Firms on Tax Compliance. *American Economic Journal: Economic Policy*,

10(1), 1–38, <https://doi.org/10.1257/pol.20160229>

- Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial. *Criminology*, 50(1), 27–69, <https://doi.org/10.1111/j.1745-9125.2011.00256.x>
- Autor, D.H., Dorn, D., Hanson, G.H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, 103(6), 2121–2168, <https://doi.org/10.1257/aer.103.6.2121>
- Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., Sant’Anna, P.H.C. (2026). Difference-in-Differences Designs: A Practitioner’s Guide. *Journal of Economic Literature*, 64(2), 498–557, <https://doi.org/10.1257/jel.20251650>
- Benon, O., Baer, K., Toro R., J. (2002). *Improving Large Taxpayers’ Compliance: A Review of Country Experience*. Washington, D.C.: International Monetary Fund.
- Borusyak, K., Jaravel, X., Spiess, J. (2024). Revisiting Event-Study Designs: Robust and Efficient Estimation. *Review of Economic Studies*, 91(6), 3253–3285, <https://doi.org/10.1093/restud/rdae007>
- Callaway, B., & Sant’Anna, P.H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230, <https://doi.org/10.1016/j.jeconom.2020.12.001>
- Carnahan, M. (2015). Taxation Challenges in Developing Countries. *Asia & the Pacific Policy Studies*, 2(1), 169–182, <https://doi.org/10.1002/app5.70>
- Coles, J.L., Daniel, N.D., Naveen, L. (2008). Boards: Does one size fit all? *Journal of Financial Economics*, 87(2), 329–356, <https://doi.org/10.1016/j.jfineco.2006.08.008>
- Cunha, J.V.A.d., Ribeiro, M.d.S., Santos, A.d. (2005). A demonstração do valor adicionado como instrumento de mensuração da distribuição da riqueza. *Revista de Contabilidade e Finanças*, 16(37), 7–23, <https://doi.org/10.1590/S1519-70772005000100001>

- DeBacker, J., Heim, B.T., Tran, A., Yuskavage, A. (2015). Legal Enforcement and Corporate Behavior: An Analysis of Tax Aggressiveness after an Audit. *The Journal of Law and Economics*, 58(2), 291–324, <https://doi.org/10.1086/684037>
- Devos, K. (2007). Measuring and analysing deterrence in taxpayer compliance research. *Journal of Australian Taxation*, 10(2), 182–222, Retrieved from <https://journalofaustralian.com.au/past-issues/volume-10-issue-2>
- Dickinson, V. (2011). Cash flow patterns as a proxy for firm life cycle. *The Accounting Review*, 86(6), 1969–1994, <https://doi.org/10.2308/accr-10130>
- Ehinger, A.C., Lee, J.A., Stomberg, B., Towery, E. (2025). IRS Enforcement and Voluntary Tax Disclosure. *The Journal of the American Taxation Association*, 47(1), 85–119, <https://doi.org/10.2308/jata-2022-009>
- Fuest, C., & Neumeier, F. (2023). Corporate taxation. *Annual Review of Economics*, 15(1), 425–450, <https://doi.org/10.1146/annurev-economics-082322-014747>
- Hanlon, M., Hoopes, J.L., Shroff, N. (2014). The effect of tax authority monitoring and enforcement on financial reporting quality. *Journal of the American Taxation Association*, 36(2), 137–170, <https://doi.org/10.2308/atax-50820>
- Hoopes, J.L., Mescall, D., Pittman, J.A. (2012). Do IRS Audits Deter Corporate Tax Avoidance? *The Accounting Review*, 87(5), 1603–1639, <https://doi.org/10.2308/accr-50187>
- Jacob, M. (2022). Real effects of corporate taxation: A review. *European Accounting Review*, 31(1), 269–296, <https://doi.org/10.1080/09638180.2021.1934055>
- Johns, A., & Slemrod, J. (2010). The distribution of income tax noncompliance. *National Tax Journal*, 63(3), 397–418, <https://doi.org/10.17310/ntj.2010.3.01>
- Klassen, K.J., Lisowsky, P., Mescall, D. (2016). The Role of Auditors, Non-Auditors, and Internal Tax Departments in Corporate Tax Aggressiveness. *The Accounting Review*, 91(1), 179–205, <https://doi.org/10.2308/accr-51137>

- Kubick, T.R., Lynch, D.P., Mayberry, M.A., Omer, T.C. (2016). The Effects of Regulatory Scrutiny on Tax Avoidance: An Examination of SEC Comment Letters. *The Accounting Review*, 91(6), 1751–1780, <https://doi.org/10.2308/accr-51433>
- Larcker, D.F., & Rusticus, T.O. (2010). On the use of instrumental variables in accounting research. *Journal of Accounting and Economics*, 49(3), 186–205, <https://doi.org/10.1016/j.jacceco.2009.11.004>
- Linck, J.S., Netter, J.M., Yang, T. (2008). The determinants of board structure. *Journal of Financial Economics*, 87(2), 308–328, <https://doi.org/10.1016/j.jfineco.2007.03.004>
- Martini, M.H. (2022). A Review of Brazil Approaches to Cooperative Compliance in Light of International Tax Practice and the OECD Concept. *Intertax*, 50(2), 177–195, <https://doi.org/10.54648/TAXI2022016>
- Mason, P., & Williams, B. (2022). Does IRS Monitoring Deter Managers From Committing Accounting Fraud? *Journal of Accounting, Auditing & Finance*, 37(3), 700–722, <https://doi.org/10.1177/0148558X20939720>
- Mills, L.F. (1998). Book-tax differences and Internal Revenue Service Adjustments. *Journal of Accounting Research*, 36(2), 343–356, <https://doi.org/10.2307/2491481>
- Mintz, J. (1995). The corporation tax: A survey. *Fiscal Studies*, 16(4), 23–68, <https://doi.org/10.1111/j.1475-5890.1995.tb00232.x>
- Nessa, M., Schwab, C.M., Stomberg, B., Towery, E.M. (2020). How do IRS Resources Affect the Corporate Audit Process? *The Accounting Review*, 95(2), 311–338, <https://doi.org/10.2308/accr-52520>
- Okunogbe, O., & Tourek, G. (2024). How Can Lower-Income Countries Collect More Taxes? The Role of Technology, Tax Agents, and Politics. *Journal of Economic Perspectives*, 38(1), 81–106, <https://doi.org/10.1257/jep.38.1.81>
- Oliveira, P.T.P.D. (2022). Improving the Relationship Between Tax Authorities and Taxpayers in Brazil. *Intertax*, 50(3), 218–230, <https://doi.org/10.54648/taxi2022021>

- Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics*, 37(2), 187–204, <https://doi.org/10.1080/07350015.2016.1227711>
- Plumley, A.H. (2002). *The impact of the IRS on voluntary tax compliance: Preliminary empirical results* (Paper presented at the 95th Annual Conference on Taxation, National Tax Association). Orlando, Florida: Internal Revenue Service, National Headquarters Office of Research. Retrieved from <https://www.irs.gov/pub/irs-soi/irsvtc.pdf>
- Pomeranz, D. (2015). No taxation without information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*, 105(8), 2539–2569, <https://doi.org/10.1257/aer.20130393>
- Sant’Anna, P.H.C., & Zhao, J. (2020). Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1), 101–122, <https://doi.org/10.1016/j.jeconom.2020.06.003>
- Scholz, J.T. (1997). Enforcement policy and corporate misconduct: The Changing Perspective of Deterrence Theory. *Law and Contemporary Problems*, 60(3), 253–268, <https://doi.org/10.2307/1192014>
- Slemrod, J. (2004). Are corporate tax rates, or countries, converging? *Journal of Public Economics*, 88, 1169–1186, [https://doi.org/10.1016/S0047-2727\(03\)00061-6](https://doi.org/10.1016/S0047-2727(03)00061-6)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryitaxpf.pdf](#)