Audit Rule Disclosure and Tax Compliance*

Enrico Di Gregorio Matteo Paradisi Elia Sartori

August 7, 2026

Abstract

Tax authorities typically concentrate enforcement resources on large businesses, yet small business evasion generates tax revenue losses that often exceed those from multinational profit shifting. We show that authorities can improve small business compliance by strategically disclosing audit-relevant information. We study audit rules that inform taxpayers that audit risk drops discontinuously above a threshold based on predicted revenues. Under empirically plausible conditions, our model shows that the tax base is concave in the size of this discontinuity. Consequently, if widening an existing discontinuity raises the tax base, the pre-reform disclosed rule must already outperform a flat, undisclosed benchmark. We test this implication using more than 26 million tax files (2007–2016) from Italy's Sector Studies. Taxpayers bunch sharply at the threshold, and bunching correlates with evasion proxies and evasion technologies. Exploiting a staggered reform that increased the audit risk discontinuity, we find that compliance rises below the threshold and falls above it, yet average reported profits, the relevant tax base, grow by 16.2% in treated sectors over six years. Our theoretical result therefore implies that disclosure outperforms nondisclosure. Structural estimates further show that strengthening incentives at the threshold could more than halve the audit budget without reducing compliance, freeing resources to target larger firms, and that the reform brought the policy close to the optimum.

JEL Classification: D04, D22, H24, H25, H26, H32.
Keywords: tax compliance, enforcement, evasion, audit, disclosure, firm, bunching.

*Contact information: Enrico Di Gregorio, International Monetary Fund, HQ21900 Pennsylvania Avenue NWWashington, DC, edigregorio@imf.org; Matteo Paradisi, EIEF, via Sallustiana 62, 00187 Rome, matteo.paradisi@eief.it; Elia Sartori, CSEF and Università Federico II, Via Cintia complesso Monte S. Angelo, Edificio 3, 80126 Napoli, elia.sartori@unina.it. We thank seminar and conference participants at various institutions for helpful comments and discussions. We thank Agenzia delle Entrate for granting access to the Sector Study database, and SOSE for their support in extracting and interpreting the raw data. This project was made possible through Research Agreement R.U. 0521883.20-12-2018-U with Agenzia delle Entrate. Di Gregorio acknowledges the support of NBER and the Peter G. Peterson Foundation. All views expressed are solely our own.

1 Introduction

Ensuring tax compliance among small businesses and the self-employed is historically a central challenge for tax agencies in developed and developing countries (Alm, Martinez-Vazquez, and Wallace, 2004). In the U.S., imperfect compliance among small businesses results in at least a 7-8% shortfall in total tax revenues.[1] Compared to profit shifting of U.S. multinationals, which amounts to 15% of corporate taxes and 1.32% of total revenues (Wier and Zucman, 2022), small businesses' evasion reduces government revenues by almost one order of magnitude more. Yet, tax authorities tend to skew their audit resources toward large firms so that more than 70 countries adopted special enforcement units for large taxpayers in the last 20 years (Almunia and Lopez-Rodriguez, 2018; Bachas, Fattal Jaef, and Jensen, 2019; Basri et al., 2019).[2] This might reflect a cost-effectiveness principle of tax administration, as auditors expect a relatively higher yield from auditing a large business rather than several small ones for any given budget (Boning et al., 2025). To the extent that tax agencies are unwilling or unable to distribute enforcement efforts equally across firm types, the identification of low cost strategies to promote small firms' voluntary tax compliance becomes essential to tax collection.

This paper provides the first evaluation of audit rule disclosure as a viable strategy to improve the incentives for tax base reporting by small businesses. We refer to audit rules as the criteria that tax agencies routinely adopt to guide audit case selection. Tax authorities seem to value the secrecy of these rules, which they generally keep from the public. Indeed, the choice of disclosure comes with a trade-off. On one hand, revealing what behaviors trigger a tax audit might nudge some taxpayers away from evasion. On the other, those who learn to be at lower risk of an audit might end up reporting a lower tax base. On the net, the impact of disclosure is ambiguous, and this paper aims to characterize and quantify this trade-off using evidence from a real-world policy.

We examine a specific case of audit rule disclosure in which the tax authority announces, prior to tax reporting, the precise threshold above which audit risk declines discontinuously. This threshold corresponds to a predicted revenue level for each firm, creating a firm-specific compliance target. A prominent example of such a rule is the Sector Studies (*Studi di Settore*,

---

[1] We sum the estimated yearly underreporting and non-filing among individual business income earners, self-employed, and small corporations, and divide by the total collected central government revenues for 2019 Internal Revenue Service (2022). Our inability to break down other tax gap items implies our estimates are lower bounds. We provide further details in section 3.

[2] In a 2019 survey from Italy, our empirical setting, the share of firms reporting any tax inspections over the previous 12 months was 9.9% among firms with less than 20 employees, and 18% among those with more than 100 (The World Bank, 2019).

henceforth SeS), an Italian audit system applied on small businesses and the self-employed.[3] To examine the compliance dynamics that result from this system, we access a novel, confidential database of over 26.6 million SeS files from the 2007-2016 tax period, including the previously unexploited universe of 2007-2010 files. This rich dataset includes small businesses earning less than 5.2 million euros in revenue each year.

Our focus on disclosed threshold rules is motivated by their widespread adoption across countries,[4] their capacity to leverage tax authorities' data and forecasting power to tailor compliance incentives, theoretical findings suggesting that threshold-based rules can maximize tax collection (Reinganum and Wilde, 1985; Sánchez and Sobel, 1993), and the intuition that tax liability notches create strong incentives even in the presence of low structural elasticities (Kleven and Waseem, 2013).[5] Despite their theoretical appeal, limited empirical attention has been devoted to the compliance effects of disclosed threshold rules. To the best of our knowledge, this paper provides the first evidence in this direction.[6]

We leverage the design of the SeS and a theoretical framework to evaluate whether disclosed audit rules based on statistical predictions can enhance tax compliance. We begin by developing a model of audit rule disclosure and deriving a test that assesses the desirability of threshold-based audit rules by examining firms' tax base responses to marginal changes in audit-related incentives. We then exploit a natural experiment introducing a *reward regime* that increased the audit risk discount at the disclosed threshold to empirically implement this test. Finally, we estimate the model and explore avenues for improving the current audit rule design.

Our theoretical framework evaluates the trade-offs involved in disclosing threshold-based audit rules. In the baseline model, a continuum of risk-neutral firms within a category – such as a sector and location – differ in income and choose how much to report facing a convex manipulation cost. Under the undisclosed benchmark, firms perceive a flat audit probability. Disclosure introduces a reporting threshold: firms declaring below it face a higher audit probability, while those declaring at or above it receive an audit discount. Firms then sort by income into four groups: (i) zero-declarers; (ii) interior declarers below the threshold;

[3] The Italian government estimates that these taxpayer categories accounted for up to 30.4% of all unpaid tax liabilities during the 2014-2016 period (Ministry of Economy and Finance, 2019).

[4] Among others, the Australia's periodic release of industry benchmarks by the Australian Tax Office (OECD, 2006), Greece's profit margin targets under its self-assessment program (Al-Karablieh, Koumanakos, and Stantcheva, 2021), Mexico's introduction of sector-specific effective tax rates in June 2021, and earlier presumptive taxation schemes like France's *forfait* and Israel's *tachshiv* (Thuronyi, 1996).

[5] Lazear (2006) show how disclosure may enhance aggregate outcomes in contexts beyond tax compliance.

[6] In this regard, our analysis diverges from studies on Large Taxpayer Units, which incentivize taxpayers to *underreport* income to avoid monitoring (Almunia and Lopez-Rodriguez, 2018; Bachas, Fattal Jaef, and Jensen, 2019; Basri et al., 2019), and from work on policies offering audit risk reductions tied to a common economy-wide threshold (Dwenger et al., 2016; Al-Karablieh, Koumanakos, and Stantcheva, 2021).

(iii) bunchers at the threshold; and (iv) interior declarers above it. We subsequently extend the model to allow for heterogeneity in revenues, costs, and manipulation costs; multiple reporting margins; a threshold that depends on reported costs; a fixed burden of being audited; and an audit schedule that varies smoothly with declarations even without a disclosed discontinuity.

We develop a test of whether disclosed threshold rules outperform undisclosed ones based on the revenue effects of marginally widening the audit discontinuity – a policy variation that closely mirrors the *reward regime* studied empirically. We show that revenue is globally concave over a nondegenerate range of audit gaps when the baseline audit probability is sufficiently low. This concavity reflects diminishing returns to attracting additional bunchers. When the gap is small, firms just below the threshold can qualify at low cost while retaining substantial evasion protected by the audit discount. As the gap widens, the marginal buncher must give up more evasion, leaving less evasion protected by the discount. Meanwhile, the higher audit probability below the threshold raises the buncher's counterfactual declaration, further reducing the revenue gain when it switches. Using estimates from the literature and our data, we verify that the conditions for concavity hold for empirically plausible parameter values. Concavity is what makes the experiment informative about the unobserved flat-rule counterfactual: it implies that the return to widening the audit gap decreases with its size. Thus, if the *reward regime* raised reported revenues and the tax base by widening an already positive gap, opening the gap from zero to its pre-reform level must also have raised them. The pre-reform threshold rule therefore outperformed the flat benchmark, which uses at least as many audits. The concavity result extends to aggregate declared revenues and profits in the richer model. Untargeted reporting margins, a fixed audit burden, and a manipulable eligibility frontier strengthen the incentive to bunch and the resulting decreasing returns, while a smooth underlying audit schedule and heterogeneous manipulation costs alter their strength without eliminating concavity.

We turn to the empirical section to validate some key predictions of the model and implement the test by leveraging the introduction of a *reward regime* as a natural experiment. In the data, we observe a significant spike in the distribution of declared revenues just above the disclosed revenue threshold. Consistent with the model's predictions, we test whether the extent of bunching correlates with various evasion margins and incentives to manipulate. We find that bunching is strongly associated with evasion in VAT, property taxes, income taxes, and with anonymous evasion reports. It is also more pronounced in upstream sectors, where evasion is more difficult due to third-party reporting, and in regions with higher personal income tax rates. Finally, in line with our theoretical predictions, bunching positively correlates with the share of firms declaring zero revenues in a given area.

To evaluate whether disclosed audit rules can improve tax compliance, we exploit a natural experiment that closely mirrors the marginal change in the audit probability jump analyzed in our theoretical model. Beginning in 2011, a staggered reform to the SeS—known as the "*reward regime*"—enhanced protections for firms that complied with SeS prescriptions and promised to devote more attention to those who did not. This reform widened the perceived audit risk gap around the presumed revenue threshold for firms subject to the new rules. Using a balanced panel from 2007 to 2016, our event-study design shows that firms in treated sectors adjusted their reported revenues closer to the SeS threshold. This behavioral shift occurred both below and above the threshold, supporting our model's prediction that disclosure-based policies create opposite incentives across taxpayers. On average, however, treated firms increased reported revenues by 12% and gross profits (the tax base) by 16.2% over six years. Interpreted through the lens of our model, this evidence suggests that disclosing audit risk discontinuities enabled the tax authority to expand the tax base among small businesses.

We then estimate a structural model to interpret firms' responses and evaluate counterfactual disclosure policies. Firms differ in revenues, costs, and their propensity to evade, and may misreport both revenues and costs. Because reported costs determine the revenue threshold, cost manipulation affects both taxable profits and eligibility for lower audit risk. We estimate the model by two-stage simulated method of moments, targeting the pre-reform distribution around the threshold and our reduced-form estimates of the reform's causal effects on reported revenues, profits, and firm-specific thresholds. The estimates closely match administrative pre-reform audit probabilities, recover the post-reform elimination of audits above the threshold, and imply that revenue and cost misreporting are close substitutes, rationalizing their simultaneous increase after the reform. Counterfactuals show that both the pre-reform rule and the *reward regime* generate more declared profits than a flat rule with the same audit budget. Widening the discontinuity further could maintain the pre-reform tax base while cutting the enforcement budget by more than half. Crucially, the Authority could redirect these substantial savings toward larger taxpayers, whose audits yield higher returns, without sacrificing compliance among small firms. Finally, the *reward regime* represented a substantial improvement over the pre-reform policy: by moving audit risk above the threshold toward zero, it brought the Authority's objective within 11% of the unconstrained optimum throughout the range of marginal audit costs for which the optimal policy features a positive audit discontinuity. Over the same range, the pre-reform rule remains 20–70% below the optimum.

Related literature: This paper contributes to the literature on tax compliance and enforcement by studying a novel strategy: the disclosure of audit rules that condition audit risk on predicted firm-level outcomes. While a large body of work analyzes how audit probabilities affect compliance, we are the first, to our knowledge, to study—both theoretically and empirically—the revenue effects of making such audit rules explicit and known to taxpayers.

Our analysis builds on the classic theory of threshold-based audit rules (Reinganum and Wilde, 1985; Sánchez and Sobel, 1993), which, in a stylized environment with one-dimensional heterogeneity and no private information for the Authority, implies that optimal audit risk typically falls above a cutoff. We extend this insight to multiple sources of heterogeneity and evasion margins and allow the Authority to observe a signal of firms' income. We show that a disclosed discontinuous rule can outperform nondisclosure and derive testable conditions under which it raises tax collection. Relatedly, Blinder and Rosen (1985) and Hungerman (2023) use numerical exercises to show that notches can outperform linear policies, including in enforcement. We complement these studies with a theoretical characterization of disclosure and of the revenue consequences of varying the audit discontinuity. This characterization applies more broadly to tax and non-tax notches (Kleven and Waseem, 2013; Pollinger, 2025; Traxler, Westermaier, and Wohlschlegel, 2018; Lockwood, 2020; Bachas and Soto, 2021; Best and Kleven, 2018; Ruh and Staubli, 2019; Liu et al., 2021; Jysmä, Benzarti, and Harju, 2026). When responses to variation in notch incentives are observable, concavity connects a local policy perturbation to the desirability of a wider discontinuity, indicating whether policymakers should introduce or enlarge a notch.

Empirically, we relate to two literatures on threshold-based audit rules that operate through distinct mechanisms. The first studies size-dependent enforcement, including large taxpayer units and regimes in which crossing a size threshold triggers more intensive enforcement (Almunia and Lopez-Rodriguez, 2018; Bachas, Fattal Jaef, and Jensen, 2019). By directing scarce enforcement toward larger firms, where its returns are highest, these policies induce firms to remain below the threshold. This evidence motivates our focus on whether rule disclosure can strengthen compliance among small businesses despite limited enforcement capacity. A second, more closely related literature studies policies that lower audit risk when taxpayers clear a reporting target (Dwenger et al., 2016; Al-Karablieh, Koumanakos, and Stantcheva, 2021). While these papers document taxpayers' responses to such policies, we develop a theory of their design, characterize when disclosure raises tax collection, and analyze the returns to widening the audit discontinuity. Crucially, we study firm-specific rather than common or category-specific thresholds. This distinction allows the Authority to use its information about individual firms and induce compliance more effectively.

Our paper also complements evidence from randomized tax-enforcement interventions, including field experiments on audit-threat letters and deterrence messages (Slemrod, Blumenthal, and Christian, 2001; Kleven et al., 2011; Pomeranz, 2015; Bérgolo et al., 2017).[7] Many of these interventions are one-off or require repeated enforcement actions and additional audit resources to sustain deterrence. We instead study a low-cost rule that can be embedded in the audit system and uses large-scale administrative data to tailor incentives without expanding audit capacity. This use of taxpayer information connects our analysis to Kuchumova (2017), who studies how information reporting generates income signals that improve audit targeting, and to recent work on data-driven targeting (Battaglini et al., 2023). We emphasize a complementary margin: disclosing how taxpayer information shapes audit risk can itself strengthen compliance and improve the effectiveness of predictive audit tools (Paradisi and Sartori, 2023).

Finally, Italy provides a suitable setting for studying small firm tax compliance due to its large share of small businesses and self-employed, which often display high tax gaps (Arachi and Santoro, 2007). While SeS taxpayers's behavior has been analyzed in both academic and policy work (Santoro, 2008; Santoro and Fiorio, 2011; Santoro, 2017; D'Agosto et al., 2017; Battaglini et al., 2020), this paper is the first to evaluate the overall compliance effects of SeS and to assess the effects of the *reward regime* reform.

2 A model of misreporting with disclosed audit rules

2.1 Firm's Problem and Audit Rules

We analyze a tax evasion model where the economy consists of a "category" of firms, which may represent a sector, location, or any combination of observable characteristics. A continuum of firms earns income y drawn from a compact support $[0, \bar{y}]$ according to CDF $F(\cdot)$, which admits a smooth single-peaked density function $f(\cdot)$ bounded away from zero and which does not vary too rapidly relative to its level.[8] Declared income d is taxed at linear rate τ, and any detected evasion is sanctioned at rate $\gamma > 1$.[9] Firms face an audit schedule $p : D \to [0, 1/\gamma]$ that maps declared income into a probability of being audited. The restriction $p < 1/\gamma$ ensures that the marginal value of evasion is positive, a necessary condition for an interior solution with positive evasion. Given their income y, firms choose their reported income d to maximize

[7] Related field experiments on tax-enforcement communications include Hallsworth et al. (2017) and Bott et al. (2020), as well as work on incentives and third-party reporting (Naritomi, 2019; Kumler, Verhoogen, and Frías, 2020).

[8] We treat income as exogenous, but we prove in Supplementary Material S.3 that production and declaration decisions are separable. Hence, one can interpret y as the optimal income of firms that are heterogeneous in productivity. Formally, we require that, $\inf_y \frac{f'(y)}{f(y)} > -\eta$ for some $\eta > 0$. Intuitively, this assumption implies that the income distribution is relatively dispersed.

[9] We assume that conditional on receiving an audit, the true income is discovered and the penalty γ is imposed without frictions.

$$V(y) = \max_d y - \tau d - \tau\gamma \cdot p(d) \cdot (y-d) - c(y-d) \tag{1}$$

for a smooth manipulation cost function $c : \mathbb{R}^+ \to \mathbb{R}^+$, which satisfies $c(0) = 0, c'(\cdot) \geq 0, c'(0) = 0, c''(\cdot) > 0$, with a stable enough convexity rate.[10] This restriction is the counterpart of the condition on $f(\cdot)$: the latter limits how rapidly the density can fall, while the former limits how rapidly cost curvature can rise. Together they are sufficient but stronger than necessary for the test on the desirability of disclosure.[11]

Since we focus on disclosed threshold-based audit rules, we consider an audit schedule that takes the following form

$$p(d, \hat{y}) = \begin{cases} p_H & d < \hat{y} \\ p_L & d \geq \hat{y} \end{cases} \tag{2}$$

with $p_H \geq p_L$. The case $p_L = p_H$ corresponds to a flat audit rule where all firms in a "category" face the same audit probability. If $p_L < p_H$ the Tax Authority grants an audit discount to all firms that declare (at or) above the threshold level $\hat{y}$. This threshold can be interpreted as a signal received by the Tax Authority regarding a firm's true income, potentially derived from a prediction model. A flat audit probability reflects a situation where audit rules and predicted incomes are not disclosed to taxpayers and kept "secret" or undisclosed. For the moment, we treat $\hat{y}$, p_H, and p_L as exogenous parameters.

To streamline the exposition, the baseline model simplifies the firm's problem by abstracting from certain adjustment margins. However, our main results remain robust under less restrictive assumptions. In Section 2.5, we extend the framework to incorporate: (i) heterogeneous costs of income manipulation; (ii) manipulation of tax base components not directly targeted by the audit rule; (iii) manipulation of the threshold $\hat{y}$; (iv) fixed costs associated with being audited; and (v) a non-constant audit schedule in the absence of disclosure.

We gather all proofs of the results presented in this Section in Appendix A.

2.2 Optimal declarations and comparative statics

If the audit rule were flat (*i.e.*, $p(d) = p$), then the maximization problem (1) would reduce to choosing the amount of tax evasion that solves

$$\max_e e\tau(1 - p\gamma) - c(e).$$

[10] The inequality $c''(\cdot) > 0$ is assumed to hold uniformly: there exists a positive constant $\delta > 0$ such that $c''(x) > \delta$ for all $x \in \mathbb{R}^+$. Convexity is a natural assumption that ensures a well-behaved declaration problem. To preserve a stable convexity rate, we assume that $\sup_x c'''(x)/c''(x) < \eta$ with $\eta > 0$, ruling out cases in which evasion is very inelastic.

[11] Indeed, the test relies on concavity of the tax base, which depends on the combination of the two restrictions rather than on each separately. Section 2.4 derives the weaker joint condition.

This holds regardless of the firm's income level y. The interior solution to this problem is

$$e^I(p) = (c')^{-1}(\tau(1-p\gamma)), \tag{3}$$

which represents the level of evasion chosen by firms for which $y - e^I(p) \geq 0$.[12] Since audit rules in (2) are non-constant, (3) does not directly apply to our setting. However, we can still use this condition to characterize the declaration behavior. Specifically, a firm's behavior is characterized by three target evasion levels, e_H, e_L and $\tilde{e}$, where $e_i = e^I(p_i)$ in equation (3) for $i = H, L$, and $\tilde{e}$ solves

$$e_H\tau(1-p_H\gamma) - c(e_H) = \tilde{e}\tau(1-p_L\gamma) - c(\tilde{e}). \tag{4}$$

Thus, e_i is optimal evasion under a flat audit probability p_i. The level $\tilde{e}$ is the evasion of the marginal buncher: the firm at the boundary of the bunching region, with income $y^{HB} = \hat{y}+\tilde{e}$, that is indifferent between bunching at $\hat{y}$, concealing $\tilde{e}$ and facing p_L, and declaring below the threshold, concealing e_H and facing p_H. We can now characterize the declarations of firms solving (1) under the audit rules in (2).

Proposition 1. *$e_L \geq e_H \geq \tilde{e}$, with strict inequalities whenever $p_H > p_L$. If $e_H < \hat{y} + \tilde{e}$, the optimal declaration strategy partitions incomes into the following declaration areas*

$$d(y) = \begin{cases} 0 & y < y^{0H} \\ y - e_H & y^{0H} \leq y < y^{HB} \\ \hat{y} & y^{HB} \leq y < y^{BL} \\ y - e_L & y \geq y^{BL} \end{cases}, \qquad e(y) = \begin{cases} y & y < y^{0H} \\ e_H & y^{0H} \leq y < y^{HB} \\ y - \hat{y} & y^{HB} \leq y < y^{BL} \\ e_L & y \geq y^{BL} \end{cases},$$

where $y^{0H} = e_H$, $y^{HB} = \hat{y} + \tilde{e}$, and $y^{BL} = \hat{y} + e_L$.

The values e_i and $\tilde{e}$, determined by the audit probabilities and cost function through (3) and (4), induce the partition in Proposition 1. Firms with $y < e_H$ would optimally declare $y - e_H < 0$ and therefore bunch at zero, forming the income region $\mathcal{O}$. Firms in income region $\mathcal{H} = [e_H, \hat{y}+\tilde{e})$ declare as under a flat p_H rule. For those below $\hat{y}$, this behavior follows because all feasible declarations face p_H. Those above $\hat{y}$ could instead bunch at the threshold, thereby reducing both evasion and audit risk; $\tilde{e}$ determines when bunching becomes preferable to the interior p_H choice. Firms in region $\mathcal{B} = [\hat{y}+\tilde{e}, \hat{y}+e_L)$ bunch at $\hat{y}$, while those in $\mathcal{L} = (\hat{y}+e_L, \bar{y}]$ declare as under a flat p_L rule, disciplined only by manipulation costs.[13]

[12] By the properties of c, we know that e^I is defined over positive domain, has positive image, and is increasing.
[13] The Proposition assumes that $e_H < \hat{y}+\tilde{e}$, which ensures that the region $\mathcal{H}$ is well-defined. If this condition

Proposition 1 allows us to analyze how firms' optimal declaration behavior responds to changes in the model's parameter, particularly the tax rate and the propensity to evade. To compare economies with varying propensities to evade, we parametrize the cost function as $c_{\kappa}(\cdot) = \kappa \cdot c(\cdot)$, where κ shifts the cost of evading at any evasion level. A lower κ implies a higher propensity to evade. Since we do not observe the evasion levels e_H, e_L and $\tilde{e}$ for each taxpayer in the data, we derive comparative statics on the masses of the declaration areas. These allow us to empirically test the model's predictions. We denote $M(i)$ the mass of area $i = \mathcal{O}, \mathcal{H}, \mathcal{B}, \mathcal{L}$. The following Proposition outlines the main comparative statics under the assumption that all regions have positive mass.

Proposition 2. *Let $\tilde{\tau} = \frac{\tau}{\kappa}$. Then, $\frac{dM(\mathcal{O})}{d\tilde{\tau}} > 0$, $\frac{dM(\mathcal{L})}{d\tilde{\tau}} < 0$. Moreover, there always exists a threshold $\bar{m}$ such that if $M(\mathcal{L}) \geq \bar{m}$, then $\frac{dM(\mathcal{B})}{d\tilde{\tau}} > 0$.*

Given the definition of $\tilde{\tau}$, the comparative statics can be interpreted both as the effect of increasing the tax rate or as the effect of decreasing the propensity to evade (*i.e.*, increasing κ). The results for the $\mathcal{O}$ and $\mathcal{L}$ areas follow directly from the analysis of the evasion behavior. An increase in $\tilde{\tau}$ raises the incentives to misreport income, resulting in $\frac{\mathrm{d}e_H}{\mathrm{d}\tilde{\tau}}, \frac{\mathrm{d}e_L}{\mathrm{d}\tilde{\tau}}, \frac{\mathrm{d}\tilde{e}}{\mathrm{d}\tilde{\tau}} > 0$. Since the $\mathcal{O}$ area includes all firms with $y < e_H$, $M(\mathcal{O})$ increases as $\tilde{\tau}$ rises. Conversely, the $\mathcal{L}$ area, which consists of firms with $y > e_L$, decreases when $\tilde{\tau}$ increases.

The behavior of the bunching region is more complex. An increase in e_L expands the upper bound of the bunching region, while an increase in $\tilde{e}$ raises the lower bound, narrowing it. To assess the net effect on $M(\mathcal{B})$, we must weigh the positive effect of increasing e_L against the negative effect of increasing $\tilde{e}$, taking into account the densities at each bound. The proof shows that when $M(\mathcal{L})$ is sufficiently large,[14] the densities at both bounds are similar. In this case, it is sufficient to show that e_L grows faster than $\tilde{e}$. This result holds because changes in $\tilde{\tau}$ have a greater effect on the optimal declaration in $\mathcal{L}$ (see (3)) than on the indifference condition (4), where both sides of the equality shift in the same direction.

2.3 Government revenues and their response to audit reforms

As an initial step in studying the desirability of disclosed rules, we write the Tax Authority's revenues as a function of the policy parameters and we study how they are affected by the jump in probability at $\hat{y}$. We set aside revenues from audits and assume that the Authority allocates a limited amount of resources to audits within the sector. These assumptions are motivated by our focus on small businesses, where audit collections contribute negligibly to government revenues due to scarce resources and where audits are less effective because

does not hold, taxpayers would switch from declaring zero to bunching at $\hat{y}$, which contradicts our data where we observe interior (non-zero) declarations below the threshold.

[14] This occurs if and only if $\tilde{\tau}$ is small.

of the limited potential gains. In this context, the Authority operates with a fixed audit threshold $\hat{y}$ and adjusts the size of the audit probability gap around it, starting from a baseline μ. To formalize this, we parametrize the audit rule (2), assuming the Authority selects $\Delta \in [0, \min\{\mu, 1-\alpha\mu\}]$ and sets $p_L = \mu - \Delta$ and $p_H = \mu + \alpha\Delta$. If μ is small, it acts as the upper-bound on Δ. The parameter α determines how much p_H increases for each unit decrease in the audit probability above the threshold. The resulting revenue function is:

$$R(\Delta) = \mathbb{E}_y [d(y, \Delta)] = \mathbb{E}_y [y] - \mathbb{E}_y [e(y, \Delta)].$$

We derive the following representation to show how widening the audit probability gap (*i.e.*, increasing Δ) impacts the Authority's objective.

Proposition 3. *The marginal revenue from a change in Δ is*

$$\frac{\mathrm{d}R(\Delta)}{\mathrm{d}\Delta} = -\frac{\mathrm{d}\tilde{e}}{\mathrm{d}\Delta}(e_H - \tilde{e}) f(\hat{y} + \tilde{e}) - M(\mathcal{H})\frac{\mathrm{d}e_H}{\mathrm{d}\Delta} - M(\mathcal{L})\frac{\mathrm{d}e_L}{\mathrm{d}\Delta}. \tag{5}$$

As $\Delta \to 0$, all terms in the sum remain bounded. In particular, $\lim_{\Delta\to 0}\frac{d\tilde{e}}{d\Delta} = -\infty$ but is stabilized by the term $e_H - \tilde{e}$.

Marginal revenue combines two intensive margins and one extensive margin. Firms remaining in $\mathcal{H}$ and $\mathcal{L}$ adjust evasion as their audit probabilities change. Each response is weighted by the region's mass. The response in $\mathcal{H}$ raises revenue, while the lower audit probability in $\mathcal{L}$ increases evasion and reduces it. On the extensive margin, a wider gap increases the audit discount from bunching, causing $\tilde{e}$—and hence the cutoff $y^{HB} = \hat{y} + \tilde{e}$—to fall and drawing additional firms into bunching. Each marginal buncher raises its declaration by $e_H - \tilde{e}$, so the revenue gain equals the cutoff movement times the density at the cutoff and this declaration gain, as captured by the first term in (5).

To understand the limit result, let $e_0 = e^I(\mu)$ and write the marginal buncher's income as $y^{HB} = \hat{y} + e_0 - \bar{g}$, where $\bar{g} = e_0 - \tilde{e}(\Delta) > 0$. Under the flat rule, this firm would declare $\hat{y} - \bar{g}$, so $\bar{g}$ is its shortfall from the threshold. Bunching requires reducing evasion by $\bar{g}$ to reach a suboptimal declaration. Because e_0 is optimal under the flat rule, the first-order payoff cost of this adjustment is zero: higher taxes are offset at the margin by lower expected sanctions and manipulation costs. The resulting loss is therefore quadratic in $\bar{g}$. By contrast, the audit discount produces a first-order benefit in Δ, equal to $\tau\gamma(1+\alpha)e_0\Delta$. The Appendix derives the marginal buncher's indifference condition equating marginal benefits and costs local to $\Delta = 0$:

$$\underbrace{\frac{1}{2}c''(e_0)\bar{g}^2}_{\text{loss from closing the shortfall}} \approx \underbrace{\tau\gamma(1+\alpha)e_0\Delta}_{\text{audit reward}}. \tag{6}$$

Hence, $\bar{g} \propto \sqrt{\Delta}$. Since $\tilde{e} = e_0 - \bar{g}$ and e_0 is constant in Δ, we obtain that $\tilde{e} \propto -\sqrt{\Delta}$. It follows that the bunching boundary moves rapidly as the audit jump opens near the flat rule because $\mathrm{d}\tilde{e}/\mathrm{d}\Delta \propto -1/\sqrt{\Delta} \to -\infty$ as $\Delta \downarrow 0$.

This rapid movement of the bunching boundary does not make marginal revenue unbounded as the declaration gain of the marginal buncher, $e_H - \tilde{e}$, vanishes at the same rate, stabilizing the product. Indeed, by smoothness of the interior problem, $\mathrm{d}e_H/\mathrm{d}\Delta$ remains bounded as $\Delta \downarrow 0$. Thus, to first order, the movement in $\tilde{e}$ determines both the rate, $1/\sqrt{\Delta}$, at which the bunching region expands, and the rate, $\sqrt{\Delta}$, at which the declaration gain contracts. Although these effects offset in the extensive-margin component of marginal revenue, they drive the concavity result in the next section.

2.4 Concavity and the reward-regime test

We use the curvature of the response in equation (5) to connect the reward-regime experiment to the flat-rule counterfactual. To make the dependence on the flat-rule benchmark explicit, write revenues as $R(\Delta; \mu)$.

Theorem 4. *There exists $\mu^\star > 0$ such that, for every $\mu < \mu^\star$, $R(\Delta; \mu)$ is strictly concave in Δ on $[0, \mu]$:*

$$0 < \Delta \leq \mu \leq \mu^\star \implies R_{\Delta\Delta}(\Delta; \mu) < 0.$$

The Theorem's concavity result makes the natural experiment of the *reward regime* informative about the unobserved flat-rule counterfactual, as formalized in the following corollary.

Corollary 5. *If $0 < \Delta_1 < \Delta_2 \leq \mu < \mu^\star$, then $R(\Delta_2; \mu) > R(\Delta_1; \mu) \implies$ implies that $R(\Delta_1; \mu) > R(0; \mu)$.*

The reform provides a natural experiment comparing a smaller pre-reform audit gap with a larger post-reform gap, both under disclosed rules. If this increase raises revenue, concavity implies that the pre-reform gap already outperformed the zero-gap benchmark.

We establish Theorem 4 by differentiating (5) and obtaining, locally around $\Delta = 0$, a representation of the form $R_{\Delta\Delta}(\Delta; \mu) = -G(\mu)/\sqrt{\Delta} + \rho_R(\sqrt{\Delta}, \mu)$. Smooth intensive responses in $\mathcal{H}$ and $\mathcal{L}$ enter the bounded remainder ρ_R, whereas the marginal buncher's response generates the component $-G(\mu)/\sqrt{\Delta}$, which diverges as $\Delta \downarrow 0$. Under our assumptions, there is a sufficiently small region $0 < \Delta \leq \mu \leq \mu^\star$ in which $G(\mu) > 0$ and ρ_R is uniformly bounded. Therefore, for each fixed μ, revenue is infinitely concave as $\Delta \downarrow 0$. Choosing $\mu^\star$ sufficiently small ensures that the negative divergent term dominates the remainder throughout this region.

Concavity is generated by two forces. As Δ grows, the bunching boundary reaches firms further below the threshold. These firms must reduce evasion by more to bunch, making the audit discount less effective. At the same time, declarations in the high-audit region rise, reducing the revenue gain when a marginal firm switches to bunching. Both forces lower the extensive-margin return to widening the audit gap.

Two forces can offset this decline. Denote $t(\mu) = \hat{y} + e^I(\mu)$ the location of the marginal buncher when $\Delta = 0$. A negative density slope $f'(t(\mu))$ means that the expanding bunching region draws in increasingly many firms. Moreover, if $c''' > 0$—an increasing convexity arising for instance from constraints on the evasion capacity—the marginal buncher faces a lower manipulation-cost curvature as Δ grows, making bunching less costly.

Collecting all these forces, we show that $G(\mu)$ has the sign of

$$S(\mu) = \frac{3}{2} + e^I(\mu)\frac{f'(t(\mu))}{f(t(\mu))} - \frac{e^I(\mu)}{2}\frac{c'''(e^I(\mu))}{c''(e^I(\mu))}. \tag{7}$$

Our restrictions on the density slope and the growth of the cost curvature ensure $S(\mu) > 0$, and thus $G(\mu) > 0$. The next subsection explains why this condition is plausible empirically.

Finally, we show that the revenue gain obtained through disclosure can be achieved while saving audit resources.

Remark 6. If μ is small and $\alpha F(t(\mu)) < [1 - F(t(\mu))]$, then the disclosed rule uses a smaller enforcement budget than the flat rule.

Holding flat-rule declarations fixed, opening an audit gap raises the audit probability by $\alpha\Delta$ for the mass $F(t(\mu))$ below the threshold and lowers it by Δ for the mass $1 - F(t(\mu))$ above it. The mechanical marginal budget effect is therefore $\alpha F(t(\mu)) - [1 - F(t(\mu))]$. Unlike revenue, the movement of the bunching boundary has no first-order effect on the enforcement budget because the difference in audit probabilities across the two regions, $(1 + \alpha)\Delta$, vanishes as $\Delta \downarrow 0$.[15] The budget-saving condition therefore requires the mass-weighted audit increase below the threshold to be smaller than the reduction above it. In our application, the *reward regime* expands audit protection above the threshold, where most firms report in our data, so the budget can fall even if $\alpha > 1$ and p_H rises more than p_L falls. Thus, stronger compliance incentives need not require greater enforcement spending: the Authority can sharpen incentives by reallocating audit risk across the threshold while using fewer audits overall. Our structural estimates show that the *reward regime* operates along this margin, raising revenues while reducing the enforcement budget.

[15] The bunching region expands at rate $\sqrt{\Delta}$, but the audit-probability difference across regions is $(1 + \alpha)\Delta$. Its budget effect is therefore of order $\Delta^{3/2}$, whose derivative vanishes as $\Delta \downarrow 0$.

2.4.1 Empirical Relevance of the Concavity Result

Theorem 4 guarantees a nondegenerate region over which revenue is concave but does not establish its empirical relevance. We therefore quantify $\mu^\star$ under standard parametric assumptions and show that concavity holds for empirically plausible baseline audit probabilities across a broad range of primitives.

Let manipulation costs be isoelastic, $c(e) = \frac{\kappa}{1+1/\varepsilon} e^{1+1/\varepsilon}$, and income lognormal, $\log y \sim N(\mu_y, \sigma^2)$. Normalize income by the threshold: $w = y/\hat{y}$, so that $\log w \sim N(m_w, \sigma^2)$, with $m_w = \mu_y - \log \hat{y}$, and the normalized threshold is one. As $\Delta \downarrow 0$, the marginal buncher is located at $\hat{t}(\mu) = 1 + \frac{e^I(\mu)}{\hat{y}}$. Condition (7) then becomes[16]

$$\underbrace{2 - \frac{1}{2\varepsilon}}_{\text{cost curvature}} + \frac{1 - \hat{t}(\mu)}{\hat{t}(\mu)} \left[\overbrace{1 + \frac{\log \hat{t}(\mu) - m_w}{\sigma^2}}^{\text{density}} \right] > 0. \tag{8}$$

The first term captures the manipulation-cost curvature and increases with ε, the elasticity of evasion with respect to its net return. When ε is low, inducing firms farther from the threshold to bunch becomes progressively cheaper, potentially undermining concavity. Available estimates suggest that such low responsiveness is unlikely. They range from 0.3 to 0.5 and 0.9 for the self-employed (Bosch and de Boer, 2019; Heim, 2010), and from 0.53 to 0.56, with a cross-country mean of 0.79, for small businesses (Devereux, Liu, and Loretz, 2014; Agostini et al., 2026).[17] Even the lowest estimate therefore exceeds the $\varepsilon > 1/4$ threshold for a positive cost-curvature term.

The second term captures how the density of firms changes as the bunching boundary moves. Because $\hat{t}(\mu) > 1$, it is positive when $\hat{t}(\mu) < \exp(m_w - \sigma^2)$, so that the flat-rule marginal buncher lies below the mode. In that region, the boundary reaches progressively fewer firms as it moves toward lower incomes. Above the mode, the number of newly induced bunchers may instead grow and weaken concavity. This concern is limited in our application: more than half of firms report at or above the threshold and must therefore have income at or above it.

Determining $\mu^\star$ also requires the divergent term in G to dominate the remainder. We assess it numerically. Figure 1 reports $\mu^\star$ across (ε, π_0), where $\pi_0 = F_Y(\hat{y} + e^I(0))$ is the share of firms whose flat-rule declarations lie below $\hat{y}$. Because more than half of the firms

[16]See the Supplementary Material for the derivation.

[17]These studies estimate the elasticity of reported taxable income with respect to $1 - \tau$, whereas ε measures the elasticity of evasion with respect to $\tau(1 - \gamma p)$. If γ and p are fixed and reported-income responses operate mainly through evasion rather than earned income, these estimates provide a lower bound for ε.

in our data report at or above their threshold, we consider $\pi_0 \in [0.20, 0.50]$.[18] We find that $\mu^\star$ is large across a broad range of elasticities and marginal-buncher locations and often reaches the feasibility ceiling $1/[\gamma(1+\alpha)]$, imposed by $p_H < 1/\gamma$. Concavity therefore holds for reasonable baseline audit probabilities. The bound becomes smaller when ε is low or the marginal buncher lies high in the income distribution, a combination inconsistent with the available evidence and our data.

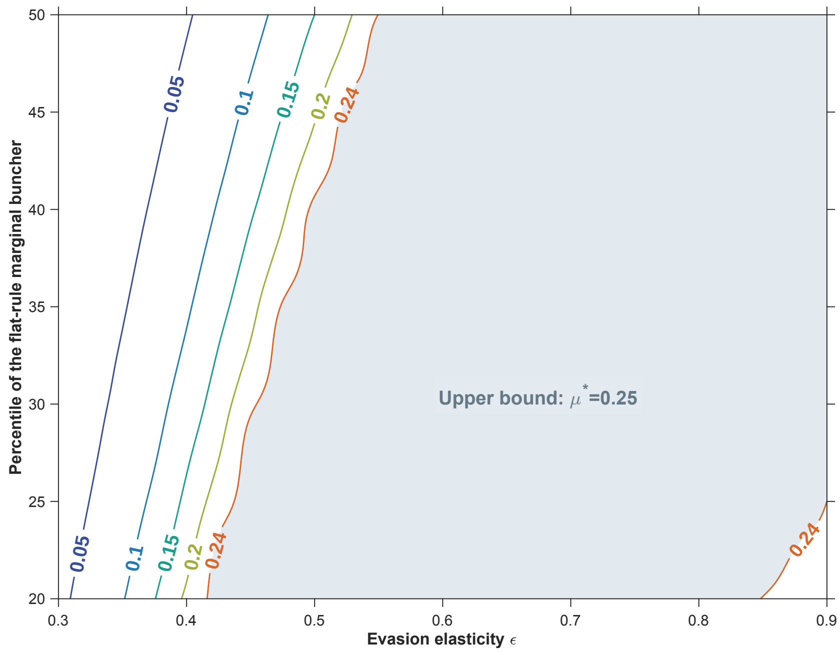


Figure 1: Radius $\mu^\star$ where concavity is ensured. Fixed values: $\hat{y} = 1$, $\sigma = 0.50$, $e^I(0)/\hat{y} = 0.30$, $\tau = 0.25$, $\gamma = 2$, and $\alpha = 1$.

2.5 Extensions and robustness

The baseline model isolates the concavity mechanism using one report and a fixed threshold. Appendix B and C extend the test to an economy that more closely matches the SeS. Firms differ in true revenues and costs, (y, x), and in their manipulation-cost type, κ. They choose evasion $\mathbf{e} = (e_y, e_x)$, generating declarations $\mathbf{d} = (d_y, d_x)$, to solve

$$\max_{\mathbf{e}} \; \tau \left[1 - \gamma p(\mathbf{d}; \Delta)\right] E(\mathbf{e}) - p(\mathbf{d}; \Delta)\omega - C(\mathbf{e}, \kappa),$$

[18] For each ε, we choose $\kappa(\varepsilon) = \tau/(0.30\hat{y})^{1/\varepsilon}$ to keep $e^I(0)/\hat{y} = 0.30$. This allows ε to vary the curvature of manipulation costs while holding flat-rule evasion fixed (at a plausible level of 30% the threshold). For each π_0, we set $m_w = \log(1.30) - \sigma\Phi^{-1}(\pi_0)$, which places the flat-rule marginal buncher at percentile π_0. Throughout, we normalize income so that $\hat{y} = 1$ and fix the dispersion of normalized log income $\sigma = 0.50$.

where $E(\mathbf{e}) = \eta'\mathbf{e}$ is the evaded tax base, C is the manipulation cost that depends on both chosen margins and the manipulation type κ, and ω is a fixed cost of being audited.[19]

The audit rule combines a smooth audit schedule, p_0, that decreases with declarations and a disclosed discontinuity:

$$p(\mathbf{d}; \Delta) = p_0(\mathbf{d}; \mu) + \begin{cases} \alpha\Delta, & d_y < \hat{Y}(d_x), \\ -\Delta, & d_y \geq \hat{Y}(d_x). \end{cases}$$

As in the SeS, the eligibility threshold depends linearly on reported costs, $\hat{Y}(d_x) = \hat{y}_0 + \hat{y}_1 d_x$. The baseline model in Section 2.1 follows by retaining only revenue heterogeneity, suppressing the cost-reporting margin, and setting $\hat{Y} \equiv \hat{y}$, $E(\mathbf{e}) = e_y$, $\omega = 0$, and $p_0 \equiv \mu$.

The extended formulation captures five features of the empirical setting. First, firms can evade profit taxes by understating revenues or overstating deductible costs. Second, because reported costs affect the threshold, firms can manipulate both their tax base and their eligibility for the audit discount. They choose not only whether to bunch but also where on the eligibility frontier to locate, so different firms may bunch at different revenue thresholds. With a fixed threshold, by contrast, cost reporting affects only the evaded base protected by the discount. Third, a fixed audit cost captures the burden of undergoing an audit and strengthens incentives to qualify for lower audit risk. Fourth, heterogeneity in κ rationalizes the absence of a missing mass immediately below the eligibility frontier in the data: high-cost firms remain below it while lower-cost firms bunch. The general cost function also accommodates substitution or complementarity between reporting margins. Finally, p_0 allows perceived audit risk to decline smoothly with declarations even without a disclosed discontinuity.

2.5.1 Concavity in the extended model.

Appendix C extends Theorem 4 to this richer economy: for sufficiently small baseline audit probabilities and gaps, aggregate declared revenues and profits remain concave in Δ. The two baseline forces remain: as Δ expands the bunching set, the new marginal buncher is costlier to attract both because it must reduce more evasion to bunch and because it already declares more in its high-audit counterfactual. Both forces reduce the marginal return to widening the gap.

Several extensions strengthen these decreasing returns.[20] An untargeted cost margin en-

[19] The profit-tax case corresponds to $\eta = (1, 1)$, while $\eta = (1, 0)$ applies when only revenues are taxed; intermediate values allow for incomplete cost deductibility. The analysis also permits smooth nonlinear E under the conditions in Appendix A.

[20] In Supplementary Material we analyze several special cases nested within the general framework of Sec-

larges the evaded base protected by the audit discount, a fixed audit burden adds a benefit of eligibility independent of that base, and a cost-dependent threshold lets firms qualify by adjusting both reports. Each makes a given gap attractive to firms farther from the frontier. Any increase in the gap therefore reaches firms that must forgo more evasion to bunch, causing the reward to decline more rapidly. A smoothly declining audit schedule weakens, but does not eliminate, this mechanism. Because additional evasion raises audit risk even without a discontinuity, firms evade less and the jump protects a smaller evaded base. This attenuation is modest when audit capacity is limited, because audit risk must decline gradually across a wide range of declarations and is therefore locally flat. With heterogeneous manipulation costs, we apply the test conditional on κ and aggregate the resulting type-specific contributions. Heterogeneity changes their relative strength but not their direction.

As in the baseline model, density slopes around the marginal buncher and changes in cost curvature may also matter and in some cases work against concavity. Appendix A derives the corresponding conditions for concavity, and our calibrated model that features most of these extensions verifies numerically that concavity continues to hold.

3 Small businesses evasion and the Italian Sector Studies

A substantial share of tax liabilities goes uncollected due to evasion by small businesses. In Italy, estimates suggest that between 40% and 70% of income earned by individual businesses is not reported (Galbiati and Zanella, 2012; MEF, 2023). Comparable levels of underreporting are observed among similar taxpayer groups in Denmark and the UK (Kleven et al., 2011; HRMC, 2021). Figure 2 compares evasion rates as a share of collected revenues for small and large businesses across four countries: the US, Italy, UK, and Australia. In all cases, small businesses exhibit significantly higher evasion rates.[21] Both Italy and the US lose around 7% of their collected revenues to small business evasion, compared to just 1–2% from larger firms. Moreover, in every country except the UK, small business evasion exceeds the revenue losses from multinational profit shifting (Wier and Zucman, 2022). In the US, for example, small business evasion is almost six times larger than profit shifting. In sum, evasion by small businesses accounts for substantial revenue losses and exceeds in magnitude more widely studied forms of tax evasion and avoidance.

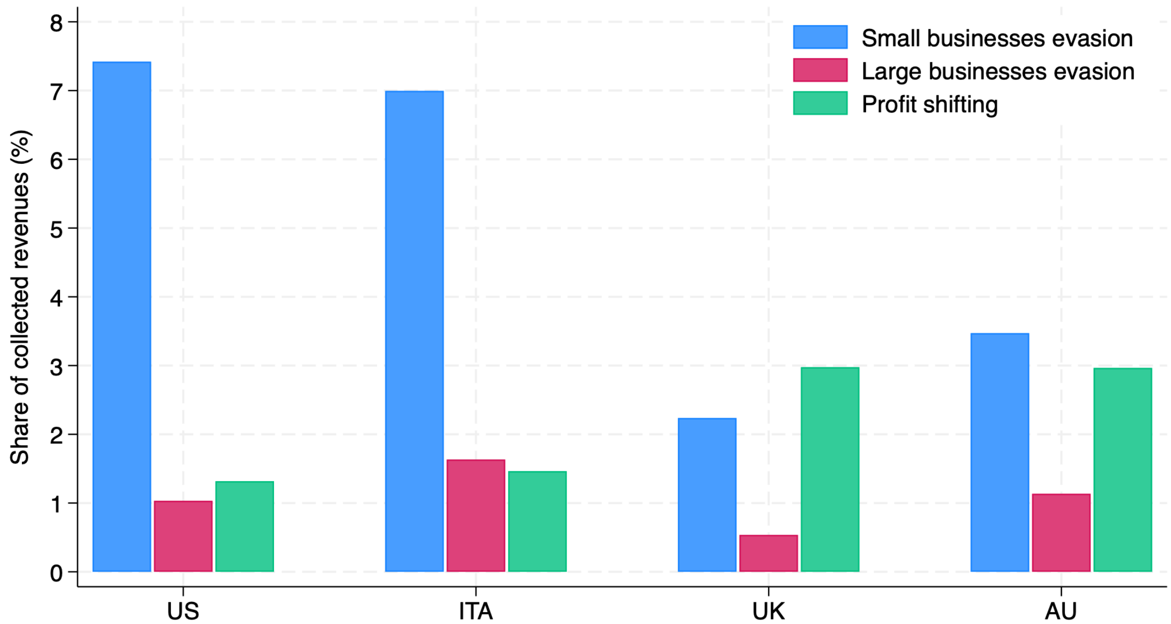
Disclosing audit rules: In response to large tax gaps, the Italian government introduced SeS in 1998, an auditing tool targeting non-employee taxpayers with revenues up to 5.2€ million. Since then, individual businesses, partnerships, and small corporations file an annual

tion 2.5.

[21] Our definition of small businesses includes the self-employed, passthrough businesses, and corporations below a certain level of assets, consistent with the taxpayer population covered by Italy's Sector Studies.

SeS tax form and are subject to audits based on the information provided.

Figure 2: Tax gaps for small and large businesses across countries



Notes: the Figure compares tax gaps between small and large businesses in four countries for 2019. Business income tax gaps are based on official reports from national tax authorities. Details on the definitions of small and large businesses, harmonization of the different sources, and methodology can be found in Supplementary Material S.4. Estimates of profit shifting by multinationals are from Wier and Zucman (2022). All quantities are expressed as a share of total collected central or federal government revenues.

SeS generate a file-specific discontinuity in audit risk based on reported revenues. Each year, the Italian Revenue Agency, in collaboration with a publicly-owned company called SOSE, estimates sector-specific linear models of presumed revenues using detailed information on turnover, costs, labor inputs, capital, and business characteristics. Taxpayers report these inputs before filing, and the model produces a firm-specific presumed revenue threshold. Under Law 146/1998, declarations below this threshold may trigger a tax assessment, as explicitly stated in the law's opening provision.[22]

The timing and transparency of SeS create strong incentives for taxpayers to align reported revenues with the presumed threshold. Production decisions for a given tax year are finalized several months before the filing season, during which taxpayers fulfill their tax and SeS obligations. Filing deadlines typically occur between June and September—roughly six months after production—leaving limited scope for real production adjustments.

[22]The opening statement of Law 146/1998 makes it explicit: *"Tax assessments based on Sector Studies [...] shall apply to taxpayers [...] when declared revenues or remunerations are less than the revenues or remunerations which may be determined on the basis of such Studies"*.

Taxpayers can learn their SeS thresholds at no cost. Each year, between February and May, the Revenue Agency releases *Gerico*, a free software that guides tax filing and computes the firm-specific presumed revenue based on the sectoral models estimated by SOSE. By entering their accounting and structural information, taxpayers observe their threshold before submitting their file and can adjust their declarations accordingly. Awareness of *Gerico* is widespread, as its release is publicly announced.[23] Detailed technical documentation on the estimation procedure is also published annually by the Revenue Agency.

SeS is just one of several compliance tools available to the administration. Taxpayers may still be audited for reasons unrelated to SeS, including through investigations by the *Guardia di Finanza*, which focuses on tax crimes.[24] Crucially for our analysis, the residual audit risk unrelated to SeS is smooth around the threshold. Moreover, crossing the presumed revenue level yields no direct fiscal benefit other than a lower audit probability, allowing us to attribute observed reporting responses to the incentives created by SeS.

***Reward regime* as a test for improved compliance:** In Section 2, we developed a test to assess whether disclosed threshold rules can improve compliance, based on perturbations to the audit probability jump at the threshold. We operationalize this test using a reform of the SeS system. Starting in 2011, Law Decree 201/2011 reinforced the incentive discontinuity associated with SeS by introducing the *regime premiale* (*reward regime*), which expanded audit protections for compliant taxpayers. Table G1 summarizes the compliance requirements and benefits before and after the reform.

The *reward regime* was rolled out gradually across SeS sectors at the start of each tax season, with the Revenue Agency publishing the list of eligible sectors annually based on observable indicators of sector performance and compliance. The reform granted exemptions from additional investigations beyond SeS and reduced the statute of limitations for audits by one year. As a result, taxpayers who complied with SeS prescriptions by declaring revenues at or above the presumed level ("congruence") and satisfying sector-specific accounting indicators ("normality" and "coherence") faced an almost zero probability of being audited. At the same time, the reform encouraged stricter enforcement among non-compliant businesses.

4 Data and Descriptive Facts

4.1 Administrative Data on Sector Studies

To analyze taxpayer behavior under SeS, we use the universe of administrative SeS files from 2007–2010, complemented by an unbalanced panel through 2016 of taxpayers filing

[23]Google searches for "gerico" peak around tax seasons (Supplementary Material S.7).

[24]Most standard file audits, however, are conducted by the Revenue Agency, which is also the only entity authorized to impose additional tax payments.

continuously between 2008 and 2010. To our knowledge, this is the first paper to exploit the full universe of SeS filings in a given year. The data cover nearly 26.7 million declarations by more than 4.7 million Italian micro-businesses and self-employed individuals, with over 3.4 million files per year between 2007 and 2010.[25]

The files contain detailed information on economic activity, including reported revenues and profits, industry codes, and business location.[26] Crucially, each file reports the exact SeS presumed revenue threshold, allowing us to measure the distance between declared revenues and the threshold disclosed prior to filing.

SeS cover firms with diverse legal statuses. Individual businesses and self-employed professionals account for 64.8% of files, partnerships for 19.5%, and corporations for 15.7%. Legal status and location determine the applicable tax regime: individuals and partnerships pay personal income taxes, while corporations are subject to corporate income taxation.

Additional sources: We complement the administrative SeS data with several additional sources.[27] These include measures of tax evasion across regions and tax bases drawn from the literature and administrative reports, over 620,000 anonymous evasion reports from *www.evasori.info* used to construct misreporting proxies for 2008–2011, local tax rates from the Ministry of Economy and Finance, and input–output tables from ISTAT to classify industries as upstream or downstream.

4.2 Bunching at the SeS threshold and tax manipulation

We begin the empirical analysis with descriptive evidence on reporting behavior in the SeS system. This section documents the extent of bunching at the disclosed revenue threshold, examines whether bunching correlates with evasion proxies and manipulation incentives as predicted by the model, and tests whether bunching reflects misreporting rather than production responses.

Measuring bunching: Figure 3 Panel A shows the distribution of reported revenues relative to the presumed threshold $\hat{y}$ using the universe of SeS files from single-sector businesses between 2007 and 2010. Reported revenues are expressed as a percentage distance from $\hat{y}$. The distribution exhibits a pronounced spike within 1 percentage point of the threshold, indicating that many taxpayers declare at or just above $\hat{y}$ to reduce audit risk.

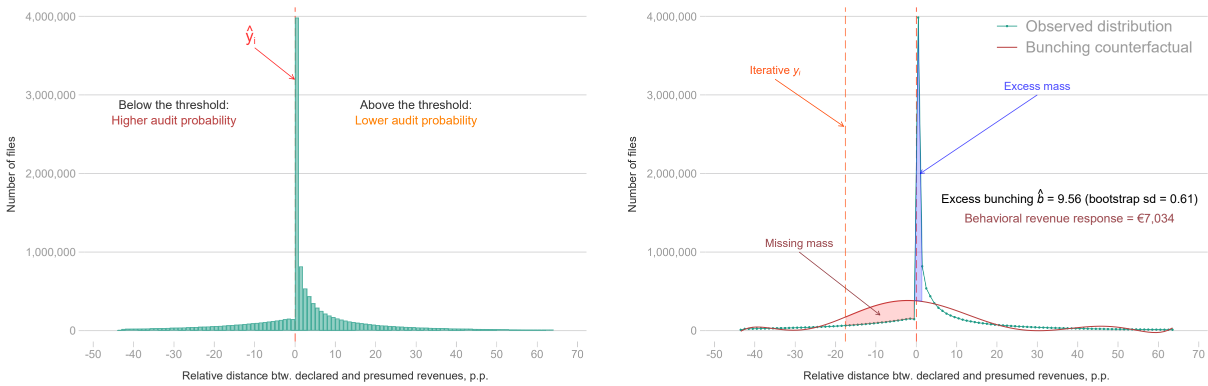
[25] See Supplementary Material S.5 for more details.

[26] Most filers are single-establishment businesses: over 98% remain in the same province throughout 2007–2010. We can assign 95% of files to one of 110 provinces and 77% of personal-income taxpayers to more than 8,000 municipalities, aggregated into 686 local labor markets defined by ISTAT.

[27] We detail these sources in Supplementary Material S.6.

Figure 3: Bunching in the universe of single-sector SeS filers, 2007-2010

(a) Reported revenues relative to SeS threshold **(b)** Bunching estimates



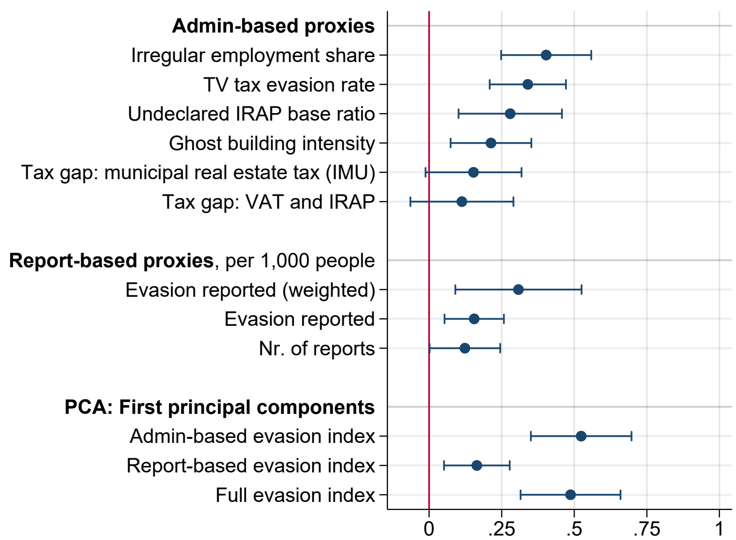
Notes: The figure shows the distribution of $(d_i - \hat{y}_i)/\hat{y}_i$, the relative distance between reported revenues d_i and presumed revenues $\hat{y}_i$, for SeS files from single-sector businesses in 2007–2010. The horizontal axis measures percentage deviations from $\hat{y}$. To avoid irregularities in the tails, we exclude observations below the 5th and above the 95th percentile of relative distance. Panel (a) displays the observed histogram. Panel (b) overlays the estimated bunching counterfactual and reports the corresponding estimates. The counterfactual density is obtained using an iterative procedure that equates excess mass above the threshold with missing mass below it, defining a lower bound y^l. The smooth fit is estimated with a 7th-order polynomial in bin order, with the upper bound set at the threshold bin (within 1 percentage point above $\hat{y}$). Excess bunching is the ratio of excess mass to the counterfactual density at the threshold. Standard errors use 1,000 bootstrap replications. Revenue responses are estimated separately using threshold distance in euros with €500 bins.

To quantify this bunching, we construct a counterfactual distribution of declarations that would arise under a constant audit probability p_L, exploiting the fact that firms declaring above $\hat{y}$ face p_L. Following Kleven and Waseem (2013), we estimate the counterfactual by fitting a polynomial to the distribution above the threshold and defining a threshold y^u that bounds the bunching region to the right of $\hat{y}$. Excess bunching is then measured as the ratio of observed to counterfactual mass within the bunching region. We obtain standard errors via a bootstrap with 1,000 replications (see Appendix D for more details).

Figure 3 Panel B illustrates the resulting estimate. The counterfactual closely tracks the empirical distribution above $\hat{y}$, while declarations below the threshold display a clear missing mass. In our baseline specification, excess bunching equals 9.56 (bootstrap s.d. $= 0.61$). The implied revenue increase relative to the counterfactual corresponds to a rightward shift of the counterfactual distribution equal to 1.13% of mean revenues and 3.05% of median revenues. Table G2 shows that these results are robust to alternative polynomial orders and choices of y^u. Our baseline estimate is conservative, as we attribute excess mass to SeS incentives only within 1 percentage point of $\hat{y}$. Overall, the magnitude of bunching indicates that the disclosed audit rule is salient and generates economically meaningful behavioral responses.

Finally, two pieces of evidence suggest that bunching reflects reporting behavior rather than production adjustments. First, bunching is extremely sharp at the threshold, which would require implausibly precise foresight about $\hat{y}$ given the several-month gap between production decisions and threshold disclosure, and would be difficult to reconcile with reasonable output elasticities (Best et al., 2015). Second, if production were adjusting to SeS incentives, declarations should increasingly align with the threshold following periodic model updates through firms' learning.[28] Exploiting sector-specific revisions to the SeS models, which occur every three years, we find no evidence of such learning dynamics (Appendix E).

Figure 4: Provincial bunching correlates positively with local evasion



Notes: The figure plots standardized coefficients β and 95% CIs from regressions of SeS bunching on evasion proxies Evasion_i^j across 110 provinces: $\text{Bunching}_i = \alpha + \beta \text{Evasion}_i^j + \gamma \log \text{VA pc}_i + \text{macroregion}_i + \varepsilon_i$. Standard errors are robust to heteroskedasticity. Bunching is computed at the province level as described in Section 4.2. Evasion proxies and sources are listed in Appendix S.6. The last three proxies correspond to the first principal components of administrative-based, report-based, and all proxies. Regressions using report-based proxies are weighted by the number of evasion reports (2008–2011).

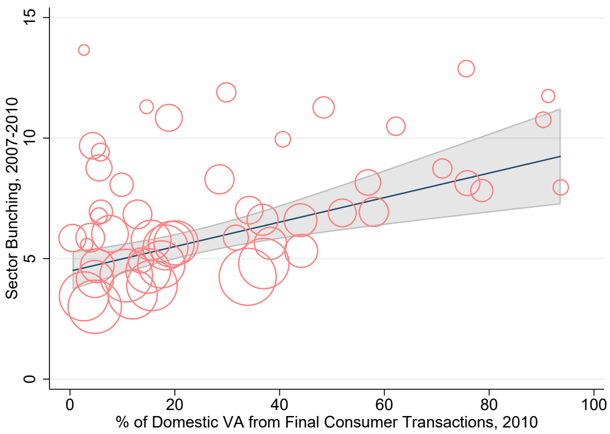
Bunching and evasion behavior: Next, we investigate whether bunching correlates with attitudes toward evasion and incentives to evade, as predicted by Proposition 2. We study the relationship between bunching in SeS files at the level of 110 Italian provinces or 686 local labor markets (LLMs) over 2007–2010 and a range of local evasion proxies across different tax bases, controlling for regional fixed effects and value added per capita.[29]

[28]Model updates for each sector involve re-estimating the presumed revenue functions using updated data, potentially impacting the selection of variables and the size of their associated coefficients.

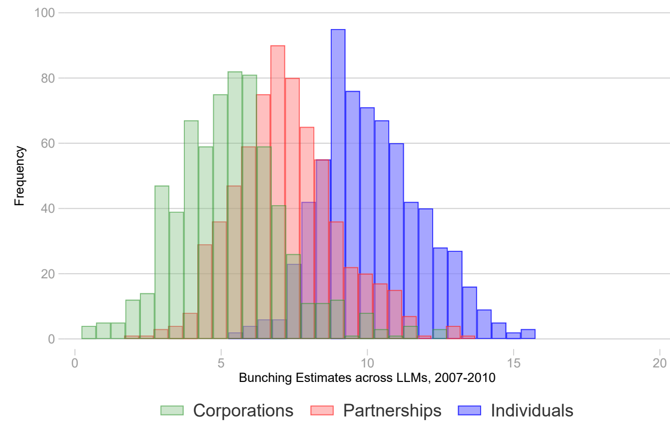
[29]Supplementary Material S.8 provides further details on this exercise and additional correlations.

Figure 5: Bunching tracks evasion potential: downstream sectors

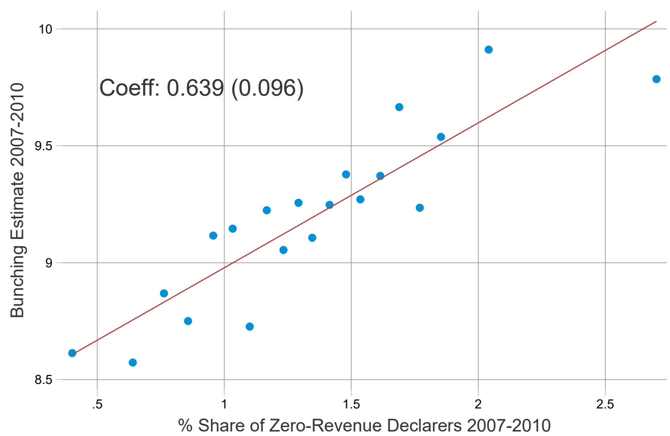
(a) Panel A: Correlation with share of sales to final consumers



(b) Panel B: Bunching across taxpayers with different legal complexity



(c) Bunching and share of zero declarers (province-level)



Notes: Panel (a) shows the sector-level relationship between 2007–2010 bunching and exposure to final consumers in 2010, defined as the share of domestic value added driven by final consumption (Appendix S.5). The sample includes 51 ATECO sectors present in both the SeS data and ISTAT 2010–2013 input–output tables. When SeS sectors aggregate multiple 2-digit sectors, bunching is computed as a file-weighted average. Regressions weight sectors by mean presumed revenues. The shaded area shows the 95% confidence interval. The slope (robust s.e.) is 5.085 (1.106). Panel (b) plots the distribution of local labor market (LLM) bunching by legal form: individuals, partnerships, and corporations, excluding 4% of estimates that are negative or above the 99th percentile. Panel (c) relates LLM-level bunching to the share of zero-revenue declarers using the specification
$\text{Bunching}_i = \alpha + \beta \text{Share zero declarers}_i^j + \gamma \log(\text{PIT base per taxpayer}_i) + \text{region}_i + \varepsilon_i$. The share of zero declarers is the 2007–2010 fraction of SeS filers reporting exactly zero revenues.

Figure 4 shows that bunching is positively and often significantly correlated with twelve distinct evasion measures, including a proxy constructed from 620,000 anonymous online evasion reports. To summarize the magnitudes, we build an evasion index using the first principal component of administrative proxies. A one–standard deviation increase in this index is associated with a 0.5–standard deviation increase in bunching, indicating a strong link between bunching intensity and evasion behavior.

These correlations may reflect both lower costs of misreporting (lower κ) and higher returns to evasion (higher τ). We provide evidence consistent with these mechanisms. Bunching is larger in environments where misreporting is easier: downstream sectors with limited third-party reporting and smaller, less complex legal forms (Figure 5). Moving from fully upstream sectors to those selling exclusively to final consumers is associated with a statistically significant 5-point increase in bunching, relative to an average just above 9 (Panel (a)). Likewise, the distribution of bunching among the self-employed—who face lighter accounting requirements—lies almost entirely to the right of that for corporations (Panel (b)).

Finally, we study the relationship between bunching and the share of zero-revenue declarers. Proposition 2 implies that when common factors such as evasion costs or tax rates affect both $M(\mathcal{O})$ and $M(\mathcal{B})$, the two masses should be positively correlated. Figure 5 Panel (c) confirms this prediction: a 1 percentage point increase in the share of zero declarers is associated with a 0.5-point increase in bunching, relative to a mean of 9.2 across LLMs.

5 Reward System: testing the effectiveness of disclosed audit rules

In this section, we exploit a natural experiment to assess the effects of disclosure-based policies through the lens of our model, and to study how audit rule disclosure affects a broader set of compliance margins. The staggered introduction of the 2011 *reward regime* closely mirrors our theoretical policy experiment, in which disclosure alters reporting incentives in opposite directions depending on a taxpayer's position relative to the threshold.

Beginning in 2011, the Italian government expanded audit exemptions for taxpayers complying with SeS prescriptions while intensifying enforcement for non-compliant firms. As a result, audit risk perceptions diverged on the two sides of the presumed revenue threshold. Taxpayers planning to report revenues at or above the threshold and satisfying the required accounting indicators received stronger audit protection, implying $p_{L,\text{Reward}} \leq p_{L,\text{Pre-Reward}}$. Conversely, non-compliant taxpayers faced higher audit risk. Taken together, these changes increased the perceived audit risk discontinuity at the threshold, so that $\Delta^{\text{Reward}} \geq \Delta^{\text{Pre-Reward}}$.

We leverage the staggered inclusion of SeS sectors into the *reward regime* between 2011 and 2016 to identify the reform's effects. Our analysis focuses on 155 treated sectors spanning manufacturing, commerce, services, and selected professional activities, and uses a balanced panel of businesses filing continuously under SeS from 2007 to 2016.[30] Since each sector s

[30] By 2016, most sectors outside the professions were covered, while only three of the twenty-four professional

entered the regime in a specific tax year t, we estimate a staggered difference-in-differences specification of the following form:

$$y_{s,t} = \lambda_s + \gamma_t + \sum_{q=-k}^{+k'} \beta_q \cdot I\left(Q_{s,t} = q\right) + \sum_{r=2007}^{2016} \delta_r \cdot X_s \cdot I\left(t = r\right) + \varepsilon_{s,t}. \quad (9)$$

For each sector-by-tax year outcome $y_{s,t}$ analyzed below, coefficients β_q quantify the effect of the *reward regime* during period q relative to the sector's entry year. These effects are identified under the assumption of parallel paths, meaning that, in the absence of the reform, outcomes in treated sectors would follow similar trends to those in not-yet-treated sectors. We include sector and tax year fixed effects (λ_s and γ_t), along with interactions between tax-year dummies and a vector X_s of pre-treatment sector characteristics capturing the Revenue Agency's criteria for selecting sectors into the *reward regime*.[31] By including the latter, we allow outcomes to follow differential time trends predicted by the criteria used to select sectors into the regime. This reduces concerns that our estimates reflect pre-existing trends correlated with the selection criteria. We further address this concern by documenting flat pre-treatment dynamics in the event study. Moreover, conditional on observables predicting sector entry, the pattern of estimated responses is consistent with the treatment effect: outcomes shift discretely in the first post-entry year and then grow monotonically over the following six years, a trajectory more consistent with gradual behavioral adaptation to the larger audit-risk discontinuity than with underlying reporting trends.

We weight sectors by the number of SeS files submitted at the start of the sample period, so that estimates reflect the behavior of the average taxpayer in our dataset. Standard errors are clustered at the sector level. Additionally, we conduct robustness checks using the de Chaisemartin and D'Haultfœuille (2020) correction, which addresses potential biases in two-way fixed-effects models when treatment effects are heterogeneous.

5.1 Distribution shifts caused by the introduction of the Reward System

Disclosure-based policies, such as the *reward regime*, shift perceived audit risk asymmetrically by lowering it above the disclosed threshold and raising it below. Consequently, reporting incentives change in opposite directions on either side of the threshold, and increased bunching may reflect taxpayers adjustments of opposite signs.

We begin by analyzing how small businesses adjust their declarations based on their position relative to the threshold. Taxpayers are grouped into six groups according to their

SeS sectors had been included (Figure G1).

[31] Controls include dummies for manufacturing, commerce, services, and professions, along with 2007–2010 averages of revenues, gross profits, employment cost incidence, and growth rates of employment costs and revenues.

distance from the presumed revenue threshold $\hat{y}$ in the year before their sector's reform: within 5, 5–10, and more than 10 percentage points below or above $\hat{y}$. Using the staggered rollout of the *reward regime*, we estimate equation (9) separately for each group, using the ratio of declared revenues to the firm-specific threshold as the outcome.[32]

Figure 6 Panel A confirms that small businesses react to the change in audit incentives, in line with our theoretical predictions. Taxpayers facing higher audit risk below the threshold increase compliance by declaring more, while those enjoying greater protection above the threshold reduce compliance. Firms originally reporting below the threshold raise their declarations by between 10 and 35 percentage points of $\hat{y}$, whereas those above reduce them by about 20 percentage points. As predicted, responses are more pronounced for firms farther from the threshold. These taxpayers are more likely to remain on their original side and can adjust more freely without crossing the threshold. Moreover, they also need to make larger adjustments to reach the threshold if they choose to bunch, further amplifying their observed response. Panel B shows that the same pattern appears for declared profits, which are the relevant tax base but are not directly targeted by the revenue-threshold rule. This is relevant because firms could in principle offset higher revenue declarations by adjusting reported costs. Instead, profits rise below the cutoff by roughly 10 to 20 percentage points of $\hat{y}$ and fall above the cutoff by about 15 percentage points. Thus, despite potential adjustment on the cost margin, the tax base displays a response similar to declared revenues.

Additional evidence of these opposing compliance responses emerges when using a compliance indicator produced by the Tax Authority and linked to each firm's tax declaration. This evidence is shown in Panel C. This binary index is designed to isolate the group of taxpayers whose declarations align with the expectations of the Tax Authority. Firms initially below the threshold increase their probability of compliance by between 20 and 40 percentage points, whereas those above the threshold see a decrease of about 30 percentage points.

5.2 Improvements over flat rules: the effects of the Reward System

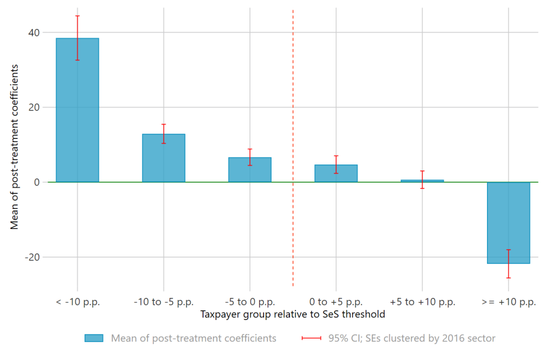
We now quantify the total period-by-period effects of the reward system on several reported margins, interpreting the results through the lens of our theoretical model. Figure 7, Panel A, displays the full set of β_q coefficients from (9), where the dependent variable is mean reported revenues by sector and tax year, expressed in logarithms. Prior to the reform, treated and control sectors follow similar reporting trends. After the reform, reported revenues in treated sectors rise sharply—by an average of 2.4% in the first year, reaching about 20.4% by year

[32] Since grouping is based on pre-reform declarations, concerns about mean reversion may arise. However, our difference-in-differences framework leverages untreated sectors as controls, allowing us to net out these dynamics.

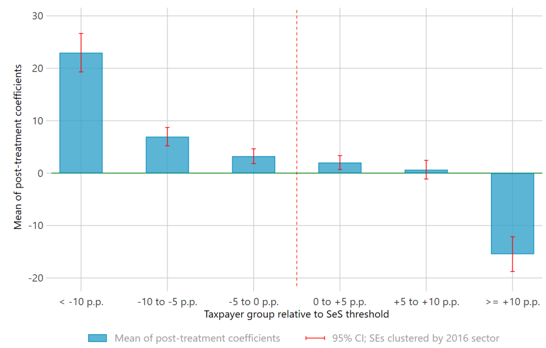
six. The steady growth of the coefficients suggests that it may take time for businesses to fully adjust to the new system.

Figure 6: *reward regime* Responses Around the Threshold

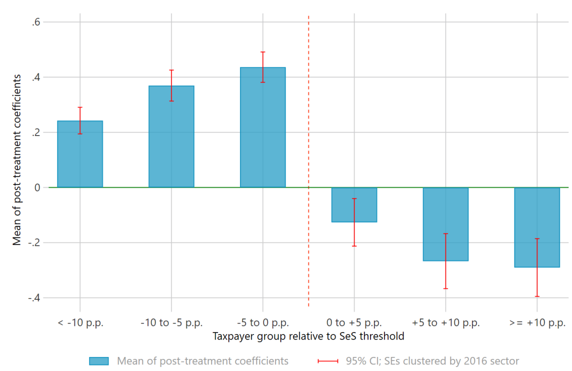
Panel A: Revenues as share of $\hat{y}$



Panel B: Profits as share of $\hat{y}$



Panel C: Compliance indicator

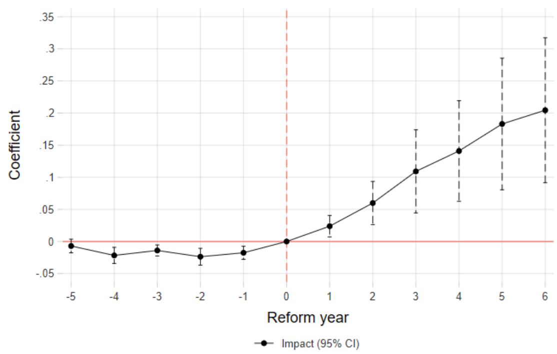


Notes: this Figure illustrates the effect of the *reward regime* on taxpayer declaration behavior above and below the presumed revenue threshold. Panels A-C show responses of taxpayers classified into six groups based on their distance from the threshold in the year prior to the regime's introduction. The dependent variables are: (i) revenues as a share of the firm-specific presumed revenues $\hat{y}$ (Panel A), (ii) profits as a share of $\hat{y}$ (Panel B), and (iii) a compliance indicator reflecting the alignment of an individual declaration with benchmarks set by the Tax Authority (Panel C). In all panels, bars represent the average of six group-specific post-treatment coefficients from a staggered difference-in-differences based on the specification in (9). Whiskers represent 95% CIs of these linear combinations of coefficients. Standard errors are clustered at the sector level. The regressions are estimated on the sample of all SeS files from single-sector taxpayers continuously filing over the 2007-2016 period, aggregated by sector-year. Only sectors accessing the *reward regime* by 2016 are considered. Number of sector-years: 1550. Declared revenues as a share of $\hat{y}$ are winsorized at the 99th percentile, while profits as a share of $\hat{y}$ are winsorized at the 1st and 99th percentiles.

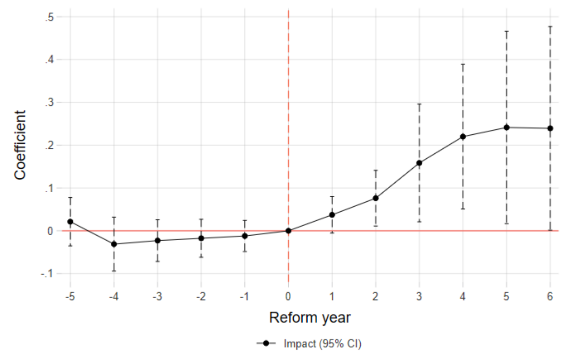
Panel B shifts the focus to gross profits, capturing the net effect on the tax base. As with revenues, the stronger audit incentives introduced by the *reward regime* appear to have substantially increased the reporting. On average, firms in treated sectors report gross profits 16.2% higher each year than those in sectors yet to adopt the reform. Cumulatively, this corresponds to an additional €33,671.77 in taxable profits per average treated business.[33]

Figure 7: *reward regime* effects on mean revenues and profits

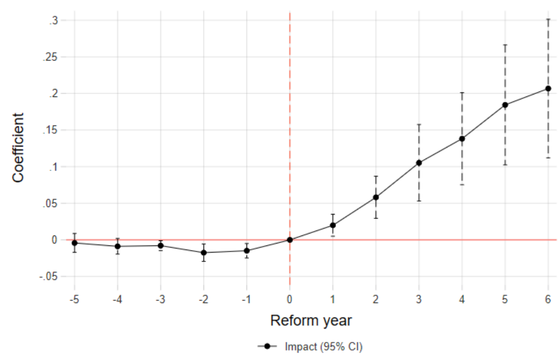
(a) Log-Mean reported revenues



(b) Log-Mean reported income



(c) Log-Mean reported log-costs



Notes: The figure shows the effects of the *reward regime*'s introduction on mean reported revenues (Panel (a)), mean gross profits (Panel (b)), and costs (Panel (c)), defined as revenues minus profits. Outcomes are in logs. Whiskers denote 95% confidence intervals. Effects are measured relative to the year before sector entry into the regime (year 0, red dashed line). Estimates follow specification (9) with standard errors clustered at the sector level. The sample includes all SeS files from single-sector taxpayers filing continuously between 2007 and 2016, aggregated by sector-year. Only sectors entering the regime by 2016 are included (1,550 sector-years). Reported revenues are winsorized at the 99th percentile.

The increase in profits is smaller than the rise in revenues because the policy targets only the revenue declaration margin while taxpayers can adjust other margins, such as costs, to offset the impact on the tax base. This behavior is consistent with evidence from Carrillo, Pomeranz, and Singhal (2017) in a different context. Our model extension shows that the concavity test remains valid even in this setting. To assess the extent to which inflated costs neutralize part of the increase in revenues, Panel C of Figure 7 examines the difference between revenues and profits, an aggregate proxy for reported costs. Treated sectors report

[33]Figure G2 presents the estimates with outcomes expressed in euros.

average costs 2% higher than control sectors in the first year post-reform, increasing to 20.7% by the final observed year. This suggests that while the reward system boosted reported revenues, it also spurred a proportional increase in reported costs, keeping profit rates stable. Nonetheless, because revenues rose overall, the net effect on the tax base remains positive.

Our results show that both revenues and profits—the relevant tax base—increased after the *reward regime*'s introduction. This has two key implications. First, disclosed audit rules could be locally improved by increasing the probability jump at the threshold. Second, by Theorem 4, the pre-*reward regime* rule outperformed a flat audit rule conducting a strictly larger number of audits. Overall, our evidence supports the desirability of SeS over flat, undisclosed rules in this context.

For consistency with the analysis in 5.1, we use a classical two-way fixed effects model as our baseline specification. To test the robustness of our findings, we also apply the estimator of de Chaisemartin and D'Haultfœuille (2020), which corrects for potential biases from heterogeneous treatment effects across cohorts and time. The results remain consistent, with similar patterns and magnitudes.

6 Checking for improvements and optimal disclosed rules

In this Section, we investigate how to improve on the existing disclosed rules in SeS and consider non-budget neutral policies. To this end, we estimate the primitives of the model using data moments derived from the natural experiment discussed in Section 5.

6.1 Model estimation

We first define the set of primitives to estimate, then we discuss the data moments that we employ in our estimation. Appendix G provides further details on the structural model and the estimation procedure.

Model primitives. Our structural model extends the baseline framework to capture key features of the SeS system: heterogeneous evasion propensities, joint misreporting of revenues and costs, and a disclosed, cost-dependent eligibility threshold.

Each firm has true revenue $y \geq 0$, true cost $x \geq 0$, and evasion propensity $1/\kappa > 0$. We model their joint distribution as lognormal, with seven parameters collected in Σ that allow revenue–cost dependence and correlation between revenue and evasion propensity (Appendix G). Firms choose reported revenue d_y and costs d_x, yielding a reported tax base $\pi_d = d_y - d_x$, taxed at rate τ, with no subsidy for reported losses.

Audit probability drops discontinuously at the threshold $\widehat{y}(d_x) = \widehat{y}_0 + \widehat{y}_1 d_x$. The case $\widehat{y}_1 = 0$ nests the exogenous-threshold baseline, while $\widehat{y}_1 > 0$ links the two reporting margins: inflating costs lowers reported profits but raises the revenue required to qualify for the low-audit regime.

To match the simultaneous increase in reported revenues and costs following the introduction of the *reward regime*, we allow misreporting costs to be non-separable across the two margins:

$$\mathcal{C}(e_y, e_x; \kappa) = (1 - b_2)c(e_y, \kappa) + (b_1 - b_2)c(e_x, \kappa) + b_2 c(e_y + e_x, \kappa),$$

where $c(e, \kappa) = \frac{(\kappa e)^{1+1/\varepsilon}}{\kappa(1+1/\varepsilon)}$. The parameter ε governs the common elasticity, b_1 scales the cost of cost misreporting relative to revenue misreporting, and b_2 captures complementarity or substitutability between the two margins. We impose $b_1 > 0$ and $0 \leq b_2 < \min\{1, b_1\}$.[34]

To capture the policy change induced by the *reward regime*, we allow audit probabilities p_L and p_H to differ before and after the reform. Using the notation of Theorem 4, the two regimes are characterized by (μ, α, Δ) and (μ, α, Δ'), which together with the threshold parameters $(\hat{y}_0, \hat{y}_1)$ define the vector of policy parameters.

The full vector of structural primitives is therefore

$$\theta = (\underbrace{\mathbf{\Sigma}, b_1, b_2, \varepsilon}_{\text{Economy parameters}}, \underbrace{\hat{y}_0, \hat{y}_1, \mu, \Delta, \alpha, \Delta'}_{\text{SeS parameters}}),$$

for a total of 16 parameters. Given these primitives, we can derive firms' evasion behavior under each audit regime, and aggregate across the distribution of firms to compute the moments used to estimate the model.

Theoretical moments and data counterparts. We target moments describing the distribution of declarations around the eligibility threshold and their causal changes under the *reward regime*. For each firm, we compute the implied threshold $\widehat{y}(d_x)$ and normalized declared revenue $\bar{d}_y = d_y/\widehat{y}(d_x)$. In the regions below, around, and above the bunching threshold, we match the pre-reform share of firms and the pre- and post-reform means of $\bar{d}_y$, normalized profits $\bar{\pi}_d = (d_y - d_x)/\widehat{y}(d_x)$, and $\widehat{y}(d_x)$.[35] This yields 20 moments: three region shares and 17 pre- and post-reform means of the three quantities in the three areas.[36]

We construct data counterparts using observed declarations and their distribution around the threshold. Firms are assigned to the three regions based on pre-policy declarations. From these we compute region-level masses and pre-reform averages of declarations. Post-reform moments are obtained by adding the reduced-form treatment effects estimated in Section 5

[34] At the optimum revenues are never over-reported, $e_y \geq 0$. Costs instead can be over-reported ($e_x \geq 0$) to reduce taxable profit, or under-reported to lower the audit threshold and increase eligibility ($e_x < 0$). Hence, we consider $\mathcal{C}(e_y, |e_x|; \kappa)$.

[35] We express quantities relative to $\hat{y}$ to compare firms with different thresholds within a sector.

[36] We exclude the mean of $\bar{d}_y$ in the bunching region because it equals one by construction. Therefore, we have six moments each for $\bar{\pi}_d$ and $\widehat{y}(d_x)$, but only five for $\bar{d}_y$.

to the corresponding pre-reform levels.[37]

Estimation: We estimate the model focusing on downstream sectors and using a simulated method of moments (SMM) in two stages. To reduce the dimensionality of the search, we fix the tax and penalty rates at externally calibrated values. We estimate the remaining primitives by minimizing the mean squared distance between simulated and empirical moments. In the first stage, a diagonal weighting matrix weights reduced-form moments by the inverse of their standard errors and assigns additional weight to changes in declarations. At the resulting estimates, we use repeated simulations to estimate each moment's simulation variance. The second stage uses a diagonal inverse-variance weighting matrix, downweighting moments with greater simulation noise. Details on the estimation procedure are provided in Appendix G.

Table 1: SMM Estimates

Parameter	Value	Description
Manipulation-cost primitives		
ε	0.7426	Manipulation-cost elasticity.
b_1	1.0151	Relative weight of cost-margin manipulation.
b_2	0.9844	Joint / interaction cost weight.
Threshold function		
$\hat{y}_1$	0.6797	Slope of the threshold function.
SeS policy, pre-Reward		
p_H^{pre}	0.1073	Audit probability below $\hat{y}$.
p_L^{pre}	0.0730	Audit probability above $\hat{y}$.
SeS policy, post-Reward		
p_H^{post}	0.2380	Audit probability below $\hat{y}$.
p_L^{post}	0.0053	Audit probability above $\hat{y}$.

Notes: The table reports the structural parameters estimated by SMM for downstream sectors. The manipulation-cost primitives are common across regimes. The audit policy parameters are reported separately before and after the *reward regime*. p_L and p_H denote the audit probabilities above and below the SeS threshold, respectively.

Estimates: Table 1 reports the parameter estimates and key theoretical moments. The estimated manipulation-cost elasticity ε is below one, implying that misreporting responds less than proportionally to audit incentives. Firms therefore adjust to the *reward regime*, but less than proportionally to the decline in the effective tax rate. The estimate is close to the

[37] For any region $\mathcal{A}$, post-reform averages for a given outcome o are $\bar{o}_{\mathcal{A}}^{\text{Post}} = \bar{o}_{\mathcal{A}}^{\text{Pre}} + \delta\bar{o}_{\mathcal{A}}$, where $\delta\bar{o}_{\mathcal{A}}$ is the causal effect of the *reward regime* on quantity o.

cross-country taxable-income elasticity of 0.79 reported by Agostini et al. (2026) and lies within the range of other estimates in the literature.

The cost parameters imply a non-separable reporting technology in which revenue and cost misreporting are almost perfect substitutes: firms offset lower evasion on one margin with greater evasion on the other. This pattern rationalizes the simultaneous increases in reported revenues and costs in Figure 7 and is consistent with Carrillo, Pomeranz, and Singhal (2017), who document that Ecuadorian firms offset enforcement-induced revenue increases through cost inflation. Finally, the estimated slope of the eligibility threshold with respect to reported costs is positive but below one, so higher costs substantially raise the required revenue threshold, though less than one-for-one.

The estimated pre-reform audit probabilities, $p_H = 0.107$ and $p_L = 0.073$, closely match the audit probabilities reported in administrative sources of 0.105 below the threshold and 0.071 above it.[38] Relative to the pre-reform rule, the estimated *reward regime* generates higher declared tax base with a smaller enforcement budget. The estimated post-reform rule sets $p_L = 0$. Although this boundary solution was not imposed in estimation, it aligns with the provision of the *reward regime* granting audit immunity to declarations above the threshold.

Figure F2 verifies numerically that reported revenues and profits are concave in Δ at the estimated parameters, confirming that the conditions underlying Theorem 4 hold in the calibrated model.

Model fit: Appendix Figure F3 summarizes the fit of the calibrated model. The model closely reproduces the pre-reform distribution of firms below, near, and above the eligibility threshold, and it captures the post-*reward regime* response: declared revenues and profits rise for firms initially below the threshold and fall for firms initially above it. The main discrepancy is a simulated revenue response near the threshold that is more muted than its empirical counterpart, while the profit responses remain within the empirical confidence intervals.

6.2 Policy counterfactuals

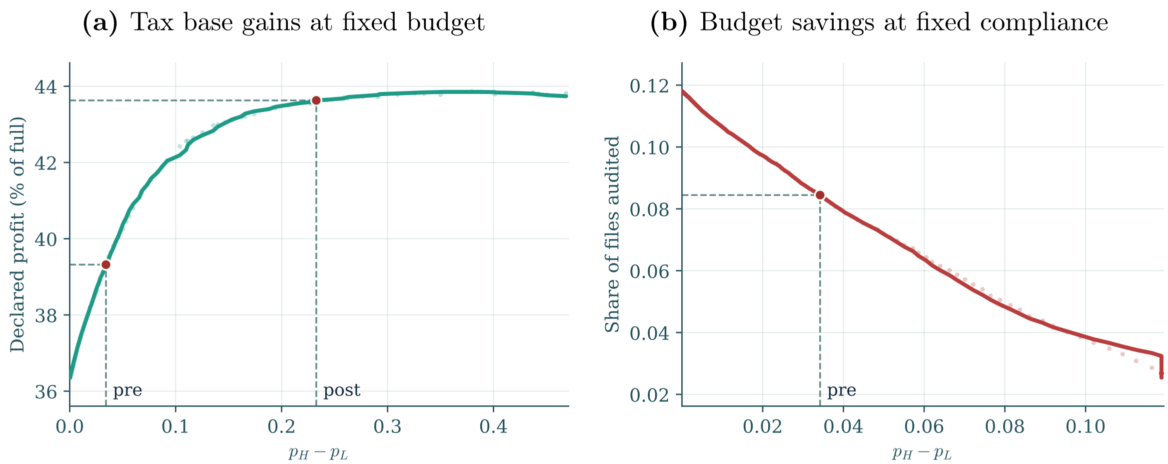
Budget- or tax-base-neutral modifications of the disclosed rule Our natural experiment reveals that the introduction of the *reward regime* led to higher declared revenues and profits. In our structural estimates, we further show that this improvement was accompanied by a reduction in audit costs. Using the calibrated model, we disentangle these two channels by conducting counterfactual policy exercises under alternative constraints. Specifically, we consider (i) policies that keep the total audit budget constant while optimizing the rule

[38] These numbers come from our computations based on data on SeS from D'Agosto et al. (2017).

for tax base collection, and (ii) policies that maintain tax-base performance constant while minimizing audit budget. The first exercise measures how much additional tax base can be collected with the same enforcement resources, while the second quantifies potential budget savings that would deliver the same declared profit target.

Figure 8, Panel A, shows that widening the audit discontinuity can substantially increase declared profits without raising the total audit budget. We construct the curve by varying p_L from its estimated pre-reform level and choosing p_H so that the equilibrium audit budget remains fixed at its pre-reform level. The horizontal axis reports the resulting gap. Declared profits raise in Δ at a decreasing rate. By steepening the audit jump, declared profits can be improved by 9.5% relative to the pre-*reward regime*. At the estimated pre-reform gap of about 0.03, the curve also confirms that the disclosed rule generates more declared profits than a flat rule with the same budget.

Figure 8: Policy counterfactuals: fixing budget and compliance



Notes: The figure shows the revenue gains from alternative policies holding the audit budget at the pre-policy level (Panel (a)), and the budget savings achievable while keeping declared profits constant (Panel (b)). The horizontal axis reports the resulting Δ when fixing p_L and choosing p_H to satisfy the budget constraint (Panel (a)) or maintain profits (Panel (b)). The vertical axis reports average profits as a share of $\hat{y}$ (Panel (a)) and the share of audited taxpayers (Panel (b)). The horizontal dashed lines indicate the pre-policy levels of revenues (Panel (a)) and audits (Panel (b)); their intersection with the curves identifies the pre-policy Δ.

Conversely, Figure 8, Panel B, quantifies potential budget savings while holding declared profits at their pre-reform level. For each audit probability gap, it reports the minimum audit budget required to attain this tax base. A flat rule ($p_H - p_L = 0$) would require an audited share nearly 50% larger than the pre-reform rule. More importantly, widening the gap beyond its pre-reform level could preserve the same tax base while cutting the enforcement budget by more than half. This is not merely an efficiency gain within small-business enforcement: it would allow the Authority to maintain small-business compliance with far fewer audits and redirect scarce capacity toward larger taxpayers, for whom audit returns are higher. Redesigning the audit schedule could therefore improve the allocation of enforcement across taxpayer groups without sacrificing compliance among small firms.

We next characterize the audit rule that optimally balances tax-base collection and enforcement costs and assess how it differs from the current *reward regime*.

Distance from the best disclosed policy We compare the observed and counterfactual policies with the optimal disclosed rule. Let Π denote the tax base (declared profits) and Q the enforcement budget. For each marginal enforcement cost λ, the Authority chooses p_L and p_H to solve

$$V^\star(\lambda) := \max_{p_L, p_H} \left\{ \Pi(p_L, p_H) - \lambda Q(p_L, p_H) \right\}.$$

We compare this optimum with the objective attained under the post-reform, pre-reform, and undisclosed flat rules. The parameter λ captures the marginal cost of enforcement and determines how the Authority trades off tax collection against audit expenditure. It includes both the direct cost of auditing small firms and the opportunity cost of diverting enforcement from larger taxpayers, whose audits yield higher returns.

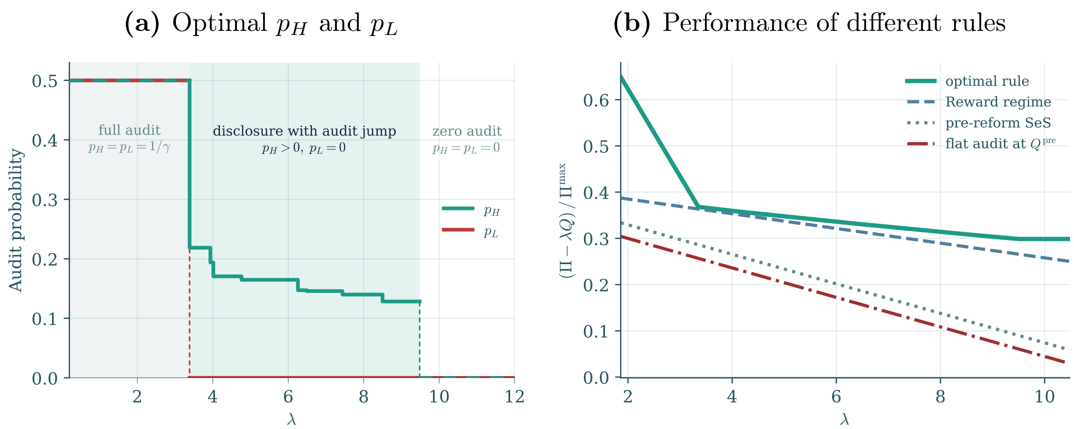
The Authority can widen the audit discontinuity through either a reward or a penalty: lowering p_L strengthens the reward for declaring above $\hat{y}$, whereas raising p_H increases the penalty below it. Both encourage declarations above the threshold, but raising p_H increases audit expenditures, while lowering p_L saves audits at the cost of greater evasion above $\hat{y}$. A higher λ therefore shifts the optimal policy toward the reward channel, freeing resources that could be redirected toward larger taxpayers.

In the unconstrained optimum, when the marginal audit cost λ is sufficiently small, the Authority sets $p_L = p_H = 1/\gamma$, the audit probability that induces truthful reporting from all firms. For intermediate and empirically relevant values of λ, the Authority sets $p_L = 0$, fully eliminating audits for firms declaring above $\hat{y}$, and lets p_H decline with λ (Figure 9, Panel A). In our estimated economy, increasing the discontinuity Δ is always beneficial: marginal revenue gains from a larger Δ dominate potential losses from the $\mathcal{L}$ region, even when audits are relatively cheap and Δ is therefore large. Once this dominance holds at a given λ, it continues to hold for all higher λ', because Δ falls with λ while losses in $\mathcal{L}$ remain roughly constant. For this reason, p_H always lies above p_L for intermediate values of λ. Finally, for very large λ, the optimal policy is to audit no firm.

When the marginal audit cost λ is sufficiently small, the unconstrained optimum sets $p_L = p_H = 1/\gamma$. At this audit probability, evasion yields no return, so firms optimally report their true incomes. For intermediate—and empirically relevant—values of λ, the Authority instead sets $p_L = 0$, eliminates audits above $\hat{y}$, and lets p_H decline with λ (Figure 9, Panel

A). Throughout this range, widening the discontinuity raises the tax base: the gains from stronger bunching incentives exceed the evasion losses in $\mathcal{L}$. As λ rises and the optimal Δ falls, concavity makes the marginal gain from widening the gap larger, while the losses in $\mathcal{L}$ remain approximately constant. Thus, once the gains dominate, they continue to do so at higher λ, implying $p_H > p_L$ throughout the intermediate range. For sufficiently large λ, audit costs dominate and the Authority optimally audits no firms.

Figure 9: Optimal policy: comparison with *reward regime* and pre-reform scenario



Notes: The figure illustrates the optimal policy for different values of the marginal audit cost λ and compares its performance with the *reward regime* and the pre-reform rule. Panel (a) shows the optimal audit probabilities (p_H and p_L) as a function of λ. Panel (b) plots the government objective (declared profits net of audit costs) as a share of $\hat{y}$. Four policies are compared: (i) the optimal rule, (ii) the *reward regime*, (iii) the pre-reward rule, and (iv) an undisclosed rule employing the same audit budget as the pre-*reward regime*.

Panel B compares the observed *reward regime* with the unconstrained optimum and shows that it performs remarkably well. The reform sets p_L at its lower bound and reduces the audit budget—the same adjustment our model identifies as improving the Authority's objective relative to the pre-reform rule. The objective attained under the *reward regime* coincides with the unconstrained optimum near the value of λ for which the *reward regime*'s p_H is optimal. This intersection can be interpreted as the marginal audit cost that rationalizes the *reward regime* as globally optimal. At other values of λ, the *reward regime* remains within 11% of the optimum, whereas the pre-reform rule falls 20–70% below it, particularly when high audit costs make a smaller budget optimal. The large advantage of the *reward regime* over the pre-reform rule in the calibrated model corroborates our reduced-form evidence of a substantial compliance response to the reform. A flat rule performs worse than both, falling 30–80% below the optimum over the displayed range. Together with our reduced-form evidence, these results show that the reform both increased compliance and moved enforcement close to the optimal policy.

7 Conclusions

Tax audits and their threat are a primary enforcement tool across developed and developing countries. The dissuasive power of audits, however, has hardly solved the long-standing problem of low compliance among micro to small businesses and the self-employed. We ask whether the strategic disclosure of audit selection criteria can improve the effectiveness of enforcement among these taxpayers. We answer our question by developing a theoretical model of audit disclosure, and implementing a test derived by the model using a quasi-experiment in the context of Sector Studies (SeS), an Italian policy informing small firms of their relative audit risk around a revenue threshold.

We develop a theory of optimal tax declarations of firms that face a discontinuous threshold-based audit probability, and we derive a test for the existence of improvements over flat audit rules. The test is based on studying the behavior of the revenue function in response to a marginal change in the audit probability jump at the threshold. Consistently with the model, the distribution of SeS files reveals that taxpayers are especially aware of and willing to adjust to clear audit risk signals. The extent of bunching is strongly related to several evasion proxies on other declaration margins, and seems to respond to the incentives to evade and the availability of evasion technologies. To implement the test for improvements over flat rules, we exploit a 2011 staggered reform that strengthened the original risk discontinuity at the disclosed SeS threshold. While taxpayers respond by bunching at the cutoff regardless of their relative position ahead of the reform, mean gross profits rise by 16.2% in treated sectors over the course of six years, suggesting that the pre-reform discontinuity performed strictly better than a counterfactual flat rule that used more audits.

Our work is encouraging as international attention grows on the importance of voluntary tax compliance and reliable tax collection for fiscal sustainability (OECD, 2017; IMF, 2021). Differently from tax lotteries and traditional tax amnesties, the disclosure framework we study grants broadly accessible and stable incentives to stimulate compliance. As tax agencies routinely define thresholds to target their audits, they might develop cost-effective communication strategies to nudge taxpayers around these cutoffs. At the same time, we are aware that net collection effects also depend on the quality of the ensuing audits once the pool of exempted taxpayers is defined. Although such effects are bounded to be of second-order importance in contexts where limited audit resources are allocated, we leave the study of realized audit collection in the presence of threshold-based rules to future research.

References

Agostini, Claudio A., Zareh Asatryan, Laurent Bach, Govindadeva Bernier, Marinho Bertanha, Katarzyna A. Bilicka, Anne Brockmeyer, Jaroslav Bukovina, Guillermo

Falcone, Pablo Garriga, Yuxuan He, et al. 2026. "The Elasticity of Corporate Taxable Income across Countries." National Bureau of Economic Research NBER Working Paper 34945.

Al-Karablieh, Y., E. Koumanakos, and S. Stantcheva. 2021. "Clearing the Bar: Improving Tax Compliance for Small Firms through Target Setting." *Journal of International Economics*, 130.

Alm, J., J. Martinez-Vazquez, and S. Wallace. 2004. *Taxing the Hard-to-tax: Lessons from Theory and Practice.* Vol. 268, Elsevier, Amsterdam.

Almunia, M. and D. Lopez-Rodriguez. 2018. "Under the Radar: The Effects of Monitoring Firms on Tax Compliance." *American Economic Journal: Economic Policy*, 10(1): 1–38.

Arachi, G. and A. Santoro. 2007. "Tax Enforcement for SMEs: Lessons from theItalian Experience?" *eJournal of Tax Research (Michigan Issue)*, 5(2): 225–243.

Australian Taxation Office. 2024. "Tax gap." Accessed August 4, 2025.

Bachas, Pierre and Mauricio Soto. 2021. "Corporate Taxation under Weak Enforcement." *American Economic Journal: Economic Policy*, 13(4): 36–71.

Bachas, P., R. Fattal Jaef, and A. Jensen. 2019. "Size-Dependent Tax Enforcement and Compliance: Global Evidence and Aggregate Implications." *Journal of Development Economics*, 140: 203–222.

Basri, M. C., M. Felix, R. Hanna, and B. Olken. 2019. "Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia." *NBER Working Paper No. 26150.*

Battaglini, Marco, Luigi Guiso, Chiara Lacava, Douglas L. Miller, and Eleonora Patacchini. 2023. "Refining Public Policies with Machine Learning: The Case of Tax Auditing." *Working paper.*

Battaglini, M., L. Guiso, C. Lacava, and E. Patacchini. 2020. "Tax Professionals and Tax Evasion."

Bérgolo, M., R. Ceni, G. Cruces, M. Giaccobasso, and R. Perez-Truglia. 2017. "Tax Audits as Scarecrows: Evidence from a Large-Scale Field Experiment." *NBER Working Paper No. 23631.*

Best, Michael Carlos and Henrik Jacobsen Kleven. 2018. "Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the U.K." *Review of Economic Studies*, 85(1): 157–193.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem. 2015. "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan." *Journal of Political Economy*, 123(6): 1311–55.

Blinder, Alan S. and Harvey S. Rosen. 1985. "Notches." *American Economic Review*, 75(4): 736–747.

Boning, William C., Nathaniel Hendren, Ben Sprung-Keyser, and Ellen Stuart. 2025. "A Welfare Analysis of Tax Audits Across the Income Distribution." *Quarterly Journal of Economics*, 140(1): 63–112.

Bosch, Nicole and Henk-Wim de Boer. 2019. "Income and Occupational Choice Responses of the Self-Employed to Tax Rate Changes: Heterogeneity across Reforms and Income." *Labour Economics*, 58: 1–20.

Bott, K. M., A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden. 2020. "You've Got Mail: A Randomized Field Experiment on Tax Evasion." *Management Science*, 66(7): 2801–2819.

Brosio, G., A. Cassone, and R. Ricciuti. 2002. "Tax Evasion across Italy: Rational Noncompliance or Inadequate Civic Concern?" *Public Choice*, 112: 259–273.

Carfora, A., R.V. Pansini, and S. Pisani. 2016. "Spatial Dynamics in Tax Gap Determinants." *Agenzia delle Entrate, Argomenti di Discussione.*

Carrillo, P. E., D. Pomeranz, and M. Singhal. 2017. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." *American Economic Journal: Applied Economics*, 9(2): 144–64.

Casaburi, L. and U. Troiano. 2016. "Ghost-House Busters: The Electoral Response to a Large Anti Tax Evasion Program." *The Quarterly Journal of Economics*, 131(1): 273–314.

Censis. 2003. "Capire il Sommerso: Supporto Conoscitivo ai Servizi per l'Impiego."

D'Agosto, E., M. Manzo, A. Modica, and S. Pisani. 2017. "Tax Audits and Tax Compliance - Evidence from Italy." *available at https://www.irs.gov/pub/irs-soi/17resconmodica.pdf.*

D'Agosto, E., M. Marigliani, and S. Pisani. 2014. "Asymmetries in the Territorial VAT Gap." *Agenzia delle Entrate, Argomenti di Discussione.*

de Chaisemartin, C. and X. D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–96.

Devereux, Michael P., Li Liu, and Simon Loretz. 2014. "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records." *American Economic Journal: Economic Policy*, 6(2): 19–53.

Dwenger, Nadja, Henrik Kleven, Imran Rasul, and Johannes Rincke. 2016. "Extrinsic and Intrinsic Motivations for Tax Compliance: Evidence from a Field Experiment in Germany." *American Economic Journal: Economic Policy*, 8(3): 203–232.

Galbiati, R. and G. Zanella. 2012. "The Tax Evasion Social Multiplier: Evidence from Italy." *Journal of Public Economics*, 96: 485–494.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev. 2017. "The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance." *Journal of Public Economics*, 148: 14–31.

Heim, Bradley T. 2010. "The Responsiveness of Self-Employment Income to Tax Rate Changes." *Labour Economics*, 17(6): 940–950.

HM Revenue and Customs. 2025. "HMRC tax receipts and National Insurance contributions for the UK: statistics table." Last published July 22, 2025.

HM Revenue & Customs. 2025. "Measuring Tax Gaps Tables 2025." Last updated June 19, 2025. Accessed July 17, 2025.

HRMC. 2021. "Measuring Tax Gaps 2021 Edition - Tax Gap Estimates for 2019 to 2020." *available at https://www.gov.uk/government/statistics/measuring-tax-gaps/measuring-tax-gaps-2021-edition-tax-gap-estimates-for-2019-to-2020.*

Hungerman, Daniel M. 2023. "Tax Evasion, Efficiency, and Bunching in the Presence of Enforcement Notches." *International Tax and Public Finance*, 30: 43–68.

IMF. 2021. *World Economic Outlook: Managing Divergent Recoveries.* Washington, DC, April.

Internal Revenue Service. 2022. "Tax gap estimates for tax years 2014–2016 (and projections for 2017–2019)." Accessed July 17, 2025.

Jysmä, Sami, Youssef Benzarti, and Jarkko Harju. 2026. "Policy Thresholds as Growth Barriers: Theory and Evidence from a Payroll Tax Notch." National Bureau of Economic Research NBER Working Paper 35484.

Kleven, H. and M. Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics*, 128(2): 669–723.

Kleven, H., M. Knudsen, C. Kreiner, S. Pedersen, and E. Saez. 2011. "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica*, 79(3): 651–692.

Kuchumova, Yulia (Paramonova). 2017. "The Optimal Deterrence of Tax Evasion: The Trade-off Between Information Reporting and Audits." *Journal of Public Economics*, 145: 162–180.

Kumler, T., E. Verhoogen, and F. Frías. 2020. "Enlisting Employees in Improving Payroll Tax

Compliance: Evidence from Mexico." *The Review of Economics and Statistics*, 102(5): 881–896.

Lazear, E. 2006. "Speeding, Terrorism, and Teaching to the Test." *The Quarterly Journal of Economics*, 121(3): 1029–1061.

Liu, Li, Ben Lockwood, Miguel Almunia, and Eddy H. F. Tam. 2021. "VAT Notches, Voluntary Registration, and Bunching: Theory and U.K. Evidence." *Review of Economics and Statistics*, 103(1): 151–164.

Lockwood, Ben. 2020. "Malas Notches." *International Tax and Public Finance*, 27(4): 779–804.

MEF. 2023. "Relazione sull'Economia Non Osservata e sull'Evasione Fiscale e Contributiva."

Ministero dell'Economia e delle Finanze. 2022. "Relazione sull'economia non osservata e sull'evasione fiscale e contributiva: Aggiornamenti per gli anni 2015–2020 a seguito della revisione dei conti nazionali apportata dall'ISTAT." Accessed July 17, 2025.

Ministry of Economy and Finance. 2019. "Relazione sull'Economia Non Osservata e sull'Evasione Fiscale e Contributiva."

Naritomi, J. 2019. "Consumers as Tax Auditors." *American Economic Review*, 109: 3031–3072.

OECD. 2006. "Strengthening Tax Audit Capabilities: Innovative Approaches to Improve the Efficiency and Effectiveness of Indirect Income Measurement Methods."

OECD. 2017. "Italy's Tax Administration: A Review of Institutional and Governance Aspects."

OECD. 2025. "Global Revenue Statistics." Last updated July 8, 2025. Accessed August 4, 2025.

Paradisi, Matteo and Elia Sartori. 2023. "Using Prediction Models to Design Tax Enforcement: Incentives vs Targeting." *Working paper.*

Pisani, S. and C. Polito. 2006. "Metodologia di Integrazione tra i Dati IRAP e Quelli di Contabilità Nazionale." *Agenzia delle Entrate, Documenti di Lavoro dell'Ufficio Studi.*

Pollinger, Stefan. 2025. "Kinks Know More: Estimating Intensive and Participation Margin Responses Using Nonlinear Budget Sets." Sciences Po Department of Economics Sciences Po Economics Discussion Paper 2025-09. Revise and resubmit at the Review of Economic Studies.

Pomeranz, D. 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review*, 105: 2539–2569.

Reinganum, J. and L. Wilde. 1985. "Income Tax Compliance in a Principal-Agent Framework." *Journal of Public Economics*, 26(1): 1–18.

Ruh, Philippe and Stefan Staubli. 2019. "Financial Incentives and Earnings of Disability Insurance Recipients: Evidence from a Notch Design." *American Economic Journal: Economic Policy*, 11(2): 269–300.

Sánchez, I. and J. Sobel. 1993. "Hierarchical Design and Enforcement of Income Tax Policies." *Journal of Public Economics*, 50(3): 345–369.

Santoro, A. 2008. "Taxpayers' Choices under Studi di Settore: What Do We Know and How Can We Interpret It?" *Giornale Degli Economisti e Annali Di Economia, Nuova Serie, 67, 2: 161-84.*

Santoro, A. 2017. "Do Small Businesses Respond to a Change in Tax Audit Rules? Evidence from Italy." *Public Finance Review*, 45: 792–814.

Santoro, A. and C. Fiorio. 2011. "Taxpayer Behavior When Audit Rules Are Known: Evidence from Italy." *Public Finance Review, 39:103–123.*

Slemrod, J., M. Blumenthal, and C. Christian. 2001. "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota." *Journal of Public Economics*, 79(3): 455–483.

The World Bank. 2019. "Enterprise Surveys." *available at http://www.enterprisesurveys.org.*

Thuronyi, V. 1996. *Presumptive Taxation.* in Tax Law Design and Drafting, International Monetary Fund.

Traxler, Christian, Franz G. Westermaier, and Ansgar Wohlschlegel. 2018. "Bunching on the Autobahn? Speeding Responses to a "Notched" Penalty Scheme." *Journal of Public Economics*, 157: 78–94.

Vallanti, G. and G. Gianfreda. 2020. "Informality, Regulation and Productivity: Do Small Firms Escape EPL through Shadow Employment?" *Small Business Economics*.

Wier, Ludvig and Gabriel Zucman. 2022. "Global Profit Shifting, 1975-2019." *NBER Working Paper No. 30673*.

Online Appendix

A Proofs of Propositions in Main Text

A.1 Proof of Proposition 1

A firm with income y solves

$$\max_d \left[\max\left\{u_H\left(y,d\right), u_L\left(y,d\right)\cdot \mathbb{I}\left(d > \hat{y}\right)\right\}\right]$$

where

$$u_H\left(y,d\right) = y - \tau d - \tau\gamma p_H\left(y-d\right)^+ - \kappa c\left(y-d\right)$$
$$u_L\left(y,d\right) = y - \tau d - \tau\gamma p_L\left(y-d\right)^+ - \kappa c\left(y-d\right)$$

First notice that for all y, $u_i\left(y,d\right)$ is decreasing in d for $d \geq y$, so no firm over-reports. Maximizing $u_i\left(y,d\right)$ is the same as maximizing $\tilde{u}_i\left(e\right) = \tau\left(1-\gamma p_i\right)e - \kappa c\left(e\right)$. By concavity of the objective, an interior maximum is characterized by the FOC

$$c'\left(e\right) = \frac{\tau\left(1-\gamma p_i\right)}{\kappa}.$$

We denote e_i the solutions to these equations and use convexity of c to conclude that $e_L > e_H > 0$. We need to deal with corner cases. First, notice that $\tilde{u}_L\left(e\right)$ is valid only for declarations $d = y - e$ that lie above $\hat{y}$. Since, for all e, $\tilde{u}_H\left(e\right) < \tilde{u}_L\left(e\right)$, then whenever e_L is feasible (i.e., $y - e_L > \hat{y}$ and therefore $y > \hat{y} + e_L$) it solves the firm's problem.

For firms with $y < \hat{y}$ all candidate declarations (recall that over-reporting is suboptimal) are in the H region. As $\tilde{u}_H\left(e\right)$ is increasing below e_H, whenever e_H is not feasible (*i.e.* when $y - e_H < 0$, with 0 being the lower bound on feasible declarations), then $e = y$ (*i.e.* $d = 0$) is optimal. When instead $y > e_H$, then $d = y - e_H$ is optimal.

We are only left with solving the problem for firms with income $y \in [\hat{y}, \hat{y} + e_L]$. For y in that range, $\hat{y} = \arg\max_d u_L\left(y,d\right)\cdot \mathbb{I}\left(d \geq \hat{y}\right)$ since the unconstrained maximum $y - e_L$ occurs at a point where the function already dropped to 0 and the objective is decreasing in the feasible domain. The maximum of $\tilde{u}_H\left(e\right)$ is instead e_H as this is feasible (recall we assumed $e_H < \hat{y}$). To find the global optimum we therefore need to compare the utility of evading $y - \hat{y}$ and facing p_L and of evading e_H and facing p_H. The former dominates iff

$$\tilde{u}_L\left(y-\hat{y}\right) > \tilde{u}_H\left(e_H\right) \equiv V_H$$

Notice that the RHS is flat in y, while the LHS is contiunous and increasing from $\tilde{u}_L\left(0\right) < V_H$ to $V_L \equiv \tilde{u}_L\left(e_L\right) > V_H$ (the latter inequality following by a simple envelope argument). Therefore there is a unique crossing, which is characterized by $V_H = \tilde{u}_L\left(\tilde{e}\right)$, which yields to the condition

$$\frac{\tau}{\kappa}\left[e_H\left(1-p_H\gamma\right) - \tilde{e}\left(1-p_L\gamma\right)\right] = c\left(e_H\right) - c\left(\tilde{e}\right).$$

Summarizing, firms declare 0 if $y < e_H$, declare $y - e_H$ if $y \in [e_H, \hat{y}+\tilde{e}]$, bunch at $\hat{y}$ if $y \in [\hat{y}+\tilde{e}, \hat{y}+e_L]$, and declare $y - e_L$ above $\hat{y} + e_L$. This solution is valid if $e_H < \hat{y} + \tilde{e}$. Otherwise, the relevant deviation in the H region is to declare 0, there is no interior declarers in the H region and $\tilde{e}$ is defined by

$$\left(\hat{y}+\tilde{e}_0\right)\tau\left(1-p_H\gamma\right) - \kappa c\left(\hat{y}+\tilde{e}_0\right) = \tilde{e}_0\tau\left(1-p_L\gamma\right) - \kappa c\left(\tilde{e}_0\right).$$

A.2 Proof of Proposition 2

We derive first the comparative statics for the three evasion levels $e_H, e_L, \tilde{e}$. The FOC for interior evasions is $c'(e_i) = \tilde{\tau}(1-\gamma p_i)$ for $i = H, L$, from which we have

$$\frac{\mathrm{d}e_i}{\mathrm{d}\tilde{\tau}} = \frac{1-\gamma p_i}{c''(e_i)} = \frac{c'(e_i)}{\tilde{\tau} c''(e_i)} > 0.$$

Manipulating the equation that determines $\tilde{e}$ we obtain

$$\tilde{\tau}\left[e_H(1-p_H\gamma) - \tilde{e}(1-p_L\gamma)\right] = c(e_H) - c(\tilde{e}).$$

Using an envelope argument and rearranging we derive the following

$$\frac{\mathrm{d}\tilde{e}}{\mathrm{d}\tilde{\tau}} = \frac{e_H(1-p_H\gamma) - \tilde{e}(1-p_L\gamma)}{\tilde{\tau}(1-p_L\gamma) - c'(\tilde{e})} = \frac{c(e_H) - c(\tilde{e})}{\tilde{\tau}\left(c'(e_L) - c'(\tilde{e})\right)} > 0,$$

where the second equality uses the FOC for e_L and the definition of $\tilde{e}$, and the last inequality follows from the fact that $e_L > e_H > \tilde{e}$ and from the fact that $c(\cdot)$ and $c'(\cdot)$ are increasing.

Notice that $M(0) = F(e_H)$, $M(\mathcal{L}) = 1 - F(\hat{y} + e_L)$, $M(\mathcal{B}) = F(\hat{y} + e_L) - F(\hat{y} + \tilde{e})$. The comparative statics of e_H, e_L then imply that $M(0)$ is decreasing and $M(\mathcal{L})$ is increasing in $\tilde{\tau}$, strictly if the areas are non-degenerate which requires, respectively, that e_H and $\hat{y} + e_L$ are below $\bar{y}$. This proves the first two statements in the comparative statics part of the Proposition.

For $M(\mathcal{B})$, a quantitative assessment is needed since both the upper bound and the lower bound of the region decrease in $\tilde{\tau}$. The change in the size of the bunching region caused by a change in $\tilde{\tau}$ is

$$\frac{\mathrm{d}M(\mathcal{B})}{\mathrm{d}\tilde{\tau}} = f(\hat{y} + e_L)\frac{\mathrm{d}e_L}{\mathrm{d}\tilde{\tau}} - f(\hat{y} + \tilde{e})\frac{\mathrm{d}\tilde{e}}{\mathrm{d}\tilde{\tau}}.$$

If $\hat{y} + e_L > \bar{y}$ (*i.e.* $M(\mathcal{L}) = 0$) then $f(\hat{y} + e_L) = 0$ and $\frac{\mathrm{d}M(\mathcal{B})}{\mathrm{d}\tilde{\tau}} < 0$ as only the lower-bound of the bunching area increases. Otherwise, the condition $\frac{\mathrm{d}M(\mathcal{B})}{\mathrm{d}\tilde{\tau}} > 0$ is equivalent to $\frac{f(\hat{y}+e_L)}{f(\hat{y}+\tilde{e})} > \frac{\frac{\mathrm{d}\tilde{e}}{\mathrm{d}\tilde{\tau}}}{\frac{\mathrm{d}e_L}{\mathrm{d}\tilde{\tau}}}$. Since $M(\mathcal{L})$ is monotonically decreasing in $\tilde{\tau}$ the condition $M(\mathcal{L}) > \bar{m}$ (stated in the Proposition) is equivalent to $\tilde{\tau} \leq \bar{\tau}$ for $\bar{\tau}$ such that $M(\mathcal{L})(\bar{\tau}) = \bar{m}$. Hence, we equivalently prove that the inequality is satisfied as $\tilde{\tau}$ vanishes, which is

$$\lim_{\tilde{\tau}\to 0}\frac{f(\hat{y} + e_L)}{f(\hat{y} + \tilde{e})} > \lim_{\tilde{\tau}\to 0}\frac{\frac{\mathrm{d}\tilde{e}}{\mathrm{d}\tilde{\tau}}}{\frac{\mathrm{d}e_L}{\mathrm{d}\tilde{\tau}}}.$$

Using the FOC we get that $\lim_{\tilde{\tau}\to 0} c'(e_i) = 0$, implying that e_L, e_H (and a fortiori $\tilde{e}$ which is bounded above by both) converge to zero. This implies that $\lim_{\tilde{\tau}\to 0}\frac{f(\hat{y}+e_L)}{f(\hat{y}+\tilde{e})} = 1$. Regarding the right hand side, we show that its limit converges to a number below 1 by contradiction. We know that for all $\tilde{\tau} > 0$, $0 < \tilde{e}(\tilde{\tau}) < e_L(\tilde{\tau})$. Given its definition, $\frac{\mathrm{d}e_L}{\mathrm{d}\tilde{\tau}}$ is bounded as long as $c''(0) > 0$. Since $e_L > \tilde{e}$, by the fundamental theorem of calculus, $\frac{\mathrm{d}\tilde{e}}{\mathrm{d}\tilde{\tau}}$ is also bounded. Because both $\frac{\mathrm{d}e_L}{\mathrm{d}\tilde{\tau}}$ and $\frac{\mathrm{d}\tilde{e}}{\mathrm{d}\tilde{\tau}}$ are bounded and well-behaved, $\frac{\frac{\mathrm{d}}{\mathrm{d}\tilde{\tau}}\tilde{e}}{\frac{\mathrm{d}}{\mathrm{d}\tilde{\tau}}e_L}$ has a limit q. Suppose $q > 1$, then there would exist $\underline{\tilde{\tau}} > 0$ such that $\forall\tilde{\tau} \in [0, \underline{\tilde{\tau}}]$, $\frac{\frac{\mathrm{d}}{\mathrm{d}\tilde{\tau}}\tilde{e}}{\frac{\mathrm{d}}{\mathrm{d}\tilde{\tau}}e_L} > 1$, which in turns implies that

$$\tilde{e}(\underline{\tilde{\tau}}) = \tilde{e}(0) + \int_0^{\underline{\tilde{\tau}}}\frac{\mathrm{d}}{\mathrm{d}\tilde{\tau}}\tilde{e}(\tilde{\tau})\,\mathrm{d}\tilde{\tau} > e_L(0) + \int_0^{\underline{\tilde{\tau}}}\frac{\mathrm{d}}{\mathrm{d}\tilde{\tau}}e_L(\tilde{\tau})\,\mathrm{d}\tilde{\tau} = e_L(\underline{\tilde{\tau}}).$$

This is however a contradiction. Hence, it must be that $q < 1$ and that $\lim_{\tilde{\tau}\to 0} \frac{\frac{\mathrm{d}\tilde{e}}{\mathrm{d}\tilde{\tau}}}{\frac{\mathrm{d}e_L}{\mathrm{d}\tilde{\tau}}} < 1$.

A.3 Proof of Proposition 3

Using the thesholds on income defined in Proposition 1, we define total revenues as follows

$$R(\Delta) = \int_{y^{0H}}^{y^{HB}} d_H(y) f(y)\,\mathrm{d}y + \int_{y^{HB}}^{y^{BL}} \hat{y} f(y)\,\mathrm{d}y + \int_{y^{BL}}^{\bar{y}} d_L(y) f(y)\,\mathrm{d}y$$

so marginal revenues are

$$\begin{aligned}
\frac{\mathrm{d}R(\Delta)}{\mathrm{d}\Delta} &= -\frac{\mathrm{d}y^{0H}}{\mathrm{d}\Delta}\left[d_H\left(y^{0H}\right)\right] f\left(y^{0H}\right) + \frac{\mathrm{d}y^{HB}}{\mathrm{d}\Delta}\left[d_H\left(y^{HB}\right) - \hat{y}\right] f\left(y^{HB}\right) \\
&\quad + \frac{\mathrm{d}y^{BL}}{\mathrm{d}\Delta}\left[\hat{y} - d_L\left(y^{BL}\right)\right] f\left(y^{BL}\right) + \int_{y^{0H}}^{y^{HB}} \frac{\mathrm{d}}{\mathrm{d}\Delta} d_H(y) f(y)\,\mathrm{d}y + \int_{y^{BL}}^{\bar{y}} \frac{\mathrm{d}}{\mathrm{d}\Delta} d_L(y) f(y)\,\mathrm{d}y \\
&= \overbrace{\underbrace{\frac{\mathrm{d}y^{HB}}{\mathrm{d}\Delta}}_{<0} \left[\underbrace{d_H\left(y^{HB}\right) - \hat{y}}_{<0}\right]}^{>0} f\left(y^{HB}\right) + M(\mathcal{H}) \underbrace{\frac{\mathrm{d}}{\mathrm{d}\Delta} d_H(y)}_{>0} + M(\mathcal{L}) \underbrace{\frac{\mathrm{d}}{\mathrm{d}\Delta} d_L(y)}_{<0}
\end{aligned}$$

where the last equality exploits the fact that $d_H\left(y^{0H}\right) = 0$ and $d_L\left(y^{BL}\right) = \hat{y}$ by definition, that for a given cost function $\frac{\mathrm{d}}{\mathrm{d}\Delta} d_H(y)$ and $\frac{\mathrm{d}}{\mathrm{d}\Delta} d_L(y)$ are constant across ys, and defines $M(\mathcal{H}) = F\left(y^{HB}\right) - F\left(y^{0H}\right)$ and $M(\mathcal{L}) = 1 - F\left(y^{BL}\right)$. To obtain the expression in the statement, notice further that

$$d_H\left(y^{HB}\right) - \hat{y} = y^{HB} - e_H - \hat{y} = \hat{y} + \tilde{e} - e_H - \hat{y} = \tilde{e} - e_H,$$

that $\frac{\mathrm{d}d_i(y)}{\mathrm{d}\Delta} = -\frac{\mathrm{d}e_i}{\mathrm{d}\Delta}$ for $i = H, L$, and $\frac{\mathrm{d}y^{HB}}{\mathrm{d}\Delta} = \frac{\mathrm{d}\tilde{e}}{\mathrm{d}\Delta}$. The limit results of the Proposition are derived in Corollary 8 within the proof of Theorem 4.

A.4 Proof of Theorem 4

We prove that there exists $\mu^\star > 0$ such that $R(\Delta;\mu)$ is concave in Δ throughout the triangular region $\mathcal{Z}_{\mu^\star} = \{(\Delta,\mu) : 0 \le \Delta \le \mu \le \mu^\star\}$. The proof proceeds in two steps. First, for each fixed μ, we expand the second derivative of revenue around the flat rule, $\Delta = 0$. Its leading term is negative and diverges as $\Delta \downarrow 0$, establishing infinite concavity locally. Second, we show that the remainder of this expansion is uniformly bounded as both Δ and μ vary. Choosing $\mu^\star$ sufficiently small then makes the negative divergent term dominate throughout $\mathcal{Z}_{\mu^\star}$.

Representation of the marginal buncher. The singularity in the second derivative is generated by the movement of the marginal buncher. To characterize this movement, let $e_0 = e^I(\mu)$ denote evasion under the flat rule and define a firm's flat-rule declaration shortfall from the threshold as

$$g(y;\mu) = \hat{y} - (y - e_0) = \hat{y} + e_0 - y.$$

Thus, $g = 0$ identifies the firm whose flat-rule declaration equals $\hat{y}$, while $g > 0$ means that the declaration falls below the threshold. Importantly, g depends on the firm type and μ, but not on Δ. When the audit gap opens, firms with sufficiently small positive shortfalls bunch at the threshold. With this notation, the bunching region can be represented in the discrepancy space as $0 \le g \le \bar{g}(z,\mu)$, where $\bar{g}$ is the discrepancy of the marginal buncher. That is $\bar{g} = e_0 - \tilde{e}$ and therefore $\bar{g} \to 0$ as $\Delta \to 0$.

Proof overview. The marginal-revenue expression in equation (5) depends on the interior evasion choices e_H and e_L, the marginal-buncher choice $\tilde{e}$, and their derivatives. The interior choices are smooth in Δ and have bounded derivatives at the flat rule. We will show that the marginal-buncher choice instead inherits the square-root behavior of $\bar{g} = e_0 - \tilde{e}$. We therefore write $z = \sqrt{\Delta}$, with $\frac{d}{d\Delta} = \frac{1}{2z}\frac{d}{dz}$, and express all local expansions in z.

Throughout the proof, ρ_i denotes a generic remainder that is uniformly bounded over the feasible policy region $\mathcal{Z}_{1/[\gamma(1+\alpha)]}$.[39] Uniform boundedness follows because the evasion levels and the relevant derivatives of c and f are evaluated on compact sets. The only potentially singular terms involve derivatives of $\tilde{e}$, which we isolate explicitly through the expansion in z.

The proof now proceeds in three steps. Lemma 7 first derives uniform expansions for e_H, e_L, and $\bar{g}$. We then substitute these expansions into the derivative of equation (5) and collect the terms generated by the moving bunching boundary. This yields a representation of the form $R_{\Delta\Delta}(\Delta;\mu) = -\frac{G(\mu)}{\sqrt{\Delta}} + \rho_R(\sqrt{\Delta},\mu)$. Finally, we show that $G(\mu) > 0$ for sufficiently small μ and that ρ_R is uniformly bounded. The negative singular term therefore dominates the remainder on a sufficiently small triangular region, establishing the result.

To simplify notation, define

$$C_2 = c''(e_0), \qquad C_3 = c'''(e_0), \qquad f_0 = f(\hat{y} + e_0), \qquad f_1 = f'(\hat{y} + e_0).$$

Lemma 7. ***(Uniform local expansions).*** *There is $m_\rho > 0$ such that on $\mathcal{Z}_{m_\rho}$ the interior choices and the marginal buncher satisfy*

$$e_H - e_0 = -\frac{\alpha\tau\gamma}{C_2}z^2 + z^3\rho_H, \qquad e_L - e_0 = \frac{\tau\gamma}{C_2}z^2 + z^3\rho_L,$$

and

$$\bar{g}(z,\mu) = e_0 - \tilde{e} = Kz + Lz^2 + z^3\rho_{\bar{g}},$$

with

$$K = \sqrt{\frac{2\tau\gamma(1+\alpha)e_0}{C_2}}, \qquad L = \frac{-\tau\gamma + C_3K^2/6}{C_2},$$

and the remainders are uniformly bounded.

Proof. The interior choices solve smooth first-order conditions: from $c'(e_L) = \tau(1-\gamma p_L) = \tau(1-\gamma\mu) + \tau\gamma\Delta$, expanding c' around e_0 gives $e_L - e_0 = (\tau\gamma/C_2)\Delta + \Delta^2\rho_L^0$, and likewise $e_H - e_0 = -(\alpha\tau\gamma/C_2)\Delta + \Delta^2\rho_H^0$; with $\Delta = z^2$ these are the stated forms. To obtain the expression for $\bar{g}$, let $u(e,p) = \tau(1-\gamma p)e - c(e)$ and $V(p) = \max_e u(e,p)$. The marginal buncher is indifferent between the high-audit interior policy and bunching at the threshold (evading $e_0 - \bar{g}$ at $p_L = \mu - \Delta$):

$$V(\mu + \alpha\Delta) = u(e_0 - \bar{g}, \mu - \Delta). \tag{A.1}$$

By the envelope theorem $V'(\mu) = -\tau\gamma e_0$, so

$$V(\mu + \alpha\Delta) = V(\mu) - \tau\gamma e_0\,\alpha\Delta + \Delta^2\rho_V$$

The function ρ_V is bounded because V is twice continuously differentiable on the relevant compact interval of audit probabilities; equivalently, by the envelope theorem $V''(p) = -\tau\gamma e_p(p)$, and $e_p(p)$ is

[39] This upper bound ensures that the relevant flat audit probabilities remain below $1/\gamma$, preserving interior evasion choices.

bounded by the interior first-order condition and the lower bound on c''. Expanding the right-hand side of (A.1) around (e_0, μ) with $u_e(e_0, \mu) = 0$, $u_p = -\tau\gamma e_0$, $u_{ee} = -C_2$, $u_{ep} = -\tau\gamma$, $u_{eee} = -C_3$ gives

$$u(e_0 - \bar{g}, \mu - \Delta) = V(\mu) + \tau\gamma e_0 \Delta - \frac{1}{2}C_2\bar{g}^2 - \tau\gamma\bar{g}\Delta + \frac{1}{6}C_3\bar{g}^3 + \rho_u(\bar{g}, \Delta, \mu) \cdot \left(\Delta^2 + \bar{g}^4 + \bar{g}^2\Delta\right).$$

The function ρ_u is bounded because the remainder in the Taylor expansion can be written as a finite sum of integral remainders. Each integral remainder averages a derivative of u evaluated at points on the line segment joining (e_0, μ) and $(e_0 - \bar{g}, \mu - \Delta)$. All these derivatives are bounded.[40] Equating the two expansions, the indifference condition (A.1) becomes

$$\frac{1}{2}C_2\bar{g}^2 = \tau\gamma(1+\alpha)e_0\Delta + \frac{1}{6}C_3\bar{g}^3 - \tau\gamma\bar{g}\Delta + \rho_\Psi(\bar{g}, \Delta, \mu) \cdot \left(\Delta^2 + \bar{g}^4 + \bar{g}^2\Delta\right), \tag{A.2}$$

where ρ_Ψ is bounded. Indeed, ρ_Ψ is obtained by collecting the ρ_V and ρ_u terms after rearranging the indifference condition, so it is a finite combination of bounded Taylor-remainder functions. This equation shows that $\bar{g}$ is of order $z = \sqrt{\Delta}$. In the concavity calculation below, the terms involving $\tilde{e}_\Delta$ and $\tilde{e}_{\Delta\Delta}$ require the marginal-buncher expansion through order z^2, with a z^3 remainder bounded uniformly on the cone.[41] We therefore write

$$\bar{g} = Kz + Lz^2 + z^3\rho_{\bar{g}}(z, \mu), \tag{A.3}$$

and prove this representation by first regularizing the indifference equation. Putting $\bar{g} = zk$ in (A.2) and dividing by z^2 gives

$$\Psi(k, z, \mu) \equiv \tfrac{1}{2}C_2k^2 - \tau\gamma(1+\alpha)e_0 - \tfrac{1}{6}C_3zk^3 + \tau\gamma zk + z^2\rho_\Psi^d = 0. \tag{A.4}$$

At $z = 0$ the positive root is $K = \sqrt{2B/A}$, where $A = C_2$ and $B = \tau\gamma(1+\alpha)e_0$ are the marginal cost and marginal benefit of the marginal buncher, respectively. Indeed, A is the shadow price of moving away from the threshold, while B is the marginal benefit from the probability reduction obtained when crossing the threshold.[42] Moreover, $\Psi_k(K, 0, \mu) = AK$ is bounded away from zero on a small μ-interval. The implicit-function theorem therefore implies that Ψ defines a smooth, uniformly bounded function $k(z, \mu)$ with bounded derivatives. To identify L, we replace the definition of $\bar{g}$ in (A.3) into (A.2). Matching the z^3 coefficients gives $AKL = \frac{1}{6}C_3K^3 - K\tau\gamma$ that can be rearranged to obtain the stated L.[43] The numerator in L is the sum of two terms. First, $C_3K^2/6$ denotes a *cost-curvature*: reaching the threshold means undoing manipulation, so, with $c''' > 0$, successive units of discrepancy g are cheaper to close. Second, $-\tau\gamma$ denotes a *tax-base-erosion* term: each unit

[40] The derivatives of u that appear in these averages are bounded on that compact set: apart from the linear dependence of u on p, the nonzero higher derivatives come from derivatives of c, and these are bounded by the primitive regularity assumptions. Factoring the polynomial scale $\Delta^2 + \bar{g}^4 + \bar{g}^2\Delta$ therefore leaves a bounded coefficient, which we denote by $\rho_u(\bar{g}, \Delta, \mu)$.

[41] Notice $\bar{g}$ and $\tilde{e}$ differ only by a constant and have opposite signs. This constant will be relevant for the indifference conditions expressed in levels (e.g., (A.2)), but immaterial for the derivative. Hence, we use derivatives of $\bar{g}$ and $\tilde{e}$ interchangeably.

[42] The cost of bunching is an opportunity cost. Indeed, reducing evasion by r sacrifices $\tau(1 - \gamma\mu)r$ but saves $c(e_0) - c(e_0 - r) = c'(e_0)r - \frac{1}{2}c''(e_0)r^2 + o(r^2)$. Because $c'(e_0) = \tau(1 - \gamma\mu)$ at the flat-rule optimum, the first-order terms cancel, leaving a loss $\frac{1}{2}c''(e_0)r^2 + o(r^2)$. At $r = g$, this is $\frac{1}{2}Ag^2 + o(g^2)$ where A is the shadow price of moving away from the threshold.

[43] Full computations are in the Supplementary Appendix.

of discrepancy closed gives up one unit of evaded base, so the audit discount rewards the marginal buncher less the farther it starts. The remainders remain bounded and are collected into $\rho_{\bar{g}}$.

We conclude the proof of the lemma by characterizing m_ρ. We take m_ρ to be any radius small enough that, on $\mathcal{Z}_{m_\rho}$, all evasion choices stay interior, the arguments of c and f remain in compact sets, c'' and K stay bounded away from zero, and (A.4) still defines implicitly the marginal buncher condition—so that every remainder $\rho_H, \rho_L, \rho_{\tilde{e}}$ (and their derivatives) is uniformly bounded on $\mathcal{Z}_{m_\rho}$. Economically, m_ρ ensures that interior choices remain close to the flat-rule case, and that the marginal buncher generated by a small audit jump remains a local perturbation of the firm with $g = 0$. A conservative sufficient characterization is the following. Let

$$\Xi(k, z, \mu) = -\frac{1}{6}C_3 k^3 + \tau\gamma k + z\rho^d_\Psi.$$

collect lower-order terms in (A.4) so that

$$\Psi(k, z, \mu) = \tfrac{1}{2}Ak^2 - B + z\Xi(k, z, \mu).$$

Let $\underline{A} = \inf A > 0$, $\underline{K} = \inf K > 0$, and $\bar{\Xi}_k = \sup |\frac{\partial \Xi}{\partial k}| < \infty$ on the relevant compact neighborhood. Then any m_ρ with

$$\sqrt{m_\rho}\,\bar{\Xi}_k < \tfrac{1}{2}\,\underline{A}\,\underline{K}$$

keeps Ψ_k bounded away from zero near the positive branch, so the implicit-function construction applies on $\mathcal{Z}_{m_\rho}$. □

Corollary 8. ***(Derivative expansions).*** *On $\mathcal{Z}_{m_\rho}$,*

$$\tilde{e}_\Delta = -\frac{K}{2z} - L + z\rho_{\tilde{e}_\Delta}, \qquad \tilde{e}_{\Delta\Delta} = \frac{K}{4z^3} + \frac{1}{z}\rho_{\tilde{e}_{\Delta\Delta}}, \qquad e_{H,\Delta} = -\frac{\alpha\tau\gamma}{C_2} + z\rho_{H,\Delta},$$

$$e_H - \tilde{e} = Kz + \left(L - \frac{\alpha\tau\gamma}{C_2}\right)z^2 + z^3\rho_{H,\tilde{e}}, \qquad f(\hat{y}+\tilde{e}) = f_0 - Kf_1 z + z^2\rho_f, \quad f'(\hat{y}+\tilde{e}) = f_1 + z\rho_{f'},$$

with uniformly bounded remainders.

Proof. We obtain the derivatives of $\tilde{e}$ differentiating wrt Δ the $\bar{g}$ definition in Lemma 7. The main terms are immediate, while the remainders are functions of the derivatives of $\rho_{\bar{g}}$, namely $\rho_{\tilde{e}_\Delta} = \frac{3}{2}\rho_{\bar{g}} + \frac{z}{2}\partial_z\rho_{\bar{g}}$ and $\rho_{\tilde{e}_{\Delta\Delta}} = \frac{3}{4}\rho_{\bar{g}} + \frac{5}{4}z\partial_z\rho_{\bar{g}} + \frac{z^2}{4}\partial_z^2\rho_{\bar{g}}$. Since $\rho_{\bar{g}}$ and its derivatives wrt z are bounded on the cone, the remainders $\rho_{\tilde{e}_\Delta}$ and $\rho_{\tilde{e}_{\Delta\Delta}}$ are also bounded. The formula for $e_{H,\Delta}$ follows in the same way from the expansion of e_H, denoting $z\rho_{H,\Delta}$ the the derivative wrt Δ of remainder of order z^4 established in the proof of Lemma 7. The expansion for $e_H - \tilde{e}$ follows by subtracting the two expansions in Lemma 7. Finally, the density expansions follow from Taylor expansion around $\hat{y} + e_0$ and boundedness of the derivatives of f on the relevant compact set. Notice that this Corollary also establishes the limit results of Proposition 3 as it shows $\lim_{\Delta\to 0}\tilde{e}_\Delta = -\infty$ and $\lim_{\Delta\to 0}\tilde{e}_\Delta(e_H - \tilde{e}) = -\frac{K^2}{2}$, where K is bounded. □

A.4.1 Revenue curvature

From Proposition 3, aggregate marginal revenue decomposes as $R_\Delta = B_R + L_R + H_R$ over the bunching, low-, and high-audit regions, with

$$B_R = -\tilde{e}_\Delta(e_H - \tilde{e})f(\hat{y}+\tilde{e}), \quad L_R = -e_{L,\Delta}\{1 - F(\hat{y}+e_L)\}, \quad H_R = -e_{H,\Delta}\{F(\hat{y}+\tilde{e}) - F(\hat{y}+e_H)\}.$$

We compute $R_{\Delta\Delta}$ (full details on computations in Supplementary Materials) and decompose it as $R_{\Delta\Delta} = R^{\text{sing}}_{\Delta\Delta} + \rho_{R_{\Delta\Delta}}$, where $\rho_{R_{\Delta\Delta}}$ remains bounded as $\Delta \to 0$ and $R^{\text{sing}}_{\Delta\Delta}$ is its singular component. Because e_H, e_L and their first two Δ-derivatives are bounded, the only terms in $R_{\Delta\Delta}$ carrying negative powers of z, and therefore driving the singularity at $\Delta \to 0$, come from $\tilde{e}_\Delta, \tilde{e}_{\Delta\Delta}$. Hence, $R^{\text{sing}}_{\Delta\Delta}$ only includes the derivatives of the terms in $\tilde{e}$ and $\tilde{e}_\Delta$. We get[44]

$$-R^{\text{sing}}_{\Delta\Delta} = T_1 + T_2 + T_3 + T_4,$$

$$\begin{aligned} T_1 &= \tilde{e}_{\Delta\Delta}(e_H - \tilde{e}) f(\hat{y} + \tilde{e}), \quad T_2 = -\tilde{e}_\Delta^2 f(\hat{y} + \tilde{e}), \\ T_3 &= \tilde{e}_\Delta^2 (e_H - \tilde{e}) f'(\hat{y} + \tilde{e}), \quad T_4 = 2\tilde{e}_\Delta e_{H,\Delta} f(\hat{y} + \tilde{e}). \end{aligned}$$

Substituting the expressions in Corollary 8 and rearranging,

$$T_1 = \frac{K^2 f_0}{4z^2} + \frac{K}{4z}\Big[\Big(L - \frac{\alpha\tau\gamma}{C_2}\Big) f_0 - K^2 f_1\Big] + \rho_{T_1}, \qquad T_2 = -\frac{K^2 f_0}{4z^2} - \frac{1}{z}\Big[-\frac{K^3 f_1}{4} + K L f_0\Big] + \rho_{T_2},$$

$$T_3 = \frac{K^3 f_1}{4z} + \rho_{T_3}, \qquad T_4 = \frac{K\alpha\tau\gamma f_0}{C_2 z} + \rho_{T_4}.$$

The z^{-2} terms of T_1 and T_2 are equal and opposite and cancel. Adding the z^{-1} parts,

$$\frac{K}{4}\Big[\Big(L - \frac{\alpha\tau\gamma}{C_2}\Big) f_0 - K^2 f_1\Big] - \Big[-\frac{K^3 f_1}{4} + K L f_0\Big] + \frac{K^3 f_1}{4} + \frac{K\alpha\tau\gamma f_0}{C_2} = K f_0\Big[-\frac{3L}{4} + \frac{3\alpha\tau\gamma}{4C_2} + \frac{K^2 f_1}{4 f_0}\Big].$$

Substituting K and L, this is G, where

$$G = \frac{K}{2}\,\frac{\tau\gamma(1+\alpha) f_0}{C_2}\, S, \qquad S = \frac{3}{2} + e_0 \frac{f_1}{f_0} - \frac{e_0}{2}\frac{C_3}{C_2}. \tag{A.5}$$

Hence

$$R_{\Delta\Delta} = -\frac{G}{\sqrt{\Delta}} + \rho_R(\sqrt{\Delta}, \mu), \qquad \sup_{\mathcal{Z}_{m_\rho}} |\rho_R| < \infty, \tag{A.6}$$

where ρ_R is the sum of all the ρ_Ts with $\rho_{R_{\Delta\Delta}}$.

We are now ready to state the main result. So far, we have not made explicit that G, S, K, L are all smooth functions of μ. Using this, we derive the following result.

Proposition 9. ***(Baseline concavity).*** *If $S(0) > 0$, there is $\mu^\star > 0$ with $R_{\Delta\Delta}(\Delta; \mu) < 0$ for all $0 < \Delta \le \mu \le \mu^\star$.*

Proof. The proof amounts to identifying a triangular region $\mathcal{Z}_{\mu^\star}$ in which i) G is positive and ii) ρ_R is uniformly bounded, so that $R_{\Delta\Delta}$ remains negative. Under our maintained assumptions $S(\mu) > 0$, which guarantees that $G(\mu) > 0$ because the factor multiplying $S(\mu)$ in $G(\mu)$ is positive at $\mu = 0$ and all terms are continuous. Therefore there exists a $\tilde{\mu} > 0$ for which $G(\mu) > 0$ in the interval $[0, \tilde{\mu}]$. To identify two conservative bounds on the conditions i) and ii), define the two auxiliary functions

$$\underline{G}(m) := \inf_{0 \le \mu \le m} G(\mu), \qquad \overline{\rho}_R(m) := \sup_{\mathcal{Z}_m} |\rho_R(z, \mu)| < \infty.$$

[44] To collect the four terms, expand $-\tilde{e}_\Delta(e_{H,\Delta} - \tilde{e}_\Delta) f = -\tilde{e}_\Delta e_{H,\Delta} f + \tilde{e}_\Delta^2 f$ in $B_{R,\Delta}$ and add the singular term $-e_{H,\Delta}\tilde{e}_\Delta f$ of $H_{R,\Delta}$. All densities are evaluated at $\hat{y} + \tilde{e}$.

Our desired region is defined by

$$\mu^{\star} = \sup\left\{m \in (0, \min\{\tilde{\mu}, m_{\rho}\}] : \sqrt{m}\,\overline{\rho}_R(m) < \underline{G}(m)\right\},$$

m_{ρ} is the expansion radius where the marginal buncher condition remains bounded (Lemma 7), $\tilde{\mu}$ defines the region where $G(\mu) > 0$, and $\mu^{\star}$ is the smaller interval on which the singular negative term dominates the bounded remainder ρ_R. To see that $\mu^{\star} > 0$, notice the following. Since $G(0) > 0$, continuity gives $\underline{G}(m) > 0$ for small m. Moreover, by monotonicity $\overline{\rho}_R(m) \le \overline{\rho}_R(m_{\rho})$ for $m \le m_{\rho}$, so $\sqrt{m}\,\overline{\rho}_R(m) \to 0$ as $m \to 0$. Hence $\mu^{\star} > 0$. Therefore, for $(\Delta, \mu) \in \mathcal{Z}_m$ with $m < \mu^{\star}$, it holds

$$R_{\Delta\Delta}(\Delta;\mu) \le -\frac{\underline{G}(m)}{\sqrt{\Delta}} + \overline{\rho}_R(m) \le -\frac{\underline{G}(m)}{\sqrt{m}} + \overline{\rho}_R(m) < 0.$$

Thus revenue is concave on the stated interval. □

Hence $R(\cdot;\mu)$ is strictly concave on $(0,\mu]$ and the endpoint follows from continuity of R on $[0,\mu]$. Since a concave function has decreasing secant slopes, if revenues increase when the audit gap rises from Δ_1 to $\Delta_2 > \Delta_1$, then the secant slope from the flat rule to Δ_1 is positive. Thus the disclosed rule at Δ_1 raises revenue relative to the flat rule.

A.5 Proof of Remark 6

Define the budget function Q as the mapping from an audit rule to the share of audited taxpayers, given by

$$Q(\Delta;\mu) = (\mu + \alpha\Delta)\, F\left(\hat{y} + \tilde{e}\,(\Delta)\right) + (\mu - \Delta)\left[1 - F\left(\hat{y} + \tilde{e}\,(\Delta)\right)\right]. \tag{A.7}$$

Here $\hat{y} + \tilde{e}(\Delta)$ is the true-revenue cutoff separating high-audit interior reporters from firms that bunch at the threshold. Differentiating (A.7),

$$\begin{aligned} Q_{\Delta}(\Delta;\mu) = {} & \alpha F\left(\hat{y} + \tilde{e}\,(\Delta)\right) - \left[1 - F\left(\hat{y} + \tilde{e}\,(\Delta)\right)\right] \\ & + (\alpha + 1)\,\Delta \tilde{e}_{\Delta}\,(\Delta)\, f\left(\hat{y} + \tilde{e}\,(\Delta)\right). \end{aligned}$$

By Lemma 7, $\tilde{e} = e_0 - Kz - Lz^2 - z^3 \rho_{\bar{g}}(z,\mu)$, with $z = \sqrt{\Delta}$. Therefore $\tilde{e}_{\Delta} = -K/(2z) + O(1)$ and $\Delta\tilde{e}_{\Delta} \to 0$ as $\Delta \to 0$. Since $\tilde{e} \to e_0$,

$$\lim_{\Delta \to 0} Q_{\Delta}(\Delta;\mu) = \alpha F\left(\hat{y} + e_0\right) - \left[1 - F\left(\hat{y} + e_0\right)\right]. \tag{A.8}$$

The right-hand side of (A.8) is negative exactly when $\alpha < 1/F(\hat{y} + e_0) - 1$. Continuity of Q_{Δ} away from $\Delta = 0$ then implies $Q(\Delta;\mu) < Q(0;\mu)$ for all sufficiently small positive Δ.

B Extension to heterogeneous costs

Our result accommodates *heterogeneity in the cost of manipulation.* The extensions in the rest of the appendix each change the firm's problem and are handled by recomputing A, B and the coefficient G. Cost heterogeneity does not affect any of this given a model specification. It only changes the *population* over which the boundary expansion is averaged. For this reason we do not restate the propositions with a cost type attached; we give the aggregation argument once, here, and read every later result as holding type by type.

Let firms differ in a cost type κ, with manipulation cost $C(\cdot;\kappa)$ and a joint distribution $F(\cdot,\kappa)$ of firm and cost types. The different model specifications can be thought as each conditional on κ, so the concavity proof applies to that type and delivers, for each κ, the representation

$$R_{\kappa,\Delta\Delta}(\Delta;\mu) = -\frac{G_\kappa(\mu)}{\sqrt{\Delta}} + \rho_\kappa(\sqrt{\Delta},\mu), \tag{B.1}$$

where the distribution of firm's income that enters the G_k and ρ_k is conditional on κ.

Denote F_κ the marginal of $F(\cdot,\kappa)$ on κ. Aggregate declared revenue is the average over types, $R(\Delta;\mu) = \int R_\kappa(\Delta;\mu)\,\mathrm{d}F_\kappa(\kappa)$. Under the assumption that we can exchange (conditional) integral and differentiation (granted by the smoothness assumptions on F), we obtain the representation of the curvature as

$$R_{\Delta\Delta}(\Delta;\mu) = -\frac{\bar{G}(\mu)}{\sqrt{\Delta}} + \bar{\rho}(\sqrt{\Delta},\mu), \qquad \bar{G}(\mu) = \int G_\kappa(\mu)\,\mathrm{d}F_\kappa(\kappa), \qquad \bar{\rho} = \int \rho_\kappa\,\mathrm{d}F_\kappa\,(k)\,.$$

This allows to interpret the concavity of aggregate revenues as an average of the second derivatives conditional on each cost type. The analogous of Proposition 9 for a given model specification then delivers concavity once the *averaged* sign condition $\bar{G}(0) = \int G_\kappa(0)\,\mathrm{d}F_\kappa(\kappa) > 0$ holds. A sufficient (but not necessary) condition for this to hold is that it holds for all types κ.

C Other Extensions to the baseline model

The general model lets firms manipulate two reports jointly and lets the threshold depend on what they report. It contains all the reporting extensions as special cases. We work directly with a linear evaded base and a linear threshold, which carry all the economics of the general case; Supplementary material Section S.2.0.2 explains why the nonlinear extension is immediate.

C.1 Setup

A firm has true revenue and cost items (y,x), distributed with a density $f(y,x)$ smooth and bounded on the relevant region, and reports (d_y,d_x), equivalently choosing manipulation amounts $e_y = y - d_y$ and $e_x = d_x - x$. The audit discount protects an evaded base $E(e_y,e_x) = \eta'\mathbf{e}$, linear in the two amounts, $\eta = (\eta_y,\eta_x)'$. Manipulation is costly according to a smooth, strictly convex $C(e_y,e_x)$. The firm's objective is

$$\max_{\mathbf{e}}\ \tau\,(1-\gamma p(\mathbf{d};\Delta))\,E(\mathbf{e}) - C(\mathbf{e},\kappa).$$

Eligibility for the low-audit region is $d_y \geq \hat{Y}(d_x)$, with $\hat{Y}(d_x) = \hat{y}_0 + \hat{y}_1 d_x$, with $\hat{y}_0, \hat{y}_1 \geq 0$, linear and weakly increasing, so that

$$p(\mathbf{d};\Delta) = \mu + \alpha\Delta\cdot\mathbf{1}\{d_y < \hat{Y}(d_x)\} - \Delta\cdot\mathbf{1}\{d_y \geq \hat{Y}(d_x)\}.$$

and the firm solves $\max_{\mathbf{e}} u(\mathbf{e},p(\mathbf{e}))$ where $u(\mathbf{e},p(\mathbf{e})) = \tau(1-\gamma p(\mathbf{e}))\eta'\mathbf{e} - C(\mathbf{e})$. Write C_e for the gradient and C_{ee} for the Hessian of C in $\mathbf{e} = (e_y,e_x)$, and $\mathbf{e}_0 = (e_{y0},e_{x0})$ for the flat-rule choice, solving $C_e(\mathbf{e}_0) = \tau(1-\gamma\mu)\eta$. Because E is linear, the objective Hessian collapses to the cost Hessian, $C_2 := C_{ee}(\mathbf{e}_0)$, which we assume positive definite uniformly on the local μ-region: this is the only curvature that matters for reaching the threshold, as $c''(e_0)$ is the only curvature that matters in the baseline.

Besides C_2, we will need the third derivative of the cost at $\mathbf{e}_0$. Let T denote the symmetric third-derivative tensor with entries $C_{ijk} = C_{ijk}(\mathbf{e}_0)$, $i,j,k \in \{y,x\}$, and use it in two forms. As a scalar, contracting it three times with a direction $a = (a_y, a_x)'$,

$$T[a,a,a] = \sum_{i,j,k} C_{ijk} a_i a_j a_k = \left. \frac{d^3}{dt^3} C(\mathbf{e}_0 + ta) \right|_{t=0} = C_{yyy} a_y^3 + 3C_{yyx} a_y^2 a_x + 3C_{yxx} a_y a_x^2 + C_{xxx} a_x^3; \quad \text{(C.1)}$$

this is just the third derivative of the one-dimensional function $t \mapsto C(\mathbf{e}_0 + ta)$, and will enter the cubic term of a firm's value loss. As a vector, leaving one slot open (a vector with two entries indexed by i),

$$\big(T[\,\cdot\,, a, a]\big)_i = \sum_{j,k} C_{ijk} a_j a_k, \qquad \text{so that} \qquad T[a,a,a] = a' T[\,\cdot\,, a, a].$$

The scalar form describes how much the cubic term of the cost matters for the firm's value loss along a given direction (relevant to characterize the marginal buncher indifference condition). The vector form describes how the cost gradient C_e itself changes along a given direction, via the second-order Taylor expansion $C_e(\mathbf{e}_0 + r) = C_e(\mathbf{e}_0) + C_2 r + \frac{1}{2} T[\,\cdot\,, r, r] + O(\|r\|^3)$ (relevant to characterize the marginal buncher's contribution to revenues).

As in the baseline model, it is convenient to work with a transformation of the firm (x, y) into a type (g, s) that summarizes its position on the bunching frontier: g denotes how far it is and s is the point where it would land. Formally, at the flat rule the firm's declarations are $d_y^0 = y - e_y^0$ and $d_x^0 = x + e_x^0$. We have

$$s = d_x^0, \qquad g = \hat{Y}(d_x^0) - d_y^0.$$

In the baseline it would be $s = \hat{y}$ and only $g = \hat{y} + e_0 - y$ would depend on the firm's type. Denote $\mathcal{S}$ as the domain of s (a compact interval due to continuity of d_x). The change of coordinates is convenient because it allows to express the set of marginal bunchers as the firms with $g = \bar{g}$, where $\bar{g} \to 0$ only as $\Delta \to 0$. Therefore, the bunching region is the set $\{(g, s) : s \in \mathcal{S},\ 0 \le g \le \bar{g}\}$, with the audit policy entering only through the upper limit. As in the baseline, the singularity of revenues will be driven by the behavior of marginal bunchers with the difference that in this extension they are a set rather than a single type.

Using the inverse map $x = s - e_{0x}$ and $y = \hat{Y}(s) + e_{0y} - g$, we express the density in (g, s) coordinates as $h(g, s, \mu) = f(\hat{Y}(s) + e_{0y} - g,\ s - e_{0x})$, which inherits the smoothness properties of $f(\cdot)$. It expands as $h(g, s, \mu) = h_0(s, \mu) + h_1(s, \mu) g + O(g^2)$, with $h_0(s, \mu) = f(\hat{Y}(s) + e_{0y}, s - e_{0x})$ and $h_1(s, \mu) = -f_y(\hat{Y}(s) + e_{0y}, s - e_{0x})$. The intercept $\hat{y}_0$ matters only through such level objects; every local expansion below involves the slope $\hat{y}_1$ alone.

C.2 Interior choices

Open a gap $\Delta = z^2$. The two audit probabilities are $p_H = \mu + \alpha\Delta$, $p_L = \mu - \Delta$. For any p near μ, the interior choice $e^I(p)$ solves $C_e(e^I(p)) = \tau(1 - \gamma p)\eta$. The Jacobian of this system in $\mathbf{e}$ is C_{ee}, positive definite, so the implicit-function theorem defines a smooth $\mathbf{e}^I(p)$ and differentiating the first-order condition in p gives $\partial_p \mathbf{e}^I = -\tau\gamma C_{ee}(\mathbf{e}^I)^{-1}\eta$. Evaluating the derivative at $\Delta = 0$,

$$\mathbf{e}_H = \mathbf{e}^I(p_H) = \mathbf{e}_0 - \alpha\tau\gamma C_2^{-1}\eta\,\Delta + O(\Delta^2), \qquad \mathbf{e}_L = \mathbf{e}^I(p_L) = \mathbf{e}_0 + \tau\gamma C_2^{-1}\eta\,\Delta + O(\Delta^2). \quad \text{(C.2)}$$

These are the two-margin analogues of e_H, e_L in the baseline, and $\mathbf{e}_H$ is the relevant deviation for the marginal buncher.

C.3 Bunching choice

Contrary to the baseline, here bunching is defined on the whole space where $d_y = \hat{Y}(d_x)$. Therefore, firms need to choose both whether to bunch, and where to bunch. To characterize where a firm bunches, we derive the solution to a cost minimization problem of a firm that has to close its discrepancy from the threshold. A firm with discrepancy $g > 0$ reaches the frontier by reducing revenue underreporting by $r_y \geq 0$, reducing cost overreporting by $r_x \geq 0$, or a combination. Local to $\Delta = 0$, these reductions close the discrepancy according to $r_y + \hat{y}_1 r_x = g$. The following Lemma characterizes this solution and its Taylor expansion to the order that is relevant to the calculation of concavity (i.e. g^2): r^B will represent the optimal reporting adjustment conditional on bunching (where the firm locates), while A will determine the marginal cost of bunching (whether the firm bunches). ℓ will be the first order expansion of r^B along the discrepancy g: the least-cost split of one unit of distance g between the revenue and the cost margin.

Lemma 10. ***(Potential bunching).*** *Consider a firm with flat-rule shortfall $g > 0$, and let $r^B(g, \Delta, \mu) = (r_y^B, r_x^B)'$ be the reductions in revenue underreporting and cost overreporting that place it exactly on the bunching frontier facing $p_L = \mu - \Delta$. Then*

$$r^B(g, \Delta, \mu) = \ell g + \tfrac{1}{2} b g^2 - \tau\gamma t_\eta \Delta + O(g^3 + g\Delta + \Delta^2), \tag{C.3}$$

where, writing $w = (1, \hat{y}_1)'$,

$$\ell = \frac{C_2^{-1} w}{w' C_2^{-1} w}, \qquad b = C_2^{-1} T[\,\cdot\,, \ell, \ell] - \big(w' C_2^{-1} T[\,\cdot\,, \ell, \ell]\big)\ell, \qquad t_\eta = C_2^{-1}\eta - C_2^{-1} w \,\frac{w' C_2^{-1}\eta}{w' C_2^{-1} w}.$$

At $\Delta = 0$, the minimized value loss from reaching the frontier expands as

$$\mathcal{L}(r^B(g, 0, \mu); \mu) := u(\mathbf{e}_0, \mu) - u(\mathbf{e}_0 - r^B(g, 0, \mu), \mu) = \tfrac{1}{2} A g^2 + C g^3 + O(g^4), \tag{C.4}$$

with $A = (w' C_2^{-1} w)^{-1}$, and $C = -\frac{1}{6} T[\ell, \ell, \ell]$.

Proof. Reaching the frontier from (d_y^0, d_x^0) means raising declared revenue by r_y and lowering declared cost by r_x, so that $d_y^0 + r_y = \hat{Y}(d_x^0 - r_x)$, i.e., $w'r = g$ with $w = (1, \hat{y}_1)'$ (the intercept $\hat{y}_0$ cancels when considering a marginal change): w is the gradient of the frontier constraint in the deviation coordinates, and the constraint is linear in r.

Expansion of r^B. Bunching takes place facing $p_L = \mu - \Delta$, so $r^B(g, \Delta, \mu)$ solves $\max_r u(\mathbf{e}_0 - r, p_L)$ subject to $w'r = g$. We expand the solution in g and Δ, deriving first the g-terms at $\Delta = 0$, then the Δ-term. The optimal reporting adjustment conditional on bunching r^B satisfies:

$$C_e(\mathbf{e}_0 - r^B) - \tau(1 - \gamma p_L)\eta + \lambda w = 0, \qquad w' r^B = g. \tag{C.5}$$

To identify the coefficients in g, we now evaluate (C.5) at $\Delta = 0$ so that $p_L = \mu$. In this case the system reduces to $C_e(\mathbf{e}_0 - r^B) - \tau(1 - \gamma\mu)\eta + \lambda w = 0$ and $w' r^B = g$. Using the vector form of the tensor (C.1) to expand the cost gradient, $C_e(\mathbf{e}_0 - r) = C_e(\mathbf{e}_0) - C_2 r + \frac{1}{2} T[\,\cdot\,, r, r] + O(\|r\|^3)$, and the FOC $C_e(\mathbf{e}_0) = \tau(1 - \gamma\mu)\eta$, the system becomes

$$-C_2 r^B + \tfrac{1}{2} T[\,\cdot\,, r^B, r^B] + \lambda w = 0, \qquad w' r^B = g. \tag{C.6}$$

Write the expansion of r^B in g when $\Delta = 0$ as $r^B = \ell g + \frac{1}{2} b g^2 + O(g^3)$, and expand λ as $\lambda = \lambda_1 g + \lambda_2 g^2 + O(g^3)$. We match coefficients.

At order g: $-C_2 \ell + \lambda_1 w = 0$ and $w'\ell = 1$, so $\ell = \lambda_1 C_2^{-1} w$ and $\lambda_1 w' C_2^{-1} w = 1$, giving $\lambda_1 = A$ and $\ell = A C_2^{-1} w$ as claimed (hence $C_2 \ell = A w$).

At order g^2: the tensor term in (C.6) contributes $\frac{1}{2}T[\,\cdot\,,\ell,\ell]$,[45] and the constraint $w'r^B = g$ being linear forces its second g-derivative to vanish, so

$$-C_2 b + T[\,\cdot\,,\ell,\ell] + 2\lambda_2 w = 0, \qquad w'b = 0.$$

Solving the first equation for b and imposing $w'b = 0$ pins down $\lambda_2 = -\frac{1}{2}A\,w'C_2^{-1}T[\,\cdot\,,\ell,\ell]$, and substituting back gives the displayed b (using $AC_2^{-1}w = \ell$). This b satisfies $w'b = 0$, since $w'\ell = 1$.

For the Δ-term of r^B, differentiate (C.5) with respect to Δ at fixed g and evaluate at $(g,\Delta) = (0,0)$. Since $p_L = \mu - \Delta$, $r^B(0,0,\mu) = 0$, and $C_2 = C_{ee}(\mathbf{e}_0)$, this gives

$$-C_2 r_\Delta - \tau\gamma\eta + \lambda_\Delta w = 0, \qquad w'r_\Delta = 0.$$

Solving in r_Δ and λ_Δ, $\lambda_\Delta = \tau\gamma\, w'C_2^{-1}\eta / w'C_2^{-1}w$ and $r_\Delta = -\tau\gamma t_\eta$ with t_η as displayed in the Lemma.

The terms not displayed in the expansion include a mixed contribution in $g\Delta$, which will not be important in the next steps and is included in the remainder $O(g^3 + g\Delta + \Delta^2)$ in the statement of the lemma. This proves the stated expansion of r^B.

Expansion of the value loss. The value loss from bunching relative to the interior under a flat rule is $\mathcal{L}(r;\mu) = u(\mathbf{e}_0,\mu) - u(\mathbf{e}_0 - r,\mu)$. Its expression follows from a third-order Taylor expansion of $u(\mathbf{e}_0 - r,\mu)$ around $\mathbf{e}_0$: the linear term vanishes by the flat-rule first-order condition $C_e(\mathbf{e}_0) = \tau(1-\gamma\mu)\eta$, and, because E is linear, the second and third derivatives of u in $\mathbf{e}$ are $-C_2$ and $-T$, so that

$$\mathcal{L}(r;\mu) = \tfrac{1}{2}r'C_2 r - \tfrac{1}{6}T[r,r,r] + O(\|r\|^4).$$

Take the r^B expansion in g, $r^B = \ell g + \frac{1}{2}bg^2 + O(g^3)$, and substitute it in $\mathcal{L}(r;\mu)$. The quadratic form expands as $r^{B\prime}C_2 r^B = g^2\,\ell'C_2\ell + 2\,(g\ell)'C_2(\frac{1}{2}g^2 b) + O(g^4)$. Using $C_2\ell = Aw$ and $w'\ell = 1$, the quadratic part gives $\frac{1}{2}g^2\ell'C_2\ell = \frac{1}{2}Ag^2$. The candidate at order g^3 is the cross term $\frac{1}{2}g^3\,\ell'C_2 b$, which vanishes because $C_2\ell = Aw$ and $w'b = 0$ give $\ell'C_2 b = Aw'b = 0$. For the cubic term in $\mathcal{L}(r;\mu)$, the definition of T in (C.1) gives

$$T[r^B,r^B,r^B] = T[\ell g,\ell g,\ell g] + 3\,T[\tfrac{1}{2}bg^2,\,\ell g,\,\ell g] + O(g^5) = g^3\,T[\ell,\ell,\ell] + \tfrac{3}{2}\,g^4\,T[b,\ell,\ell] + O(g^5).$$

So b only appears in terms of order at least $g^2\cdot g\cdot g = g^4$, implying it does not enter the cubic expansion, whose coefficient is therefore only $C = -\frac{1}{6}T[\ell,\ell,\ell]$.[46] Collecting terms gives the displayed expansion of the value loss and completes the proof of the Lemma. □

With these expressions at hand, we can now characterize the condition of the marginal buncher through an analogue of the one in the baseline model.

[45] Using the multi-linearity of the tensor (and ignoring the $O(g^3)$ in r^B:

$$\begin{aligned}
\tfrac{1}{2}T[\,\cdot\,,r^B,r^B] &= \tfrac{1}{2}T[\,\cdot\,,\ell g + \tfrac{1}{2}bg^2,\ell g + \tfrac{1}{2}bg^2] \\
&= \tfrac{1}{2}T[\,\cdot\,,\ell g,\ell g + \tfrac{1}{2}bg^2] + \tfrac{1}{2}T[\,\cdot\,,\tfrac{1}{2}bg^2,\ell g + \tfrac{1}{2}bg^2] \\
&= \tfrac{1}{2}T[\,\cdot\,,\ell g,\ell g] + \tfrac{1}{2}T[\,\cdot\,,\ell g,\tfrac{1}{2}bg^2] + \tfrac{1}{2}T[\,\cdot\,,\tfrac{1}{2}bg^2,\tfrac{1}{2}bg^2] + \tfrac{1}{2}T[\,\cdot\,,\tfrac{1}{2}bg^2,\ell g] \\
&= \tfrac{1}{2}g^2 T[\,\cdot\,,\ell,\ell] + O\left(g^3\right)
\end{aligned}$$

[46] None of these objects depends on the identity of the firm: every firm shares the flat-rule choice $\mathbf{e}_0$ and the frontier slope is constant, so ℓ, A, b and C are common across firms. The identity of the firm will matter only through the density of types over the bunching strip.

Indifference. Define the indifference gap by

$$\mathcal{I}(g,z,\mu) = u(\mathbf{e}_H, p_H) - u(\mathbf{e}_0 - r^B(g,z^2,\mu), p_L), \qquad \Delta = z^2. \tag{C.7}$$

The marginal buncher is characterized by $\mathcal{I}(\bar{g},z,\mu) = 0$. Add and subtract $u(\mathbf{e}_0,\mu)$:

$$\begin{aligned}\mathcal{I}(g,z,\mu) =& \{u(\mathbf{e}_H, p_H) - u(\mathbf{e}_0,\mu)\} \\ &+ \{u(\mathbf{e}_0,\mu) - u(\mathbf{e}_0 - r^B(g,z^2,\mu), p_L)\}.\end{aligned} \tag{C.8}$$

Because both utilities are evaluated at their respective optima, the first bracket in (C.8) is the change in value $V(p_H) - V(\mu)$ when the flat rule moves from μ to $p_H = \mu + \alpha\Delta$. By the envelope theorem,

$$u(\mathbf{e}_H, p_H) - u(\mathbf{e}_0,\mu) = -\alpha\tau\gamma\,\eta'\mathbf{e}_0\,\Delta + O(\Delta^2). \tag{C.9}$$

We split the second bracket in (C.8) using the loss $\mathcal{L}(r^B(g,\Delta,\mu);\mu) = u(\mathbf{e}_0,\mu) - u(\mathbf{e}_0 - r^B(g,\Delta,\mu),\mu)$ characterized in the previous Lemma and the definition of p_L, obtaining

$$u(\mathbf{e}_0,\mu) - u(\mathbf{e}_0 - r^B, p_L) = \mathcal{L}(r^B(g,\Delta,\mu);\mu) - \tau\gamma\Delta\,\eta'(\mathbf{e}_0 - r^B(g,\Delta,\mu)). \tag{C.10}$$

We now expand the loss accounting also for Δ, since the marginal buncher adjusts by $r^B(g,\Delta,\mu)$. Since Δ enters $\mathcal{L}(\,\cdot\,;\mu)$ only through r^B, substitute (C.3) into $\mathcal{L}(r;\mu) = \frac{1}{2}r'C_2r - \frac{1}{6}T[r,r,r] + O(\|r\|^4)$. The pure-$g$ terms are those of (C.4). The only candidate at order $g\Delta$ is the cross product, in the quadratic form $\frac{1}{2}r'C_2r$, between the g-term ℓg and the Δ-term $-\tau\gamma t_\eta\Delta$ of r^B, namely $-\tau\gamma\,\ell'C_2t_\eta\,g\Delta$.[47] This term vanishes as $\ell'C_2t_\eta = A\,w't_\eta = 0$, by $C_2\ell = Aw$ and $w't_\eta = 0$ (immediate from the definition of t_η in the Lemma). All remaining terms are of order $g^2\Delta$ or Δ^2: in particular the $g\Delta$ response of r^B, which (C.3) leaves uncomputed in the remainder, enters the loss only against the leading term ℓg, i.e. at order $g^2\Delta$. Hence the value-loss part of (C.10) is

$$\mathcal{L}(r^B(g,\Delta,\mu);\mu) = \tfrac{1}{2}Ag^2 + Cg^3 + O(g^4 + g^2\Delta + \Delta^2). \tag{C.11}$$

Replacing the expansion of r^B from Lemma (C.3), and keeping only the term in Δ and the one in $g\Delta$, the second term in (C.10) becomes

$$-\tau\gamma\Delta\,\eta'(\mathbf{e}_0 - r^B(g,\Delta,\mu)) = -\tau\gamma\,\eta'\mathbf{e}_0\Delta + \tau\gamma\,\eta'\ell\,g\Delta + O(g^2\Delta + \Delta^2). \tag{C.12}$$

Combining (C.9), (C.10), (C.11), and (C.12), and setting $\Delta = z^2$,

$$\mathcal{I}(g,z,\mu) = \tfrac{1}{2}Ag^2 - Bz^2 + Cg^3 + Dgz^2 + O(g^4 + g^2z^2 + z^4), \tag{C.13}$$

where A and C are the value-loss coefficients from Lemma 10, while

$$B = (1+\alpha)\tau\gamma\,\eta'\mathbf{e}_0, \qquad D = \tau\gamma\,\eta'\ell,$$

denote, respectively, the value of the audit-probability gap at the flat-rule evaded base, and the

[47] Substituting (C.3) into the quadratic form, the squares and double products give

$$r^{B\prime}C_2r^B = \underbrace{g^2\,\ell'C_2\ell}_{\ell g\times\ell g}\underbrace{-2\tau\gamma\,g\Delta\,\ell'C_2t_\eta}_{\ell g\times(-\tau\gamma t_\eta\Delta)} + \underbrace{g^3\,\ell'C_2b}_{\ell g\times\frac{1}{2}bg^2} + \underbrace{\tau^2\gamma^2\Delta^2\,t_\eta'C_2t_\eta}_{(-\tau\gamma t_\eta\Delta)\times(-\tau\gamma t_\eta\Delta)} + \dots,$$

where the dots collect all remaining products, of order g^4, $g^2\Delta$ or $g\Delta^2$.

first-order adjustment in the evaded base when the firm moves toward the frontier.

Equation (C.13) defines implicitly the marginal buncher as the $\bar{g}$ that sets $\mathcal{I}(\bar{g}, z, \mu) = 0$. Write $\bar{g} = zk(z, \mu)$. Substituting this expression in (C.13) and dividing by z^2 gives

$$\tfrac{1}{2}Ak^2 - B + z(Ck^3 + Dk) + O(z^2) = 0. \tag{C.14}$$

At $z = 0$ the positive root is

$$K = \sqrt{2B/A}.$$

Since the derivative of the left-hand side of (C.14) with respect to k is $AK > 0$ at $(K, 0)$, the implicit-function theorem gives a unique local root $k(z, \mu)$. Write $k(z, \mu) = K + Lz + O(z^2)$. Substituting this expansion in (C.14) and matching the coefficient on z gives $AKL + CK^3 + DK = 0$. Hence

$$\bar{g}(z, \mu) = Kz + Lz^2 + O(z^3), \qquad L = -\frac{CK^2 + D}{A}. \tag{C.15}$$

The differences from the baseline are the definitions of A, B, C, D functions. A and B have already been interpreted: they are respectively the marginal cost of the distance from the frontier, and the value of the audit jump at the flat-rule evaded base. C — the *cost-curvature* term — is how the marginal cost of distance changes for larger shortfalls: moving to the threshold means undoing manipulation, so with a convex marginal cost successive units of gap are cheaper to close. Compared with baseline, this cost saving now depends on both margins as the tensor is evaluated in the direction ℓ including adjustments on revenue and cost declarations. D — the *tax-base-erosion* term — is the rate at which bunching erodes its own reward: each unit of gap closed sacrifices $\eta'\ell$ of evaded base, so the audit discount protects less the farther the buncher starts. Together they determine L, the second-order movement of the marginal buncher discrepancy $\bar{g}$: a softening cost widens it, the erosion of the reward narrows it.

C.4 Revenue curvature

In the coordinates (g, s), total declared revenue sums the declarations of three groups of firms: those already eligible at the flat rule ($g \leq 0$), which choose the low-audit interior $\mathbf{e}_L$; the bunchers ($0 < g \leq \bar{g}(z, \mu)$), which declare on the frontier; and those too far from it ($g > \bar{g}(z, \mu)$), which choose the high-audit interior $\mathbf{e}_H$. Writing $d_y^H = y - e_{H,y}$, $d_y^L = y - e_{L,y}$ and, for a buncher, $d_y^B = d_y^0 + r_y^B(g, z^2, \mu)$

$$R(z, \mu) = \int_{\mathcal{S}} \Big\{ \int_{g \leq 0} d_y^L \, h \, dg + \int_0^{\bar{g}(z,\mu)} d_y^B \, h \, dg + \int_{g > \bar{g}(z,\mu)} d_y^H \, h \, dg \Big\} \, ds. \tag{C.16}$$

where $h(g, s, \mu)$ is the joint density of firms in the (g, s) coordinates introduced in the Setup, and the inner integrals run over the shortfall g at fixed s. Add and subtract the bunchers' counterfactual high-audit declarations, $\int_{\mathcal{S}} \int_0^{\bar{g}} d_y^H \, h \, dg \, ds$, and define $q \equiv d_y^B - d_y^H$, then we can write:

$$R(z, \mu) = R^{\text{ref}}(z, \mu) + B_R(z, \mu), \tag{C.17}$$

where

$$R^{\text{ref}}(z, \mu) = \int_{\mathcal{S}} \int_{g > 0} d_y^H \, h \, dg \, ds + \int_{\mathcal{S}} \int_{g \leq 0} d_y^L \, h \, dg \, ds, \tag{C.18}$$

$$B_R(z,\mu) = \int_{\mathcal{S}} \underbrace{\int_0^{\bar g(z,\mu)} q(g,z,\mu)\, h(g,s,\mu)\, dg}_{B_{R,s}}\, ds. \tag{C.19}$$

B_R is the bunchers' correction relative to the counterfactual in which they kept the high-audit interior choice. The decomposition in (C.17) is convenient as it separates the bounded from the singular components of $R_{\Delta\Delta}$. R^{ref} contributes only bounded terms to $R_{\Delta\Delta}$ as the domains of R^{ref} are fixed and its integrands are smooth in Δ with bounded first and second derivatives by (C.2). The singularity can therefore only come from B_R, i.e. from the bunching strip $0 \le g \le \bar g(z,\mu)$.

Using $d_y = y - e_y$ and (C.2), the high-audit interior declaration is $d_y^H = d_y^0 + \alpha\tau\gamma(C_2^{-1}\eta)_y z^2 + O(z^4)$, and substituting (C.3) into the buncher's declaration $d_y^B = d_y^0 + r_y^B$ gives, for a buncher, $d_y^B = d_y^0 + \ell_y g + \frac{1}{2} b_y g^2 - \tau\gamma(t_\eta)_y z^2 + O(g^3 + gz^2 + z^4)$. Throughout, for a two-vector $v = (v_y, v_x)'$ the subscript y extracts the revenue component: $(C_2^{-1}\eta)_y$, ℓ_y, b_y and $(t_\eta)_y$ denote the revenue components of $C_2^{-1}\eta$, ℓ, b and t_η. The revenue gain q from (C.19) therefore expands as

$$q(g,z,\mu) = \ell_y g + \tfrac{1}{2} b_y g^2 - \tau\gamma\{(t_\eta)_y + \alpha(C_2^{-1}\eta)_y\} z^2 + O(g^3 + gz^2 + z^4), \tag{C.20}$$

the two-margin analogue of the baseline gap $e_H - \tilde e$.

Substitute (C.20), the density expansion $h(g,s,\mu) = h_0(s,\mu) + h_1(s,\mu)g + O(g^2)$, and the marginal-buncher expansion (C.15) into (C.19). Since $\bar g = O(z)$, only terms up to order g^2 in the integrand can contribute to order z^3, which is the order that determines the singular curvature in the expansion. We therefore ignore terms in the integrand that are of order g^3 or above (including the terms in $z^2 g$). Recognizing that the inner integrals are $\int_0^{\bar g} g\, dg = \frac{1}{2}\bar g^2$, $\int_0^{\bar g} g^2 dg = \frac{1}{3}\bar g^3$ and $\int_0^{\bar g} dg = \bar g$ and, by (C.15), that $\bar g^2 = K^2 z^2 + 2KLz^3 + O(z^4)$ and $\bar g^3 = K^3 z^3 + O(z^4)$, we write the inner integral $B_{R,s}$ in (C.19) as the sum of the following terms:

$$\int_0^{\bar g} \ell_y h_0 g\, dg = \tfrac{1}{2}\ell_y h_0 K^2 z^2 + \ell_y h_0 K L z^3 + O(z^4),$$

$$\int_0^{\bar g} \Big(\tfrac{1}{2} b_y h_0 + \ell_y h_1\Big) g^2\, dg = \Big(\tfrac{1}{6} b_y h_0 + \tfrac{1}{3}\ell_y h_1\Big) K^3 z^3 + O(z^4),$$

$$\int_0^{\bar g} -\tau\gamma\{(t_\eta)_y + \alpha(C_2^{-1}\eta)_y\} h_0 z^2\, dg = -\tau\gamma\{(t_\eta)_y + \alpha(C_2^{-1}\eta)_y\} h_0 K z^3 + O(z^4).$$

We rewrite (C.19) by integrating over $s \in \mathcal{S}$, and factoring out the z^2 and z^3 terms that do not depend on s, obtaining

$$B_R(z,\mu) = a_2(\mu) z^2 + a_3(\mu) z^3 + O(z^4), \qquad a_2 = \int_{\mathcal{S}} \tfrac{1}{2}\ell_y h_0 K^2\, ds, \tag{C.21}$$

$$a_3 = \int_{\mathcal{S}} K h_0 \left[\ell_y L + \left(\tfrac{1}{6} b_y + \tfrac{1}{3}\ell_y \frac{h_1}{h_0}\right) K^2 - \tau\gamma\{(t_\eta)_y + \alpha(C_2^{-1}\eta)_y\}\right] ds. \tag{C.22}$$

Thus a_3 is the $z^3 = \Delta^{3/2}$ coefficient of the bunching contribution to revenue, and will therefore sign the singular component of the second derivative of the revenue function.

Relative to the baseline, (C.22) contains three new objects: the revenue component ℓ_y of the least-cost mix—in the baseline the whole adjustment is declared revenue; b_y, which captures that the optimal adjustment on revenues depends on the size of the adjustment: larger adjustments rely relatively more on revenues if their marginal cost rises more slowly; and $(t_\eta)_y$, the response of the mix to the audit discount. The last two are identically zero with one margin. Moreover, since there is a set of marginal bunchers rather than a single type, the coefficients are aggregated over the continuum of landing points s, weighted by the frontier density.

Differentiate (C.17) twice in Δ. We already argued that R^{ref} contributes only bounded terms. By (C.21), the z^2 term of B_R is linear in Δ and so has zero second Δ-derivative, while, since $z^3 = \Delta^{3/2}$, $\partial_{\Delta\Delta}(a_3 z^3) = \frac{3}{4} a_3/\sqrt{\Delta}$,[48] so

$$R_{\Delta\Delta}(\Delta;\mu) = -\frac{G(\mu)}{\sqrt{\Delta}} + O(1), \qquad G(\mu) = -\tfrac{3}{4} a_3(\mu). \tag{C.23}$$

Proposition 11. ***(Concavity in the extended model).*** *If $G(0) > 0$, there is $\mu^\star > 0$ with $R_{\Delta\Delta}(\Delta;\mu) < 0$ for all $0 < \Delta \le \mu \le \mu^\star$.*

Proof. By (C.23), write $R_{\Delta\Delta}(\Delta;\mu) = -G(\mu)/\sqrt{\Delta} + \rho(\sqrt{\Delta},\mu)$, with ρ bounded on a sufficiently small cone. Since $G(0) > 0$ and G is continuous, choose $m > 0$ such that $\underline{G}(m) := \inf_{0\le\mu\le m} G(\mu) > 0$. Let $\overline{\rho}(m)$ bound the remainder on $0 < \Delta \le \mu \le m$. Shrink m so that $\sqrt{m}\,\overline{\rho}(m) < \underline{G}(m)$. Then, for every $0 < \Delta \le \mu \le m$,

$$R_{\Delta\Delta}(\Delta;\mu) \le -\frac{\underline{G}(m)}{\sqrt{\Delta}} + \overline{\rho}(m) \le -\frac{\underline{G}(m)}{\sqrt{m}} + \overline{\rho}(m) < 0.$$

Taking $\mu^\star = m$ proves the claim. □

Remark 12. **(Profit versus revenue tax base).** If the planner's objective is declared profit rather than declared revenue, the firm side is unchanged — firms do not care which base the planner records — so the bunching strip, K and L, and the $1/\sqrt{\Delta}$ form of the singularity are those of the revenue case. What changes is the object integrated over the strip $[0,\bar{g}]$: replace $q = d_y^B - d_y^H$ by $q^\Pi = (d_y^B - d_x^B) - (d_y^H - d_x^H)$. Subtract the cost declarations: a buncher adjusts cost overreporting, $d_x^B = d_x^0 - r_x^B$, while under the high-audit interior choice $d_x^H = d_x^0 - \alpha\tau\gamma(C_2^{-1}\eta)_x z^2 + O(z^4)$, so that

$$q^\Pi = q - (d_x^B - d_x^H) = q + r_x^B - \alpha\tau\gamma(C_2^{-1}\eta)_x z^2 + O(z^4).$$

Replacing q with (C.20) and r_x^B with the x-component of (C.3):

$$q^\Pi = (\ell_y + \ell_x)g + \tfrac{1}{2}(b_y + b_x)g^2 - \tau\gamma\{(t_\eta)_y + (t_\eta)_x + \alpha[(C_2^{-1}\eta)_y + (C_2^{-1}\eta)_x]\}z^2 + O(g^3 + gz^2 + z^4).$$

Each vector now enters through the sum of its components: the cost adjustment passes through declared profit one for one and with a plus sign, so the planner records the buncher's entire adjustment, not just its revenue component ℓ_y. Substituting q^Π into (C.19) gives, by the same steps,

$$a_3^\Pi = \int_{\mathcal{S}} \Big[(\ell_y+\ell_x)h_0 KL + \Big(\tfrac{1}{6}(b_y+b_x)h_0 + \tfrac{1}{3}(\ell_y+\ell_x)h_1\Big)K^3 - \tau\gamma\{(t_\eta)_y+(t_\eta)_x+\alpha[(C_2^{-1}\eta)_y+(C_2^{-1}\eta)_x]\}h_0 K\Big]\,ds.$$

Proposition 2 applies verbatim with $G^\Pi = -\frac{3}{4}a_3^\Pi$: concavity now turns on $G^\Pi(0) > 0$, a different condition from $G(0) > 0$.[49]

[48] The $O(z^4)$ remainder of (C.21) is of the form $\rho(z,\mu)z^4 = \rho\,\Delta^2$ with ρ and its first two Δ-derivatives bounded on the cone, as in the baseline, so $\partial_{\Delta\Delta}(\rho\Delta^2) = O(1)$; together with $R_{\Delta\Delta}^{\text{ref}}$ it is absorbed in the $O(1)$ term of (C.23).

[49] Supplementary Material Section S.2.0.2 discusses how the concavity result extends to the case where $\hat{Y}$ and E are non-linear.

C.5 A declining undisclosed schedule

Throughout, the no disclosure counterfactual was piecewise flat. In principle the audit rule can depend smoothly on the declaration, and firms may perceive audit risk falling as reported revenues rise. We therefore open the disclosed jump within a declining schedule. Define $p_H(d) = p_0(d) + \alpha\Delta$ below the cutoff and $p_L(d) = p_0(d) - \Delta$ at or above it, with p_0 smooth, decreasing, and normalized to $p_0(0) = \mu$. For illustrative purposes, consider the linear case $p_0(d) = \mu - \beta d$ (in general, read β as $-p_0'(d^0)$ at the flat-rule declaration).[50] A smooth schedule, by itself, generates no bunching but only affects interior choices that in this case solves $\tau\big(1 - \gamma p_0(d)\big) - \tau\gamma\beta e = c'(e)$. A marginal unit of evasion, by lowering the declaration, raises the audit probability by β on the inframarginal units of evasion e, generating the wedge $\tau\gamma\beta e$. Differentiating the FOC, the effective curvature of the firm's objective is $A = c''(e_0) + 2\tau\gamma\beta$: the slope enters once by eroding the marginal benefit along the schedule, and once through the inframarginal penalty. Since one unit of discrepancy from the threshold is closed by one unit of evasion, this is also the curvature in g of the value loss from bunching, $\frac{1}{2}Ag^2 + O(g^3)$. The coefficient A keeps its meaning of shadow price of the distance from the threshold. B, and with a linear schedule C and D, are those of the baseline.[51]

Intuitively, the flexibility of the non-flat rule becomes immaterial at low levels of baseline audit probability μ (as β shrinks with μ), thus making the inframarginal effect irrelevant. Under the same sign condition $G(0) > 0$, the proof of Proposition 11 applies verbatim. The linearity of the schedule is not essential: for any smooth family scaling with the baseline level, $p_0(d;\mu) = \mu\,\pi(d)$ with π decreasing and $\pi(0) = 1$, every derivative of the schedule is $O(\mu)$ (including the curvature that would correct C) so all coefficients still converge to their baseline values and the same argument applies.

D Bunching Estimation

We define bunching as the excess mass in the observed revenue declaration distribution relative to a counterfactual distribution that would arise with a constant p_L probability. Since businesses declaring above $\hat{y}$ face probability p_L, our strategy uses their distribution to infer this counterfactual. Empirically, we follow the approach in Kleven and Waseem (2013), which relies on a flexible polynomial and excludes an area $[y_l, y_u]$ around $\hat{y}$ from the density distribution estimation. We bin the data in segments whose length is 1 percentage point of $\hat{y}$ and run the following regression for the number of SeS files c in each bin j

$$c_j = \sum_{i=1}^{K} \beta_i \left(y_j\right)^i + \sum_{h=y_l}^{y_u} \gamma_h \mathbb{1}\left(y_j = h\right) + \varepsilon_j, \tag{D.1}$$

where i indicates the polynomial degree in the first sum. We use a 7$^{\text{th}}$ degree polynomial in our baseline estimates. To avoid irregularities coming from the far tails of the distribution, we exclude files with reported revenues below the 5$^{\text{th}}$ percentile or above the 95$^{\text{th}}$ percentile of relative distance from $\hat{y}$. These restrictions automatically drop files with zero reported revenues, which account for slightly less than 2% of all 2007-2010 files. The excluded segment $\left[y^l, y^u\right]$ is the area affected by bunching responses. Bin dummies for $y \in \left[y^l, y^u\right]$ ensure that the excess mass at $\hat{y}$ does not affect

[50] Positivity of p_L on the range of declarations $[0, \bar{d}]$ requires $\beta\bar{d} + \Delta \leq \mu$, which reduces to $\beta \leq \mu/\bar{d}$ when $\Delta = 0$. This restriction binds only at declarations far above the threshold, where no bunching occurs, and therefore does not affect the concavity argument.

[51] With p_0 nonlinear, $p_0''(d^0)$ adds $\tau\gamma p_0''(d^0)e_0$ to A and the corresponding third-derivative term to C. B and D only depend on the audit jump and are unchanged.

the counterfactual distribution fit. While our preferred choice is to set y^u visually at the first bin above $\hat{y}$, we choose y^l using an iterative procedure. The latter searches for the bin that generates an estimated counterfactual with a missing mass below $\hat{y}$ equal to the excess mass above $\hat{y}$. Using the estimated counterfactual, we can compute the excess mass as the ratio between the excess (relative to counterfactual) observed number of SeS files and the average level of the counterfactual in the segment $[\hat{y}, y^u]$. We will refer to this relative excess mass as a bunching estimate, or $\hat{B}$.

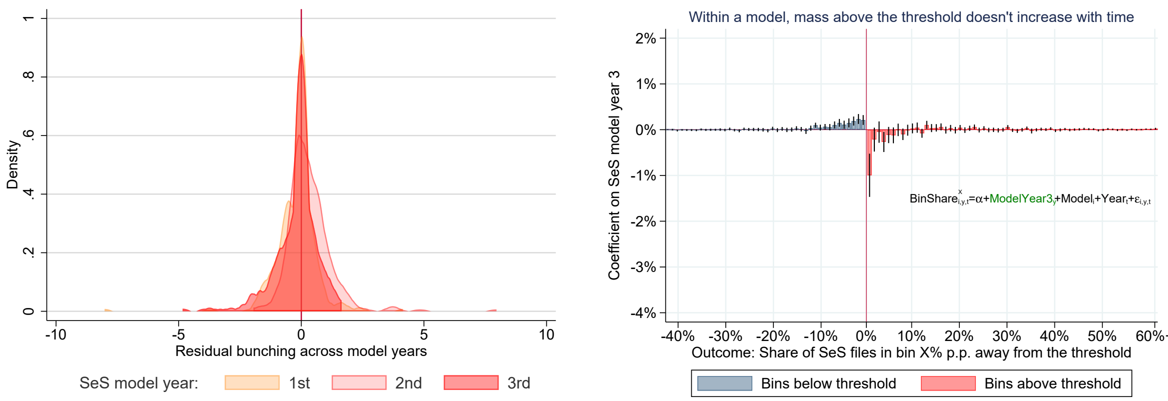
We compute standard errors to bunching estimates using a semi-parametric bootstrap procedure. Equation (D.1) provides the structure for our routine. In every bootstrap iteration we draw with replacement from the residuals $\hat{\varepsilon}_j = c_j - \hat{c}_j$, where $\hat{c}_j = \sum_{i=1}^{K} \hat{\beta}_i (y_j)^i + \sum_{i=y_l}^{y_u} \hat{\gamma}_i \mathbb{1}(y_j = i)$, and $\left(\hat{\beta}, \hat{\gamma}\right)$ are the estimated coefficients from the specification in (D.1). We use the residuals to build a new number of taxpayers in each bin j so that in iteration r the number of taxpayers in bin j is $c_j^r = \hat{c}_j + \varepsilon_j^r$ and ε_j^r is the residual drawn for bin j in iteration r. We use the new vector $\left(c_j^r\right)_{j \in J}$ as the dependent variable when re-estimating (D.1) and we employ the resulting $\left(\hat{c}_j^r\right)_{j \in J}$ as the counterfactual needed to compute a bunching quantity $\hat{B}^r$. We repeat this routine for 1,000 iterations. Confidence intervals on $\hat{B}$ can be computed by taking the 2.5$^{\text{th}}$ and the 97.5$^{\text{th}}$ percentiles of the bunching estimate distribution across all iterations, while the bunching standard deviation is simply the standard deviation of the same empirical distribution.

E Reporting VS Production Responses

We test the learning-to-adjust hypothesis discussed in Section 4.2 in two ways. First, we estimate sector-year bunching and residualize it by sector and year fixed effects. Figure D1 Panel (a) plots the residual distributions for the first, second, and third year of application of each sector's revenue prediction model. Despite potential learning, bunching does not increase in later years of the same model. Second, we split SeS files into one-percentage-point bins of distance from the presumed revenues in each sector-year. If firms gradually adjust production, mass should increase in bins just above the threshold. For each bin, we regress its file share on a dummy for the model's final year, controlling for sector and calendar-year fixed effects. Figure D1 Panel (b) shows the coefficients for bins around the threshold. We find no evidence of mass shifting above the threshold over time.

Figure D1: Bunching evolution within SeS models, 2007-2010

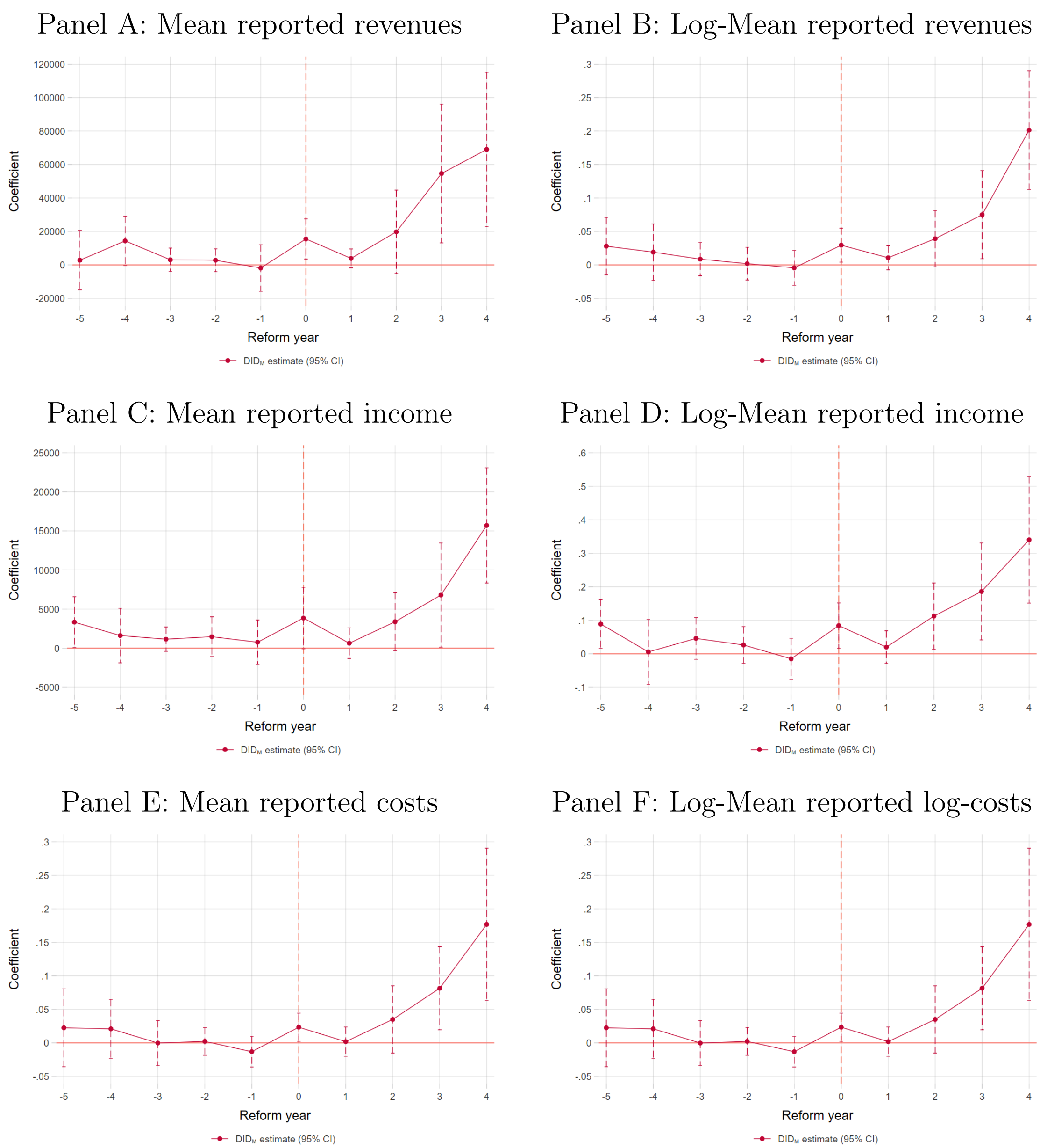
(a) Bunching evolution within SeS models, 2007-2010 **(b)** Effect of the last year of application of a SeS model



Notes: The figure shows the evolution of bunching within SeS models. Panel (a) plots the distribution of 2007–2010 bunching residuals from the regression bunching$_{i,t} = \alpha + \beta_i + \gamma_t + \varepsilon_{i,t}$, where observations are SeS model-years and β_i and γ_t denote model and year fixed effects. A SeS model corresponds to the three-year application of a sector-specific estimation model. Residuals are shown separately for the first, second, and third year of a model's application. The sample includes 762 model-years and excludes negative bunching estimates. Panel (b) plots coefficients from regressions of bin shares on a dummy for the third (final) year of a SeS model's application. Observations are SeS model-years. For each one–percentage-point bin of distance X from the presumed revenue threshold, we regress the share of files in that bin on the third-year dummy, controlling for SeS model and calendar-year fixed effects and clustering standard errors by model. Coefficients are shown for bins below (blue) and above (red) the threshold with 95% confidence intervals. Bin shares are computed using SeS taxpayers continuously filing between 2007 and 2016.

F Robustness of Event Study Estimates

Figure F1: Robustness: de Chaisemartin and D'Haultfoeuille (2020) correction



Notes: this Figure shows the effects of the reward regime's introduction in a sector from the estimator proposed by de Chaisemartin and D'Haultfoeuille (2020) (DID_M). The regressions are estimated on the sample of all Sector Study files from single-sector taxpayers continuously filing over the 2007-2016 period, aggregated by sector-year. Only sectors accessing the reward regime by 2016 are considered. Number of sector-years: 1550. Reported revenues are winsorized at the 99th percentile. Controls and weighting are as defined in the discussion of Eq. (9). We mark the relative year before the reform with the red dashed vertical line at year 0. Estimation is performed using the *dynamic* and *placebo* options of the authors' supplied Stata package *did_ multiplegt*. Estimation requirements allow us to compute only up to four post-treatment effects. Standard errors are computed with a bootstrap procedure with 50 replications.

G Structural Estimation

G.1 Environment, eligibility threshold, and audit probabilities

We calibrate the following extension of our model. A firm has true (pre-tax) revenue $y \geq 0$, true (pre-tax) cost $x \geq 0$, and manipulation capacity $1/\kappa > 0$. The firm chooses reported revenue $d_y \geq 0$ and reported costs $d_x \geq 0$, implying declared profit (tax base) $\pi_d = d_y - d_x$. The tax system features a proportional tax rate $\tau \in (0,1)$ and no subsidy for losses, so total due taxes are $\tau(d_y - d_x)^+$, where $z^+ := \max\{z, 0\}$. Audit probability depends on whether the firm satisfies the eligibility rule based on reported revenues and a cost-adjusted threshold:

$$d_y \geq \widehat{y}(d_x), \qquad \widehat{y}(d_x) = \widehat{y}_0 + \hat{y}_1 d_x, \qquad \widehat{y}_0 > 0,\ \hat{y}_1 \geq 0.$$

The parameter $\hat{y}_1$ is the slope of the eligibility threshold in reported costs: it determines how much required reported revenue increases when reported costs rise. Audit probability is a step function,

$$p(d_y, d_x) = \begin{cases} p_H & \text{if } d_y < \widehat{y}(d_x), \\ p_L & \text{if } d_y \geq \widehat{y}(d_x), \end{cases} \qquad 0 \leq p_L < p_H \leq 1.$$

Denote true profit $\pi = y - x$, so the maximum taxable base that can be reduced is π^+. We define the reduction in the taxable base relative to truthful reporting as

$$e_\tau(d_y, d_x; y, x) := \min\Big\{(\underbrace{y - x}_{\pi} - \underbrace{(d_y - d_x)}_{\pi_d})^+,\ \pi^+\Big\}.$$

If audited, the firm pays a proportional punishment $\gamma > 1$ per unit of avoided taxes; expected penalties equal $\tau\gamma\, p(d_y, d_x)\, e_\tau(d_y, d_x; y, x)$.

The firm chooses (d_y, d_x) to maximize expected after-tax profit net of expected penalties and manipulation costs:

$$\max_{d_y \geq 0,\ d_x \geq 0} (y - x) - \tau(d_y - d_x)^+ - \tau\gamma\, p(d_y, d_x)\, e_\tau(d_y, d_x; y, x) - \mathcal{C}(y - d_y, d_x - x; \kappa).$$

Define revenue misreport $e_y := y - d_y \geq 0$ and cost misreport $e_x := d_x - x \in [-x, \infty)$. An optimality condition ensures revenues are never over-reported, $e_y \geq 0$, while costs can be inflated ($e_x \geq 0$) to reduce taxable profit, or under-reported to lower the audit threshold and increase eligibility. We discipline $\mathcal{C}$ as follows. For $m \geq 0$ we use an isoelastic manipulation cost

$$c(m, \kappa) = \frac{\varepsilon}{1 + \varepsilon}\, \kappa^{1/\varepsilon}\, m^{1+1/\varepsilon}, \qquad \varepsilon > 0,$$

so higher κ increases marginal manipulation costs. To allow for non-separability across margins we assume

$$\mathcal{C}(e_y, e_x; \kappa) = (1 - b_2)c(e_y, \kappa) + (b_1 - b_2)c(|e_x|, \kappa) + b_2 c(e_y + |e_x|, \kappa), \qquad 0 \leq b_2 < \min\{1, b_1\}.$$

The parameter $b_1 > 0$ scales the difficulty of cost manipulation relative to revenues. When $b_2 = 0$ margins are separable; when $b_2 > 0$ joint manipulation is more expensive than the sum of its parts. The term $c(e_y + |e_x|, \kappa)$ captures situations in which inflating costs facilitates revenue misreporting and vice versa, so that a common concealment technology supports manipulation on both sides of the profit equation. This feature allows the model to match patterns in the data where both margins increase in response to Δ.

G.2 Characterization of the solution

We can show that the optimal report of each type (y, x, κ) falls into one of four categories: interior (in either regime), zero declared profits, and bunchers. See supplemental appendix S.9 for details.

Optimal reporting is obtained by comparing a small set of candidate policies. Interior candidates in the low- and high-audit regions are characterized by equating the marginal benefit of evasion to the corresponding marginal manipulation costs. The key difference from the baseline model is that now both revenue and cost are optimally set by equating the common marginal benefit to the (interdependent, when $b_2 > 0$) marginal misreporting costs. Joint manipulation may also imply negative reported profits. Since losses are not subsidized, such reports are never optimal: if an interior candidate implies negative declared profits, the firm instead chooses the least costly zero-profit declaration while respecting feasibility. Bunching arises when the interior policy under p_L violates eligibility, i.e. $d_y < \widehat{y}(d_x)$, in which case the relevant candidate imposes $d_y = \widehat{y}(d_x)$ and solves the resulting one-dimensional maximization problem.

G.3 Calibration and simulation of the economy

To simulate the economy we draw (y, x, κ) from a joint distribution ensuring all are positive with positive correlation between revenues and costs:

$$\log y = \mu_y + \sigma_y Z_1, \qquad x = \vartheta\, y\, \exp(\sigma_\eta Z_2), \qquad \log \kappa = \mu_\kappa + \sigma_\kappa Z_3,$$

where $(Z_1, Z_2, Z_3) \sim \mathcal{N}(0, \Sigma_Z)$, and $\mathrm{Corr}(Z_1, Z_3) = \rho$ so that firm size and manipulation costs may be correlated. Z_2 is assumed independent. Correlation between y and x is induced by the multiplicative structure; $\vartheta \in (0,1)$ controls the typical cost share. Specifically, $\mathbb{E}\left[\log\left(\frac{x}{y}\right)\right] = \log \vartheta$, and negative-profit firms arise with probability $\Pr(x > y) = \Pr\left(\vartheta e^{\sigma_\eta Z_2} > 1\right) = 1 - \Phi\left(\frac{-\log \vartheta}{\sigma_\eta}\right) > 0$. The parameters of the joint distribution are $\mathbf{\Sigma} = (\mu_y, \sigma_y, \vartheta, \sigma_\eta, \mu_\kappa, \sigma_\kappa, \rho)$. The parameters (μ_y, σ_y) determine the distribution of firm scale, (ϑ, σ_η) determine average profitability and dispersion in cost shares, and $(\mu_\kappa, \sigma_\kappa, \rho)$ determine how manipulation costs are distributed and their covariance with size.

G.4 Matching empirical and theoretical moments

For each simulated firm we compute optimal declarations (d_y, d_x) under the pre- and post-reform audit rules, form the implied threshold $\widehat{y}(d_x) = \widehat{y}_0 + \hat{y}_1 d_x$, and compute $\bar{d}_y = d_y / \widehat{y}(d_x)$. We divide firms into 3 bins of this ratio in the pre-policy economy, where the bunching region is defined in $(0.999, 1.05]$, and compute averages and policy responses of threshold-adjusted declarations $(\bar{d}_y)$, threshold-adjusted profits $(\pi_d / \hat{y}\,(d_x))$, and the threshold itself in the same three regions.

Measuring the masses of taxpayers in the data is straightforward since the three regions are defined by observed declared revenues before the policy. We similarly measure the average threshold and normalized declarations and profits conditional on being in each area. To quantify outcomes after the Reward Regime, we exploit the fact that for outcome $o_{\mathcal{A}}$ in area $\mathcal{A}$, $o_{\mathcal{A}}^{Post} = o_{\mathcal{A}}^{Pre} + \delta o_{\mathcal{A}}$, where $\delta o_{\mathcal{A}}$ is the causal effect estimated using the identification strategy in Section 5. We summarize the time-varying effects from equation (9) into a single $\delta o_{\mathcal{A}}$ per dependent variable by averaging across the post-reform coefficients.

G.5 Estimation and constraints

We estimate $\theta = (\Sigma, \widehat{y}_0, \hat{y}_1, \mu, \alpha, \Delta^{pre}, \Delta^{post}, \varepsilon, b_1, b_2)$ and calibrate (τ, γ) externally. Estimation uses a two-step simulated method of moments.[52] The first step minimizes the quadratic distance between simulated and empirical moments using an identity weighting matrix. The second step repeatedly simulates the model at the first-step estimate $\hat{\theta}_1$ and uses the simulated variance of each moment to construct a diagonal weighting matrix that down-weights more variable moments. Specifically, if $\widehat{\text{Var}}(m_j)$ denotes the simulated variance of moment j across replications, the second-step weighting matrix is[53]

$$W_2 \propto \text{diag}\left(\frac{1}{\widehat{\text{Var}}(m_1)}, \ldots, \frac{1}{\widehat{\text{Var}}(m_J)}\right).$$

We impose bounds on audit probabilities and manipulation-cost parameters to ensure a well-behaved environment, including $\Delta^t \geq 0$ and an admissible range for b_2 given b_1, and restrict the range of the cost share of revenues. Finally, we impose an enforcement-feasibility restriction based on equilibrium audit intensity under optimal reporting. Let $\mathrm{Q}^t(\theta) = \mathbb{E}\big[p_L^t \mathbf{1}\{d_y \geq \widehat{y}(d_x)\} + p_H^t \mathbf{1}\{d_y < \widehat{y}(d_x)\}\big]$ for $t \in \{\text{pre}, \text{post}\}$. We require $Q^{pre}(\theta), Q^{post}(\theta) \in [\underline{Q}, \overline{Q}]$, ruling out implausibly low or high enforcement intensity. These restrictions do not impose budget neutrality across regimes or constrain the relative ranking of Q^{pre} and Q^{post}.

Table F1: Additional calibrated primitives

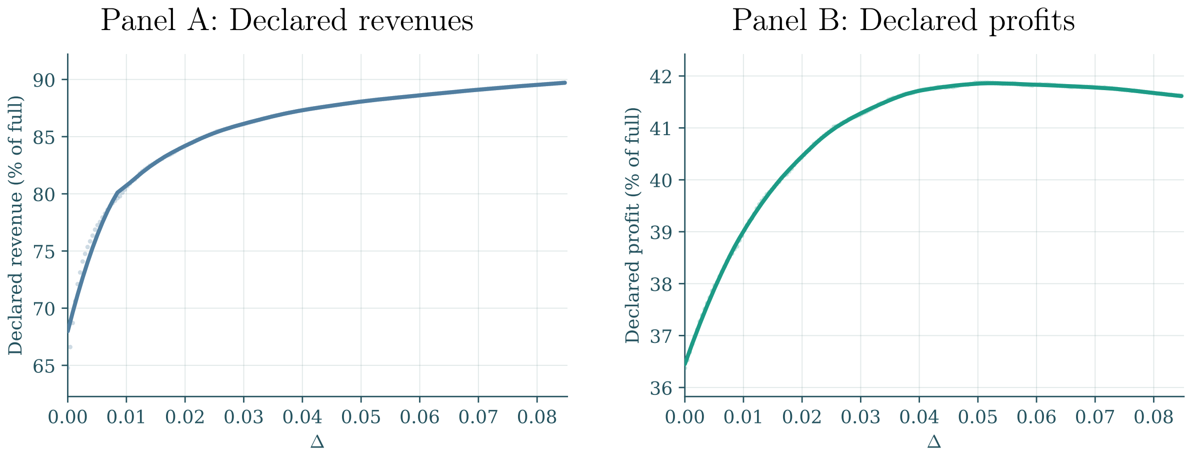
Parameter	Value	Description
μ_y, σ_y	1.0376, 0.3305	Mean and standard deviation of $\log y$.
$E[x/y]$, SD(x/y)	0.1109, 0.0058	Mean and standard deviation of the true cost-to-revenue ratio.
μ_κ, σ_κ	1.4985, 0.7305	Mean and standard deviation of $\log \kappa$.
ρ	0.8239	Correlation between $\log y$ and $\log \kappa$.
$\hat{y}_0$	1.5997	Eligibility intercept.
$\hat{y}_1$ (δ)	0.6797	Eligibility slope on declared cost.

Notes: The table reports the auxiliary primitives estimated in the stage-1 downstream calibration. The moments for x/y are reported in levels and are implied by the estimated log-normal cost-share distribution. Higher κ corresponds to higher marginal manipulation costs.

[52] Across evaluations of the objective function we use a common seed, so that changes in the objective reflect changes in the parameter vector rather than changes in the simulation draw.

[53] For numerical stability, we normalize diagonal weights, cap extreme values, and shrink the weighting matrix toward the identity by assigning the identity a small weight. We also allow for block-specific weights on the four groups of moments—shares, $\bar{d}_y$, $\bar{\pi}_d$, and $\widehat{y}(d_x)$—to control their relative influence in the criterion.

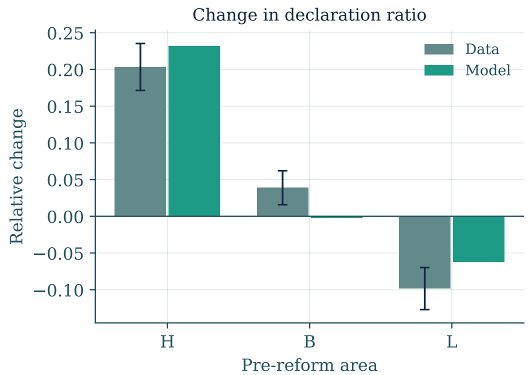
Figure F2: Concavity in Δ of declared revenues and profits



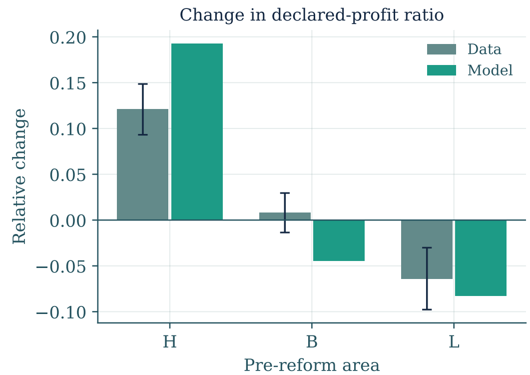
Notes: The figure numerically investigates the concavity in Δ of aggregate declared revenues (Panel A) and profits (Panel B) in the calibrated economy.

Figure F3: Model fit

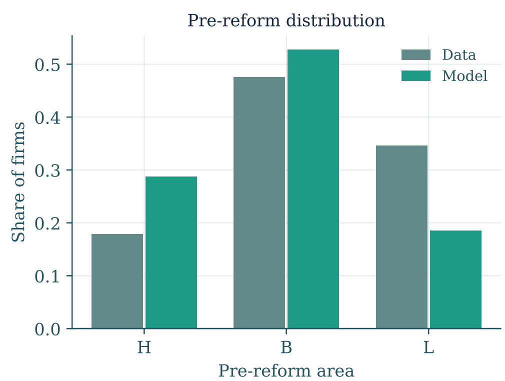
(a) Declared revenues



(b) Declared profits



(c) Pre-reform distribution of declared revenues

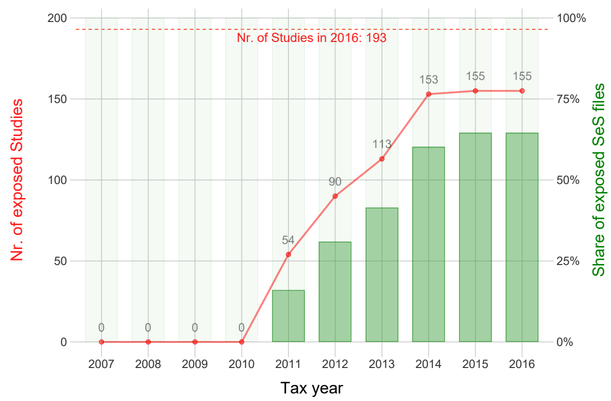


Notes: The figure compares empirical post-reform responses with simulations from the calibrated downstream model. Panels A and B report changes in declared revenues, $\Delta d_y/\widehat{y}$, and declared profits, $\Delta[(d_y - d_x)/\widehat{y}]$, respectively. Light bars show simulated moments; dark bars show empirical estimates with 95% confidence intervals. Groups $\mathcal{H}$, $\mathcal{B}$, and $\mathcal{L}$ are based on firms' pre-reform positions. Panel C compares their empirical and simulated pre-reform shares.

H Additional Tables and Figures

H.1 Additional Figures

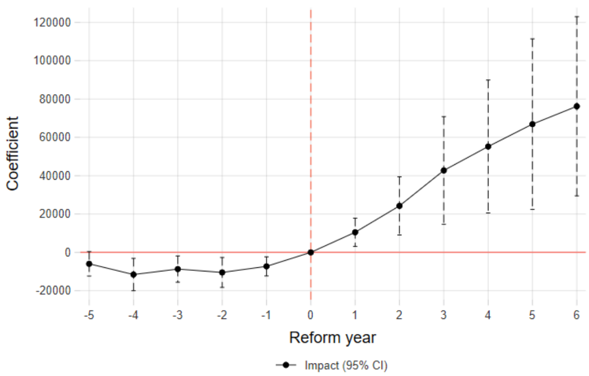
Figure G1: Reward regime: staggered introduction, 2011-2016



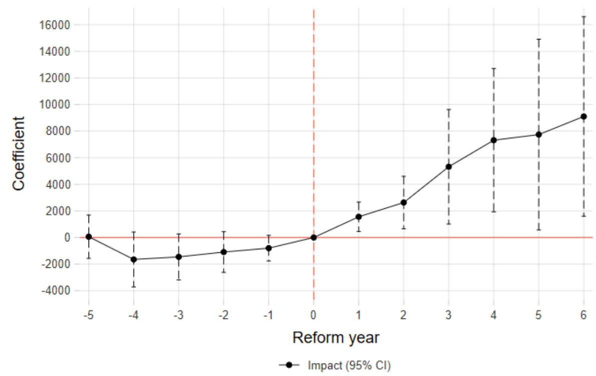
Notes: The figure shows the staggered introduction of the 2011 reward regime across SsS. The red line reports the number of sectors with access to the regime each year up to 2016 (left axis), while the dark green bars show the share of SeS files covered in each year (right axis), computed over single-sector taxpayers filing continuously between 2007 and 2016. For simplicity, five sectors with partial access are treated as fully covered. The horizontal dashed line indicates the total number of sectors in 2016.

Figure G2: Reward regime effects on mean revenues and profits (in Euros)

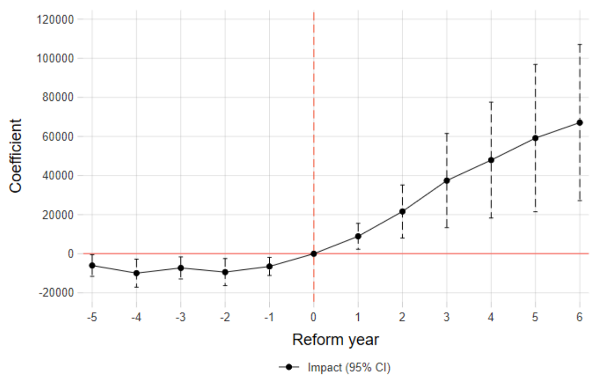
(a) Panel A: Mean reported revenues



(b) Panel B: Mean reported income



(c) Panel C: Mean reported costs



Notes: this Figure shows the effects of the reward regime's introduction in a sector on mean reported revenues (Panel A), mean gross profits (Panels B), and costs (Panel C) defined as the difference between reported revenues and gross profits. Dependent variables are expressed in Euro terms. Whiskers represent 95% CIs. Effects are relative to the year before the advent of the reform in each sector, marked at year 0 by the red dashed vertical line. Estimates are based on our specification in (9). Standard errors are clustered at the sector level. The regressions are estimated on the sample of all Sector Study files from single-sector taxpayers continuously filing over the 2007-2016 period, aggregated by sector-year. Only sectors accessing the reward regime by 2016 are considered. Number of sector-years: 1550. Reported revenues are winsorized at the 99th percentile.

H.2 Additional Tables

Table G1: Sector Studies compliance benefits, before and after 2011

<table>
<tr><th colspan="3">SeS required condition</th><th colspan="2">Audit exemption benefits</th></tr>
<tr><th>Congruence</th><th>Normality</th><th>Coherence</th><th>Before 2011</th><th>Since 2011</th></tr>
<tr><td>✓</td><td></td><td></td><td colspan="2">No SeS audits (revenues)</td></tr>
<tr><td></td><td>✓</td><td>✓</td><td colspan="2">No SeS audits (costs, inputs)</td></tr>
<tr><td>✓</td><td>✓</td><td></td><td>No analytic-inductive audits
up to $e \leq 40\% y$, $e \leq €50{,}000$</td><td></td></tr>
<tr><td>✓</td><td>✓</td><td>✓</td><td></td><td>1. No analytic-inductive audits
up to any amount
2. No synthetic audits
up to $\pi(s)-\pi \leq 33\% \cdot \pi(s)$
3. Shorter statute of limitation</td></tr>
</table>

Notes: the Table reports the main tax audit and assessment benefits from being congruous, coherent, and normal by the definitions provided by Sector Studies, before and after the introduction of the 2011 reward regime. Congruence refers to the condition of reporting revenues at or above the level presumed by *Gerico*. Normality and coherence refer to the condition of reporting a number of accounting and economic indicators within sector-specific acceptable ranges as determined by *Gerico*. Notation: e refers to undeclared amounts, y to revenues, π to gross profits or income, and $\pi(s)$ to synthetically-determined income. The statute of limitation to inspect an eligibile taxpayer's file drops by one year since 2011.

Table G2: Bunching estimates by polynomial order and upper bound

	POLYNOMIAL ORDER							
UPPER BOUND	3	4	5	6	7	8	9	10
0	15.38	14.95	11.53	11.42	**9.56**	9.15	8.47	7.69
1	18.57	18.77	14.45	14.01	12.27	11.63	10.74	9.95
2	21.36	20.66	16.86	16.25	14.6	13.03	12.75	11.45
3	23.73	23.91	19.55	18.33	16.85	14.57	14.77	13.24
4	15.07	24.78	21.83	19.93	18.91	16.39	16.32	14.91
5	17.17	26.59	24.34	21.31	20.4	17.99	17.43	16.71

Notes: the Table reports various bunching estimates computed with the SeS files submitted by the universe of single-sector businesses in the 2007-2010 tax years. Estimates are reported for each combination of two parameters choices. First, the upper bound y_u of the area affected by excess bunching, identified by the floor of the relevant bin in percentage points of presumed revenues. For reference, upper bound 0 indicates we limit the bunching area to the one-percentage-point bin including the presumed revenues threshold. Second, the polynomial order, that is the degree of the polynomial in bin order used to estimate the smooth bunching counterfactual. We select 0 for the upper bound and 7 for the polynomial order in our baseline estimate, highlighted in orange. In all estimations, bin width is fixed at one percentage point of presumed revenues.

Supplementary Material

S.1 Omitted Computations from Main Propositions

S.1.1 Derivation of L in Lemma 7

To identify L, substitute $\bar{g} = Kz + Lz^2 + z^3\rho_{\bar{g}}$ into equation (A.2) with $\Delta = z^2$, and match powers of z. The left-hand side becomes

$$\frac{1}{2}C_2\bar{g}^2 = \frac{1}{2}C_2\left(K^2z^2 + 2KLz^3 + z^4\rho_x(z,\mu)\right),$$

where ρ_x includes all the cross-products of $\rho_{\bar{g}}$ and the L^2 term, which are bounded. The right-hand side is

$$\tau\gamma(1+\alpha)e_0z^2 + \frac{1}{6}C_3K^3z^3 - K\tau\gamma z^3 + z^4\rho^1_\Psi(z,\mu),$$

with bounded ρ^1_Ψ. Here we used $\bar{g}^3 = K^3z^3 + O(z^4)$, $\bar{g}\Delta = Kz^3 + O(z^4)$ and $\Delta^2 + \bar{g}^4 + \bar{g}^2\Delta = O(z^4)$. Matching the coefficients on z^2 recovers $\frac{1}{2}C_2K^2 = \tau\gamma(1+\alpha)e_0$, the expression for K in the Lemma. Matching coefficients on z^3 gives

$$C_2KL = \frac{1}{6}C_3K^3 - K\tau\gamma.$$

Thus

$$L = \frac{-\tau\gamma + C_3K^2/6}{C_2}.$$

S.1.2 Derivations of the singular terms of $R_{\Delta\Delta}$

From Proposition 3, aggregate marginal revenue can be decomposed as

$$R_\Delta = B_R + L_R + H_R,$$

where

$$B_R = -\tilde{e}_\Delta(e_H - \tilde{e})f(\hat{y} + \tilde{e}),$$

and

$$L_R = -e_{L,\Delta}\{1 - F(\hat{y} + e_L)\}, \qquad H_R = -e_{H,\Delta}\{F(\hat{y} + \tilde{e}) - F(\hat{y} + e_H)\}.$$

We differentiate the three components and isolate the singular terms (*i.e.*, those that can contain negative powers of z).

The derivative of the bunching component is

$$\begin{aligned} B_{R,\Delta} = &- \tilde{e}_{\Delta\Delta}(e_H - \tilde{e})f(\hat{y} + \tilde{e}) - \tilde{e}_\Delta(e_{H,\Delta} - \tilde{e}_\Delta)f(\hat{y} + \tilde{e}) \\ &- \tilde{e}^2_\Delta(e_H - \tilde{e})f'(\hat{y} + \tilde{e}). \end{aligned}$$

The derivatives of the high- and low-region components are

$$H_{R,\Delta} = -e_{H,\Delta\Delta}\{F(\hat{y} + \tilde{e}) - F(\hat{y} + e_H)\} - e_{H,\Delta}\{f(\hat{y} + \tilde{e})\tilde{e}_\Delta - f(\hat{y} + e_H)e_{H,\Delta}\},$$

and

$$L_{R,\Delta} = -e_{L,\Delta\Delta}\{1 - F(\hat{y} + e_L)\} + e^2_{L,\Delta}f(\hat{y} + e_L).$$

By Lemma 7 and Corollary 8, the first and second derivatives of e_H and e_L are bounded on the cone. Thus the singular terms in $R_{\Delta\Delta}$ can only come from $B_{R,\Delta}$ and from the term $-e_{H,\Delta}f(\hat{y} + \tilde{e})\tilde{e}_\Delta$ in

$H_{R,\Delta}$. The remaining terms contain only bounded derivatives of e_H and e_L, bounded density terms, and interval differences that can be written as interval lengths times bounded density averages. So we can write

$$R_{\Delta\Delta} = R_{\Delta\Delta}^{\text{sing}} + \rho_{R_{\Delta\Delta}}, \tag{S.1.1}$$

where $\rho_{R_{\Delta\Delta}}$ is uniformly bounded in $\mathcal{Z}_{m_\rho}$ and

$$-R_{\Delta\Delta}^{\text{sing}} = T_1 + T_2 + T_3 + T_4,$$

with

$$\begin{aligned}
T_1 &= \tilde{e}_{\Delta\Delta}(e_H - \tilde{e}) f(\hat{y} + \tilde{e}), \\
T_2 &= -\tilde{e}_\Delta^2 f(\hat{y} + \tilde{e}), \\
T_3 &= \tilde{e}_\Delta^2 (e_H - \tilde{e}) f'(\hat{y} + \tilde{e}), \\
T_4 &= 2\tilde{e}_\Delta e_{H,\Delta} f(\hat{y} + \tilde{e}).
\end{aligned}$$

We now compute the four terms in $-R_{\Delta\Delta}^{\text{sing}}$ using Corollary 8. The first term is

$$\begin{aligned}
T_1 &:= \tilde{e}_{\Delta\Delta}(e_H - \tilde{e}) f(\hat{y} + \tilde{e}) \\
&= \left(\frac{K}{4z^3} + \frac{\rho_{\tilde{e},\Delta\Delta}}{z}\right)\left(Kz + \left(L - \frac{\alpha\tau\gamma}{C_2}\right) z^2 + z^3 \rho_{H\tilde{e}}\right)\left(f_0 - Kf_1 z + z^2 \rho_f\right) \\
&= \frac{K^2 f_0}{4z^2} + \frac{K}{4z}\left\{\left(L - \frac{\alpha\tau\gamma}{C_2}\right) f_0 - K^2 f_1\right\} + \rho_{T_1}(z, \mu).
\end{aligned}$$

The second term is

$$\begin{aligned}
T_2 &:= -\tilde{e}_\Delta^2 f(\hat{y} + \tilde{e}) \\
&= -\left(\frac{K^2}{4z^2} + \frac{KL}{z} + \rho_{\tilde{e},\Delta^2}\right)\left(f_0 - Kf_1 z + z^2 \rho_f\right) \\
&= -\frac{K^2 f_0}{4z^2} - \frac{1}{z}\left\{-\frac{K^3 f_1}{4} + KLf_0\right\} + \rho_{T_2}(z, \mu).
\end{aligned}$$

The third term is

$$\begin{aligned}
T_3 &:= \tilde{e}_\Delta^2 (e_H - \tilde{e}) f'(\hat{y} + \tilde{e}) \\
&= \left(\frac{K^2}{4z^2} + \frac{KL}{z} + \rho_{\tilde{e},\Delta^2}\right)\left(Kz + z^2 \rho_{H\tilde{e}}^0\right)\left(f_1 + z\rho_{f'}\right) \\
&= \frac{K^3 f_1}{4z} + \rho_{T_3}(z, \mu).
\end{aligned}$$

The fourth term is

$$\begin{aligned}
T_4 &:= 2\tilde{e}_\Delta e_{H,\Delta} f(\hat{y} + \tilde{e}) \\
&= 2\left(-\frac{K}{2z} - L + z\rho_{\tilde{e},\Delta}\right)\left(-\frac{\alpha\tau\gamma}{C_2} + z\rho_{H,\Delta}\right)\left(f_0 - Kf_1 z + z^2 \rho_f\right) \\
&= \frac{K\alpha\tau\gamma f_0}{C_2 z} + \rho_{T_4}(z, \mu).
\end{aligned}$$

Each remainder ρ_{T_j} is bounded on $\mathcal{Z}_{m_\rho}$. Indeed, after the displayed z^{-2} and z^{-1} terms are removed, every remaining term is a finite product of bounded primitive terms and bounded $\rho_\cdot$-type remainders, multiplied by a nonnegative power of z.

Adding the four expressions, the z^{-2} terms cancel. The coefficient of z^{-1} in $T_1+T_2+T_3+T_4$ is

$$\frac{K}{4}\left\{\left(L-\frac{\alpha\tau\gamma}{C_2}\right)f_0-K^2f_1\right\}-\left\{-\frac{K^3f_1}{4}+KLf_0\right\}+\frac{K^3f_1}{4}+\frac{K\alpha\tau\gamma f_0}{C_2}$$
$$=Kf_0\left[-\frac{3L}{4}+\frac{3\alpha\tau\gamma}{4C_2}+\frac{K^2f_1}{4f_0}\right].$$

Substituting $L=(C_3K^2/6-\tau\gamma)/C_2$ into the bracket,

$$-\tfrac{3}{4}L+\tfrac{3}{4}\alpha\tau\gamma/C_2+\tfrac{1}{4}K^2f_1/f_0=\tfrac{3}{4}(1+\alpha)\tau\gamma/C_2-\tfrac{1}{8}C_3K^2/C_2+\tfrac{1}{4}K^2f_1/f_0.$$

Using $\tau\gamma(1+\alpha)/C_2=K^2/(2e_0)$, this equals $\frac{K^2}{4e_0}\big(\frac{3}{2}-\frac{e_0}{2}C_3/C_2+e_0f_1/f_0\big)$, so the coefficient of z^{-1} is

$$Kf_0\cdot\tfrac{K^2}{4e_0}\big(\tfrac{3}{2}-\tfrac{e_0}{2}C_3/C_2+e_0f_1/f_0\big)=\tfrac{K}{2}\,\tfrac{\tau\gamma(1+\alpha)f_0}{C_2}\,\big(\tfrac{3}{2}-\tfrac{e_0}{2}C_3/C_2+e_0f_1/f_0\big).$$

We therefore write

$$T_1+T_2+T_3+T_4=\frac{G}{z}+\rho_T(z,\mu),\qquad \sup_{\mathcal{Z}_{m_\rho}}|\rho_T(z,\mu)|<\infty,$$

where

$$G=\frac{K}{2}\frac{\tau\gamma(1+\alpha)f_0}{C_2}S,$$

and

$$S=\frac{3}{2}+e_0\frac{f_1}{f_0}-\frac{e_0}{2}\frac{C_3}{C_2}.$$

This is the desired expression for G and S.

S.1.3 Derivation of the lognormal-isoelastic concavity condition (8)

Under the flat rule, interior evasion satisfies $c'(e^I(\mu))=\tau(1-\gamma\mu)$. Since $c'(e)=\kappa e^{1/\varepsilon}$, this choice is $e^I(\mu)=[\tau(1-\gamma\mu)/\kappa]^\varepsilon$. Moreover, $c''(e)=(\kappa/\varepsilon)e^{1/\varepsilon-1}$ and $c'''(e)=(\kappa/\varepsilon)(1/\varepsilon-1)e^{1/\varepsilon-2}$, so $e^I(\mu)c'''(e^I(\mu))/c''(e^I(\mu))=1/\varepsilon-1$. For the income distribution, differentiating the logarithm of the lognormal density gives $f_y'(y)/f_y(y)=-y^{-1}[1+(\log y-\mu_y)/\sigma^2]$. The income of the marginal buncher at the flat-rule is $\hat{y}+e^I(\mu)$. Letting $\hat{t}(\mu)=1+e^I(\mu)/\hat{y}$ and $m_w=\mu_y-\log\hat{y}$, we have $e^I(\mu)/(\hat{y}+e^I(\mu))=[\hat{t}(\mu)-1]/\hat{t}(\mu)$ and $\log(\hat{y}+e^I(\mu))-\mu_y=\log\hat{t}(\mu)-m_w$.

Substituting these relationships and the cost-curvature ratio into equation (7) yields

$$2-\frac{1}{2\varepsilon}+\frac{1-\hat{t}(\mu)}{\hat{t}(\mu)}\left[1+\frac{\log\hat{t}(\mu)-m_w}{\sigma^2}\right].$$

which is expression (8).

To numerically compute $\mu^\star$ in Figure 1 we proceed as follows. For each candidate radius m, define $\underline{G}(m)=\inf_{\mu\le m}G(\mu)$ and evaluate, at the exact marginal buncher, $\rho_R(\sqrt{\Delta},\mu)=R_{\Delta\Delta}(\Delta;\mu)+G(\mu)/\sqrt{\Delta}$. We set $\mu^\star$ to the largest m such that $\underline{G}(m)>0$ and $\sqrt{m}\sup_{\mathcal{Z}_m}|\rho_R|<\underline{G}(m)$. This sufficient bound is conservative because the extrema need not occur at the same policy and the remainder may reinforce concavity. Thus, exceeding $\mu^\star$ does not imply that concavity fails; it must instead be assessed pointwise.

S.2 Extensions: Omitted discussions and interpretations

S.2.0.1 Reduction of extended model to baseline

The extended model Proposition 11 modifies the proof of the baseline result for two main reasons. First, the marginal buncher is indexed by the landing point s on the frontier (bunching is a whether *and where* problem in the extension), so the coefficient a_3 is integrated over $\mathcal{S}$ with the transformed density h_0. Second, the revenue brought by a buncher is no longer mechanically one unit of revenue per unit of discrepancy: it depends on the revenue component ℓ_y of the least-cost adjustment, on the second-order adjustment b_y, and on the way the constrained location moves with the audit discount, $(t_\eta)_y$. Notice that the baseline (where the frontier collapses to a single point) is obtained from (C.22) letting $w = 1$, $\ell = 1$, $A = C_2$, $b = 0$, $t_\eta = 0$, $h_0 = f_0$, and $h_1 = -f_1$. This gives $a_3 = -\frac{4}{3}G$ with K, L and G reducing to the values in Lemma (10) and (A.5), respectively.

S.2.0.2 Nonlinear $\hat{Y}$ and E

Consider the setting where the eligibility threshold may depend nonlinearly on d_x, and the tax base may weigh the two margins nonlinearly. The analysis extends replacing the vector $w = (1, \hat{y}_1)'$ with the frontier gradient $(1, \hat{Y}'(s))'$, evaluated at the firm's landing point, and η with the gradient $E_e(\mathbf{e}_0)$. The interior choices change only through their first-order condition: $\mathbf{e}^I(p)$ now solves $C_e(\mathbf{e}^I) = \tau(1-\gamma p)E_e(\mathbf{e}^I)$, and (C.2) goes through with C_2 replaced by $\mathcal{H}_0 = C_{ee} - \tau\,(1-\gamma p)\,E_{ee}$ and η by $E_e(\mathbf{e}_0)$. Notice that this adds no firm-dependence: the condition still does not involve the firm's type, so $\mathbf{e}_0$, $\mathbf{e}_H$, $\mathbf{e}_L$ and their expansions remain common across firms. Bunching choices instead become firm-specific. Since $\hat{Y}'(s)$ potentially varies with the firm, the least-cost direction of adjustment $\ell(s)$ is firm-specific, and with it $A(s)$ and $b(s)$. The curvature $\hat{Y}''(s)$ also enters $b(s)$ through the frontier constraint at order g^2. T continues to denote the third-derivative tensor (of the entire objective instead of the cost only, when E is nonlinear); C_2 remains common across firms, being evaluated at the common flat-rule choice $\mathbf{e}_0$; $C(s) = -\frac{1}{6}T[\ell(s), \ell(s), \ell(s)]$ is the same as before but evaluated along the firm-specific direction $\ell(s)$, and so does $D(s) = \tau\gamma\, E_e(\mathbf{e}_0)'\ell(s)$; B involves $\mathbf{e}_0$ only and remains firm-invariant. Firm-wise, the analysis of Lemma 10 and of the subsequent indifference condition carries through unchanged: at each landing point s, the expansions hold with the firm's own coefficients, and $\bar{g}(z, s, \mu) = K(s)z + L(s)z^2 + O(z^3)$. Care is needed only when aggregating into a_2 and a_3: $A(s)$, $C(s)$ and $D(s)$—hence $K(s)$ and $L(s)$—are firm-specific and no longer come out of the integral over $\mathcal{S}$, so conditions ensuring their uniform regularity must be added: $\hat{Y}'$ bounded and bounded away from zero, $\mathcal{H}_0$ uniformly positive definite, and third derivatives uniformly bounded in s.

S.2.0.3 Special economies: detailed discussion

The framework accommodates, as special cases, several economies that are directly relevant to our application. In each, the analysis of the extended model goes through unchanged. Only the coefficients of the indifference condition move: the marginal cost of reaching the threshold A, the marginal benefit of bunching B, the cost-curvature term C, and the tax-base-erosion term D. The test goes through with the appropriately adjusted G. The summary table in the main section collects the expressions for A, B, C, and D.

An untargeted manipulation margin In our application the tax base is profits. Firms can misreport both sides of the income statement: underreport revenues (e_y) and overreport deductible costs (e_x), so that, with full deductibility of costs, the evaded base becomes $e_y + e_x$ (i.e., $\eta = (1, 1)$). The additional margin does not affect the disclosed threshold: $\hat{Y}(d_x) = \hat{y}_0$, i.e. $\hat{y}_1 = 0$, so cost misreporting does not affect eligibility and the rule leaves that margin untargeted.

With a separable manipulation cost, $C(e_y, e_x) = c_y(e_y) + c_x(e_x)$, the frontier normal is $w = (1,0)'$: the only way to reach the threshold is to declare more revenue, so $\ell = (1,0)'$ and $A = c_y''(e_{y0})$, as in the baseline with $c = c_y$. The additional margin generates two effects. First, it enlarges the base that the audit discount protects, $B = (1+\alpha)\tau\gamma\,\eta'\mathbf{e}_0$ rather than $(1+\alpha)\tau\gamma\, e_{y0}$: bunching is more attractive because eligibility shields the cost evasion too. Second, this margin responds to the discount: $t_\eta = (0, \eta_x/c_x'')'$, so a buncher, while standing at the threshold with its revenue report, evades more on costs as the audit jump Δ widens. This second channel makes bunching more attractive, but does not affect the singularity properties of the revenue function. Because the threshold is independent of the cost margin, the cost-curvature and tax-base-erosion terms are the same as in the baseline and only involve the revenue margin. In particular, $C = -\frac{1}{6}c_y'''(e_{y0})$ and $D = \tau\gamma\,\eta_y$: per unit of gap closed the buncher forgoes only the revenue component of the evaded base.

A cost-based threshold In practice the cutoff is often computed from reported characteristics rather than fixed. For example, in our empirical application, the disclosed threshold is a prediction of revenue based partly on reported costs. Reporting then moves the threshold itself: $\hat{Y}(d_x) = \hat{y}_0 + \hat{y}_1 d_x$ with $\hat{y}_1 > 0$, so a firm that overreports costs faces a higher revenue threshold. Cost misreporting now has two roles: it evades the tax base and it tightens eligibility. This model is exactly the extended economy studied above: closing the gap optimally combines the two reports, and $A = \left[(1,\hat{y}_1)C_2^{-1}(1,\hat{y}_1)'\right]^{-1}$, which with separable costs is below the one-margin curvature: the possibility of adjusting the threshold through the additional reporting margin makes it cheaper to reach. $B = (1+\alpha)\tau\gamma\,\eta'\mathbf{e}_0$ as with the untargeted margin. The tax-base-erosion term now draws on both evasions: with separable costs, $D = \tau\gamma\, A\big(\eta_y/c_y'' + \hat{y}_1\eta_x/c_x''\big)$, each margin contributing its weight on the evaded base (η_i), scaled by its own curvature and by its effectiveness on the threshold. A similar reasoning adjusts the cost-curvature term, which becomes $C = -\frac{A^3}{6}\left[c_y'''/(c_y'')^3 + \hat{y}_1^3 c_x'''/(c_x'')^3\right]$.

A fixed cost of being audited Suppose that, irrespectively of its outcome, an audit causes a disutility ω, which can arise from administrative and psychological costs. The value of facing a flat probability p becomes $V(p) = \max_e\{\tau(1-\gamma p)e - c(e) - p\omega\}$. Because ω does not depend on the amount evaded, interior choices in both regions are unchanged and $A = c''(e_0)$. The fixed cost enters only the indifference condition that defines the marginal buncher $\bar{g}$, as it raises the value of being in the low-audit side. The marginal benefit from bunching becomes $B = (1+\alpha)\{\tau\gamma\, e_0 + \omega\}$. The cost-curvature C and the tax-base-erosion D are unchanged. The reason is that the fixed cost opens no manipulation channel: it leaves the adjustment path— hence C —unchanged, and it is escaped in full by every buncher, near or far, so it does not erode with distance, leaving also D unchanged.

A declining counterfactual schedule Throughout, we have worked with piecewise flat rules: within each audit region the probability is a constant, and the declaration matters only through the side of the cutoff on which it falls. In principle, however, the audit rule can depend smoothly on the declaration, with the perceived audit risk falling gradually as reported revenues rise. We therefore consider an extension where the audit probability jump is opened within an otherwise smoothly declining schedule.

Formally, consider the schedule

$$p(d) = \begin{cases} p_H(d) = p_0(d) + \alpha\Delta & d < \hat{y} \\ p_L(d) = p_0(d) - \Delta & d \geq \hat{y} \end{cases},$$

where $p_0(d)$ is a declining and smooth function of d normalizing the audit probability at the lowest declaration as $p_0(0) = \mu$. To fix ideas, suppose p_0 declines linearly at rate β, $p_0(d) = \mu - \beta d$ (in general, read β as $-\partial_d p_0$ evaluated at the relevant declaration). Admissibility makes the slope small when μ is small: for p_0 to remain a nonnegative probability on the range of declarations $[0, \bar{d}]$, it must be that $\beta \leq \mu/\bar{d}$.[54] Note that a non-flat schedule, by itself, generates no bunching but only alters the first-order condition for the interior choice that becomes:

$$\tau\big(1 - \gamma p_0(d)\big) - \tau\gamma\beta e = c'(e). \tag{S.2.1}$$

A marginal unit of evasion, by lowering the declaration, raises the audit probability by β on the inframarginal units of evasion e, generating the wedge $\tau\gamma\beta e$. Differentiating (S.2.1), the effective curvature of the firm's objective is $A = c''(e_0) + 2\tau\gamma\beta$: the slope enters once by eroding the marginal benefit along the schedule, and once through the inframarginal penalty. Since one unit of discrepancy from the threshold is closed by one unit of evasion, this is also the curvature in g of the value loss from bunching, $\frac{1}{2}Ag^2 + O(g^3)$. The coefficient A keeps its meaning of shadow price of the distance from the threshold. The condition $c''(e_0) + 2\tau\gamma\beta \geq \underline{a} > 0$ — automatic for $\beta \geq 0$ — keeps it positive and bounded away from zero, so the interior choices remain smooth in Δ with bounded derivatives. The marginal benefit from bunching is unchanged, $B = (1 + \alpha)\tau\gamma\, e_0$, since the slope is present on both sides of the cutoff and already in the flat-rule benchmark. With a linear schedule, C and D are those of the baseline. If p_0 is not linear, read $\beta = -p_0'(d^0)$ where d^0 is the optimum at the flat-rule at μ. In this case, we must carry the local curvature of p_0 in the same way we dealt with $\hat{Y}''$ in the nonlinear case discussed in Supplementary Material Section S.2.0.2. So $p_0''(d^0)$ adds $\tau\gamma p_0''(d^0)e_0$ to A and the corresponding third-derivative term to C. Conversely, B and D that depend on the audit jump are unchanged. The indifference condition then has the same form as in the baseline and the marginal buncher's shortfall grows at rate $\sqrt{\Delta}$. The concavity result extends in this parametrization: since the admissible slope vanishes with μ, $\beta \leq \mu/\bar{d}$, every coefficient of the indifference condition is continuous in μ and converges to its baseline value as $\mu \to 0$. In particular, the recomputed $G(0)$ coincides with the baseline one. Intuitively, the flexibility of the non-flat rule becomes immaterial at low levels of baseline audit probability μ (as β shrinks with μ), thus making the inframarginal effect irrelevant. Under the same sign condition $G(0) > 0$, the proof of Proposition 11 applies verbatim. The linearity of the schedule is not essential: for any smooth family scaling with the baseline level, $p_0(d; \mu) = \mu\, \pi(d)$ with π decreasing and $\pi(0) = 1$, every derivative of the schedule is $O(\mu)$ (including the curvature that would correct C) so all coefficients still converge to their baseline values and the same argument applies.

S.2.0.4 Summary of the extensions

The following table summarizes the values of the coefficients A, B, C, D across all discussed extensions. The marginal benefit of bunching B differs from its baseline value when the economy changes what bunching protects. The cost-curvature C differs when the economy changes how the threshold is reached, specifically when the manipulation margins can affect the eligibility threshold. A and D instead differ even if additional margins do not directly affect the threshold, as they represent the value of a problem (closing the discrepancy) that has additional margins of adjustment.

[54] Strictly, once the gap opens the binding requirement is on the low-probability schedule, $p_L(d) = p_0(d) - \Delta \geq 0$, i.e. $\beta\bar{d} + \Delta \leq \mu$: at $\beta = \mu/\bar{d}$ no gap could be opened without negative audit probabilities near the top of the declaration range. We therefore restrict feasibility to $\beta\bar{d} + \Delta \leq \mu$.

Table S1: Summary of the drivers of concavity

Economy	**Curvature** A	**Jump value** B	**Cost curvature** C	**Tax Base erosion** D
Baseline	$c''(e_0)$	$(1+\alpha)\tau\gamma\, e_0$	$-\frac{1}{6}c'''(e_0)$	$\tau\gamma$
Untargeted margin	$c''_y(e_{y0})$	$(1+\alpha)\tau\gamma\, \eta'\mathbf{e}_0$	$-\frac{1}{6}c'''_y(e_{y0})$	$\tau\gamma\, \eta_y$
Cost-based threshold	$\left[(1,\hat{y}_1)C_2^{-1}(1,\hat{y}_1)'\right]^{-1}$	$(1+\alpha)\tau\gamma\, \eta'\mathbf{e}_0$	$-\frac{1}{6}T[\ell,\ell,\ell]$	$\tau\gamma\, \eta'\ell$
Fixed audit cost	$c''(e_0)$	$(1+\alpha)\{\tau\gamma\, e_0+\omega\}$	$-\frac{1}{6}c'''(e_0)$	$\tau\gamma$
Declining schedule	$c''(e_0)+2\tau\gamma\beta$	$(1+\alpha)\tau\gamma\, e_0$	$-\frac{1}{6}c'''(e_0)$	$\tau\gamma$

S.3 Additional Theoretical Results

Separability of Production and Declaration Decisions

We assume that y is exogenous, which is without loss of generality given that the declaration decision is separable from production. A model with endogenous production is equivalent to ours once y is interpreted as the optimal profit in a setting where firms differ in productivity. Formally, let $\pi(x,z)$ denote the profit of a firm choosing input x, with productivity z drawn from a distribution f_z. The firm solves:

$$V(z)=\max_{x,d}\pi(z,x)-\tau d-\tau\gamma\cdot p(d)\cdot(\pi(z,x)-d)-c(\pi(z,x)-d),$$

which delivers the following FOCs

$$x:\left[1-\tau\gamma p-c'(\pi(z,x)-d)\right]\pi'(z,x)=0$$

$$d:\tau(1-\gamma p)-c'(\pi(z,x)-d)=0.$$

Notice that the first factor in the FOC for x cannot be zero due to the FOC for d. It follows that the firm optimally chooses x such that $\pi'_x(z,x)=0$. Let $x^*(z)$ be the solution to this FOC. The variable y used in our analysis corresponds to the resulting profit, $y=\pi(z,x^*(z))$, and its distribution is given by $f(y)=f_z(z:\pi(z,x^*(z))=y)$.

A general test for the desirability of disclosed rules

We consider a test involving a local perturbation of the policy, which yields the following result.

Theorem 13. *If*

$$e^I(\mu)>\frac{1-F\left(\hat{y}+e^I(\mu)\right)}{f\left(\hat{y}+e^I(\mu)\right)}\tag{S.3.1}$$

then a threshold-based audit rule exists that outperforms a flat audit rule by generating higher revenues with a smaller budget.

Proof. Putting the marginal-buncher expansion and the limiting bunching component from the proof of Theorem 4 together, we obtain

$$\begin{aligned}\lim_{\Delta\to 0}R'(\Delta)&=\frac{\tau\gamma(1+\alpha)e^I(\mu)}{c''(e^I(\mu))}f\left(y^{HB}\right)-\left(1-F\left(\hat{y}+e^I(\mu)\right)\right)\frac{\tau\gamma}{c''(e^I(\mu))}\\&\quad+\left[F\left(\hat{y}+e^I(\mu)\right)-F\left(e^I(\mu)\right)\right]\frac{\tau\gamma\alpha}{c''(e^I(\mu))}\\&=\frac{\tau\gamma}{c''(e^I(\mu))}\left[(1+\alpha)\left[e^I(\mu)f\left(\hat{y}+e^I(\mu)\right)-\left(1-F\left(\hat{y}+e^I(\mu)\right)\right)\right]+\alpha\left[1-F\left(e^I(\mu)\right)\right]\right]\end{aligned}$$

which, in the case $\alpha = 0$ simplifies to $e^I(\mu) f\left(\hat{y} + e^I(\mu)\right) - \left(1 - F\left(\hat{y} + e^I(\mu)\right)\right)$. Therefore, if

$$e^I(\mu) > \frac{1 - F\left(\hat{y} + e^I(\mu)\right)}{f\left(\hat{y} + e^I(\mu)\right)}$$

then marginally decreasing the audit probability for declarations above $\hat{y}$ improves revenues. Since the number of audits run also decreases (by quantity $1 - F\left(\hat{y} + e^I(\mu)\right)$), we have our desideratum. □

This Theorem tests whether the marginal revenues from equation (5) are positive when $\Delta = 0$ in an environment where $\alpha = 0$. This scenario corresponds to the Authority marginally reducing the audit probability above the threshold $\hat{y}$, while keeping the probability below the threshold constant at μ. This policy clearly saves budget, as no firm faces a higher number of audits. To assess its impact on revenues, note that $\alpha = 0$ eliminates the second term in the decomposition of (5) (since $\frac{\mathrm{d}e_H}{\mathrm{d}\Delta} = 0$). As a result, the effect on revenues depends on the balance between losses in the $\mathcal{L}$ region and gains from the bunching margin. Near $\Delta = 0$, $\tilde{e}$ decreases with infinite pace from its original level $e^I(\mu)$. Consequently, the bunching region grows quickly for a small audit premium, which offsets the fact that $\tilde{e} \to e^I(\mu)$, leading to a finite, positive limit on marginal revenues from the extensive margin.[55] Condition (S.3.1) ensures that this finite limit more than compensates for the revenue loss from the marginal increase in evasion above the threshold.

The interpretation of condition (S.3.1) is straightforward. The hazard rate reflects the ratio between losses and gains as $\Delta \to 0$. The term $1 - F\left(\hat{y} + e^I(\mu)\right)$ represents the proportion of firms that reduce their declarations when Δ increases, while $f\left(\hat{y} + e^I(\mu)\right)$ measures the mass of marginal bunchers who increase their declarations on the extensive margin. The level of evasion, on the other hand, indicates how much bunching can enhance compliance: when evasion is low, there is less room for firms to adjust their reported revenues. Since $e^I(\mu)$ decreases with μ, this condition is satisfied when μ is sufficiently low for a fixed $\hat{y}$. This result implies that threshold-based audit discounts are particularly effective when the number of audits assigned to a firm class is small.

Although conceptually insightful, Theorem 13 is of limited practical interest. Condition (S.3.1) is challenging to quantify empirically, as it requires knowledge of the true income distribution and $e^I(\mu)$, which is typically unobservable. Moreover, the assumption that $\alpha = 0$ may be restrictive, as any reduction in p_L will likely be accompanied by an increase in audits in the $\mathcal{H}$ region.

The test in (S.3.1) is conducted with $\hat{y}$ held constant, but if the authority could also choose $\hat{y}$, an improvement would certainly be achievable. Condition (S.3.1) only requires that evasion under the flat rule exceeds the hazard rate at the "degenerate bunching point" $\hat{y} + e^I(\mu)$. By moving this point to the upper end of the support, where the hazard rate approaches zero, we ensure that this condition is satisfied. Indeed, the hazard rate approaches zero as the cdf converges to 1 (*i.e.*, as y approaches $\bar{y}$.), while the pdf remains bounded away from zero, as per our initial assumption. This reasoning leads to the following corollary.

Corollary 14. *If the authority can set $\hat{y}$ arbitrarily large, then they can always improve over any flat rule using less budget.*

Proof. The Corollary follows from the fact that $h(y) = \frac{1-F(y)}{f(y)}$ decreases to zero as y approaches $\bar{y}$. This implies that, if the authority is given a flat rule μ, they can always set a threshold such that condition (S.3.1) is satisfied and a marginal increase in the probability of audit above the threshold raises revenues. □

[55] The limiting bunching term in the proof of Theorem 4 provides a formal proof.

S.4 Details on Tax Gaps and Profit Shifting in Figure 2

This section documents the methodology used to construct Figure 2 quantifying tax revenue losses due to evasion and avoidance in four countries: the United States, Italy, United Kingdom, and Australia. The estimates refer to the year 2019 (the most recent available) and distinguish three sources of revenue loss: (i) tax evasion by small businesses, (ii) tax evasion by large businesses, and (iii) profit shifting by multinational firms. All values are expressed as a share of total central government tax revenue, allowing for a consistent cross-country comparison despite differences in tax systems and reporting methods.

Small and Large Business Income Tax Evasion: The standard metric used to measure tax evasion is the *tax gap*, defined as the difference between the total theoretical tax liability (under full compliance) and the actual amount collected. We focus on the gross tax gap, which is the gap before amendments, enforced collections, and late payments. We retrieve the gross tax gap and, where necessary, isolate the underreporting component, excluding the underpayments (*i.e.*, reported taxes that are not collected).

Because no harmonized international estimates of tax gaps exist, we rely on official national sources, which vary across countries in tax structure, estimation method, and reporting conventions. For each country and category (*e.g.*, Italian small businesses), we compute a gross income tax gap measure and scale it by total central government tax revenue in 2019 (OECD (2025)), to allow for cross-country and cross-statistic comparability.

Table S2: Definitions and sources: small and large business gross income tax gaps

Country	Source	Notes	Small Business Tax Gap	Large Business Tax Gap
United States	IRS (2022)	No yearly estimates available, only average projections for 2017–2019.	CIT gap for "small" firms (assets $< \$10M$) PIT gap from business income.	CIT gap for "large" firms (assets $> \$10M$).
Italy	MEF (2022)	No firm-size disaggregation for CIT (IRES) is available.	– PIT (IRPEF) gap for self-employed and individual business income earners; – CIT (IRES) gap not included → lower bound.	CIT (IRES) gap for all firms (small and large combined) → upper bound.
United Kingdom	HMRC (2025) Fiscal year 19-20	We disregard HMRC's aggregate customer segmentation and isolate only the relevant components.	– CIT gap for "small" firms (turnover <£10M; < 20 empl.) – PIT (Self-Assessment)[a] for business taxpayers and partnerships.	CIT gap for "medium" and "large" firms (turnover >£10M; > 20 empl.)
Australia	ATO (2024) Fiscal year 19-20	Estimates are segmented by customer group, not by legal form.	ATO "Small business income" tax gap (<A$10M turnover; includes sole traders, small companies, trusts, and partnerships).[b]	ATO "Medium business income" + "Large corporate groups" tax gap (>A$10M turnover).

Notes: this Table presents an overview of the sources, definitions and methodological choices behind the estimates of Figure 2.

[a] Under the Self-Assessment (SA) system, HMRC reports a combined tax gap covering income tax, national insurance contributions (NICs), and capital gains. To isolate the income tax component, we scale the reported estimate by the share of SA income tax receipts in total SA receipts (i.e., income tax plus NICs). While separate data for SA capital gains are unavailable, their contribution is likely negligible. In the fiscal year 2019–2020, this share is approximately 90% (HMRC, 2025).

[b] Reported on Schedules C (sole proprietorship), E (rents, S-corps, partnerships), and F (farming) in IRS Form 1040.

To harmonize definitions, we adopt a consistent breakdown between "small" and "large" businesses. The small business income tax gap includes both the corporate income tax (CIT) gap for small firms and the personal income tax (PIT) gap on business income earned by sole proprietors, partnerships, and self-employed individuals. The large business tax gap includes only the CIT gap attributable to large firms. Importantly, thresholds for "small" and "large" businesses are not uniform: they depend on country-specific classifications and may differ substantially. When no disaggregation between small and large is available, we make conservative choices and attribute the full corporate tax evasion to large businesses. Conceptual and methodological differences across countries are detailed in Table S2, which reports, for each country, the relevant data sources, the definitions adopted for "small" and "large" business taxpayers, and any additional notes.

Profit shifting: The data source for profit shifting is Wier and Zucman (2022), which provides comparable estimates across countries. They define profit shifting as "*a tax-motivated and artificial transfer of paper profits within a multinational firm from high-tax countries to low-tax locales*". Based on this definition, they measure profit shifting to tax havens as "*the amount of multinational profits booked by companies in these havens above and beyond what can be explained by real economic activity (such as capital, labour, country characteristics, industry composition, and research and development [R&D] spending)*" (p. 3). Their estimates of revenue losses due to profit shifting are reported as a share of corporate income tax (CIT) revenue.[56] To ensure consistency with the rest of our analysis, we convert these losses into a share of total central government tax revenue by multiplying the original estimate by the ratio of CIT revenue to total central government revenue in each country in 2019 (OECD (2025)).

[56] Appendix Table B in Wier and Zucman (2022).

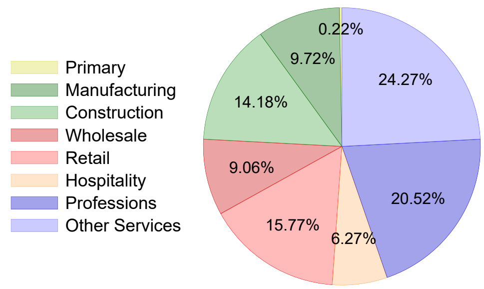
S.5 Statistics on SeS Data

Figure S1: SeS dataset overview

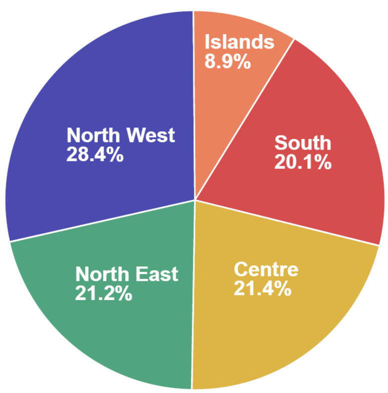
Panel (a): Dataset Structure

Tax Year	Observations
2007	3,753,997
2008	3,475,482
2009	3,470,191
2010	3,482,862
2011	2,472,183
2012	2,318,416
2013	2,142,884
2014	2,016,286
2015	1,849,767
2016	1,700,551

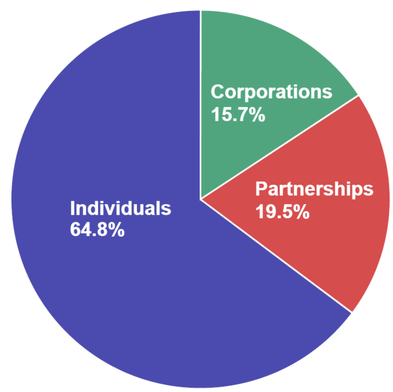
Panel (b): 2007-2010 macro-sectors



Panel (c): 2007-2010 geographic composition



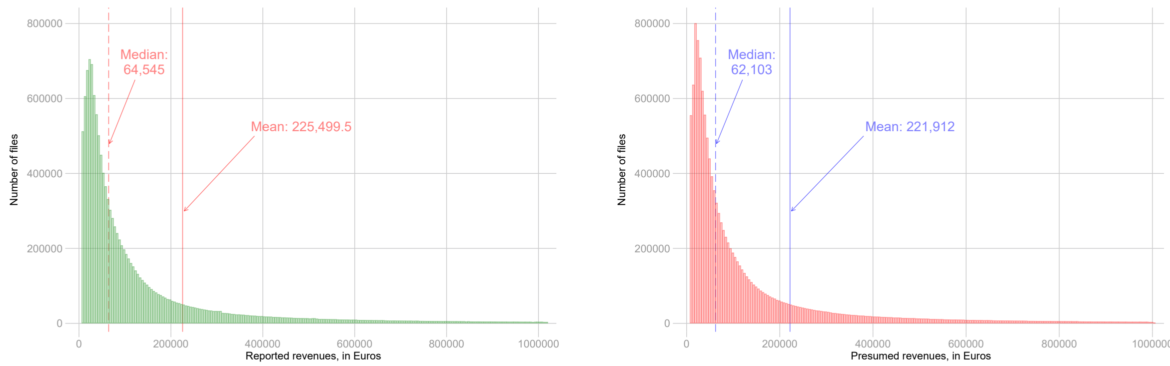
Panel (d): 2007-2010 legal status



Notes: The figure summarizes our SeS database. Italian businesses and self-employed individuals file SeS if annual revenues do not exceed €5.2 million. Panel (a) shows the number of files for each tax year from 2007 to 2016. The first four years (blue) cover the universe of SeS filings, while later years (green) include only taxpayers who filed continuously between 2008 and 2010, resulting in a declining sample size. Panels (b)–(d) describe the composition of the 2007–2010 universe: Panel (b) reports the distribution across eight macro-sectors defined by the authors; Panel (c) across the five Italian NUTS-1 macroregions; and Panel (d) across legal forms—individuals, partnerships, and corporations. Individuals and partnerships are subject to personal income tax, while corporations pay corporate income tax.

Figure S2: Distribution of reported revenues and presumed revenues, 2007-2010

Panel (a): Reported revenues   Panel (b): Presumed revenues



Notes: The figure shows the distribution of revenues reported in SeS files (Panel (a)) and revenues presumed by *Gerico* (Panel (b)). The sample includes the universe of SeS filings for 2007–2010, trimmed at the 5th and 95th percentiles of each distribution; in Panel (a) this excludes about 2% of files reporting zero revenues. Reported revenues include spontaneous adjustments to the presumed SeS level.

S.6 Additional Data Sources

S.6.1 Local evasion proxies

We assemble a dataset of local evasion proxies at the regional, provincial, and municipal levels. Because evasion is difficult to measure directly, we combine indicators from the literature and administrative sources with citizen-reported incidents from the platform *evasori.info*. Below we summarize the variables sources and their original level of aggregation.

Irregular employment share: Average irregular employment share in 1999–2000 for 103 provinces (Table 3 in Censis (2003)). Estimates derive from ISTAT regressions relating regional irregular employment to contextual factors and applying the coefficients at the provincial level.

TV tax evasion rate: Ratio of TV subscriptions (2014) to the number of households in the 2011 Census. Municipal estimates come from the Italian public TV service *RAI* subscription records (*twig.carto.com*); provincial and LLM values are household-weighted averages across municipalities.

Undeclared IRAP base ratio: Ratio of undeclared to declared IRAP tax bases, 1998–2002 (Table A1 in Pisani and Polito (2006)). Estimates compare ISTAT value added at factor prices with reported IRAP bases. We also compute a regional IRAP base gap using Table 31 in the same source as the ratio between the undeclared IRAP base and the sum of the declared and undeclared IRAP base. The declared base stems from dividing the undeclared base by the reported intensity of underreporting.

Ghost-building intensity: Share of land registry parcels with unregistered buildings. Municipal estimates are based on a 2007 aerial survey conducted by the *Agenzia del Territorio* (Casaburi and Troiano (2016)). Provincial and LLM values are parcel-weighted averages.

Tax gap: municipal real estate tax (IMU): Ratio of the tax gap to the potential tax base for the 2012 IMU (*Imposta Municipale Unica*) property tax. Provincial estimates (108 provinces) are provided by the Ministry of Economy and Finance based on municipal data. Provinces in Trentino–Alto Adige are excluded as they are subject to a different real estate tax.

Tax gap: VAT and IRAP: Combined VAT and IRAP gaps for 2007–2010 at the provincial level (Table 3 in Vallanti and Gianfreda (2020)). Gaps are defined as the difference between expected and reported revenues relative to expected revenues.

Concealed income share: Difference between audited and reported taxable income relative to audited income in 1987 based on the universe of audits of individual businesses and the self-employed (Table 2 in Galbiati and Zanella (2012)).

PIT evasion index: Ratio of taxed to taxable income (late 1980s) from Brosio, Cassone, and Ricciuti (2002) Table 1.

VAT evasion index: Ratio of taxed to taxable value added (late 1980s), also from Brosio, Cassone, and Ricciuti (2002) Table 1.

VAT base gap: Ratio of the VAT base gap to theoretical liability averaged over 2007–2010 (Table B.3 in D'Agosto, Marigliani, and Pisani (2014)).

Total tax gap ratio: Median ratio (2001–2011) of potential minus actual revenues to voluntary returns for major taxes administered by the Italian Revenue Agency (VAT, PIT, CIT, IRAP) from Carfora, Pansini, and Pisani (2016) Table 1.

Evasion reports from *evasori.info*: Through this platform, business customers can anonymously report the location, amount, and sector of any evasion instance they encounter in their daily life in Italy. Most commonly these are missing receipts for modest amounts, but they might reflect more sizable underreporting, as in the case of salaries paid out to irregular workers. We download all anonymous reports submitted between 2008 and 2011 using the platform's public API available at `evasori.info/api`. From these data we construct two province-level measures: (i) the number of reports per 1,000 inhabitants and (ii) the reported evasion volume per 1,000 inhabitants (population from the 2011 Census).

S.6.2 Other data sources

Personal income tax data: Data on the national PIT schedule and municipal PIT surcharges for 2007–2010 are from the Ministry of Economy and Finance (`finanze.gov.it`). The same source provides the number of PIT filers and total reported PIT base at the municipal level. Regional surcharges are obtained from the PIT return instruction tables.

For the correlational analysis, we construct an LLM-level PIT surcharge as the 2007–2010 average of yearly LLM rates, where yearly rates are computed as municipality-level surcharges weighted by the number of PIT filers.

Local value added and population data: Province-level value added per capita comes from ISTAT national accounts tables (*Principali aggregati territoriali di Contabilità Nazionale*) available at `dati.istat.it`. We average yearly values over 2007–2010. Provincial population data are from the 2011 Census (`dati-censimentopopolazione.istat.it`).

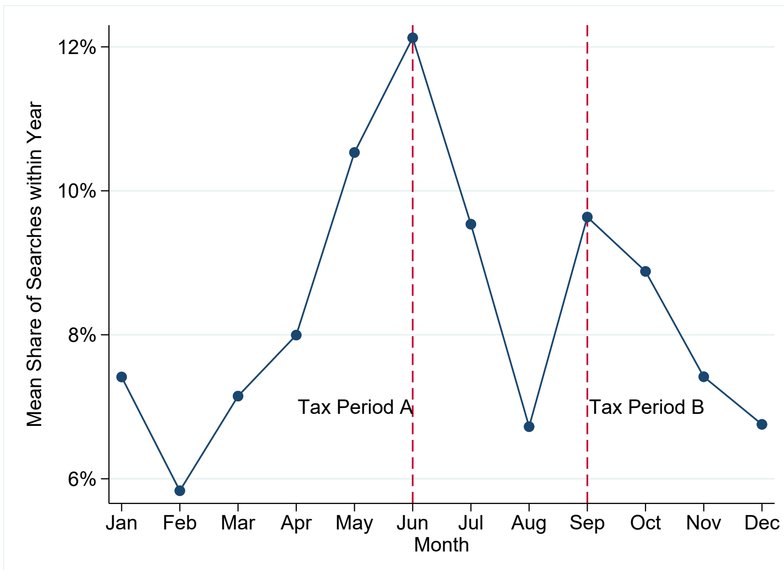
Input-output tables: Sector exposure to final consumers is computed using ISTAT input–output tables for 2010–2013 (`https://www.istat.it/it/archivio/195028`). We use the symmetric table for 63 sectors, matching 51 sectors with the SeS database. Exposure is defined as the ratio of final consumer spending to domestic uses (total uses minus exports), using values at 2010 current prices.

Tax litigation: To proxy the cost of interacting with the tax administration, we use the average duration of cases in provincial tax courts. Data come from annual reports on tax litigation published by the Ministry of Economy and Finance (`finanze.gov.it`). For each province, we compute the mean case duration over 2009–2012, defined by the Ministry as total adjudication days divided by the number of adjudicated cases. Provincial litigation durations exceed those at higher courts (regional courts and the Court of Cassation) by roughly one-third to one-half during this period.

S.7 Knowledge of the threshold: Gerico's software Google Searches

Audit rules are effectively disclosed if taxpayers understand how they operate. In the SeS context, firms can learn the presumed threshold at no cost through *Gerico*, a free software released by the Revenue Agency before each tax season to assist with SeS filings. To gauge awareness of this tool, we examine Google searches for "*gerico*" in Italy between 2004 and 2017. Figure S3 shows clear spikes in June and September, the two main filing periods.

Figure S3: Google searches for "gerico" spike in tax periods, 2004-2017



Notes: this Figure shows the month-by-month average intensity of Google searches for "*gerico*" over the 2004-2017 period in Italy. This time frame fully includes our SeS sample period, which stretches over the 2007-2016 tax years and the 2008-2017 filing years. Month-level data come from trends.google.com. Searches in off-peak months are partly explained by the fact that the actual filing deadlines are postponed in some years due to administrative constraints.

S.8 Bunching Correlation With Evasion and Tax Incentives

In Figure 5, we study the correlation between bunching in SeS files and local proxies of evasion across several tax bases. Figure S4 summarizes the geographic patterns: Panel (a) reports province-level bunching across the 110 Italian provinces, while Panel (b) shows local labor market (LLM) patterns. At both levels of aggregation, bunching is sizable and heterogeneous across locations. In Figure 5, we regress local bunching estimates for area i on one evasion proxy Evasion^j at a time using the following specification:

$$\text{Bunching}_i = \alpha + \beta \text{Evasion}_i^j + \gamma \log \text{VA pc}_i + \text{macroregion}_i + \varepsilon_i,$$

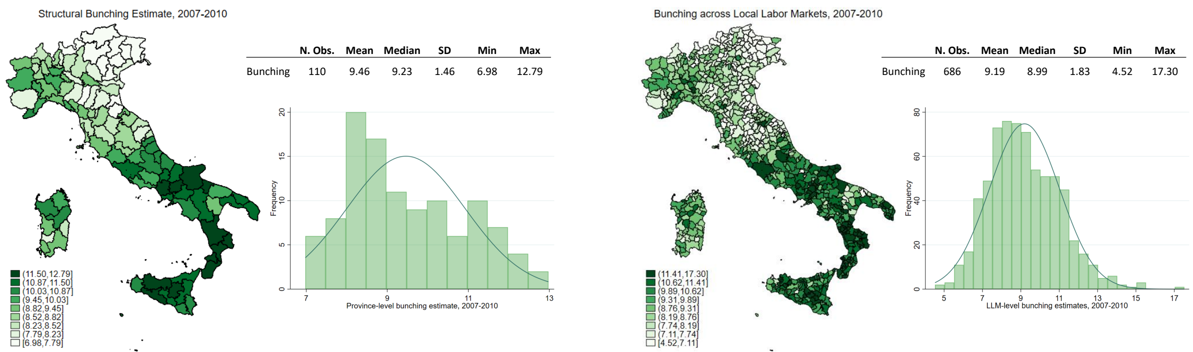
where we include fixed effects for the five NUTS-1 macroregions (North West, North East, Center, South, and Islands) and the logarithm of value added per inhabitant to control for provincial prosperity. The definition of i varies with the level of observation of each evasion proxy.

Bunching positively correlates with incentives to underreport. Figure S5 shows a positive and significant conditional correlation between LLM bunching and the weighted average of municipal PIT surcharge rates. Municipalities can impose a surcharge rate of less than 1% on top of the national personal income tax schedule.

Figure S4: Local heterogeneity in bunching estimates, 2007-2010

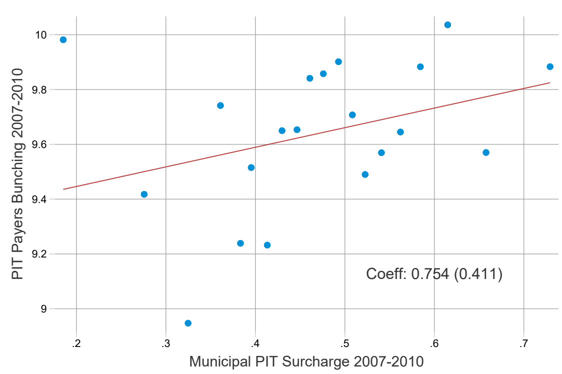
(a) Bunching across 110 provinces

(b) Bunching across 686 local labor markets (LLMs)



Notes: this Figure plots and summarizes our estimates of bunching at the SeS presumed revenues at the level of the Italian provinces (Panel (a)) and LLMs (Panel (b)). The sample includes each SeS file in the universe of single-sector businesses in the 2007-2010 tax years, except the top and bottom 5% in each province- or local-labor-market-level distribution that we trim to avoid irregularities in the estimation of the counterfactual. Bunching is computed at the local level following the procedure outlines in Section 4.2.

Figure S5: Bunching tracks evasion incentives: municipal taxes



Notes: The figure shows a binned scatterplot and linear fit of SeS bunching among PIT taxpayers against the weighted average municipal PIT surcharge at the LLM level (2007–2010). The reported slope (robust s.e.) comes from a regression of bunching on PIT surcharges controlling for regional fixed effects, log PIT base per taxpayer, and average provincial tax court litigation length (2009–2012). Municipal surcharges do not exceed national PIT rates by more than 0.8%, while regional variation is absorbed by fixed effects. The sample includes all SeS files for single-sector firms in 2007–2010, trimming the top and bottom 5% of each LLM distribution. Bunching is computed as detailed in Section 4.2.

S.9 Structural Model: Optimal declarations

This appendix formally derives the declaration regions discussed in Section G. We write the declaration problem in terms of two margins of misreport: revenue understatement $e_y := y - d_y \geq 0$ and cost misreport $e_x := d_x - x \in [-x, \infty)$. Let $\pi = y - x$ and define $A := y - \widehat{y}_0 - \hat{y}_1 x$, so that eligibility for the low-audit regime, $d_y \geq \widehat{y}(d_x) = \widehat{y}_0 + \hat{y}_1 d_x$, is equivalent to the linear constraint $e_y \leq A - \hat{y}_1 e_x$. Throughout we assume $B(p) := \tau(1 - \gamma p) > 0$ for $p \in \{p_L, p_H\}$, so that evasion has a positive net marginal benefit before the zero-profit tax kink binds.

Negative-profit firms are honest. If $\pi = y - x \leq 0$, then $\pi^+ = 0$, taxes and penalties are zero for all reports, and the firm minimizes manipulation costs by setting $e_y = e_x = 0$, i.e. $d_y = y$ and $d_x = x$. Henceforth suppose $\pi > 0$. Because costs are convex, even firms with very high evasion costs optimally engage in an infinitesimal amount of evasion, so negative-profit firms are the only fully compliant ones in the model.

Interior optima. Consider interior solutions within a fixed audit regime $p \in \{p_L, p_H\}$ where the zero-profit tax kink is not binding, i.e. $0 < e_y + e_x < \pi$, and focus on $e_x \geq 0$. The first-order conditions equate the marginal benefit of evasion $B(p)$ to marginal manipulation costs along each margin:

$$B(p) = (1 - b_2)c'(e_y, \kappa) + b_2 c'(e_y + e_x, \kappa), \qquad B(p) = (b_1 - b_2)c'(e_x, \kappa) + b_2 c'(e_y + e_x, \kappa).$$

Subtracting yields a constant optimal mix independent of $B(p)$,

$$(1 - b_2)c'(e_y, \kappa) = (b_1 - b_2)c'(e_x, \kappa).$$

Using $c'(m, \kappa) = \kappa^{1/\varepsilon} m^{1/\varepsilon}$, interior levels are

$$e_x^{\text{int}}(p) = \frac{1}{\kappa}\left(\frac{B(p)}{(b_1 - b_2) + b_2(1 + r)^{1/\varepsilon}}\right)^{\varepsilon}, \qquad e_y^{\text{int}}(p) = r\, e_x^{\text{int}}(p),$$

where $r := \left(\frac{b_1 - b_2}{1 - b_2}\right)^{\varepsilon}$. When $b_2 = 0$ this reduces to the separable cost benchmark.

Zero-profit optimum. If the interior candidate implies $e_y^{\text{int}}(p) + e_x^{\text{int}}(p) \geq \pi$, declared profits are non-positive and taxes are zero. Since further increases in $e_y + e_x$ cannot reduce taxes or penalties, the firm solves

$$\min_{e_y \geq 0,\ e_x \geq 0} \mathcal{C}(e_y, e_x; \kappa) \qquad \text{s.t.} \qquad e_y + e_x = \pi.$$

The constraint fixes $b_2 c(e_y + e_x, \kappa) = b_2 c(\pi, \kappa)$, so the cost-minimizing split depends only on the separable part of the cost function. The optimal split again satisfies $e_y / e_x = r$, giving

$$e_y^{\text{cap}} = \frac{r}{1 + r}\pi, \qquad e_x^{\text{cap}} = \frac{1}{1 + r}\pi.$$

We refer to these as the zero-profit policies; they may or may not satisfy eligibility for the audit risk discount.

Bunching. Eligibility introduces a kink in tax liability through the constraint $e_y \leq A - \hat{y}_1 e_x$. When the interior solution at p_L violates this constraint, the best eligible policy lies on the boundary

$$e_y = A - \hat{y}_1 e_x \qquad \Longleftrightarrow \qquad d_y = \widehat{y}(d_x).$$

Substituting into the objective yields a one-dimensional maximization over feasible e_x:

$$\max_{e_x \in [-x, +\infty]} \; \pi - \tau(\pi - (e_y + e_x)) - \tau\gamma\, p_L\, (e_y + e_x) - \mathcal{C}(e_y, e_x; \kappa),$$

$$\text{s.t.} \quad e_y = A - \hat{y}_1 e_x, \quad 0 \leq e_y \leq y.$$

When $\hat{y}_1 > 0$ there is generally no closed form for the bunching maximizer, but strict concavity ensures a unique solution.

Figure 7 shows that both reported revenues and costs rise after the introduction of the reward regime. The mechanism generating joint movements in declarations is driven by the value of eligibility. A larger audit gap $p_H - p_L$ increases the gain from declaring in the low-audit regime, making bunching more attractive for firms near the threshold. For such firms, inflating reported costs ($d_x \uparrow$, equivalently $e_x \uparrow$) reduces declared profits and taxes, but since higher d_x raises the revenue threshold $\widehat{y}(d_x)$, the firm must also raise d_y to remain eligible, reducing revenue understatement e_y. Thus, when the audit gap widens, firms near the threshold optimally shift toward a bunching policy with lower e_y and higher e_x, increasing both declarations.

This mechanism is most relevant for firms initially in the high-audit region but not far below the threshold, or for firms in the low-audit region with little slack. For these firms, the reward from reaching or remaining in the low-audit regime grows as $p_H - p_L$ rises. By contrast, firms far from the threshold are governed primarily by the interior trade-off between marginal tax savings and manipulation costs, and are therefore less likely to display the joint upward movement in declarations characteristic of a bunching response.