Optimal Taxation When Firms Can Underreport and Split

Faheem Gilani

October 25, 2026

Abstract

Firms can lower their taxes by underreporting their tax base and by splitting their business among separately taxed entities. Theories of optimal taxation with avoidance and evasion generally do not consider tax-motivated splitting. In this paper, a tax authority chooses a marginal tax schedule and enforcement levels, while firms with heterogeneous tax bases and costs choose a sequence of underreports and splits in order to minimize the sum of their tax payments and manipulation costs. Optimal rates follow a Ramsey rule where the mass of entities reporting above a threshold plays the role of quantity demanded. Under certain conditions, the optimal schedule in the presence of splitting has flat top rates and higher lower rates than the optimal schedule with only underreporting. The ratio of the mass of entities reporting above a threshold to the mass of entities whose base is above that threshold ranks small reforms to the current schedule without requiring manipulation costs. This ratio can be estimated through audit and ownership data.

1 Introduction

Firms can lower their taxes through real responses, legal avoidance or evasion (Slemrod, 1995; Slemrod and Yitzhaki, 2002; Saez et al., 2012). I treat all of these as costly manipulation of a firm's *potential tax base*, the base it would report without taxes or enforcement. I distinguish between two basic forms of manipulation: *underreporting* and *splitting*. Underreporting lowers the base that an entity reports, by concealment or by real cuts in activity. Splitting divides the base among separately taxed entities that the firm controls. Firms do both, sometimes at once (Section 1.1), but, with few exceptions, models of optimal taxation with avoidance or evasion, of tax administration and of size-based business taxes

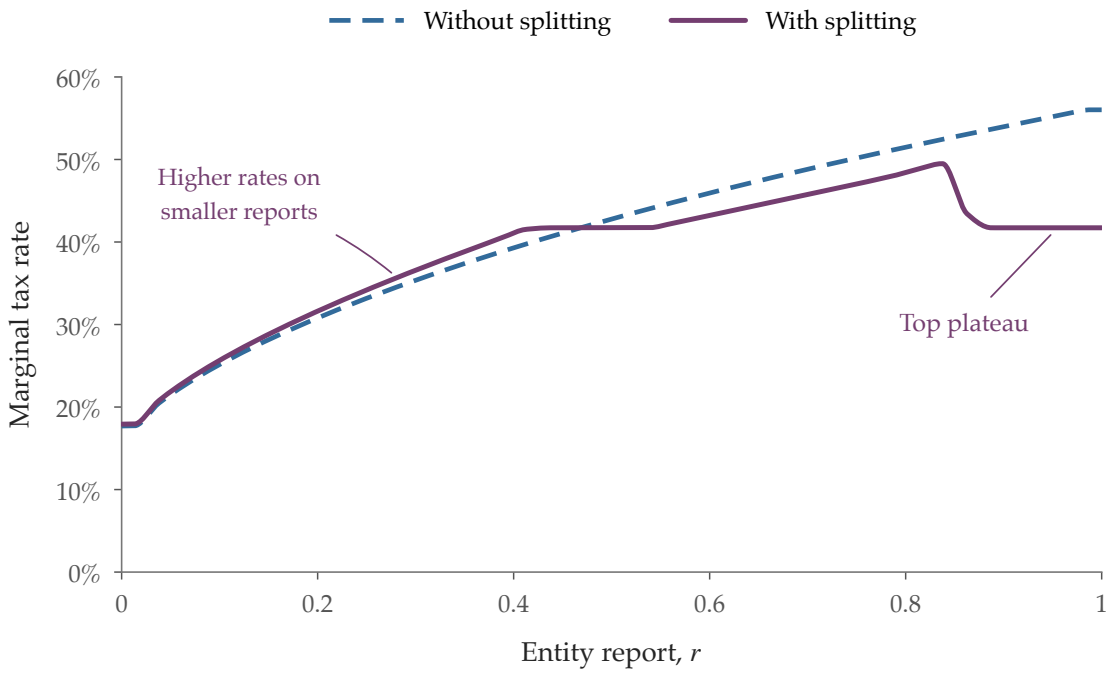


Figure 1: Splitting creates a top plateau and raises lower rates. Firms have potential bases $x = i/40$, $i = 1, \ldots, 40$, with population shares proportional to $1/x$. A firm with cost parameter c, uniform on $[0, 1]$, pays private manipulation cost $c[0.6(1 + 4x)(x - q) + 0.3(n - 1)]$ when it reports a total q through n entities. Reports are continuous, divisions are unrestricted, and any finite number of entities is feasible when splitting is available. The authority maximizes average firm surplus net of taxes and manipulation costs, subject to the same revenue requirement of approximately 0.06 in both economies. It chooses marginal rates between zero and 60%; there is no fee, enforcement or exit. Blue and plum show the optima without and with splitting. Smooth curves interpolate the computed step rates without overshooting them. Appendix N.4 gives the calibration and certifies optimality over the full schedule class.

treat each taxpayer as a single reporting unit (Section 1.2). I add splitting to this theory and ask how a tax authority should set marginal tax rates and enforcement in order to raise a given revenue while maximizing aggregate firm surplus. Accounting for splitting changes optimal rates even when few firms are observed splitting. Figure 1 illustrates both responses in a many-base version of the benchmark technology. When firms cannot split, optimal marginal rates rise throughout the base distribution. Allowing splitting raises rates on smaller reports and creates a plateau at the top. Few firms split at the new optimum, because the schedule adjusts to deter it. The observed incidence of splitting therefore understates its influence on tax design.

The authority chooses a common marginal tax schedule, a fee per entity, and enforcement against underreporting and splitting. Firms differ in their potential tax base and each chooses a costly manipulation strategy that minimizes its *burden*, i.e. the sum of its tax

payments and manipulation costs. A manipulation strategy is composed of underreporting and splitting steps and culminates in a number of entities controlled by the firm and the base reported by each entity. A private cost parameter scales each firm's manipulation costs and enforcement raises the cost of each step linearly. The paper assumes that the cost parameter follows a power-law distribution, which gives explicit policy and measurement formulas[1]. The authority maximizes firm welfare subject to a revenue requirement net of enforcement spending.

The first contribution is a global characterization of optimal taxes and enforcement for general manipulation technologies. Marginal rates follow a Ramsey rule in which report thresholds are the goods and the *entity base*, the number of entities reporting above each threshold, is the quantity demanded (Ramsey, 1927; Diamond, 1975). At an interior rate, the entity base's elasticity to a proportional cut in the whole tax schedule, including the fee, equals $1 - w_r/\mu$. Here w_r averages the welfare weights of firms with entities above the threshold, counting each firm once per entity and μ is the multiplier on the revenue constraint. When the highest-cost firms comply under all admissible schedules and welfare weights are not too dispersed, an accounting identity makes the authority's Lagrangian concave, so the first-order conditions are sufficient for global optimality. Access to splitting cannot improve attainable welfare. Splitting imposes a strict welfare loss when every optimum without it gives a positive share of firms a strict incentive to split.

For report-dependent underreporting costs and a fixed cost per split, I derive a recursive solution to the firm's problem. It gives exact tests for whether splitting changes firms' burdens. When marginal underreporting costs rise with the entity's current report, splitting first is optimal; when they fall with it, underreporting first is optimal.

The second contribution explains how the threat of splitting shapes the tax schedule. Consider an interior top range where large reports come from complying firms. The Ramsey rule fixes the share of firms that comply at each potential base. With equal welfare weights and a common cost exponent, this share is constant. When the marginal noncompliers split their full base at a common cost, their indifference then fixes the tax saving from splitting. An additional unit of base must consequently face the same marginal tax rate before and after the split. Top rates follow the rates on the split entities' smaller reports. A progressive optimum must have a plateau extending from those reports to the top. Splitting can also induce low rates to rise as top rates fall. A split firm pays each rate below its smallest report through several entities, making the entity base at those thresholds less responsive to taxes. The paper establishes conditions under which access to splitting raises the rates paid by small firms and increases their burdens. In the three-base

[1]Section 10 discusses the scope of this assumption.

running example, average burdens rise by 37.84 percent for small firms and fall by 8.56 percent for large firms. Aggregate welfare falls by about 2.95×10^{-3} at fixed enforcement. The welfare cost includes the concessions made to deter splitting, as well as the cost of the splitting that remains.

The third contribution connects the policy conditions to empirics. The *coverage ratio* compares the entity base above a threshold with the number of firms whose potential base exceeds it. When the exponent in the power law is constant, this ratio determines the revenue raised per dollar of burden by the marginal rate at that threshold. Under equal welfare weights, higher coverage ratios identify less costly sources among rate increases that raise revenue. The ratios therefore rank small reforms at the current schedule without requiring knowledge of firms' manipulation technologies. Entity counts, audits of potential bases, and ownership information supply the relevant data.

Proportional tax reforms provide a complementary measurement strategy. At fixed enforcement, changes in tax payments trace out firms' private manipulation costs. With a common exponent, one reform and an observed response, such as the change in the splitting share, identify their level. Reported bases alone can give the wrong signal. Enforcement against splitting can raise receipts while lowering total reports. Lower-bracket relief can raise total reports yet lose more revenue than predicted at unchanged behavior. Firms create entities to claim the relief repeatedly. Measuring tax payments across a firm's entities captures both responses.

1.1 Empirical evidence

In 2018, Brazil's administrative council of tax appeals found that a clothing retailer had split its revenue among companies nominally owned by relatives and employees, so that each company qualified for the simplified regime for small businesses (CARF, 2018). The group had also declared only about 40 percent of the sales recorded in its accounts. In the Netherlands, the number of companies per owner rose significantly and persistently once the tax advantage of separate assessment for related companies seemed permanent and minimum capital requirements were abolished (Massenz, 2025). Japanese corporations cluster below the sales threshold for a simplified value-added tax, consistent with larger firms incorporating parts of their businesses separately (Onji, 2009). Before Russia raised the revenue threshold of its simplified regime in 2017, eligible firms were overrepresented far below the threshold compared with later years, and Kuchumova et al. (2026) attribute this mass to splitting. Spanish firms report revenue just below a threshold for closer monitoring, while their reported inputs suggest undeclared sales (Almunia and Lopez-

Rodriguez, 2018).

Manipulation is costly, and it responds to enforcement. French firms just below a regulatory threshold of 50 employees appear to forgo profitable growth (Garicano et al., 2016). Creating an entity has setup, accounting and filing costs (Tazhitdinova, 2020; Massenz, 2025). Chilean firms paid more value-added tax when told that they faced more monitoring. The response was largest for sales to final consumers, which leave no paper trail (Pomeranz, 2015). When Ecuadorian firms were told that third-party reports showed more revenue than they had declared, they raised their reported revenue but offset most of the increase with higher reported costs, which are harder to verify (Carrillo et al., 2017).

1.2 Related literature

I add splitting to the theory of optimal taxation with costly manipulation of the tax base. In that theory, the elasticity of the reported base guides tax rates, and enforcement can change that elasticity (Slemrod and Kopczuk, 2002). The costs of avoidance and evasion shape optimal nonlinear schedules (Grochulski, 2007; Casamatta, 2021, 2023), the use of enforcement (Kaplow, 1990; Chander and Wilde, 1998; Keen and Slemrod, 2017) and the case for redistribution (Kopczuk, 2001). The one-entity model nests Kopczuk's costly underreporting and quasilinear labor supply. More general costs can reproduce labor-supply responses in the style of Mirrlees (1971) (Section 2.3). Because costs can depend on the order of steps, splitting can make underreporting cheaper. Whether enforcement against one response raises or lowers the other then follows from monotone comparative statics (Topkis, 1978; Milgrom and Shannon, 1994).

Models of the optimal design of size-based business taxes weigh revenue against administrative and compliance costs (Keen and Mintz, 2004; Dharmapala et al., 2011; Kanbur and Keen, 2014; Wei and Wen, 2019, 2023; Sharma et al., 2025), and Kanbur and Keen (2014) set splitting aside explicitly. Liu et al. (2021) explain bunching and voluntary registration at thresholds. Dabla-Norris et al. (2018) let a fixed share of firms split in a quantitative model of size-dependent regulation. Kuchumova et al. (2026) take the revenue threshold of a simplified regime as given and model firms that both underreport and split in response to it. They argue that models without splitting may recommend thresholds and tax rates that are not optimal.

Splitting is the tax counterpart of multiple purchases in nonlinear pricing, where a subadditive tariff removes the gain from buying several small bundles (Wilson, 1993; Alger, 1999). A subadditive tax schedule likewise removes the direct tax saving from dividing a base, but it need not deter splitting that makes later underreporting cheaper. In

axiomatic taxation, progressivity implies that taxpayers cannot gain by merging (Ju and Moreno-Ternero, 2011). Under rising marginal rates, dividing a report lets several entities each use the lower tax brackets. At an optimum where the firms closest to complying split, the schedule departs from subadditivity at the top by exactly what splitting costs them (Proposition 4).

Unlike agents in models of costly verification and falsification, firms here can change the number of entities as well as their reports. Townsend (1979) studies optimal contracts when verifying a privately known state consumes resources. Lacker and Weinberg (1989) allow an agent to falsify the observable state at a cost. Perez-Richet and Skreta (2022, 2026) study tests and non-market allocation rules when agents can falsify their measured characteristics. In their setting, preventing falsification can benefit every agent while reducing the designer's allocative payoff.[2] Here, removing the option to split never lowers the welfare attainable at a given net revenue requirement (Theorem 2).

The rule for optimal rates adapts the demand-profile approach to nonlinear pricing (Goldman et al., 1984; Wilson, 1993) and the many-person Ramsey rule of Diamond (1975). As in Saez (2001), optimal rates can be expressed through behavioral responses, but the relevant response is that of the entity base.

By the envelope logic of Feldstein (1999), changes in tax payments measure changes in private manipulation costs (Section 9). Malik (1990) shows that resources spent avoiding apprehension can make maximal fines undesirable. Here stronger enforcement instead lowers the resources that firms spend on manipulation. Saez et al. (2012) emphasize that shifting between tax bases requires accounting for revenue across all affected bases. In this model, the affected bases are a firm's entities. Because they face the same schedule, a firm's taxes and manipulation costs can change while its total reported base does not. The elasticity of taxable income therefore need not measure the efficiency cost of taxation. Chetty (2009) obtains a similar distinction when part of the private cost of avoidance is a transfer.

2 The model

2.1 Policies and firm choices

There is a unit mass of firms, each with characteristics $s \in \mathcal{S}$ and a private cost parameter $c \in [0,1]$. The marginal distribution of characteristics is H. Conditional on s, the cost

[2] Proposition 4 of Perez-Richet and Skreta (2026) establishes this comparison under their increasing-differences condition. It compares rules that prevent falsification with unrestricted allocation rules.

parameter has distribution

$$F_s(c) = c^{\gamma_s}. \tag{1}$$

A firm with characteristics s has *potential tax base* $x(s)$, where $x : \mathcal{S} \to [0, X]$. This is the firm's unmanipulated tax base. Characteristics also determine the manipulation technology, represented by the correspondence

$$\mathcal{A} : \mathcal{S} \rightrightarrows \bigsqcup_{n=0}^{\infty} \left(\mathbb{R}_+^n \times \mathbb{R}_+^3 \right).$$

It assigns to each $s \in \mathcal{S}$ a set $\mathcal{A}(s)$ of feasible strategies in this space. A generic element $\alpha(s) \in \mathcal{A}(s)$ has the form[3]

$$\alpha(s) = \left(r_1^\alpha, \ldots, r_n^\alpha, k_0^\alpha, k_R^\alpha, k_S^\alpha\right),$$

where $r^\alpha := (r_1^\alpha, \ldots, r_n^\alpha)$ is a vector of nonnegative entity reports, and $k^\alpha := (k_0^\alpha, k_R^\alpha, k_S^\alpha)$ is a vector of nonnegative effort coefficients. Expectations are over the joint distribution of (s, c).

The authority commits to a policy $\pi = (T, e)$, where T is a common tax schedule and $e = (e_R, e_S)$ are the enforcement levels against underreporting and splitting. Let $\mathcal{T}$ be the class of absolutely continuous schedules

$$T(r) = \tau + \int_0^r t(v)\, \mathrm{d}v, \qquad 0 \le \tau \le F, \qquad 0 \le t(v) \le L \quad \text{almost everywhere.} \tag{2}$$

Here τ is a fee paid by each entity and t is the marginal tax rate. A schedule is *progressive* if its marginal rate is nondecreasing. For a fixed partition $\mathcal{P}^N$ of $[0, X]$, let $\mathcal{T}_{\mathcal{P}^N}$ denote the subclass with a constant marginal rate on each bracket. The authority can choose any schedule $T \in \mathcal{T}$. For the finite-partition results, its choice is restricted to $\mathcal{T}_{\mathcal{P}^N}$. Enforcement is chosen from $E = [0, \bar{e}_R] \times [0, \bar{e}_S]$. Its public cost is $A(e)$, where $A : E \to \mathbb{R}_+$.

Writing $r_j = r_j^\alpha$ and $k_i(\alpha) = k_i^\alpha$ for $i \in \{0, R, S\}$, the tax payment and manipulation effort of strategy α are

$$T[\alpha] = \sum_{j=1}^{n} T(r_j), \qquad K(\alpha; e) = k_0(\alpha) + e_R k_R(\alpha) + e_S k_S(\alpha). \tag{3}$$

Private manipulation cost is $cK(\alpha; e)$, a real loss of surplus. Exit, when feasible, is repre-

[3]The disjoint union of sets Z_n is $\bigsqcup_{n \ge 0} Z_n = \{(n, z) : n \ge 0,\ z \in Z_n\}$. The index n distinguishes strategies with different numbers of entities. An entity reporting zero is distinct from an absent entity. For $n = 0$, the report vector is empty.

sented by $n = 0$; it entails no tax payment but can have a positive private cost. Distinct strategies can produce the same reports at different costs.

In a *sequential technology*, firms manipulate through finite sequences of underreporting and splitting operations. Consecutive reductions with no intervening operation have additive costs. The order of splitting and underreporting can nevertheless affect total effort. In the *benchmark technology*, order does not matter:

$$K(\alpha; e) = (p_s + e_R)\Big(x(s) - \sum_j r_j\Big) + (\sigma_s + e_S)(n-1), \qquad n \geq 1,$$

where $p_s > 0$ is the marginal effort of underreporting at zero enforcement and $\sigma_s > 0$ is the effort required for each additional entity.

Given policy π, each firm minimizes its burden:

$$P_s(c; \pi) = \min_{\alpha \in \mathcal{A}(s)} \{T[\alpha] + cK(\alpha; e)\}. \tag{4}$$

Write $R_s(c; \pi)$ for taxes paid at a minimizing strategy and $D_s(c; \pi) = P_s(c; \pi) - R_s(c; \pi)$ for private manipulation costs.

The authority evaluates firm surplus using functions U_s. It chooses policy to maximize welfare subject to a net revenue requirement $G > 0$:

$$\begin{gathered} W(\pi) = \mathbb{E}U_s\big(x(s) - P_s(c; \pi)\big), \qquad Q(\pi) = \mathbb{E}R_s(c; \pi) - A(e), \\ V(G) = \max_{\pi \in \mathcal{T} \times E,\ Q(\pi) \geq G} W(\pi). \end{gathered} \tag{5}$$

For fixed bracket results, the maximization is over $\mathcal{T}_{\mathcal{P}^N} \times E$. We suppress s and other arguments when they are understood.

2.2 Assumptions

Assumption 1 (Regularity)**.**

(i) *Characteristics and costs.* The space $\mathcal{S}$ is compact and metrizable, and H is a Borel probability measure on $\mathcal{S}$. The maps x and $s \mapsto \gamma_s$ are continuous, with $1 \leq \gamma_s \leq \bar{\gamma} < \infty$.

(ii) *Manipulation technology.* Each set $\mathcal{A}(s)$ is compact,[4] and the correspondence $s \mapsto \mathcal{A}(s)$

[4]The strategy space carries the disjoint-union topology: each component $\mathbb{R}^n_+ \times \mathbb{R}^3_+$ is open and has its usual Euclidean topology. A sequence of strategies converges precisely when its entity count eventually equals that of the limit and its reports and effort coefficients converge coordinatewise. Compactness and hemicontinuity are understood in this topology.

is upper and lower hemicontinuous. Every feasible strategy satisfies $\sum_j r_j^\alpha \leq x(s)$. Every firm can comply by reporting its potential base through one entity at zero effort.

(iii) *Bounds and enforcement costs.* The bounds $X, F, L, \bar{e}_R, \bar{e}_S$ are finite and nonnegative. The function A satisfies $A(0) = 0$ and admits a convex, continuously differentiable extension to an open neighborhood of E.

(iv) *Welfare.* The functions U_s are increasing and concave. For every s, the function U_s is continuously differentiable on $[x(s) - (F + LX), x(s)]$. The map $(s, y) \mapsto U_s(y)$ is jointly measurable on these intervals, and

$$\int_{\mathcal{S}} |U_s(x(s))| \, \mathrm{d}H(s) < \infty.$$

When a technology is specified by unrestricted finite manipulation plans, $\mathcal{A}(s)$ may be a compact representation of those plans, provided that it preserves the minimum burden for every firm and every admissible policy.

Assumption 2 (Economic restrictions)**.**

(i) *No manipulation at the top.* Firms with $c = 1$ can optimally comply under every admissible policy. Equivalently, for every s and $\alpha \in \mathcal{A}(s)$,

$$k_0(\alpha) \geq \sup_{T \in \mathcal{T}} \{T(x(s)) - T[\alpha]\}. \tag{6}$$

For fixed brackets, the supremum is taken over $\mathcal{T}_{\mathcal{P}^N}$.

(ii) *Welfare weights.* There are constants $0 < w_- \leq w_+ < \infty$ such that, for every s,

$$w_- \leq U_s'(y) \leq w_+ \qquad \text{for } y \in [x(s) - (F + LX), x(s)],$$

and

$$w_-(1 + \gamma_s) \geq w_+ \qquad \text{for } H\text{-almost every } s. \tag{7}$$

(iii) *Strict feasibility.* Some admissible policy satisfies $Q(\pi) > G$.

The first restriction gives $P_s(1; \pi) = T(x(s))$. It requires sufficiently costly manipulation and is not a normalization of c. The welfare restriction holds whenever the largest marginal welfare weight is at most twice the smallest. Strict feasibility is Slater's condition.

2.3 Special Cases of the Model

With one entity and no exit, a strategy is a reduction $h \in [0, x]$ and the reported base is $x - h$. The burden problem becomes

$$P_x(c; \pi) = \min_{0 \le h \le x} \{T(x - h) + cK_x(h; e_R)\}. \tag{8}$$

Costly underreporting. Fix labor supply in the monetary-cost specification of Kopczuk (2001, p. 56). With earnings x, no transfer, and avoidance cost $cK_x(h; e_R)$, consumption is

$$C = x - T(x - h) - cK_x(h; e_R).$$

Maximizing any utility increasing in consumption is equivalent to minimizing burden. Applying the same social evaluation to attained consumption gives $W = \mathbb{E}U_x(x - P_x)$, and receipts are $\mathbb{E}T(x - h)$. Thus the individual and planner problems coincide when their policy classes agree. Kopczuk's proportional-tax case is obtained with $T(r) = tr$ and $\tau = 0$. A positive transfer would require a negative fee.

Labor supply. In the labor-supply model of Mirrlees (1971), a reduction in the tax base can represent lower earnings. The following quasilinear specification fits our burden problem exactly. Set the fee to zero and make enforcement unavailable. Let a worker have wage $x > 0$, hours $q \in [0, 1]$, consumption $C = xq - T(xq)$, and preferences

$$u(C, q) = C + (1 - 0.6c)x(1 - q) - 0.1cx(1 - q)^2. \tag{9}$$

Leisure is desirable and has diminishing marginal utility. Full-time earnings x are an untaxed optimum. Writing $h = x(1 - q)$ gives

$$u = x - \{T(x - h) + cK_x(h)\}, \qquad K_x(h) = 0.6h + \frac{0.1}{x}h^2. \tag{10}$$

The worker therefore solves the burden problem exactly. Here cK_x is earnings forgone net of the utility from additional leisure, measured in consumption units. With social evaluation U_x of attained utility, welfare and receipts again equal $\mathbb{E}U_x(x - P_x)$ and $\mathbb{E}T(x - h)$. For $L \le 0.6$, $K_x(h) \ge Lh \ge T(x) - T(x - h)$, so the highest-cost worker does not reduce earnings under any admissible schedule.

More generally, the flexibility of K_x lets the model reproduce labor-supply responses without duplicating the preferences in Mirrlees (1971). Fix $c > 0$ and consider a continuous,

strictly increasing reduction response to a proportional tax. Setting $cK'_x(h)$ equal to its inverse gives a convex cost that reproduces choices along that tax path. The quasilinear construction above nests both choices and welfare exactly. Matching responses alone does not identify welfare or establish no manipulation at the top. Appendix C verifies the two exact examples. Online Appendix S.2.2 uses the general-density extension to match aggregate earnings and revenue in Mirrlees's logarithmic example.

2.4 Numerical Toy Example

Example 1 (Reusing lower tax brackets)**.** Firms have three potential bases. Their population shares and marginal underreporting efforts are

	Small	Medium	Large
Potential base x	0.3	0.6	1
Population share	0.5	0.3	0.2
Marginal effort p	0.6	2.4	6

Within each group, c is uniform on $[0,1]$ and welfare weights equal one. Any firm can reduce reports by any amount, divide its base in any proportions, and repeat these operations any finite number of times. Effort is

$$K = (p + e_R)\Big(x - \sum_j r_j\Big) + (0.3 + e_S)(n-1). \tag{11}$$

The marginal effort p belongs to the firm and applies to all its entities. There is no exit. Enforcement lies in $[0,2]^2$ and costs $A(e) = (0.05e_R^2 + 0.02e_S^2)/2$. The authority requires net revenue $G = 0.15$ and chooses

$$T(r) = \tau + t_1 \min\{r, 0.3\} + t_2(r - 0.3)^+, \qquad 0 \le \tau \le 0.01, \quad 0 \le t_1, t_2 \le 0.6.$$

With increasing rates, the lowest tax on a total report q through n entities is

$$n\tau + t_1 q + (t_2 - t_1)(q - 0.3n)^+.$$

Adding linear reduction costs makes the burden piecewise affine in q. An optimum is therefore at an endpoint or the kink. Counts above $\lceil x/0.3 \rceil$ cannot help, since fewer entities can already place the whole report in the lower bracket. With decreasing rates, merging dominates splitting. Thus one, two and four entities suffice for the three sizes,

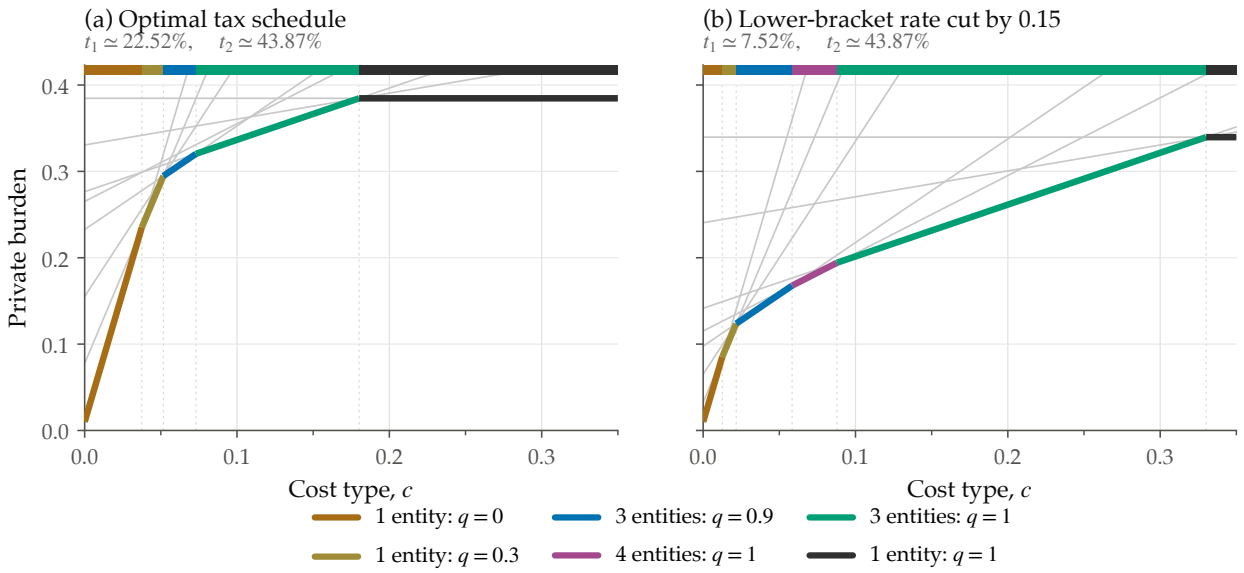


Figure 2: Firm choices and burdens before and after a lower-bracket tax cut. The figure shows firms with potential base 1 in Example 1. A strategy with n entities and total report q has private cost $c[6(1-q)+0.3(n-1)]$, where c is the firm's cost parameter. Burden is this cost plus taxes. Each entity pays a fee of 0.01 and faces one marginal rate below report 0.3 and another above it. Panel (a) uses the optimal rates of approximately 22.52% and 43.87%. Panel (b) cuts the bottom rate by 15 percentage points and holds the top rate and fee fixed. Enforcement is zero in both panels. Thin lines give candidate burdens; colored segments and the upper strips identify the minimizing strategy. Legend entries give entity counts and total reports. Arbitrary finite splits and continuous reports are feasible; the candidates shown are obtained by optimization.

respectively. These bounds follow from optimization. Appendix N gives the proof and verifies the policy calculations.

Figure 2 shows large-firm choices at the optimum with splitting and zero enforcement, and after a cut in the bottom rate. At the optimum, as costs rise, firms first report zero, then report 0.3 through one entity, report 0.9 through three entities, report their whole base through three, and finally comply. A deeper bottom-rate cut also induces four-entity firms. Some divisions give the same taxes and effort: $(0.3, 0.3, 0.4)$ and equal thirds are an example. The figure uses one representative of each such choice. Entity counts, total reports and bracket exposures are invariant across these divisions.

The example satisfies the assumptions. The inequalities $p \geq L$ and $0.3 \geq Lx/2$ ensure no manipulation at the top. A flat rate of 0.4 at zero fee and enforcement raises $58/375 > 0.15$, establishing strict feasibility (Appendix N).

3 Optimal taxes and enforcement

We first express revenue in terms of firms' minimized burdens and use this identity to obtain a concave reformulation of the authority's problem. Convex duality under Slater's condition and subdifferential optimality then give global conditions, which take the form of a Ramsey rule for entity counts. The optimization procedure is standard. Its application here requires the accounting identity and a representation of policy effects through firms' reports and manipulation costs.

3.1 Burden and revenue

A firm's strategy can jump as policy or its cost parameter changes. Its minimized burden nevertheless has the regularity needed to recover tax receipts and private costs.

Proposition 1 (Burden and revenue)**.** *Suppose that Assumption 1 and Assumption 2(i) hold.*

(i) *For each firm, $P(c;\pi)$ is continuous in (c,π) and concave in policy. It is nondecreasing, concave and Lipschitz in c. At a fixed policy, chosen effort weakly decreases and taxes paid weakly increase with c.*

(ii) *At almost every cost parameter, every minimizing strategy satisfies*

$$K = P_c, \qquad R = P - cP_c, \qquad D = cP_c. \tag{12}$$

Taxes and private costs are therefore almost surely independent of how ties are resolved.

(iii) *At every admissible policy,*

$$\mathbb{E}R = \mathbb{E}[(1+\gamma)P - \gamma T(x)], \qquad \mathbb{E}D = \mathbb{E}[\gamma\{T(x) - P\}]. \tag{13}$$

Net revenue Q is continuous and concave in policy.

Proof sketch. Under the regularity assumptions, Berge's maximum theorem gives continuity, and the envelope theorem gives $K = P_c$ almost everywhere. Concavity follows because burden is an infimum of affine functions. Integration by parts against the conditional cost density, with boundary value $P(1) = T(x)$, gives (13). Appendix F provides the details and proves the general-density counterpart, Proposition F.1.

Reports and entity counts can differ across minimizing strategies even when their taxes and costs agree. Taxes can also jump when a firm switches strategy. In the running example, a large firm switches from three entities to compliance at $c \simeq 0.18$: its tax jumps from approximately 0.28 to 0.38 while its burden remains continuous (Figure 2).

With a common exponent γ, a dollar reduction in average burden at unchanged $T(x)$ raises private costs by γ dollars and lowers receipts by $1+\gamma$ dollars. The linear boundary term comes from the highest-cost firm, whose burden is $P(1) = T(x)$. When exponents differ across firms, they remain inside the population average.

At the running example's optimum with splitting and zero enforcement, total burden per 100 firms is approximately 16.42 and their taxes under compliance would be 17.84. Actual receipts are 15.00.

Corollary 1 (Private costs and the tax gap)**.** *Suppose that Assumption 1 and Assumption 2(i) hold. At every admissible policy, conditional on almost every bundle of firm primitives, let $\mathbb{E}_c$ denote expectation over its cost parameter. Then*

$$\mathbb{E}_c D = \frac{\gamma}{1+\gamma}\big(T(x) - \mathbb{E}_c R\big). \tag{14}$$

Private manipulation costs are therefore between one half and $\bar{\gamma}/(1+\bar{\gamma})$ of the tax gap $\mathbb{E}T(x) - \mathbb{E}R$, and they equal $\gamma/(1+\gamma)$ of it when the exponent is a common γ.

The tax gap includes losses from splitting as well as underreporting, at any policy. At the same optimum in the running example, the tax gap per 100 firms is approximately 2.84 and private costs are 1.42, exactly half of it.

3.2 Concavity and global optimality

Proposition 1 establishes that net revenue $Q(\pi)$ is concave in policy π. Welfare need not be concave. Consider the Lagrangian

$$\begin{aligned} \mathcal{L}_\mu &= W + \mu(Q - G) \\ &= \mathbb{E}[U(x-P) + \mu(1+\gamma)P - \mu\gamma T(x)] - \mu A(e) - \mu G \end{aligned} \tag{15}$$

For $\mu \geq w_-$, the burden-dependent integrand is concave and nondecreasing in P: its derivative is $\mu(1+\gamma) - U'(x-P) \geq 0$. Thus $\mathcal{L}_\mu$ is concave in policy.

This yields an exact concave reformulation,

$$V(G) = \max_{\pi \in \mathcal{T}\times E,\ Q(\pi)\geq G} \mathcal{L}_{w_-}(\pi).$$

The original optima are precisely the solutions of this program whose budgets bind. Strict feasibility is Slater's condition. The convex Lagrange multiplier theorem then supplies a supporting multiplier $\mu \geq w_-$. It satisfies $\max_\pi \mathcal{L}_\mu(\pi) = V(G)$. Every optimum

raises exactly G and maximizes this Lagrangian. Subdifferential optimality then gives necessary and sufficient conditions. Proposition F.2 and Theorem F.1 verify these steps using Luenberger (1969, §8.3, Theorem 1) and the Moreau–Rockafellar sum rule (Barbu and Precupanu, 2012, Theorem 3.57).

To express the conditions economically, let n_r count a firm's entities reporting strictly above r and write $M_r = \mathbb{E}n_r$ for the *entity base*. Let n count all its entities, including zero reports. At a candidate policy define

$$w(c) = U'(x - P(c)), \qquad \Omega(c) = \mu(1 + \gamma) - w(c).$$

The coefficients on a marginal rate, the fee and enforcement are

$$\begin{aligned} \Xi_r &= \mathbb{E}[\Omega n_r] - \mu\mathbb{E}[\gamma \mathbf{1}\{x > r\}], \\ \Xi_\tau &= \mathbb{E}[\Omega n] - \mu\mathbb{E}\gamma, \\ g_i &= \mathbb{E}[\Omega c k_i] - \mu A_i(e), \qquad i \in \{R, S\}, \end{aligned} \tag{16}$$

respectively. Here A_i is the derivative of A with respect to e_i. When firms have multiple minimizing strategies, a selection assigns a probability distribution over them, measurably across firms. One selection must supply all these coefficients. The randomization represents convex combinations of active strategies in the Lagrangian's superdifferential. Appendix F proves the representation.

Theorem 1 (Optimal taxes and enforcement)**.** *Under Assumptions 1 and 2, a policy π^* is optimal if and only if $Q(\pi^*) = G$ and some $\mu \geq w_-$ and one selection at π^* satisfy:*

(i) *If $L > 0$, in the full rate class, almost everywhere,*

$$\Xi_r \leq 0 \text{ if } t^*(r) = 0, \quad \Xi_r = 0 \text{ if } 0 < t^*(r) < L, \quad \Xi_r \geq 0 \text{ if } t^*(r) = L. \tag{17}$$

For fixed brackets, use the bracket rate and the integral of Ξ_r over that bracket.

(ii) *If $F > 0$, Ξ_τ satisfies the same signs at the lower bound, interior and upper bound of $[0, F]$.*

(iii) *$g \cdot (e - e^*) \leq 0$ for every $e \in E$. Each available enforcement instrument therefore has coefficient nonpositive at zero, zero in its interior and nonnegative at its cap.*

Every supporting multiplier admits such a selection at every optimum.

These are the KKT conditions for the authority's optimization problem. The coefficients in (16) depend on firms' chosen reports and effort, without requiring smooth behavioral responses. The appendix gives the corresponding normal-cone condition for general convex policy classes.

The sequential example gives an exact certificate in the full schedule class (Example 2). At $T(r) = r/2$, $e_R = 1/10$ and $e_S = 1/5$, set $\mu = 9/8$. Then $\Omega = 5/4$ and $M_r = 9/10$ for $0 < r < 1$, so every rate coefficient is $(5/4)(9/10) - 9/8 = 0$. The enforcement exposures are $\mathbb{E}[ck_R] = 1/200$ and $\mathbb{E}[ck_S] = 1/100$, while $A_R = 1/180$ and $A_S = 1/90$, making both enforcement coefficients zero. The fee is fixed at zero, both enforcement instruments are interior, and the budget binds. Appendix E verifies the assumptions and the certificate.

3.3 Entity bases and the Ramsey rule

The entity base M_r counts a firm once for each entity reporting above r. Replacing a report of 1 by reports of 0.4 and 0.6 doubles its contribution below 0.4 and removes it above 0.6, without changing total reporting. Where $M_r > 0$, let $w_r = \mathbb{E}[w(c)n_r]/M_r$ be the average marginal welfare weight in this base.

For an interval I, raising its rate by one unit increases a fixed strategy's tax by $\int_I n_r \, \mathrm{d}r$. We call a policy regular for a change if, for almost every firm, the objective has the same derivative at all minimizers. At such a policy, the envelope theorem and accounting identity give, along a feasible interval-rate increase,

$$\begin{aligned} \partial_I \mathbb{E}P &= \int_I M_r \, \mathrm{d}r, \\ \partial_I \mathbb{E}R &= \int_I \{\mathbb{E}[(1+\gamma)n_r] - \mathbb{E}[\gamma \mathbf{1}\{x > r\}]\} \, \mathrm{d}r. \end{aligned} \tag{18}$$

Welfare falls by $\int_I \mathbb{E}[w(c)n_r] \, \mathrm{d}r$. The integrands describe effects per unit length at almost every threshold. Changing a rate at a single point does not change taxes. When minimizers have different exposures, the supporting selection still gives the optimality coefficients, but actual one-sided responses can differ.

Now scale all taxes to λT, with enforcement fixed and $0 < \lambda \leq 1$. Since

$$\lambda T[a] + cK = \lambda\{T[a] + (c/\lambda)K\},$$

a firm behaves like one with cost parameter c/λ under the original schedule. Select that firm's original strategy when $c < \lambda$, and compliance when $c \geq \lambda$. The power law gives

$$M_r(\lambda) = \mathbb{E}[\lambda^\gamma n_r + (1 - \lambda^\gamma)\mathbf{1}\{x > r\}], \tag{19}$$

and hence

$$M_r'(1-) = \mathbb{E}[\gamma n_r] - \mathbb{E}[\gamma \mathbf{1}\{x > r\}]. \tag{20}$$

The finite response uses this selection rule. It is independent of tie-breaking when the relevant counts are unique. Appendix G proves the formulas and their general-density counterparts.

Define $\varepsilon_r = -M_r'(1-)/M_r$ for $M_r > 0$. Under the stated constancy of exposures, the derivative densities are M_r for burden, $M_r(1-\varepsilon_r)$ for revenue and $M_r\varepsilon_r$ for private costs. Thus the entity-base elasticity under a proportional tax cut measures the share of the marginal burden absorbed by manipulation. With a common exponent, define the *coverage ratio* where $\Pr(x > r) > 0$ by

$$C_r = \frac{M_r}{\Pr(x > r)} \tag{21}$$

The denominator is the entity base under full compliance. Thus the ratio compares actual and potential entity bases; it is not a compliance share. Where $M_r > 0$, the exact elasticity is

$$\varepsilon_r = \gamma\Big(\frac{1}{C_r} - 1\Big). \tag{22}$$

Splitting can make coverage exceed one. Then a proportional tax cut shrinks the entity base as firms return to compliance, and a marginal-rate increase reduces private manipulation costs.

Substitution into (16) gives

$$\Xi_r = \mu M_r \left(1 - \frac{w_r}{\mu} - \varepsilon_r\right). \tag{23}$$

The global conditions therefore take the familiar Ramsey form.

Corollary 2 (The Ramsey rule and enforcement)**.** *Use a policy, multiplier and selection satisfying Theorem 1.*

(i) *In the full class with $L > 0$, at almost every threshold with $M_r > 0$, an interior rate satisfies*

$$\varepsilon_r = 1 - \frac{w_r}{\mu}. \tag{24}$$

At zero the inequality is $\varepsilon_r \geq 1 - w_r/\mu$; at the cap it is reversed. For an interior fixed-bracket rate with $M_r > 0$ almost everywhere on its bracket, $\int_I M_r(1 - w_r/\mu - \varepsilon_r)\,\mathrm{d}r = 0$. Otherwise use the integral of the undivided coefficient Ξ_r. The conditions for zero bases are given directly by (16).

(ii) *If $F > 0$ and $\mathbb{E}n > 0$, the fee obeys the same rules with base $\mathbb{E}n$, welfare weight $\mathbb{E}[w(c)n]/\mathbb{E}n$ and elasticity $-(\mathbb{E}[\gamma n] - \mathbb{E}\gamma)/\mathbb{E}n$.*

(iii) *At an interior level of an available instrument* $i \in \{R, S\}$,

$$\mu\{\mathbb{E}[(1+\gamma)ck_i] - A_i(e^*)\} = \mathbb{E}[w(c)ck_i]. \tag{25}$$

At zero the left side is at most the right side; at the cap it is at least the right side.

An interior instrument equates the welfare value of its net revenue with the burden it imposes on firms. The rate rule can also be written directly in counts:

$$\mathbb{E}[\gamma(\mathbf{1}\{x > r\} - n_r)] = \mathbb{E}[(1 - w(c)/\mu)n_r]. \tag{26}$$

The fee counts entities with zero reports as well as those with positive reports. Online Appendix S.1.4 develops its effects on entity creation and exit.

With equal welfare weights w, put $\beta = 1 - w/\mu$. The Ramsey rule then fixes an entity target.

Corollary 3 (Entity and total-report targets)**.** *Under Theorem 1 with* $U(y) = wy$*, use its multiplier and selection, and let* $\beta = 1 - w/\mu \in [0, 1)$.

(i) *At almost every interior rate in the full class,*

$$\mathbb{E}[(\gamma + \beta)n_r] = \mathbb{E}[\gamma \mathbf{1}\{x > r\}]. \tag{27}$$

With common exponent γ*, this becomes*

$$M_r = \frac{\gamma}{\gamma + \beta} \Pr(x > r). \tag{28}$$

For fixed brackets, integrate these equalities over each interior bracket.

(ii) *If all rates are interior almost everywhere, or every fixed bracket is interior, then*

$$\mathbb{E}[(\gamma + \beta) \sum_j r_j] = \mathbb{E}[\gamma x]. \tag{29}$$

With common exponent,

$$\mathbb{E} \sum_j r_j = \frac{\gamma}{\gamma + \beta} \mathbb{E}x. \tag{30}$$

The technology determines the policy and multiplier needed to meet these targets. With a common exponent, the unreported base is β/γ times the reported base. At the running example's optimum with splitting and zero enforcement, $\beta \simeq 0.21$, so firms report 82.81 percent of potential bases. With interior rates in the full class, the measure of

positive reports is $\gamma/(\gamma+\beta)$ times the measure of potential bases (Appendix G.1). The two distributions are therefore identical conditional on being positive. Fixed brackets impose only integrated restrictions. These count statements use the common supporting selection. They are independent of tie-breaking when almost every firm's report measure is unique.

The running example illustrates the bracket version of the rule. At its optimum with splitting and zero enforcement, expected exposures per 100 firms to the bottom and top rates are approximately 24.84 and 19.05, against no-manipulation exposures of 30.00 and 23.00. Both coverage ratios are approximately 82.81 percent. With $\gamma = 1$ and equal welfare weights, the corresponding elasticity is about 0.21. The supporting value $\mu \simeq 1.26$ gives the same elasticity through $1 - 1/\mu$ in both brackets. These bracket statistics do not depend on which tax-minimizing division a firm selects.

After a cut of 15 percentage points in the bottom rate, holding the other instruments fixed, the bracket coverage ratios become approximately 1.13 and 0.69. Total reports rise from about 0.44 to 0.50 (Section 5.3). The bottom-bracket ratio exceeds one, so raising that rate would reduce private manipulation costs. The reformed policy cannot be optimal for any revenue requirement within the same two-bracket class. An interior rate with equal welfare weights must have a coverage ratio at most one.

4 What sequential manipulation adds

The accounting identities and optimality conditions apply to any set of feasible strategies satisfying the model's assumptions. A sequential technology restricts how firms reach those strategies. It gives tests of individual behavior and, when continuation costs depend only on the current report, a recursive solution of the firm's problem. Example 2 uses this structure to solve jointly for reporting, organization, taxes and enforcement. Divisions and the number of entities are unrestricted.

4.1 Additive reductions and continuation choices

In the sequential technology of Section 2.1, reduction costs are additive along the path. With no intervening operation, reducing an entity's report from a to b costs the same as two consecutive reductions from a to d and from d to b, for $b \leq d \leq a$. The integrals over adjacent intervals add, as does group-level effort across underreports by different entities. Effort can still depend on more than terminal reports. A split changes the holdings and reports at which later reductions take place.

Sequential feasibility imposes a useful restriction on every optimal terminal state. Consider the technology of unrestricted finite plans, represented by the sufficient finite bound in Lemma D.1. At the end of an optimal plan, let entity j hold y_j, report q_j, and have underreported $\eta_j = y_j - q_j$. Write $q_G = \sum_j q_j$ and $\eta_G = x - q_G$ for the group totals. Use the entity and group marginal efforts p^E, p^G and split effort σ defined in Appendix D.1. Appending an underreport or a split cannot lower burden, so

$$T(q_j) - T(q_j - h) \le c \int_0^h \left[p^E(y_j, \eta_j + v; e) + p^G(\eta_G + v; e)\right] \mathrm{d}v, \quad 0 \le h \le q_j, \tag{31}$$

$$T(q_j) - T(q_j - z) - T(z) \le c\sigma(x, q_G, q_j, z; e), \qquad 0 \le z \le q_j. \tag{32}$$

Here the effort functions are evaluated at the fixed enforcement level, as in Appendix D. These are necessary conditions even when costs depend on the order of operations. For $q_j > 0$, differentiability of T at q_j also gives

$$t(q_j) \le c\left[p^E(y_j, \eta_j; e) + p^G(\eta_G; e)\right].$$

The conditions examine deviations from the entire terminal state. Checking only one operation at the initial state is insufficient, because a costly first operation can make a later one profitable. Example 2 exhibits this failure at a jointly optimal policy.

4.2 When the current report determines continuation costs

A stronger restriction makes the continuation problem one-dimensional. Fix a firm's characteristics and enforcement and suppress its type index. Allow only reductions and splits, with no exit. Suppose that reducing an entity's current report from a to b requires effort

$$C(a) - C(b), \qquad C(q) = \int_0^q p(v)\,\mathrm{d}v, \tag{33}$$

where p is continuous and nonnegative. Every split requires the same effort $\sigma > 0$. Costs depend only on the entity's current report, independently of its holdings, other entities' reports and earlier operations. In the original notation, this means $p^G = 0$, $p^E(y, \eta; e) = p(y - \eta)$ and a state-independent split effort. The functions may retain the affine dependence on enforcement in (3); enforcement is held fixed throughout this subsection.

Let $V_c(q)$ be the least remaining burden starting from one entity that currently reports q, with all primitives held fixed and all earlier costs sunk. This is an auxiliary continuation problem, not a comparison of firms with different characteristics. The firm's original

burden is $P(c;\pi) = V_c(x)$.

Proposition 2 (A recursive characterization)**.** *Under* (33) *and a constant split effort* $\sigma > 0$*, the following statements hold for every* $c \in [0,1]$*.*

(i) V_c *is the greatest function* v *on* $[0,x]$ *that satisfies*

$$v(q) \le T(q), \tag{34}$$
$$v(q) \le v(r) + c[C(q) - C(r)], \qquad 0 \le r \le q, \tag{35}$$
$$v(q) \le v(z) + v(q-z) + c\sigma, \qquad 0 \le z \le q. \tag{36}$$

It also satisfies the recursive representation

$$V_c(q) = \min_{0\le r\le q}\left\{c[C(q) - C(r)] + \min\left\{T(r),\ c\sigma + \min_{0\le z\le r}[V_c(z) + V_c(r-z)]\right\}\right\}. \tag{37}$$

(ii) V_c *is nondecreasing and absolutely continuous,* $V_c(0) = T(0)$*, and*

$$0 \le V_c'(q) \le cp(q) \qquad \text{for almost every } q. \tag{38}$$

(iii) *The no-manipulation strategy is optimal at every starting report* $q \in [0,x]$ *if and only if*

$$\begin{aligned} t(q) &\le cp(q) && \text{almost everywhere,} \\ T(q) &\le T(z) + T(q-z) + c\sigma && 0 \le z \le q \le x. \end{aligned} \tag{39}$$

(iv) *Let* U_c *be the burden when splitting is unavailable,*

$$U_c(q) = \min_{0\le r\le q}\{T(r) + c[C(q) - C(r)]\}.$$

Splitting leaves this burden unchanged at every starting report if and only if

$$U_c(q) \le U_c(z) + U_c(q-z) + c\sigma \qquad (0 \le z \le q \le x). \tag{40}$$

(Proof in Appendix D.2.) The recursion separates the first reduction, the choice to stop or split, and the two continuation problems after a split. Part (i) gives the same solution without enumerating plans. Any function below the tax schedule that satisfies both deviation inequalities bounds every plan's burden from below. The greatest such bound is attainable. Part (ii) bounds marginal *burden*, which can differ from the statutory marginal rate. The test in part (iii) must hold at all continuation reports. At the initial report x alone, the two deviation inequalities are necessary but need not be sufficient.

Part (iv) allows optimal underreporting. Its binary inequality applies to the burden after underreporting, rather than to the tax schedule.

The reduction rule still permits different costs for different sequences that end at the same reports. For a plan ending at $r_1, \ldots, r_n$, let (q_ℓ, z_ℓ) record the report of the splitting entity and the amount transferred at split ℓ. Its total effort is

$$K = C(x) - \sum_{j=1}^{n} C(r_j) + \sum_{\ell=1}^{n-1} \big[\sigma + C(q_\ell - z_\ell) + C(z_\ell) - C(q_\ell)\big]. \tag{41}$$

The first term depends only on terminal reports. The terms in brackets generally depend on where the splits occur, so the order of mixed operations can affect costs even with additive reductions. The two plans in Example 2 illustrate this distinction.

4.3 The order of reductions and splits

The slope of marginal reduction effort determines which order is cheapest. For a continuous nonnegative function f, define the reduction and splitting operators

$$(\mathcal{R}_c f)(q) = \min_{0 \le r \le q} \{f(r) + c[C(q) - C(r)]\}, \tag{42}$$

$$(\mathcal{S}_c f)(q) = \inf_{n \ge 1} \min_{\substack{y_j \ge 0 \\ \sum_{j=1}^{n} y_j = q}} \left\{ \sum_{j=1}^{n} f(y_j) + c\sigma(n-1) \right\}. \tag{43}$$

The first permits only reductions; the second permits only splits, followed by the continuation burden f for each resulting entity.

Corollary 4 (The cheapest order)**.** *Under the assumptions of Proposition 2:*

(i) *If p is nondecreasing, some optimal plan makes every split before any reduction, and $V_c = \mathcal{S}_c \mathcal{R}_c T$.*

(ii) *If p is nonincreasing, some optimal plan makes every reduction before any split, and $V_c = \mathcal{R}_c \mathcal{S}_c T$.*

(iii) *If p is constant, both orders are optimal and the two expressions coincide. If p is nonincreasing and T is subadditive, then $V_c = \mathcal{R}_c T$: splitting adds no valuable option.*

(Proof in Appendix D.2.) With increasing marginal effort, splitting first moves subsequent reductions to lower reports, where they cost less. With decreasing marginal effort, reducing in the parent first uses the cheaper marginal effort at higher reports. The first case specializes Proposition S.3.1(i) in the Online Appendix. The second gives a differ-

ent sufficient condition for reducing first: reductions are entity-specific here, whereas Proposition S.3.1(ii) assumes group-level reductions.

4.4 Splitting to make concealment cheaper

Subadditivity deters splitting when merging never raises costs. The following example shows why that qualification matters. Every division and every finite sequence of operations is feasible.

Example 2 (Splitting to make concealment cheaper)**.** Every firm has potential tax base $x = 1$, a cost parameter uniform on $[0, 1]$, and welfare weight one. Reducing an entity's current report from a to $b \leq a$ requires effort $C_e(a) - C_e(b)$, where

$$C_e(q) = (1 + e_R)q + 6q^2.$$

Every split requires effort $3/4 + e_S$. There is no exit, the fee is zero, and marginal rates lie in $[0, 1]$. Enforcement lies in $[0, 1] \times [0, 3]$ and has public cost $A(e) = (e_R^2 + e_S^2)/36$. The revenue requirement is $G = 323/720$.

For any flat schedule $T(q) = tq$, unrestricted choice gives

$$\begin{aligned} P(c) &= \min\{t, cB(e)\}, \\ B(e) &= \min\left\{7 + e_R, \frac{19}{4} + e_R + e_S, \frac{9}{2} + e_R + 2e_S\right\}. \end{aligned} \tag{44}$$

The three expressions are the least efforts of concealing the whole base through one, two or three entities. An optimal concealment plan has equal holdings. Increasing marginal reduction effort permits all splits to precede reductions. Each entity's objective is then concave in its report, so it reports all its holdings or zero. Convexity of C_e makes concealed holdings equal. Lemma E.1 excludes mixed reporting and concealment and solves the remaining integer choice.

The joint optimum in the full schedule class is

$$T(q) = q/2, \qquad e_R = 1/10, \qquad e_S = 1/5. \tag{45}$$

Here $B(e) = 5$: firms with $c < 1/10$ divide into three entities holding $1/3$ each and conceal the entire base; firms with higher costs comply. Gross receipts are $9/20$, public enforcement costs $1/720$, private manipulation costs $1/40$, and welfare $21/40$. The supporting value of public funds is $\mu = 9/8$. Appendix E verifies the model's assumptions and the global optimality conditions, including both enforcement conditions.

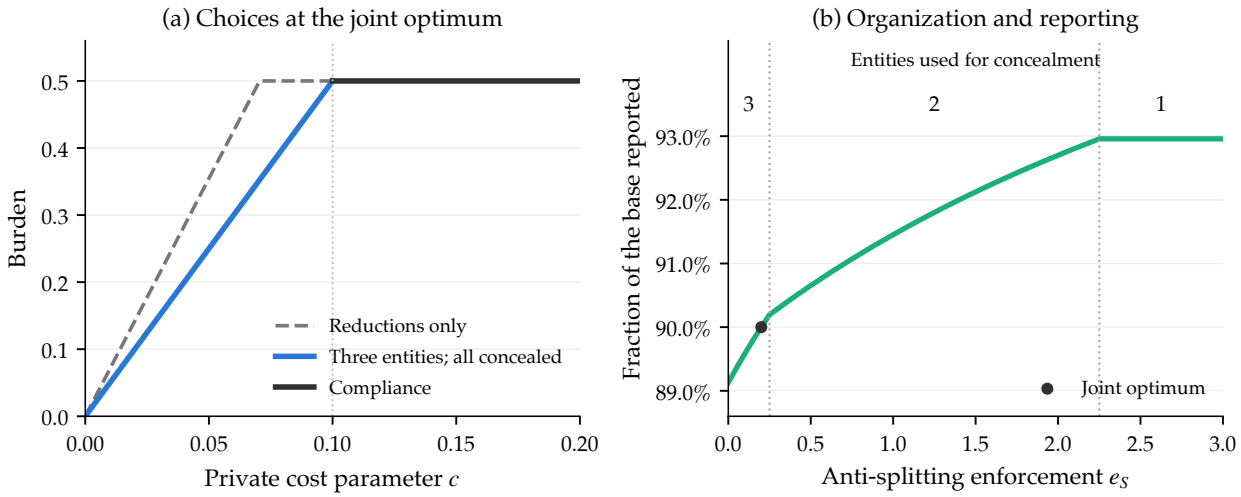


Figure 3: Splitting to reduce the cost of concealment. Every firm has potential base 1 and a private cost parameter c uniform on $[0, 1]$. Reducing an entity's report from a to b costs $c[C_e(a) - C_e(b)]$, where $C_e(q) = (1 + e_R)q + 6q^2$; each split costs $c(0.75 + e_S)$. Arbitrary finite sequences and divisions are feasible. There is no fee or exit. Panel (a) plots minimized burden, taxes plus manipulation costs, at the joint optimum of Example 2: a 50% tax rate and enforcement $e_R = 0.10$, $e_S = 0.20$. The lowest-cost tenth of firms divides into three equal entities and conceals the whole base; the others comply. The dashed curve excludes splitting at the same policy. Panel (b) holds the tax rate and e_R fixed and varies e_S. The preferred entity count for concealment falls from three to two at 0.25 and to one at 2.25. The dot marks the joint optimum. The revenue requirement is not held fixed along this curve.

The optimal linear schedule gives no tax saving from dividing a fixed total report. Splitting is useful because it moves reductions to lower reports, where marginal effort is smaller. The operation order matters even when terminal reports are positive. At zero enforcement, splitting into thirds and reducing each report to $1/6$ requires effort $7/2$. Reducing the parent's report to $1/2$ first and then dividing it into three reports of $1/6$ requires $13/2$. Both plans end at the same reports.

At the joint optimum, a firm with $c = 1/11$ cannot lower its burden by a single reduction or by a split without subsequent reductions. Its three-entity concealment plan nevertheless costs $5/11 < 1/2$. This is why the inequalities of Proposition 2 must hold at every continuation report for its exact deterrence tests to apply.

While splitting remains optimal, stronger anti-splitting enforcement raises total reporting here by making concealment more costly (Figure 3(b)). This differs from preventing tax-motivated splitting when firms respond by underreporting instead. At the same net revenue, the optimum without splitting has a flat rate of 48.16 percent and underreporting enforcement of about 0.05. Anti-splitting enforcement is zero, and welfare is about 0.53. Allowing splitting lowers attainable welfare by about 9.87×10^{-3}, even though both optimal

schedules are linear. The optimum raises the rate and enforcement to meet the revenue requirement. The exact comparison and enforcement responses are in Appendix E.

5 The threat of splitting

A small observed incidence of splitting can coexist with a substantial effect on optimal taxes. This section gives conditions under which the threat of splitting determines top marginal rates and shifts burdens toward smaller firms. The top-rate result describes a single optimum. The burden comparison concerns optima with and without splitting.

Throughout this section, the cost exponent is a common γ. Schedules belong to the class (2) with a rate cap $L > 0$. Welfare weights are linear and depend on characteristics only through the potential tax base: $U(y) = w(x)y$ with w continuous. As in Section 2.1, a firm complies if it uses the no-manipulation strategy (64), reporting its whole potential tax base through one entity. For a supporting value of public funds μ, write $\beta(x) = 1 - w(x)/\mu$. The Lagrangian value of a unit of burden is then $\Omega = \mu(\gamma + \beta(x))$. By Theorem 1, the Ramsey rule for an interior rate at r becomes the entity target

$$\mathbb{E}\left[(\gamma + \beta(x))\, n_r\right] = \gamma \Pr(x > r). \tag{46}$$

With equal weights, β is a constant and (46) is the target (28). At almost every threshold with an interior rate, the number of entities that report above the threshold is then $\gamma/(\gamma + \beta)$ times the number of firms whose potential tax base exceeds it.

5.1 Top rates follow split rates

With equal welfare weights, two behavioral conditions link top marginal rates to the rates on reports after a tax-minimizing split. First, noncomplying firms near the top put every report below a common threshold. Second, firms on the margin of compliance split without underreporting. The Ramsey rule fixes the share complying at each top base. Indifference then fixes the tax saving from splitting. A constant saving equates top and split marginal rates.

Say that $(\underline{x}, X]$ is a *top range* for a policy and a selection if every strategy that the selection assigns to a firm whose potential tax base exceeds $\underline{x}$ either is the no-manipulation strategy or reports at most $\underline{x}$ through each entity. The potential tax bases in a top range are *top bases*. Above any threshold in a top range, the only entities are those of firms that comply.

Proposition 3 (Compliance at the top)**.** *Maintain the assumptions of Theorem 1 and the restrictions of this section. Let π^* be optimal, and let μ and a selection satisfy Theorem 1 at π^*. Suppose that $(\underline{x}, X]$ is a top range for π^* and that selection, with $0 < t^*(r) < L$ at almost every $r \in (\underline{x}, X)$. Then, for almost every potential tax base $x \in (\underline{x}, X]$ under the distribution of potential tax bases, the share of the firms with base x that comply is*

$$\frac{\gamma}{\gamma + \beta(x)}. \tag{47}$$

If all firms with the same base share one technology, then for almost every such x, and up to a null set of cost parameters, the firms with base x comply exactly when $c > \kappa(x)$, where $\kappa(x) = \big(\beta(x)/(\gamma + \beta(x))\big)^{1/\gamma}$.

(Proof in Appendix H.) In a top range, the weighted entity target is a restriction on the tail distribution of complying firms. Equality of these tails gives the compliance share at almost every base, including atoms. With equal weights and one technology at each base, the cutoff κ is common across top bases. Call the firm at $\kappa(x)$ the *marginal complier*. In the next result, saying that the *marginal non-compliers* use a specified strategy means that such strategies are optimal along a sequence of cost parameters increasing to $\kappa(x)$ from below.

For a schedule T and a number of entities $n \geq 2$, the tax saved by dividing a base x among n entities that report it in full is

$$S(x) = T(x) - \min\Big\{ \sum_{j=1}^{n} T(r_j) : r \in \mathbb{R}^n_+,\ \sum_{j=1}^{n} r_j = x \Big\}. \tag{48}$$

A *minimizing division* of x is a vector of reports $r = (r_1, \ldots, r_n)$ that attains the minimum. The marginal rate $t(r_j)$ at a positive report r_j where T is differentiable is a *split rate* at x. Where T is differentiable, t denotes its derivative.

Proposition 4 (Top rates follow split rates)**.** *Maintain the assumptions of Proposition 3, and let the distribution of potential tax bases have full support on $(\underline{x}, X]$, with $\beta > 0$ there. Suppose that, for some $n \geq 2$ and $\sigma > 0$, all firms with the same base $x \in (\underline{x}, X]$ share one technology. A firm can divide its whole base among n entities in any proportions at effort σ at the enforcement of π^*. No cheaper strategy with n entities has reports summing to x. Suppose also that, for almost every such x, the marginal non-compliers split into n entities and report their whole base. Then, with the saving and the minimizing divisions computed at T^* and this n:*

(i) *At every $x \in (\underline{x}, X]$, the tax saved by splitting equals the marginal complier's cost of splitting, $S(x) = \sigma\kappa(x)$.*

(ii) *With equal weights, the top rates equal the split rates. At Lebesgue-almost every $x \in (\underline{x}, X)$,*

$t^*(x) = t^*(r_j)$ *for every minimizing division* r *of* x *and every positive report* r_j *at which* T^* *is differentiable. In particular,* $t^*(x) = t^*(x/n)$ *if the equal division is minimizing and* T^* *is differentiable at* x/n.

(iii) *With general weights,* κ *is Lipschitz on the top range and* $t^*(x) = t^*(r_j) + \sigma\kappa'(x)$ *for Lebesgue-almost every top base and the same divisions and reports.*

(Proof in Appendix H.) The marginal complier's indifference gives part (i). Adding or removing a unit from any positive report of a minimizing division gives parts (ii) and (iii), wherever the corresponding derivatives exist. Falling welfare weights make κ nondecreasing, so the top rate exceeds the split rate by $\sigma\kappa'(x)$ almost everywhere. With finitely many bases, the corresponding restriction bounds average top rates between average split rates (Lemma S.4.1 in the Online Appendix), without equating pointwise rates.

Corollary 5 (A progressive optimum has a plateau)**.** *Under Proposition 4, suppose welfare weights are equal and* T^* *is convex. Then* T^* *is affine on* $(\underline{x}/n, X)$.

Convexity makes equal division tax-minimizing. Equality of the marginal rates at x/n and x then forces the rate to be constant between them. As x ranges over the top bases, these intervals overlap and cover $(\underline{x}/n, X)$ (Appendix H).

Figure 1 shows the top-rate mechanism in a many-base version of the benchmark technology. Potential bases range from $1/40$ to 1 in increments of $1/40$, with masses proportional to $1/x$. Underreporting effort is $p(x) = 0.6(1 + 4x)$ per unit removed, and each additional entity costs effort 0.3. Cost types are uniform and welfare weights equal one. The authority chooses the entire marginal-rate profile subject to a cap of 0.6. The fee and enforcement are zero. With splitting, the optimum has matching rate plateaus on smaller reports and at the top, separated by a hump. Without splitting, rates rise throughout the range. The plateau is an outcome of the optimization. The intermediate hump shows that the optimum need not be progressive. Appendix N.4 verifies the unrestricted choices and the full-class welfare bound.

For the remaining policy comparisons, return to the three-base, two-bracket economy of Example 1. Appendix N derives the finite candidate set from unrestricted manipulation choices and provides the numerical certificates.

With enforcement fixed at zero, allowing splitting raises the bottom rate from 13.98 to 22.52 percent and lowers the top rate from 52.43 to 43.87 percent (Table 1). Only 2.57 percent of firms split at the new optimum. If the authority kept the schedule optimized without splitting, 13.15 percent would split and receipts would be about 0.14, short of the 0.15 requirement. The tax adjustment substantially reduces the splitting that is observed.

Table 1: Optimal policies in the unrestricted three-base example. All rows raise net revenue 0.15 and have fee $\tau = 0.01$. The rates, shown as percentages, apply below and above the fixed threshold 0.3. "Splitting" enforcement means that only e_S can be chosen; "Both" permits both instruments. The last column is the percentage of original firms that split. Appendix N gives the computation and exact certificates.

Splitting	Enforcement	t_1 (%)	t_2 (%)	e_R	e_S	Split (%)
Unavailable	None	13.98	52.43	0	0	0
Available	None	22.52	43.87	0	0	2.57
Available	Splitting	19.48	46.87	0	0.21	1.79
Unavailable	Both	15.22	50.10	0.13	0	0
Available	Both	21.07	43.50	0.17	0.15	1.48

This comparison illustrates rate compression in a fixed-bracket class. Pointwise equality of top and split rates requires the support and behavioral assumptions of Proposition 4.

5.2 Who pays for the threat of splitting

The comparison with an economy without splitting requires additional restrictions. Below every report of a split firm's entities, that firm pays the marginal rate once per entity. These extra exposures can support higher low rates, but the comparison also depends on the value of public funds and on the shape of the two optimal schedules.

Let T^1 and T^2 be optimal schedules for the same requirement without and with splitting. Let values of public funds μ_1 and μ_2 and selections satisfy Theorem 1 at T^1 and T^2. Impose equal welfare weights and no exit. Fix enforcement at a common level and require the two schedules to have equal fees. Suppose also that underreporting has a constant marginal effort, so that a firm can make any one-entity report $q \leq x$ at effort $p(x - q)$, and at no less. Let r_0 be the essential infimum of the reports of the entities of split firms under the selection at T^2, that is, the smallest report of a split firm's entity. The comparison also requires that the value of public funds not fall when splitting becomes possible, $\mu_1 \leq \mu_2$. At requirements where both welfare value functions are differentiable, this ordering holds when the welfare loss from splitting increases with the requirement.

Proposition 5 (Small firms pay for the threat of splitting)**.** *In the comparison just described, suppose that $\mu_1 \leq \mu_2$, that a positive mass of firms splits at T^2 and $r_0 > 0$, and that both schedules have interior rates almost everywhere on $(0, r_0)$. Suppose also that T^1 is progressive, that T^2 is progressive on $(0, r_0)$, and that, up to null sets, no rate of T^2 above r_0 is below a rate of T^2 below r_0. Then:*

(i) *$t^2(r) > t^1(r)$ at almost every $r \in (0, r_0)$.*

(ii) *Almost every firm with potential tax base at most r_0 bears a weakly higher burden under T^2 than under T^1, and a strictly higher one if its report under T^2 is positive.*

(Proof in Appendix H.) Below r_0, split firms contribute more exposures than they would through one entity. The maintained rate-shape restrictions make one-entity exposure strictly decreasing in the local rate. Together with $\mu_2 \geq \mu_1$, the entity targets therefore require higher rates under T^2. Integrating the rate inequality gives the burden comparison. Access to splitting alone implies neither the multiplier ordering nor the required rate shapes. Online Appendix S.11 gives related decompositions and comparisons in a two-bracket specialization.

The unrestricted example exhibits this redistribution (Figure 4). With enforcement held at zero, access to splitting raises average burden by 37.8 percent for the smallest firms and 1.6 percent for the middle group, and lowers it by 8.6 percent for the largest firms. Burden includes tax payments and manipulation costs. The smallest firms never split; their higher burden comes entirely from the higher bottom rate. Large firms benefit from the lower top rate as well as from the option to split.

With fixed brackets, each optimal rate can also be compared with the *one-entity rate*. This is the rate prescribed by the Ramsey rule when each firm reports through one entity, evaluated at the optimal enforcement and value of public funds. Suppose that the optimum is progressive, weights are equal, no firm exits and one-entity underreporting costs are convex. Consider a bracket whose one-entity rate lies strictly between zero and the cap. If splitting raises its average entity base relative to one-entity reporting at the same policy, its optimal rate exceeds the one-entity rate. If splitting lowers that average, the optimal rate is lower. This comparison holds for any splitting technology (Proposition S.11.1 in Online Appendix S.11).

With two brackets, interior rates and $\beta > 0$, suppose splitting gives firms more access to the lower bracket without changing the cost of underreporting a given total amount. The ratio of the top rate to the bottom rate is then at most the ratio of the one-entity rates. The inequality is strict if some firms split and incur a fee or effort even without enforcement, or if the authority spends on enforcement against splitting (Proposition S.11.2 and inequality (S.56)). Further conditions place the ratio below its optimal value without splitting (Corollary S.11.1).

The ratio of the top rate to the bottom rate falls from 3.75 to 1.95 when splitting becomes available. Optimizing enforcement against splitting raises it to 2.41 and reduces the share that splits to 1.79 percent. The bottom rate falls to 19.48 percent, the top rate rises to 46.87 percent, and the redistribution of burdens toward small firms is smaller. When both instruments are available in the example, the authority chooses positive enforcement

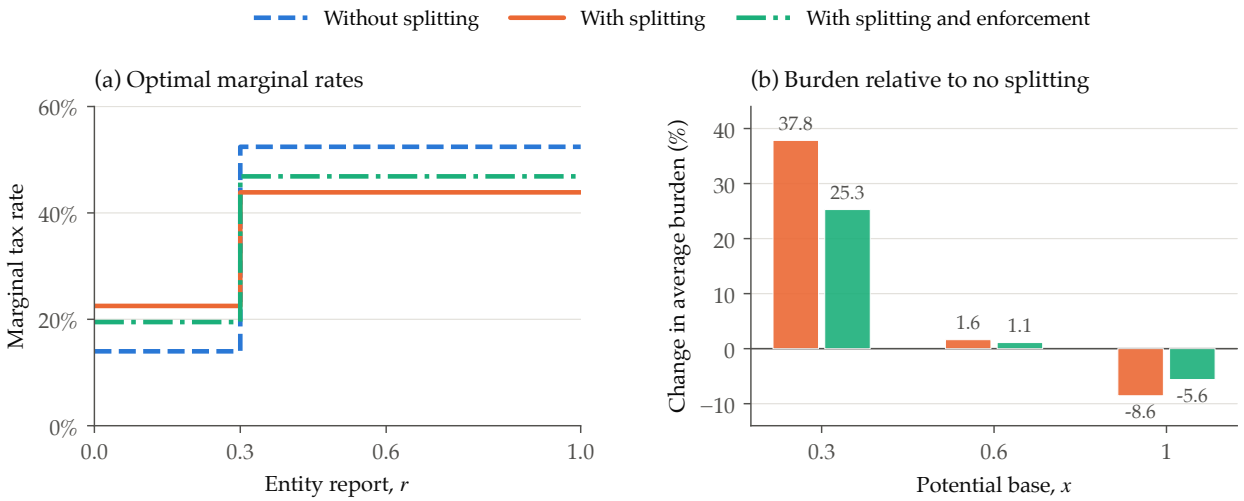


Figure 4: Optimal rates and the distribution of burdens when firms can split. Potential bases are 0.3, 0.6 and 1, with population shares 0.5, 0.3 and 0.2. Private manipulation cost is $c[p(x-q)+(0.3+e_S)(n-1)]$ for total report q through n entities; p equals 0.6, 2.4 and 6 at the three bases, and c is uniform on $[0,1]$. Divisions and reports are unrestricted. The authority chooses two marginal rates capped at 60%, separated at report 0.3, and a fee capped at 0.01. It maximizes average firm surplus net of taxes and manipulation costs, subject to net revenue 0.15. Underreporting enforcement is zero. When available, anti-splitting enforcement $e_S \in [0,2]$ costs $0.01e_S^2$. Panel (a) compares the three optimal policies. Panel (b) gives the percentage change in average burden, taxes plus private costs, at each potential base relative to the optimum without splitting. There is no exit.

against each form of manipulation and 1.48 percent of firms split (Table 1).

5.3 Lower-bracket relief

Splitting also raises the cost of relief aimed at small firms. The lower rates apply separately to the first part of each entity's report. Consider a cut of the same amount in every marginal rate below a threshold z. Let the cut be a, and hold the rates above z, the fee and enforcement fixed. The schedule becomes $T(r) - a\min\{r, z\}$. A firm that keeps its strategy saves a times its *exposure* $B_z^0 = \sum_j \min\{r_j, z\}$. Exposure sums each entity's report up to the threshold. The *mechanical estimate* of the revenue loss is therefore $a\mathbb{E}B_z^0$. It holds reports and entity counts at their initial values.

Proposition 6 (Lower-bracket relief)**.** *Maintain Assumption 1 and no manipulation at the top, (6), and suppose every schedule along the cut is admissible.*

(i) *If firms can operate at most one entity, with or without the option to exit, the revenue loss from any admissible cut is at most its mechanical estimate.*

(ii) *With splitting, the loss can exceed the mechanical estimate even when the initial policy is optimal and no firm initially splits.*

(iii) *Gross receipts are concave in the amount of the cut. Once a cut loses more than its mechanical estimate, every larger admissible cut does too.*

(Proof in Appendix J; Online Appendix S.5 gives the exact decomposition of the loss.) Without splitting, a firm's response to the cut can only raise its exposure. A larger report under the new schedule pays at least the tax predicted from its old report. With splitting, a sufficiently large cut induces firms to create entities that claim the lower rates. A firm that divides its report between two entities that each report at least z doubles its exposure without reporting less.

In Example 1, cutting the optimal bottom rate by 15 percentage points, from 22.52 to 7.52 percent, raises the share of firms that split from 2.57 to 13.75 percent. Firms now use two, three and four entities, all chosen from the unrestricted technology. Receipts fall by about 4.28×10^{-2}, about 15 percent more than the mechanical estimate of 3.73×10^{-2}, although total reports rise from about 0.44 to 0.50. Figure 5 traces these responses as the cut grows. The revenue loss exceeds its mechanical estimate once the cut exceeds about 8.79 percentage points.

A rise in total reports therefore does not show that relief partly pays for itself. Attributing its excess cost to splitting requires records that link entities to their owners. The ratio of the actual to the mechanical loss can be arbitrarily large. Online Appendix S.5 constructs

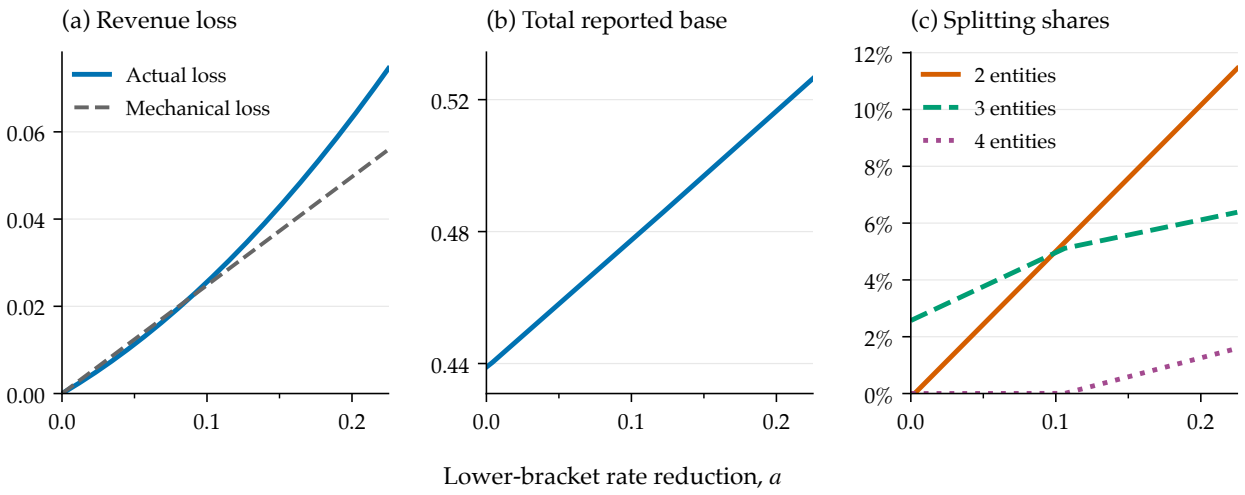


Figure 5: Revenue and reporting after a cut in the lower marginal rate. Potential bases 0.3, 0.6 and 1 have population shares 0.5, 0.3 and 0.2. Private manipulation cost is $c[p(x-q)+0.3(n-1)]$ for total report q through n entities, with $p = 0.6$, 2.4 and 6 at the three bases and c uniform on $[0,1]$. Arbitrary finite splits and continuous reports are feasible. Enforcement is zero and there is no exit. The initial schedule raises revenue 0.15, charges a fee of 0.01, and has rates of approximately 22.52% below report 0.3 and 43.87% above it. The reform reduces the bottom rate by a, holding the top rate and fee fixed. Panel (a) compares the actual revenue loss with the mechanical estimate that holds all initial reports and entity counts fixed. Panel (b) gives total reports per original firm. Panel (c) gives the shares of original firms using two, three and four entities. Outcomes integrate exactly over cost parameters.

economies with increasing numbers of entities, decreasing incremental splitting costs and thresholds that fall as the number of entities rises.

6 The welfare cost of splitting

The welfare cost of splitting includes the cost of changing policy to discourage it. The relevant comparison therefore lets the authority redesign taxes and enforcement while holding its revenue requirement fixed. Splitting may lower attainable welfare even when little of it remains after that adjustment.

6.1 Adding the option to split

Compare two economies with the same population, welfare functions and policy instruments. Their strategy sets satisfy $\mathcal{A}^1 \subseteq \mathcal{A}^2$. Shared strategies have the same reports and costs, and the second economy adds splitting strategies, which can also make subsequent underreporting cheaper. Exit, if available, is shared by the two economies. At a fixed policy, the private gain from the added options is

$$\Delta(c;\pi) = P^1(c;\pi) - P^2(c;\pi) \geq 0.$$

The comparison also applies to other expansions of the strategy set when both economies satisfy the maintained assumptions.

Theorem 2 (Manipulation options and welfare)**.** *Suppose that $\mathcal{A}^1 \subseteq \mathcal{A}^2$ for every firm and that both economies satisfy Assumptions 1 and 2, with the same population, welfare functions, admissible schedules, enforcement set and public cost.*

(i) *At every policy, the larger set of strategies weakly lowers revenue and weakly raises manipulation costs:*

$$\mathbb{E}R^1 - \mathbb{E}R^2 = \mathbb{E}\big[(1+\gamma)\Delta\big], \qquad \mathbb{E}D^2 - \mathbb{E}D^1 = \mathbb{E}[\gamma\Delta]. \tag{49}$$

(ii) *The larger set does not raise attainable welfare, $V^2(G) \leq V^1(G)$. For every supporting multiplier μ_1 of $V^1(G)$ and every optimal policy π_2 with the larger set,*

$$V^1(G) - V^2(G) = V^1(G) - \mathcal{L}^1_{\mu_1}(\pi_2) + \mathbb{E}\int_{P^2(c;\pi_2)}^{P^1(c;\pi_2)} \big\{\mu_1(1+\gamma) - U'(x-p)\big\}\,\mathrm{d}p. \tag{50}$$

Both terms are nonnegative. In particular,

$$V^1(G) - V^2(G) \;\geq\; \mathbb{E}\Big[\big(\mu_1(1+\gamma) - w_+\big)\Delta(c;\pi_2)\Big] \;\geq 0. \tag{51}$$

(iii) *Attainable welfare is the same in both economies, $V^2(G) = V^1(G)$, if and only if some optimal policy with the smaller set has $\Delta = 0$ for almost every firm. In that case, the optimal policies with the larger set are exactly the optimal policies with the smaller set that have $\Delta = 0$ for almost every firm. Hence, if a positive mass of firms gains from the added strategies at some optimal policy with the larger set, then $V^2(G) < V^1(G)$.*

The first term in (50) evaluates the policy chosen with splitting in the economy without it, valuing any revenue surplus at μ_1. It is a Lagrangian gap, not a welfare comparison at a fixed revenue requirement. The second term values the revenue lost to the remaining splitting opportunities and subtracts the owners' private gain.

(Proof in Appendix I.) The equality criterion tests the policies that would be optimal without splitting. If at least one leaves no strict private gain from the added strategies, it remains optimal. Otherwise attainable welfare falls.

At a fixed policy, added strategies reduce burdens while leaving $T(x)$ unchanged. The accounting identities convert each dollar of burden reduction into $1 + \gamma$ dollars of lost receipts and γ dollars of private costs when the exponent is common. At a supporting multiplier, the restriction on welfare weights makes the lost revenue worth at least the private gain. This comparison of Lagrangians ranks welfare after policy adjusts. Both the weight restriction and no manipulation at the top matter; heterogeneous avoidance opportunities can improve redistribution in other settings (Kopczuk, 2001).

With equal linear welfare weights and a binding budget, the interpretation simplifies to

$$W = w\{\mathbb{E}x - G - A(e) - \mathbb{E}D\}.$$

The welfare loss is the additional public and private resources required to raise the same revenue. With unequal weights, the distribution of burdens across owners also matters, so resource totals alone do not measure the welfare change.

In the running example, allowing splitting lowers attainable welfare by about 2.95×10^{-3} when enforcement is fixed at zero and by 1.45×10^{-3} when both economies can optimize enforcement (Table 1). The corresponding split shares at the optima with splitting are 2.57 and 1.48 percent. In the sequential example, the loss is about 9.87×10^{-3} even though both economies have an optimal linear schedule: splitting lowers the effort required for concealment (Example 2).

In the benchmark technology of Section 2.1, the effort σ spent creating an entity could instead underreport σ/p units through the parent. Under a progressive schedule, Online Appendix S.12.1 gives an exact test that compares taxes after splitting with those paid through one entity at that alternative report. Online Appendix S.12.2 derives further implications for policy in this technology.

6.2 Bounds on the welfare cost

A schedule is *subadditive* if $T(r_1 + r_2) \leq T(r_1) + T(r_2)$ whenever $r_1 + r_2 \leq X$. Dividing a report among entities then never lowers taxes. Proposition S.12.3 in Online Appendix S.12.3 characterizes these schedules. Say that *merging never raises costs* if every operating strategy of the larger set has a one-entity strategy in the smaller set that reports the same total with no larger effort coefficients.

Proposition 7 (The subadditivity bound)**.** *Suppose that the assumptions of Theorem 2 hold and merging never raises costs. Let the smaller set consist of the strategies without splitting, together with exit if available, and let the larger set add strategies with splitting. Let $V^1_{\text{sub}}(G)$ be the supremum of welfare attainable without splitting when the admissible schedules are restricted to those that are subadditive. Then*

$$V^1(G) - V^2(G) \leq V^1(G) - V^1_{\text{sub}}(G). \tag{52}$$

Every linear schedule with a nonnegative fee is subadditive, so the right side is at most the gain from the best schedule over the best linear schedule in the economy without splitting. If an optimum without splitting is subadditive, it remains optimal with splitting and $V^2(G) = V^1(G)$.

(Proof in Appendix I.3.) Subadditivity and the merging condition ensure equal burdens in both economies. Thus every feasible subadditive policy without splitting remains available with splitting at the same welfare and revenue. The bound is informative when such a policy raises G. The benchmark technology admits a tighter bound based on the splitting test of Proposition S.12.1 in the Online Appendix.

The running example satisfies the merging condition. Consolidating a given total report leaves its reduction effort unchanged and saves the effort of splitting. The sequential example does not: smaller entities make concealment cheaper, so firms split even under its optimal linear tax. Subadditivity alone cannot eliminate the welfare loss.

Applying the argument behind (51) at the optimum without splitting gives an upper bound. The two bounds relate the welfare cost to the revenue effects of splitting at the two optima.

Proposition 8 (Revenue bounds on the welfare cost)**.** *Suppose that the assumptions of Theorem 2 hold. Let π_1 be an optimal policy with the smaller set and μ_2 a supporting multiplier of $V^2(G)$. Then*

$$V^1(G) - V^2(G) \leq \mathbb{E}\big[(\mu_2(1+\gamma) - w_-)\,\Delta(c;\pi_1)\big]. \tag{53}$$

If, in addition, $U(y) = wy$ for every firm and the exponent is a common γ, then for every supporting multiplier μ_1 of $V^1(G)$ and every optimal policy π_2 with the larger set,

$$\frac{\mu_1(1+\gamma) - w}{1+\gamma}\,\Delta R(\pi_2) \leq V^1(G) - V^2(G) \leq \frac{\mu_2(1+\gamma) - w}{1+\gamma}\,\Delta R(\pi_1), \tag{54}$$

where $\Delta R(\pi) = \mathbb{E}R^1(\pi) - \mathbb{E}R^2(\pi)$ is the revenue that the added strategies cost at the policy π.

(Proof in Appendix I.3.) In the running example with enforcement fixed at zero, splitting reduces receipts by approximately 1.27×10^{-2} at the optimum without splitting and 1.65×10^{-3} at the optimum with splitting. Equation (54) converts these into welfare bounds of 1.14×10^{-3} and 9.71×10^{-3}, bracketing the actual loss of 2.95×10^{-3}. Both policies meet the same requirement $G = 0.15$ in their respective economies.

6.3 Why optimal policy can leave firms splitting

When splitting strictly lowers attainable welfare, complete deterrence is inconsistent with the common selection supporting an optimum. If that selection used only the original strategies, the same policy, multiplier and exposures would satisfy the optimality conditions in both economies.

Proposition 9 (Residual splitting and enforcement)**.** *Suppose that the assumptions of Theorem 2 hold, that the smaller set contains all one-entity and exit strategies of the larger set, and that the added strategies have at least two entities. If $V^2(G) < V^1(G)$, then every selection supporting an optimum in the larger economy assigns positive population mass to splitting. If splitting status is unique for almost every firm at that optimum, a positive mass of firms actually splits.*

Suppose also that welfare weights are equal and linear, and that enforcement against splitting has $k_S = 0$ on one-entity and exit strategies and $k_S > 0$ on every split strategy. If a small increase from $e_S = 0$ is feasible and $A_S(e_R, 0) = 0$ at every admissible e_R, every optimum in the larger economy has $e_S > 0$.

The proof in Appendix I also covers general conditional densities. Without the uniqueness condition on splitting status, positive mass in a supporting selection need not imply positive observed splitting: firms may resolve ties differently.

If merging a firm's reports into one entity weakly reduces every effort coefficient, positive enforcement makes merging strictly cheaper. An optimally chosen split must then retain a strictly positive tax saving. Corollary S.4.1 in Online Appendix S.4 gives this result and bounds the remaining tax saving in terms of marginal enforcement expenditure.

6.4 The value of linking entities

Ownership information allows the authority to tax each firm's consolidated report and charge the fee once. If splitting never reduces the cost of achieving a given total report, costless consolidation restores the possibilities of the economy without splitting.

Proposition 10 (Linking entities)**.** *Suppose that the assumptions of Theorem 2 hold. Let the smaller strategy set consist of the one-entity strategies and any exit options, and let the larger set add operating strategies with several entities. For every operating strategy in the larger set, suppose that the smaller set contains a one-entity strategy with the same total report and no larger effort coefficient for any enforcement instrument or at baseline. Hold the original firms, their welfare functions, exit options and policy costs fixed. Under costless group consolidation, apply an admissible common schedule to each firm's total report and charge the fee once per operating firm. Attainable welfare is then $V^1(G)$. The gain relative to separate entity taxation is $V^1(G) - V^2(G)$, bounded below by (51).*

Proof. It suffices to show that consolidation reproduces the smaller economy's burdens at every policy. Under consolidation, an operating strategy with total report q pays $T(q)$, as does its available one-entity counterpart. The latter has no larger cost at any enforcement level, so it weakly dominates the strategy with several entities. One-entity strategies and exit retain their taxes and costs. Minimized burdens therefore coincide in the two economies. The accounting identities (13) give the same aggregate receipts, and the unchanged welfare functions give the same welfare. Optimizing over the common policy set proves the result. □

In the running example, costless consolidation raises welfare by about 1.45×10^{-3} when both economies optimize enforcement. Under the proposition's conditions, the revenue bounds in Proposition 8 also bound the value of consolidation. The cost of collecting and using ownership information must be assessed separately. Consolidation need not reproduce the no-splitting economy when operating several entities lowers underreporting costs. Linked records also help identify private costs from proportional reforms, since statistics such as the split share count firms rather than entities (Section 9).

7 Coverage ratios as sufficient statistics

With a common cost exponent, coverage ratios determine the revenue raised per dollar of additional burden at the current policy. With equal welfare weights, they also rank revenue-raising instruments by welfare cost. We use these properties to evaluate reforms and derive an optimization rule based on firms' policy exposures.

Throughout this section, cost parameters follow the power law with a common exponent γ. Assumption 1 and no manipulation at the top (Assumption 2(i)) hold. Enforcement is fixed except in Section 7.3. Regular policies, exposures to a policy change and the average welfare weight w_r of the entity base are as in Section 3.3.

7.1 The welfare cost of a dollar raised by each rate

Proposition 11 (The welfare cost of a dollar raised by each rate)**.** *Maintain the assumptions of this section.*

(i) *In the full class, at a regular policy and almost every threshold r with $M_r > 0$, the marginal welfare loss is $w_r M_r$ and the marginal revenue gain is $(1+\gamma)M_r - \gamma \Pr(x > r)$. These are the densities of the interval effects in (18). If the coverage ratio C_r of (21) exceeds $\gamma/(1+\gamma)$, each dollar of revenue that the increase raises therefore costs*

$$\frac{w_r}{1+\gamma-\gamma/C_r} \tag{55}$$

in welfare.

(ii) *At a policy regular for raising the fee, with $\mathbb{E}n > \gamma/(1+\gamma)$, a rise in the fee obeys (55) with the number of entities per firm, $\mathbb{E}n$, in place of C_r and the average marginal welfare weight of firms' entities, $\mathbb{E}[w(c)n]/\mathbb{E}n$, in place of w_r.*

(iii) *Let I be a positive-length interval whose rate can be changed, and suppose the policy is regular for this change. Define*

$$C_I = \frac{\int_I M_r \, \mathrm{d}r}{\int_I \Pr(x > r) \, \mathrm{d}r}, \qquad w_I = \frac{\int_I \mathbb{E}[w(c)n_r] \, \mathrm{d}r}{\int_I M_r \, \mathrm{d}r}.$$

If the denominators are positive and $C_I > \gamma/(1+\gamma)$, (55) holds with C_I, w_I in place of C_r, w_r. In a fixed-bracket class take $I = I_j$ and write $C_j = C_{I_j}$, $w_j = w_{I_j}$.

For an interval, integrate (18) and divide the welfare loss by the positive revenue gain. For the fee, the no-manipulation exposure is one and the actual exposure is n, including zero-report entities but excluding exit. These observations prove all three parts. If the

relevant coverage ratio equals $\gamma/(1+\gamma)$, the revenue derivative is zero and a finite welfare cost per dollar is undefined. Below that threshold, a rate increase reduces revenue.

By (22), the denominator is $1-\varepsilon_r$. The formula is the marginal cost of public funds weighted by the welfare weight of the exposed entities. Where splitting makes $C_r > 1$, a dollar raised costs less than w_r because the increase reduces private manipulation costs.

With equal positive welfare weights, coverage ratios rank the welfare cost of instruments that raise revenue at the margin.

Corollary 6 (Three tests of a schedule)**.** *Suppose that the policy is regular for the changes considered and that those changes are feasible. Interval statements below use C_I, w_I when their denominators are positive; the corresponding marginal statements hold for almost every threshold in the sense of* (18).

(i) *If $C_I < \gamma/(1+\gamma)$, every positive feasible increase in the rate on I, holding other instruments fixed, lowers revenue. A small feasible cut raises revenue to first order and weakly raises welfare; the welfare gain is strictly positive to first order when $\int_I \mathbb{E}[w(c)n_r]\,\mathrm{d}r > 0$. If $C_I > \gamma/(1+\gamma)$, a small feasible increase raises revenue to first order.*

(ii) *Suppose that welfare weights are equal and positive, and that $C_I, C_J > \gamma/(1+\gamma)$ for two disjoint intervals. A change that raises the rate on I, lowers it on J, and leaves revenue unchanged to first order raises welfare to first order if and only if $C_I > C_J$.*

(iii) *Suppose that welfare weights are equal and positive and that strict feasibility (Assumption 2(iii)) holds, and write $\beta = 1 - w/\mu$. At an optimum in the class* (2)*, the coverage ratio equals $\gamma/(\gamma+\beta)$ at almost every threshold with an interior rate. Each such coverage ratio therefore reveals $\beta = \gamma(1/C_r - 1)$ and the value of public funds relative to the welfare weight, $\mu/w = 1/(1-\beta)$.*

In a fixed-bracket class, the same statements hold for the rates of brackets, with the bracket coverage ratios of Proposition 11(iii).

For part (i), the revenue derivative has the sign of $(1+\gamma)C_I - \gamma$. Concavity of revenue bounds every finite increase along this same interval direction by its initially negative derivative. Lower taxes weakly reduce every firm's burden, proving the welfare assertion. Part (ii) follows because $w/(1+\gamma-\gamma/C)$ is strictly decreasing in C on its positive-revenue domain. Part (iii) restates the entity target. Thus, with equal weights, an interior optimum has a common coverage ratio in $(\gamma/(1+\gamma), 1]$. A common ratio is necessary; feasibility, boundary instruments and the remaining optimality conditions must also hold.

At the no-splitting optimum, admitting splitting raises bottom-bracket coverage to about 1.06 and lowers top-bracket coverage to about 0.67. The corresponding welfare costs per dollar are approximately 0.95 and 1.96. Shifting revenue from the top rate to the

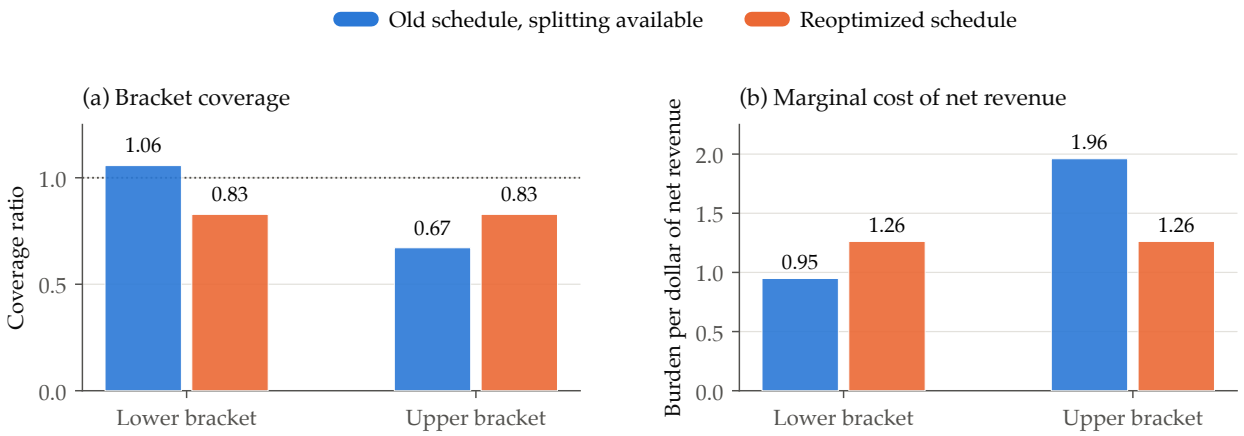


Figure 6: Reporting relative to compliance and the welfare cost of raising revenue. Potential bases 0.3, 0.6 and 1 have population shares 0.5, 0.3 and 0.2. Private manipulation cost is $c[p(x-q)+0.3(n-1)]$ for total report q through n entities, with $p = 0.6$, 2.4 and 6 and c uniform on $[0,1]$. Firms may choose arbitrary finite divisions and continuous reports. Enforcement is zero, there is no exit, and welfare weights equal one. The two tax brackets meet at report 0.3 and the fee is 0.01. Blue evaluates the optimum without splitting after splitting becomes available; orange reoptimizes for revenue 0.15. Panel (a) divides the expected amount subject to each bracket's rate by its value under full compliance: 0.3 for the bottom bracket and 0.23 for the top. Panel (b) converts each ratio C into welfare cost per additional dollar of revenue using $C/(2C-1)$. Reoptimization equalizes these costs at approximately 1.26.

bottom rate therefore raises welfare locally. Reoptimization raises the bottom rate and lowers the top rate until both costs equal $\mu \simeq 1.26$. This comparison concerns the two freely chosen bracket rates, not pointwise rates within either bracket.

7.2 Measuring coverage ratios

Tax returns give the numerator of a coverage ratio: the number of entities reporting above each threshold. The denominator counts firms whose potential base exceeds the threshold. Constructing it requires information on both potential bases and ownership.

Random audits can measure potential bases for a sample of firms when the gap between potential and reported bases is concealment. Some tax authorities already use such audits to measure the tax gap. Audits of filers miss activity forgone through real responses, the bases of firms that exit, and concealment that auditors do not detect. Each omission raises measured coverage and understates the welfare cost of a dollar raised.

Ownership links are needed to add up each firm's potential base and count it once. Registries record nominal owners, as the retailer in Section 1.1 illustrates. An audit can establish ownership for a sample even when the authority cannot do so for every entity it taxes. Measuring coverage therefore requires less ownership information than taxing by

owner (Section 6.4).

A reform that scales every tax by a common factor at fixed enforcement reveals the response of the entity base at every threshold. It therefore reveals the welfare cost of a dollar raised by each rate (Section 9). Such reforms are rare. Although the responses change with policy, the cost exponent remains fixed. A proportional reform identifies it from a nonzero responsive statistic, such as the split share, under the stability and selection conditions of Proposition 15(iii). If the population, technology and cost distribution remain stable, audit estimates of later coverage ratios give the corresponding entity-base responses through (22).

When exponents differ across characteristic groups, the revenue effect of a rate increase is $\mathbb{E}[(1+\gamma)n_r] - \mathbb{E}[\gamma \mathbf{1}\{x > r\}]$. It requires coverage ratios and exponents group by group. With general welfare weights, (55) also requires the average weight w_r of the firms behind the entities above r. At low thresholds, these include entities of large firms that split. Manipulation that does not respond to taxes falls outside the maintained cost-scale model. Applying the tests to it requires a separate model or an identified adjustment to both counts and potential bases. Except for the finite-increase conclusion along a fixed direction in part (i), the tests are local. Large reforms can activate new strategies, as lower-bracket relief does in Section 5.3.

7.3 Learning the optimal policy

Policy exposures can also be used to maximize the Lagrangian at a fixed value of public funds. At every policy, a common selection of minimizing strategies gives a *supergradient*. This vector of net values bounds the gain from any policy change through (78). With equal weights and a common exponent,

$$\Xi_r = \mu \Pr(x > r)\big[(\gamma+\beta)C_r - \gamma\big], \qquad \Xi_\tau = \mu\big[(\gamma+\beta)\mathbb{E}n - \gamma\big].$$

For an instrument with a positive revenue derivative, its score is positive exactly when its welfare cost per dollar is below μ. Integrating Ξ_r gives a bracket's score. The enforcement score is $\mu(\gamma+\beta)\mathbb{E}[ck_i] - \mu A_i(e)$. At a policy regular for that instrument, its cost exposure equals $1/(1+\gamma)$ times the derivative of gross receipts.

Proposition 12 (Learning the optimal policy)**.** *Maintain the assumptions of this section and the welfare-weight restriction in Assumption 2(ii). Let the rate class be a fixed-bracket class and $\mu \geq w_-$. Measure each instrument in fixed units and each net value per unit of its instrument, and let $|\cdot|$ be the Euclidean norm in those units. Start from an admissible policy π_0. For $k \geq 0$,*

let g_k be the vector of net values of the fee, the bracket rates and the enforcement instruments at π_k, computed from one common measurable selection of firms' minimizing strategies at that policy. Stop if $\operatorname{clip}(\pi_k + g_k) = \pi_k$, *and otherwise set*

$$\pi_{k+1} = \operatorname{clip}\big(\pi_k + \eta_k g_k/|g_k|\big), \qquad \eta_k = \eta/\sqrt{k+1}, \tag{56}$$

where $\eta > 0$ and clip *moves each instrument to the nearest point of its interval.*

(i) *Each g_k is a supergradient of $\mathcal{L}_\mu$ at π_k. If the rule stops at π_k, then π_k maximizes $\mathcal{L}_\mu$ over admissible policies.*

(ii) *Let $\mathcal{L}_\mu^*$ be the maximum of $\mathcal{L}_\mu$ over admissible policies, let D be the diameter of the set of admissible policies, and let B bound $|g_k|$ for every k. The best of $\pi_0, \dots, \pi_k$ falls short of $\mathcal{L}_\mu^*$ by at most*

$$B\,\frac{D^2 + \eta^2\big(1 + \log(k+1)\big)}{4\eta\big(\sqrt{k+2} - 1\big)},$$

which vanishes as k grows.

(iii) *If $\mathcal{L}_\mu$ has a unique maximizer, the best of $\pi_0, \dots, \pi_k$ converges to it. If μ also supports the optimum, that maximizer is the optimal policy.*

(Proof in Appendix K.) The rule allows ties because any common selection gives a valid supergradient. Empirical implementation requires that selection's exposures. With equal weights and a common exponent, counts and regular enforcement responses recover the scores. At an enforcement kink, however, the right receipt derivative uses the minimizing strategies with the smallest exposure to that instrument. Combining it with counts from a different selection need not yield a supergradient. Implementation must therefore measure a common selection's exposures or use policies regular for the relevant instruments. Heterogeneous exponents or welfare weights require the corresponding weighted exposures.

Choosing the best iterate also requires evaluating the Lagrangian. With equal weights and a common exponent, (13) recovers its value differences from receipts, enforcement spending and $\mathbb{E}T(x)$. The multiplier must be determined separately. Revenue among exact Lagrangian maximizers is nondecreasing in μ, which permits bracketing a supporting value (Online Appendix S.6.1). If its maximizer is not unique, one must select a maximizer with $Q = G$. The proposition proves neither convergence of a simultaneous multiplier update nor feasibility of intermediate policies.

Interpreting these iterations as actual reforms requires repeated adjustment and measurement, stable technologies, and no sunk costs carried between periods. Sampling error in measured scores also requires analysis. In the running example, the same concave

support bound certifies the computed policies. Exact integration and a common selection give a welfare shortfall below 10^{-11} for each full-precision policy in Table 1 (Appendix N).

8 Enforcement against underreporting and splitting

Enforcement changes both the level and the form of manipulation. Raising the cost of either response can push firms toward the other. Under the maintained assumptions, however, stronger enforcement at a fixed schedule weakly raises gross receipts whichever way reported bases move.

8.1 Receipts and manipulation costs

The accounting identity (13) gives the revenue effect directly.

Proposition 13 (Enforcement, receipts and private costs)**.** *Maintain Assumption 1 and no manipulation at the top, (6). Fix T and let $e' \geq e$ componentwise, with both enforcement vectors admissible. Put $\Delta^e = P(c;T,e') - P(c;T,e)$. Then $\Delta^e \geq 0$ and*

$$\begin{aligned} \mathbb{E}R(T,e') - \mathbb{E}R(T,e) &= \mathbb{E}[(1+\gamma)\Delta^e] \geq 0, \\ \mathbb{E}D(T,e') - \mathbb{E}D(T,e) &= -\mathbb{E}[\gamma\Delta^e] \leq 0. \end{aligned} \tag{57}$$

Welfare at the fixed tax schedule weakly falls. If the increase in gross receipts is positive, the reduction in private manipulation costs is between $1/2$ and $\bar{\gamma}/(1+\bar{\gamma})$ of that increase.

At an optimum satisfying Theorem 1, suppose small increases and decreases of instrument $i \in \{R,S\}$ are feasible and ck_i is the same at all optimal strategies for almost every firm. Then the marginal response of net revenue satisfies

$$Q_{e_i} = \mathbb{E}[(1+\gamma)ck_i] - A_i(e) = \frac{\mathbb{E}[w(c)ck_i]}{\mu}. \tag{58}$$

It is strictly positive whenever $\mathbb{E}[ck_i] > 0$.

(Proof in Appendix G.) Higher enforcement raises every strategy's cost and hence minimized burdens. The accounting identity then implies higher receipts and lower aggregate private costs. With a common exponent, the private cost saving is exactly $\gamma/(1+\gamma)$ of the receipts increase. Individual firms may still switch to strategies that pay less tax. The aggregate gain includes other firms moving toward compliance. Net receipts subtract public expenditure and need not rise away from the optimum.

At the joint optimum, (58) equates each effective interior instrument's welfare cost per net dollar to the value of public funds. At a policy bound, Theorem 1 gives the corresponding inequality. At a kink, a supporting selection need not describe the response to an increase, which is why the marginal statement requires regularity.

With equal linear welfare weights, write $\beta = 1 - w/\mu$. At a regular interior optimum, a share $(\gamma_i + \beta)/(1 + \gamma_i)$ of an instrument's marginal gross receipts pays for public enforcement. The remaining share $(1 - \beta)/(1 + \gamma_i)$ is retained as net receipts. Here γ_i is the average exponent among the firms reached by the instrument, weighted by their cost exposure ck_i (Online Appendix S.7.1). Every effective interior instrument therefore has resource cost $\beta/(1 - \beta)$ per additional net dollar, the same ratio as an interior marginal tax rate. This equality of the marginal cost of funds across instruments also holds under Assumption 3, extending Mayshar (1991); Slemrod and Yitzhaki (1996) to entity bases. With general welfare weights, (58) weights the firms reached differently.

8.2 Two motives for splitting

The response of underreporting depends on why firms split. Splitting can save taxes on a given total report, or it can make underreporting cheaper. The two specializations below separate these motives and exclude exit. Let h be the total underreported base and $q = x - h$ the total report. With n entities, define $T_n(q)$ as the smallest group tax payment at total report q. Let $C_n(h)$ be the smallest underreporting effort at total underreported base h. In the technologies below these minima can be attained together. If an additional entity costs effort $\sigma + e_S$, its gain at a given h is

$$[T_n(x - h) - T_{n+1}(x - h)] + c[C_n(h) - C_{n+1}(h)] - c(\sigma + e_S). \tag{59}$$

The first term rewards reorganizing reports, and the second rewards reorganizing underreporting.

Under the *tax motive*, the firm divides its potential tax base among its entities. Holdings lie in a feasible set of nonnegative vectors that sum to x, and each entity reports at most its holdings. Underreporting requires effort $(p + e_R)h$, with $p > 0$, regardless of the number of entities. Each additional entity requires effort $\sigma + e_S$. This is the benchmark technology of Section 2.1. Online Appendix S.12.1 gives a closed-form splitting test for this technology, exact when marginal rates are nondecreasing. Divisions are unrestricted if every such vector of holdings is feasible. Only the first saving in (59) remains. With nondecreasing marginal rates, that saving grows with the total report under either unrestricted divisions or a single feasible division into two entities. Splitting therefore raises the return to

reporting relative to underreporting.

Under the *cost motive*, each entity incurs a nonnegative convex effort $\psi(\eta_j)$ when it underreports η_j, with $\psi(0) = 0$. Underreporting also requires group effort ph with $p > 0$, and an additional entity costs $\sigma + e_S$. This is the *convex-cost technology*. For unrestricted divisions, spreading underreporting equally minimizes its cost, giving $C_n(h) = ph + n\psi(h/n)$. This falls with n, and the saving grows with h. The assumptions $p \geq L$ and $\sigma \geq LX/2$ ensure no manipulation at the top. Under linear taxation, the tax payment at a given total report changes only through the entity fee, so splitting is attractive only because it reduces the cost of underreporting.

Because strategies need not be unique, the following statement uses each firm's largest and smallest optimal underreported bases and entity counts.

Proposition 14 (Substitutes and complements)**.** *Fix a firm's primitives and cost parameter c, and let T be a schedule in the class (2) with fee τ.*

(i) (The tax motive.) *Suppose the firm has the benchmark technology, with either unrestricted divisions or $m = 2$ and a single feasible division, and let the marginal rate of T be nondecreasing. Then every optimal strategy with two or more entities underreports no more than the largest optimal underreported base among the one-entity strategies. The largest and smallest optimal underreported bases are nondecreasing, and the largest and smallest optimal entity counts nonincreasing, in the fee and in enforcement against splitting. Enforcement against underreporting has the opposite effects.*

(ii) (The cost motive.) *Suppose the firm has the convex-cost technology, with either unrestricted divisions or $m = 2$ and a single feasible division, and let T be linear. Then every optimal strategy with two or more entities underreports at least the smallest optimal underreported base among the one-entity strategies. The largest and smallest optimal underreported bases and entity counts are all nonincreasing in the fee and in enforcement against splitting.*

(Proof in Appendix L.) A more expensive entity increases underreporting under the tax motive and reduces it under the cost motive. Proposition S.7.1 in Online Appendix S.7.2 treats the case in which both motives operate. It allows taxes and costs to be jointly minimized over whole plans. When the two minima are jointly attainable, the criterion compares the marginal tax saving from an extra entity with its marginal saving in underreporting costs. Corollary S.7.1 then signs responses to cheaper splitting enforcement at fixed taxes and a fixed value of public funds. The response after all policy instruments are reoptimized need not have the same sign.

8.3 Reports and receipts

The reported base is $\mathbb{E}\sum_j r_j$, and the unreported base is $\mathbb{E}x - \mathbb{E}\sum_j r_j$. Let economy 1 permit only one-entity strategies and economy 2 also permit splitting. Write $\Delta = P^1 - P^2 \geq 0$ for the reduction in burden when splitting is added at the same policy. The next result contrasts the reporting responses under the two motives. In both cases, adding splitting weakly lowers revenue by Theorem 2(i).

Corollary 7 (Reported bases, revenue and enforcement)**.** *Maintain Assumption 1 and no manipulation at the top, (6).*

(i) *Let firms have the benchmark technology as in Proposition 14(i), and let the policy's schedule have a nondecreasing marginal rate. Compared with the economy in which only one-entity strategies are available, the option to split weakly raises the reported base, for any selections of optimal strategies in the two economies. It also lowers revenue by $\mathbb{E}[(1+\gamma)\Delta]$ and raises manipulation costs by $\mathbb{E}[\gamma\Delta]$. Raising the fee or enforcement against splitting weakly lowers the number of entities for any selections. It weakly raises the unreported base under the selections that choose each firm's largest, or each firm's smallest, optimal underreported base. Raising enforcement against underreporting weakly lowers the unreported base for any selections. It weakly raises the number of entities under the selections that choose each firm's largest, or each firm's smallest, optimal entity count.*

(ii) *Let firms have the convex-cost technology as in Proposition 14(ii) and the schedule be linear. Then the option to split weakly lowers the reported base for any selections. Raising the fee or enforcement against splitting weakly lowers the number of entities for any selections and weakly lowers the unreported base under the selections that choose each firm's largest, or smallest, optimal underreported base.*

(Proof in Appendix L.) Under the tax motive, access to splitting raises total reports but lowers revenue because new entities receive the lower portions of the schedule. Conversely, enforcement against splitting can lower reports while raising receipts. Net fiscal gains also subtract public expenditure. At a regular interior joint optimum they remain positive at the margin by (58).

The running example illustrates the tax motive after the bottom rate is cut by 0.15 from its optimum with splitting and zero enforcement (Section 5.3). Raising enforcement against splitting from zero to 0.1 at this reformed schedule lowers total reports by about 2.66×10^{-3}, yet raises gross receipts by about 5.49×10^{-3}. This increase is twice the average increase in burden, and private manipulation costs fall by half that amount. Public expenditure rises by 10^{-4}, so net receipts also increase. Raising underreporting enforcement by 0.1 instead raises total reports by about 3.36×10^{-3} and slightly increases the split share. These

comparisons hold the tax schedule fixed.

Example 2 gives the opposite reporting response through a different cost technology. Its effort depends on current reports rather than only on the amount hidden, so Proposition 14(ii) does not apply directly. Its explicit solution gives total reporting $1 - t/B(e)$ under a flat rate t, where $B(e)$ is the least effort of full concealment. Anti-splitting enforcement raises $B(e)$ while splitting remains optimal and therefore raises total reporting. Gross receipts rise and aggregate private costs fall, as (57) requires.

A proportional tax change leaves the relative costs of underreporting and splitting unchanged. Conditional on a firm's primitives, a proportional cut by a factor λ scales the shares using each form of manipulation by λ^γ, under the tie-breaking rule of Section 3.3. It therefore preserves the mix of responses within that characteristic group. With heterogeneous exponents, groups contract at different rates and the aggregate mix can change. Section 9 uses this scaling property to measure private manipulation costs.

9 Measuring the costs of manipulation

At fixed enforcement, the behavioral change in tax payments measures the change in private manipulation costs with the opposite sign. The envelope theorem gives this result whenever the mechanical tax change is the same at all optimal strategies. For a rate increase on a short interval near r, private costs rise by $M_r \varepsilon_r$ per unit of tax added above the interval, and the behavioral revenue response is $-M_r \varepsilon_r$. This extends the logic of Feldstein (1999) to payments summed across a firm's entities. Total reports alone do not suffice. They rise after the relief reform in Section 5.3 even though the revenue loss exceeds its mechanical estimate. They can also fall when enforcement raises receipts. Private manipulation costs here are real losses, whereas in Chetty (2009) they partly consist of transfers.

Proportional reforms can also identify the level of private costs. These reforms scale the whole schedule, including its fee. Empirical identification requires causal tax variation, unchanged technology and enforcement, and a common response horizon. The population is held fixed and includes firms that exit. Firm-level statistics also require ownership links; aggregate payments suffice for reforms covering the whole population.

Hold enforcement and an admissible original schedule T fixed. For a scale factor $\lambda \in (0, 1]$, let $a(\lambda)$ be the firm's chosen strategy under λT and define

$$Z(\lambda) = \mathbb{E}\, T[a(\lambda)].$$

This *tax-payment curve* values the reports chosen after each reform at the original schedule.

Actual revenue under the reform is $\lambda Z(\lambda)$. Write $\mathbb{E}_\lambda$ for outcomes under λT.

The area between this curve and $Z(1)$ is total private manipulation cost. Observing the full reform path identifies the area without a power law. A finite grid gives bounds. Under a common power-law exponent, one reform identifies costs if a responsive statistic identifies the exponent; otherwise, two reforms identify both from aggregate payments.

Proposition 15 (Identification from proportional reforms)**.** *Suppose that Assumption 3(i)–(ii) and Assumption 2(i) hold. For parts (i) and (ii), assume only that* $\Pr(c = 0) = 0$.

(i) *The function Z is nonincreasing, bounded above by $\mathbb{E}T(x)$, and converges to $\mathbb{E}T(x)$ as $\lambda \downarrow 0$. Expected burdens and manipulation costs per firm at the original schedule are*

$$\mathbb{E}P = \int_0^1 Z(\lambda)\,\mathrm{d}\lambda, \qquad \mathbb{E}D = \int_0^1 Z(\lambda)\,\mathrm{d}\lambda - Z(1). \tag{60}$$

(ii) *For $\bar{n} \geq 2$ scale factors $0 < \lambda_1 < \cdots < \lambda_{\bar{n}} = 1$ and a known bound $\bar{B} \geq \mathbb{E}T(x)$, let*

$$\underline{D} = \lambda_1 Z(\lambda_1) + \sum_{k=2}^{\bar{n}} (\lambda_k - \lambda_{k-1}) Z(\lambda_k) - Z(1),$$
$$\overline{D} = \lambda_1 \bar{B} + \sum_{k=2}^{\bar{n}} (\lambda_k - \lambda_{k-1}) Z(\lambda_{k-1}) - Z(1).$$

Then $\underline{D} \leq \mathbb{E}D \leq \overline{D}$, and $\overline{D} - \underline{D} \leq \lambda_1 \big[\bar{B} - Z(\lambda_1)\big] + \max_{2 \leq k \leq \bar{n}} (\lambda_k - \lambda_{k-1}) \big[Z(\lambda_1) - Z(1)\big]$.

(iii) *Suppose that cost parameters follow a power law with a common exponent γ and that firms follow the tie-breaking rule of Section 3.3. For every bounded statistic $Y(\alpha)$ that may depend on the firm's other primitives, is zero at the no-manipulation strategy, and has no explicit dependence on c or λ, $\mathbb{E}_\lambda Y = \lambda^\gamma \mathbb{E}_1 Y$. Consequently, for every $\lambda \in (0, 1)$,*

$$Z(\lambda) = \lambda^\gamma Z(1) + (1 - \lambda^\gamma)\,\mathbb{E}T(x), \qquad \mathbb{E}D = \frac{\gamma}{1+\gamma} \cdot \frac{Z(\lambda) - Z(1)}{1 - \lambda^\gamma}. \tag{61}$$

One reform $\lambda < 1$ identifies $\gamma = \log(\mathbb{E}_\lambda Y / \mathbb{E}_1 Y) / \log \lambda$ from any observed statistic of this kind with $\mathbb{E}_1 Y \neq 0$, such as the share of firms that split or the share that exit. Without such a statistic, two reforms $\lambda_1 < \lambda_2 < 1$ identify γ whenever $Z(\lambda_2) \neq Z(1)$, because the ratio $[Z(\lambda_1) - Z(1)] / [Z(\lambda_2) - Z(1)] = (1 - \lambda_1^\gamma)/(1 - \lambda_2^\gamma)$ is strictly decreasing in γ.

(Proof in Appendix M.) A firm's burden along the reform path is the lower envelope of $\lambda T[\alpha] + cK(\alpha; e)$. Its slope in λ is the tax payment evaluated at the original schedule. At $\lambda = 0$, the firm pays neither taxes nor manipulation costs, fixing the constant of integration in (60).

The finite-reform bounds are lower and upper sums for Z and converge as the mesh and smallest scale factor vanish. A single local reform does not identify the area without additional restrictions. Under the power law, the change of variable $c = \lambda v$ gives the exact scaling in part (iii). It also identifies the tax gap as $[Z(\lambda) - Z(1)]/(1 - \lambda^{\gamma})$. The denominator must be bounded away from zero for a stable finite-sample estimate.

When exponents differ across characteristic groups, a proportional cut scales each group's manipulation by its own factor λ^{γ}. A positive aggregate of nonnegative responses is then a mixture. Its elasticity is an average of the exponents, and its log curvature is their variance. A nonzero statistic in each group identifies its exponent, after which the groups' manipulation costs can be added. An aggregate share ratio alone identifies neither a common exponent nor total costs (Proposition M.1 in Appendix M).

In the running example at its optimum with splitting and zero enforcement, uniform cost parameters imply $\gamma = 1$ and a linear tax-payment curve. It falls from about 0.18, the taxes paid without manipulation, to 0.15 at $\lambda = 1$. The area gives private manipulation costs of about 1.42×10^{-2} per firm, half the tax gap. Halving the schedule halves the split share from 2.57 to 1.28 percent and the share manipulating in any way from 27.85 to 13.92 percent. Entities per firm fall from about 1.05 to 1.03 (Online Appendix S.8.1).

Entity records alone show about 2.57×10^{-2} entities disappearing per original firm in this example. Linking entities keeps that organizational change and the increased reports of the remaining firms in a common population.

Under independent random assignment across firms, sample means of the bounded outcomes are consistent and asymptotically normal. Online Appendix S.8 gives estimators and inference for the finite-reform bounds, which remain useful when reforms near zero are infeasible.

The elasticity in the Ramsey rule can also be estimated from a proportional reform. In the sufficient-statistic approach of Saez (2001), behavioral elasticity estimates enter optimal tax conditions. Here the relevant elasticity is that of the entity base above a threshold. Compare groups facing the original schedule T and a proportionally lower schedule λT. All entities controlled by a treated firm must face the same scale. Let $\widehat{M}_r(1)$ and $\widehat{M}_r(\lambda)$ estimate the entity base above r in the two groups, counting entities per original firm. The samples must represent comparable firm populations, count newly created entities and assign zero entities to firms that exit. The estimator is

$$\hat{\varepsilon}_{r,\lambda} = \frac{\widehat{M}_r(\lambda) - \widehat{M}_r(1)}{(1 - \lambda)\widehat{M}_r(1)}. \tag{62}$$

It divides the proportional change in the entity base by the proportional tax cut. For a

reform covering the whole population, aggregate entity counts suffice because the fixed population denominator cancels. Ownership links are needed to assign firm-specific reforms or construct firm-level and characteristic-group statistics.

The estimator is consistent for ε_r as $\lambda \uparrow 1$ and the samples grow under three conditions. The base must be positive. Firms must follow the tie-breaking rule, unless entity counts are the same at all optimal strategies. Finally, sampling errors divided by $1-\lambda$ must vanish in probability. For independent samples of firms of comparable size N, bounded entity counts give sampling errors of order $N^{-1/2}$, so $\sqrt{N}(1-\lambda) \to \infty$ suffices. Inference must respect the unit at which the reform is assigned. The left derivative in (20) exists under these conditions (Appendix G); no further smoothness of firms' choices is needed. With a known or separately identified common exponent, (19) gives the finite-reform estimator

$$\hat{\varepsilon}^{\mathrm{PL}}_{r,\lambda} = \frac{\gamma}{1-\lambda^{\gamma}} \frac{\widehat{M}_r(\lambda) - \widehat{M}_r(1)}{\widehat{M}_r(1)}. \tag{63}$$

Its population counterpart equals ε_r for any $\lambda \in (0,1)$, and the sample estimator is consistent as the samples grow.

When exponents differ across groups, the finite-reform correction must be applied within each group before aggregating. An average exponent cannot replace it (Online Appendix S.8). The local estimator (62) remains available without a common exponent. At an interior rate, the estimated elasticity identifies the ratio w_r/μ in the Ramsey rule. Separating welfare weights from μ requires additional information about the welfare function and the revenue requirement.

The bunching estimator of Saez (2010) infers a compensated reported-income elasticity from excess mass at a kink under restrictions on the counterfactual income distribution. Here the response is to the common scale of the whole schedule. Splitting can generate several reports per owner, so excess mass in the entity distribution cannot generally be mapped into movements of a fixed population of taxpayers. The reform estimator includes organizational changes directly and measures the response required by the Ramsey rule.

The power law also ties the gap between potential and reported bases to optimal rates. With equal weights, a common exponent and interior rates, the unreported base at an optimum is β/γ times the reported base (Corollary 3). By Proposition 15(iii), γ is also the elasticity of the unreported base with respect to the scale of taxes. This gives an inverse-elasticity rule for the unreported share of potential bases. Keen and Slemrod (2017) derive a rule of this form from optimal enforcement, with the elasticity of evasion with respect to enforcement in its denominator. Here the rule follows from the choice of tax rates. The same reform provides estimates of γ and the unreported base. The observed

unreported share then reveals the only value of β at which the current schedule could be optimal. The implied marginal cost of public funds, $\mu/w = 1/(1-\beta)$, can be compared with independent estimates.

Nonproportional reforms still bound changes in manipulation costs, by the revealed preference of firms' choices before and after the reform (Appendix M).

10 Scope

The main model uses a bounded conditional power law for private costs. Firms can differ in their exponents, manipulation technologies and welfare weights. Exit is allowed when its loss is scaled by the same cost parameter. Costs can depend on the whole sequence of underreports and splits. Table S.1 in Online Appendix S.9 records which conclusions survive without the power law.

Two properties produce the exact formulas. First, one cost parameter scales every manipulation cost. A firm facing proportionally lower taxes therefore chooses as a firm with proportionally higher costs would. This holds for any cost distribution. Second, the power law has the same form at every scale. A proportional tax cut then shrinks every form of manipulation within a characteristic group by the same factor. The elasticity $\ell(c) = (cf(c))'/f(c)$ of $cf(c)$ equals the group's exponent γ and is constant in c.

This elasticity weights burdens in the accounting identities. Appendix F allows it to vary with both characteristics and the cost parameter, subject to a restriction on its size and a boundary condition on the conditional density. The global characterization, Ramsey rule, welfare comparison, and signs of the enforcement and relief results extend to these densities. The entity and report targets, the fixed fraction of the tax gap spent on manipulation, the coverage formulas and identification from one reform use the power law.

Sequential structure restricts an individual firm's choices (Section 4). An optimal plan cannot gain from an elementary continuation. When reduction costs depend only on the current and final report and add across successive reductions, the firm's problem admits a recursive characterization and an elementary test for deterring all manipulation. These properties simplify behavioral analysis and computation. They do not by themselves strengthen the general welfare ranking or imply the rate shapes of Section 5, which require further restrictions.

Each fixed strategy's cost is affine in enforcement. Separate private cost parameters for underreporting and splitting, or nonlinear effects of enforcement on a fixed strategy, would require a different characterization. No manipulation at the top must hold throughout the

policy class. A smooth interior choice of untaxed labor generally has a marginal opportunity cost that starts at zero and therefore fails this condition. Section 2.3 distinguishes exact nesting of labor supply from matching responses alone. Matching responses does not establish welfare equivalence.

An arbitrary outside opportunity can add a correction to the revenue identity and destroy concavity, even under uniform costs (Online Appendix S.15). Remitted fines and recovered taxes require different accounting from real manipulation costs. Online Appendix S.14 gives a restricted extension in which the authority receives a share of those costs. General audit and penalty schemes remain outside the policy class.

Adding tax brackets weakly raises attainable welfare when the new grid retains the old thresholds. As the grid becomes arbitrarily fine, optimal welfare approaches the unrestricted value (Proposition S.13.1 in the Online Appendix). If the optimal marginal rate has bounded variation and an affine policy is strictly feasible, the welfare loss from using N freely placed brackets is $O(N^{-2})$. Online Appendix S.13 gives this bound and conditions under which its order is sharp. This connects the analysis to piecewise-linear taxation (Sheshinski, 1989; Slemrod et al., 1994) and approximation by finite menus (Wilson, 1989; Bergemann et al., 2021).

Aggregate uncertainty is also allowed if the authority commits to one policy before the aggregate state is realized and must meet the revenue requirement in expectation. Theorem 1 then applies with expectations over states and firms (Online Appendix S.9.1).

11 Conclusion

When taxes attach to entities and welfare attaches to owners, the relevant tax base counts entities at their owners' welfare weights. The optimal schedule satisfies a Ramsey rule on this base. With equal weights and a common cost exponent, the rule fixes coverage ratios at interior marginal rates and makes those ratios informative about the direction of reform away from an optimum.

The threat of splitting can matter even when few firms split. Under the conditions of Section 5, top rates equal the rates on the entities' reports in a tax-minimizing split. A progressive optimum must then have a plateau. Under additional restrictions, access to splitting raises the rates paid by smaller firms. The running example exhibits a modest aggregate welfare loss alongside a substantial redistribution of burdens. These shape and incidence results complement the welfare bounds for general manipulation technologies. In the sequential example, splitting lowers concealment costs even under a linear schedule. The optimal flat rate rises when that option is added.

Tax payments also measure resource costs that reported bases alone miss. Relief can raise total reports while losing more revenue than its mechanical estimate. Enforcement can raise receipts while reports fall. Proportional reforms recover private manipulation costs from tax payments, with exact finite-reform formulas under the power law. Ownership links and audits of potential bases would improve both measurement and policy design. Their costs, persistent organizational choices, and report-dependent audits or penalties remain outside the model.

References

Alger, I. (1999). Consumer strategies limiting the monopolist's power: Multiple and joint purchases. *RAND Journal of Economics*, 30(4):736–757.

Almunia, M. and Lopez-Rodriguez, D. (2018). Under the radar: The effects of monitoring firms on tax compliance. *American Economic Journal: Economic Policy*, 10(1):1–38.

Bergemann, D., Yeh, E., and Zhang, J. (2021). Nonlinear pricing with finite information. *Games and Economic Behavior*, 130:62–84.

Boyd, S. (2014). Subgradient methods. Notes for EE364b, Stanford University, May 2014, with help from Jaehyun Park. https://stanford.edu/class/ee364b/lectures/subgrad_method_notes.pdf.

Conselho Administrativo de Recursos Fiscais (CARF) (2018). Acórdão 1402-003.006, House Confecções Ltda. e outros. Process 15586.720522/2012-15, decided 10 April 2018.

Carrillo, P., Pomeranz, D., and Singhal, M. (2017). Dodging the taxman: Firm misreporting and limits to tax enforcement. *American Economic Journal: Applied Economics*, 9(2):144–164.

Casamatta, G. (2021). Optimal income taxation with tax avoidance. *Journal of Public Economic Theory*, 23(3):534–550.

Casamatta, G. (2023). Optimal income taxation with tax avoidance and endogenous labour supply. *Canadian Journal of Economics*, 56(3):913–939.

Chander, P. and Wilde, L. L. (1998). A general characterization of optimal income tax enforcement. *Review of Economic Studies*, 65(1):165–183.

Chetty, R. (2009). Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance. *American Economic Journal: Economic Policy*, 1(2):31–52.

Dabla-Norris, E., Jaramillo Mayor, L., Lima, F., and Sollaci, A. (2018). Size dependent policies, informality and misallocation. IMF Working Paper 18/179, International Monetary Fund. https://www.imf.org/-/media/files/publications/wp/2018/wp18179.pdf.

Dharmapala, D., Slemrod, J., and Wilson, J. D. (2011). Tax policy and the missing middle: Optimal tax remittance with firm-level administrative costs. *Journal of Public Economics*, 95(9–10):1036–1047.

Diamond, P. A. (1975). A many-person Ramsey tax rule. *Journal of Public Economics*, 4(4):335–342.

Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics*, 81(4):674–680.

Garicano, L., Lelarge, C., and Van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from France. *American Economic Review*, 106(11):3439–3479.

Goldman, M. B., Leland, H. E., and Sibley, D. S. (1984). Optimal nonuniform prices. *Review of Economic Studies*, 51(2):305–319.

Grochulski, B. (2007). Optimal nonlinear income taxation with costly tax avoidance. *Federal Reserve Bank of Richmond Economic Quarterly*, 93(1):77–109.

Ju, B.-G. and Moreno-Ternero, J. D. (2011). Progressive and merging-proof taxation. *International Journal of Game Theory*, 40(1):43–62.

Kanbur, R. and Keen, M. (2014). Thresholds, informality, and partitions of compliance. *International Tax and Public Finance*, 21(4):536–559.

Kaplow, L. (1990). Optimal taxation with costly enforcement and evasion. *Journal of Public Economics*, 43(2):221–236.

Keen, M. and Mintz, J. (2004). The optimal threshold for a value-added tax. *Journal of Public Economics*, 88(3–4):559–576.

Keen, M. and Slemrod, J. (2017). Optimal tax administration. *Journal of Public Economics*, 152:133–142.

Kopczuk, W. (2001). Redistribution when avoidance behavior is heterogeneous. *Journal of Public Economics*, 81(1):51–71.

Kuchumova, Y., Ozhegova, A., and Odud, K. (2026). Beyond the notch: Revenue manipulation and business splitting under Simplified Tax System. Working paper, SSRN 6108727. https://doi.org/10.2139/ssrn.6108727.

Lacker, J. M. and Weinberg, J. A. (1989). Optimal contracts under costly state falsification. *Journal of Political Economy*, 97(6):1345–1363.

Liu, L., Lockwood, B., Almunia, M., and Tam, E. H. F. (2021). VAT notches, voluntary registration, and bunching: Theory and UK evidence. *Review of Economics and Statistics*, 103(1):151–164.

Malik, A. S. (1990). Avoidance, screening and optimum enforcement. *RAND Journal of Economics*, 21(3):341–353.

Massenz, G. (2025). Tax-motivated firm splitting. IFN Working Paper 1539, Research Institute of Industrial Economics, Stockholm. https://www.ifn.se/wfiles/wp/wp1539.pdf.

Mayshar, J. (1991). Taxation with costly administration. *Scandinavian Journal of Economics*, 93(1):75–88.

Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601.

Milgrom, P. and Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62(1):157–180.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38(2):175–208.

Onji, K. (2009). The response of firms to eligibility thresholds: Evidence from the Japanese value-added tax. *Journal of Public Economics*, 93(5–6):766–775.

Perez-Richet, E. and Skreta, V. (2022). Test design under falsification. *Econometrica*, 90(3):1109–1142.

Perez-Richet, E. and Skreta, V. (2026). Falsification-proof non-market allocation mechanisms. Working paper, April 2.

Polyak, B. T. (1987). *Introduction to Optimization*. Optimization Software, New York.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review*, 105(8):2539–2569.

Ramsey, F. P. (1927). A contribution to the theory of taxation. *Economic Journal*, 37(145):47–61.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1):205–229.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.

Saez, E., Slemrod, J., and Giertz, S. H. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, 50(1):3–50.

Sharma, R. R., Slemrod, J., Stimmelmayr, M., Wilson, J. D., and Choi, P. (2025). Optimal dual-regime business tax systems. NBER Working Paper 34418, National Bureau of Economic Research.

Sheshinski, E. (1989). Note on the shape of the optimum income tax schedule. *Journal of Public Economics*, 40(2):201–215.

Slemrod, J. (1995). Income creation or income shifting? Behavioral responses to the Tax Reform Act of 1986. *American Economic Review*, 85(2):175–180.

Slemrod, J. and Kopczuk, W. (2002). The optimal elasticity of taxable income. *Journal of Public Economics*, 84(1):91–112.

Slemrod, J. and Yitzhaki, S. (1996). The costs of taxation and the marginal efficiency cost of funds. *IMF Staff Papers*, 43(1):172–198.

Slemrod, J. and Yitzhaki, S. (2002). Tax avoidance, evasion, and administration. In Auerbach, A. J. and Feldstein, M., editors, *Handbook of Public Economics*, volume 3, pages 1423–1470. Elsevier, Amsterdam.

Slemrod, J., Yitzhaki, S., Mayshar, J., and Lundholm, M. (1994). The optimal two-bracket linear income tax. *Journal of Public Economics*, 53(2):269–290.

Tazhitdinova, A. (2020). Are changes of organizational form costly? Income shifting and business entry responses to taxes. *Journal of Public Economics*, 186:104187.

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations Research*, 26(2):305–321.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2):265–293.

Wei, F. and Wen, J.-F. (2019). The optimal turnover threshold and tax rate for SMEs. IMF Working Paper 19/98, International Monetary Fund.

Wei, F. and Wen, J.-F. (2023). Designing a presumptive income tax based on turnover in countries with large informal sectors. IMF Working Paper 23/267, International Monetary Fund.

Wilson, R. (1989). Efficient and competitive rationing. *Econometrica*, 57(1):1–40.

Wilson, R. (1993). *Nonlinear Pricing*. Oxford University Press, New York.

Aliprantis, C. D. and Border, K. C. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, Berlin, 3rd edition.

Barbu, V. and Precupanu, T. (2012). *Convexity and Optimization in Banach Spaces*. Springer, Dordrecht, 4th edition.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*. Wiley, New York.

Appendix

A Notation

Notation	Meaning
s, H	Firm characteristics and their population distribution
$x(s), c, \gamma_s$	Potential base, private cost parameter and conditional cost exponent
$T, t, \tau;\ e, A$	Tax schedule, marginal rate and fee; enforcement and its public cost
$\mathcal{T}, \mathcal{T}_{\mathcal{P}^N}$	Full schedule class and fixed-bracket schedule class
$\alpha \in \mathcal{A}(s)$	A feasible manipulation strategy for characteristics s
$K;\ P = R + D$	Manipulation effort; burden equals taxes plus private manipulation costs
W, Q, G	Welfare, net revenue and the revenue requirement
$n_r;\ M_r = \mathbb{E}n_r$	Entities of one firm reporting above r; their population total
$w(c), w_r$	A firm's marginal welfare weight; its average over the entity base above r
$\mu;\ \Omega(c)$	Value of public funds; Lagrangian value of an additional unit of burden
Ξ_r, Ξ_τ, g_i	Lagrangian coefficients for a marginal rate, the fee and enforcement i
ε_r, C_r	Entity-base elasticity under a proportional tax cut; coverage ratio
$\beta = 1 - w/\mu$	Common elasticity target when welfare weights equal w

Table 2: Notation used in the model and optimality conditions. Firm and policy arguments are suppressed when understood.

B General setting and model details

This appendix allows general conditional cost densities, compact convex policy classes and I enforcement instruments. The main model follows by taking $f_s(c) = \gamma_s c^{\gamma_s - 1}$ and $I = 2$, with coordinates R, S. In this general setting and its proofs, $\mathcal{T}$ denotes the schedule class specified in Assumption (ii) below. The full class in (2) and each fixed-bracket class $\mathcal{T}_{\mathcal{P}^N}$ satisfy these requirements.

B.1 Regularity and the general setting

Retain the characteristics s and their distribution H from Section 2.1. Thus $x = x(s)$, $\mathcal{A} = \mathcal{A}(s)$, $\gamma = \gamma_s$ and $U = U_s$. The index records these primitives and introduces no additional choice or private information. Distinct types may have the same potential base.

Write $P_s(c)$ and $w_s(c)$ for burden and marginal welfare weights. Expectations integrate over the distribution H of s and the conditional distribution of c.

The following conditions extend Assumption 1 to general conditional cost densities and policy classes. Under the main power law, part (iii) below is replaced by continuity of $s \mapsto \gamma_s \in [1, \bar{\gamma}]$. The remaining economic restrictions are in Assumption 2.

Assumption 3 (General-density extension)**.** Put $\bar{T} = \sup_{T \in \mathcal{T}} \|T\|_\infty$.

(i) Characteristics have Borel distribution H on compact metric $\mathcal{S}$, and $x : \mathcal{S} \to [0, X]$ is continuous. Strategies lie in $\mathcal{Z} = \bigsqcup_{n=0}^{\infty}(\mathbb{R}_+^n \times \mathbb{R}_+^{1+I})$, with Euclidean topology within each component and distinct components separated. The correspondence $\mathcal{A} : \mathcal{S} \rightrightarrows \mathcal{Z}$ satisfies:

(a) Every $\mathcal{A}(s)$ contains the compliance strategy

$$\alpha^0 = ((x(s)), 0), \tag{64}$$

which reports $x(s)$ through one entity at zero effort.

(b) Reports satisfy $\sum_j r_j \leq x(s)$. All effort coefficients are nonnegative, and $K = k_0 + \sum_i e_i k_i$. Exit has $n = 0$ and cost cK.

(c) Each $\mathcal{A}(s)$ is compact.

(d) The correspondence $s \mapsto \mathcal{A}(s)$ is upper and lower hemicontinuous.

(ii) The admissible schedule class $\mathcal{T}$ consists of nonnegative continuous schedules. It is convex, compact in the uniform norm and contains zero. Enforcement belongs to a compact convex $E \subset \mathbb{R}_+^I$ containing zero. Public cost is nonnegative on E, satisfies $A(0) = 0$, and is convex and continuously differentiable on a neighborhood of E.

(iii) The map $(s, c) \mapsto f_s(c)$ is measurable. Each f_s is a positive, locally absolutely continuous probability density on $(0, 1)$, with $c f_s(c) \to 0$ at zero and a finite limit $f_s(1)$ at one. Suppose $\bar{J} = \int f_s(1)\, \mathrm{d}H(s) < \infty$ and put $\ell_s(c) = (c f_s(c))' / f_s(c)$ where defined.

(iv) Each U_s is continuously differentiable on $[x(s) - \bar{T}, x(s)]$. The map $(s, y) \mapsto U_s(y)$ is measurable, and $\mathbb{E}|U_s(x)| < \infty$. On this interval $0 < w_- \leq U_s' \leq w_+ < \infty$.

(v) Each U_s is concave, and the welfare weights and cost density satisfy

$$w_-(1 + \ell_s(c)) \geq w_+ \qquad \text{almost surely.} \tag{65}$$

(vi) No manipulation at the top, (6), holds for the present schedule class.

(vii) Some policy raises $Q > G > 0$.

The compact-image theorem (Aliprantis and Border, 2006, Lemma 17.8) makes $\mathcal{A}(\mathcal{S})$

compact. It meets finitely many components of $\mathcal{Z}$. Entity counts therefore have a finite bound m, and effort coefficients are uniformly bounded. For continuous ϕ, both $\phi[\alpha] = \sum_j \phi(r_j)$ and $k_i(\alpha)$ are continuous on this compact metric union. A compact representation must retain feasible strategies that attain every burden minimum. Lemma D.1 derives such a representation for sequential technologies.

The main power density gives $\ell_s = f_s(1) = \gamma_s$. The full class in (2) and the fixed-bracket classes $\mathcal{T}_{\mathcal{P}^N}$ are convex and compact by Arzelà–Ascoli. Thus the main assumptions imply the present ones. Compliance and the top boundary give $0 \le P \le T(x) \le \bar{T}$ and $P_s(1) = T(x)$.

For the full rate class, an operating strategy's largest tax saving is $L(x - \max_j r_j)$. Nonnegative taxes give $T[\alpha] \ge T(\max_j r_j)$. The bound is attained by setting the fee and all rates below that report to zero, and rates above it to L. Hence (6) is equivalent to $k_0 \ge L(x - \max_j r_j)$. In the benchmark, $p_s \ge L$ and $\sigma_s \ge Lx/2$ suffice: for $n \ge 2$, $\sum_j r_j - \max_j r_j \le (n-1)x/n \le (n-1)x/2$. For fixed brackets, the supremum instead runs over the restricted class.

C Verification and scope of the one-entity examples

C.1 The underreporting and labor-supply examples

We verify that the two constructions in Section 2.3 satisfy the maintained assumptions. Set $x = X = 1$, let c be uniform, take $U(y) = y$, $G = 0.1$, and allow one entity, continuous reductions $h \in [0, 1]$, and no exit. The fee is zero and $L = 0.5$. For monetary underreporting use

$$K(h; e_R) = (0.6 + e_R)h + 0.1h^2, \quad E = [0, 0.1] \times \{0\}, \quad A(e) = e_R^2/2;$$

for labor supply use $E = \{(0, 0)\}$ and $A \equiv 0$. These are the same cost function at zero enforcement.

The strategy set is the continuous image of $[0, 1]$, hence compact. The other regularity conditions follow from continuity of the primitives. Uniform costs and equal welfare weights satisfy the distribution and weight restrictions. For every admissible schedule,

$$T(1) - T(1 - h) \le 0.5h \le K(h; 0), \tag{66}$$

so the highest-cost type complies. To check strict feasibility, the policy $T(r) = 0.2r$, $e = 0$

gives

$$h(c) = \begin{cases} 1, & c \le 1/4, \\ 1/c - 3, & 1/4 < c < 1/3, \\ 0, & c \ge 1/3. \end{cases}$$

This follows by minimizing $0.2(1-h) + c(0.6h + 0.1h^2)$ on $[0,1]$. Integrating reports gives

$$\mathbb{E}r = 1 - \log(4/3), \qquad Q = \tfrac{1}{5}\{1 - \log(4/3)\} > \tfrac{2}{15} > G. \tag{67}$$

Both examples therefore satisfy the assumptions of the characterization. In the labor interpretation, the opportunity cost cK already includes lost output.

By contrast, smooth preferences $C - v(q)$, wage a, and an interior untaxed optimum q_0 give opportunity cost

$$D(h) = h - \{v(q_0) - v(q_0 - h/a)\}, \qquad D'(0) = 0.$$

Such preferences fail no manipulation at the top if an admissible schedule has a positive marginal rate at untaxed earnings aq_0.

D Sequential manipulation technologies

D.1 Operations and feasible plans

A firm builds its strategy by applying operators to a *state* (y, q, k). Each entity $j = 1, \dots, n$ holds y_j and reports $0 \le q_j \le y_j$. The vector $k \in \mathbb{R}^3_+$ records the coefficients of effort spent so far. The firm starts at $y = q = (x)$ and $k = 0$, which gives the no-manipulation strategy. Let $q_G = \sum_j q_j$, and let $\eta_j = y_j - q_j$ and $\eta_G = x - q_G$ be what entity j and all entities have underreported. The firm can apply two operators to any entity $j \le n$:

- the *underreport* U_j^h, with $0 \le h \le q_j$, lowers q_j by h and raises (k_0, k_R) by $\int_0^h [p_s^E(y_j, \eta_j + v) + p_s^G(\eta_G + v)]\,\mathrm{d}v$;
- the *split* S_j^z, with $0 \le z \le q_j$, lowers y_j and q_j by z, adds an entity that holds and reports z, and raises (k_0, k_S) by $\sigma_s(x, q_G, q_j, z)$.

Here p_s^E, p_s^G and σ_s are jointly continuous in s and their other arguments, with values in $\mathbb{R}^2_+$. A *plan* is a finite sequence of operators. Its strategy is the terminal pair (q, k). The *sequential manipulation technology* $\mathcal{A}(s)$ is the set of strategies of plans with at most m entities. Consecutive reductions with unchanged holdings cost the same as a single reduction of their combined amount. The group component also depends only on total

underreporting. Mixed sequences can have different costs. A split changes holdings, and an underreport changes the reports at which later splits occur. For fixed finite m, continuity of the primitives and $x(s)$ gives properties (a)–(d). If each split also raises k_0 by at least $\sigma_{\min} > 0$, a sufficiently large m leaves every burden unchanged relative to unrestricted finite plans (Lemma D.1).

Effort notation. For a coefficient-valued function $f = (f_0, f_i)$, write $f(\cdot; e) = f_0(\cdot) + e_i f_i(\cdot)$, where $i = R$ for reductions and $i = S$ for splits. The proofs use q_j and q_G for current entity and group reports. No manipulation at the top is a separate restriction on the effort functions.

Lemma D.1 (A compact representation of sequential choices)**.** *In the sequential technology, every plan with n entities yields the same strategy as a plan with the same $n-1$ splits and at most one reduction by each entity in each phase before, between and after the splits. For any fixed finite bound m, the resulting technology has properties (a)–(d) when $x(s)$ and the primitives are continuous.*

If each split raises k_0 by at least $\sigma_{\min} > 0$, let $\bar{K}$ bound the effort of fully underreporting through the parent, uniformly over firms and enforcement, and set $m = \lfloor 1 + \bar{K}/\sigma_{\min} \rfloor$. Restricting unrestricted finite plans to at most m entities then preserves every burden. An exit strategy with coefficients continuous in s preserves compactness and continuity.

Proof. Holdings do not change within a phase. Each entity's underreporting integrals in a phase add up to the integral over its total increase in accumulated underreporting, and the group integrals add up in the same way. Replacing each entity's underreports in the phase by a single underreport of the same total therefore leaves every effort coefficient, and the state at the next split, unchanged. This proves the first statement.

The empty plan yields α^0, giving property (a). Reports never exceed holdings, which sum to x, giving property (b). For properties (c) and (d), fix $n \leq m$ and the entity that makes each split. Express every transfer and underreport in the merged plan as a fraction in $[0, 1]$ of the entity's current report. The resulting states are continuous functions of $x(s)$ and these fractions. Continuity of the effort functions then makes reports and effort coefficients continuous in s and in fractions drawn from a fixed cube. There are finitely many entity counts and assignments of splitting entities, so $\mathcal{A}(s)$ is the image of a finite union of fixed cubes under maps that are jointly continuous in s and the fractions. It is therefore compact and varies continuously with s.

For the bound, let q^* minimize T on $[0, x]$. Taxes are nonnegative, so every plan with at least one entity pays at least $T(q^*)$, and the parent alone reports q^* at an effort of at most $\bar{K}$,

because the integrand of an underreport is nonnegative. A plan with more than m entities has more than $\bar{K}/\sigma_{\min}$ splits. Its effort therefore exceeds $\bar{K}$. For $c > 0$ its burden exceeds that of reporting q^* through the parent, and for $c = 0$ it is no smaller. Every burden is therefore attained by a plan with at most m entities. An exit strategy adds one point to each $\mathcal{A}(s)$, continuous in s, so it preserves both properties. □

D.2 Proofs for the recursive characterization

Throughout this subsection, characteristics and enforcement are fixed, $x = x(s)$, $C(q) = \int_0^q p(v)\,\mathrm{d}v$, and a split costs $\sigma > 0$. Continuation costs depend only on an entity's current report. Plans may contain any finite number of splits. The bound below gives an equivalent compact representation.

Existence and the number of entities. For $c > 0$, reducing the entire report q in one entity costs $T(0) + cC(q)$. Any plan with n terminal entities has taxes at least $T(0)$, nonnegative reduction effort, and split effort $\sigma(n-1)$. If $n - 1 > C(x)/\sigma$, its burden exceeds that of reducing to zero in one entity. Thus every minimum can be found among plans with

$$n \leq \lfloor 1 + C(x)/\sigma \rfloor .$$

Consecutive reductions on an entity between splits add, so Lemma D.1 supplies a compact representation and a minimum. For $c = 0$, every plan pays at least $T(0)$ and one entity can report zero, so $V_0(q) = T(0)$. Thus the bound preserves the unrestricted minimum at every continuation state.

Proof of Proposition 2(i). Stopping, reducing from q to r and then continuing optimally, or splitting q into $z, q - z$ and continuing optimally in each entity gives (34)–(36) for V_c. The split construction is feasible because one entity's continuation costs do not depend on the other's choices. It uses finitely many splits. If the resulting count exceeds the bound above, a plan within the bound has no larger burden.

Conversely, suppose v satisfies these inequalities and take any finite plan from q. Apply (35) at a reduction and (36) at a split. Induction along the plan gives

$$v(q) \leq cK + \sum_j v(r_j) \leq cK + \sum_j T(r_j),$$

where K is the plan's effort and r_j its terminal reports. Minimizing over plans gives $v(q) \leq V_c(q)$, proving the greatest-function statement.

Before the first split, any reductions combine into one reduction from q to some $r \leq q$. If the plan stops there it pays $T(r)$; otherwise it pays $c\sigma$ and starts independent continuation problems at z and $r - z$. An optimal plan has optimal continuations: replacing one changes neither the other nor earlier costs. This gives (37). Continuity, proved below, and compactness of the intervals justify the minima. Conversely, minimizing reduction, split and continuation plans attain the displayed value. The recursive identity may have other fixed points. Part (i) identifies the value function as the greatest function satisfying the inequalities.

Proof of Proposition 2(ii). First prove monotonicity with a bound on the number of splits. Let $V_c^{[N]}(q)$ be the least burden with at most N splits. If f is nondecreasing, then $\mathcal{R}_c f$ is nondecreasing: for $q_1 < q_2$ and any candidate $r \leq q_2$, either $r \leq q_1$, in which case its cost from q_2 is at least its cost from q_1, or $r > q_1$, in which case $f(r) + c[C(q_2) - C(r)] \geq f(q_1) \geq \mathcal{R}_c f(q_1)$. Also, for nondecreasing functions f, g, the function

$$q \longmapsto \min_{0 \leq z \leq q} \{f(z) + g(q - z)\}$$

is nondecreasing. Indeed, any division of a larger q can be reduced componentwise to a division of a smaller q, weakly reducing both terms.

Now $V_c^{[0]} = \mathcal{R}_c T$ is nondecreasing. For $N \geq 1$, a first split allocates at most $N - 1$ remaining splits between the two branches. Thus $V_c^{[N]}$ is $\mathcal{R}_c$ applied to the minimum of T and the finitely many functions

$$q \longmapsto c\sigma + \min_{0 \leq z \leq q} \{V_c^{[a]}(z) + V_c^{[b]}(q - z)\}, \qquad a + b \leq N - 1.$$

Induction proves that every $V_c^{[N]}$ is nondecreasing. Their infimum V_c is nondecreasing as well. Combining this fact with (35), for $0 \leq q_1 < q_2 \leq x$ we have

$$0 \leq V_c(q_2) - V_c(q_1) \leq c \int_{q_1}^{q_2} p(v)\, \mathrm{d}v.$$

Since p is continuous, this proves absolute continuity and (38). At $q = 0$ every entity reports zero and incurs the nonnegative fee; stopping with one entity is optimal, so $V_c(0) = T(0)$.

Proof of Proposition 2(iii). If stopping is optimal at every report, $V_c = T$, so the split inequality follows from part (i) and the marginal inequality from part (ii). Conversely, T is

absolutely continuous. Integrating $t \leq cp$ yields (35) with $v = T$. The binary inequality in (39) is (36). Part (i) then gives $T \leq V_c$, and stopping gives the reverse inequality. The equivalence concerns the same primitives at every starting report. It is stronger than optimality of stopping at the single report $x(s)$.

Proof of Proposition 2(iv). Since splitting expands the available plans, $V_c \leq U_c$. Additivity of reduction costs implies

$$U_c(q) \leq c[C(q) - C(r)] + U_c(r) \qquad (0 \leq r \leq q),$$

because an initial reduction to r followed by an optimal reduction from r is feasible without splitting. Also $U_c \leq T$. If (40) holds, U_c therefore satisfies all three inequalities in part (i), giving $U_c \leq V_c$. Conversely, if $U_c = V_c$ at every report, (36) gives (40).

The effort identity. Track the potential $\sum_j C(q_j)$ along a plan. A reduction lowers it by exactly that reduction's effort. A split at (q, z) changes it by $C(q - z) + C(z) - C(q)$. Adding these changes from the initial report x to the terminal reports, then adding σ for each split, gives (41).

Proof of Corollary 4. Operations on different existing entities commute, because their costs and feasibility depend only on their own reports. Consider a reduction from q to $q - h$, followed by a split into a, b, with $a + b = q - h$. Reversing these operations by first splitting into $a + h, b$ and then reducing the first report to a preserves the reports after the two operations and the split cost. It replaces reduction effort

$$\int_{q-h}^{q} p(v)\,\mathrm{d}v \quad \text{by} \quad \int_{a}^{a+h} p(v)\,\mathrm{d}v.$$

Because $a \leq q - h$, the replacement weakly lowers effort when p is nondecreasing. Repeated adjacent exchanges and commutation of operations on different entities move all splits before all reductions. Terminal holdings then form a division $y_1 + \cdots + y_n = q$. Each entity solves its own reduction problem, and the $n - 1$ splits require effort $(n - 1)\sigma$. Minimizing gives $\mathcal{S}_c\mathcal{R}_c T$.

For the reverse order, consider a split from q into a, b, followed by a reduction of the first entity from a to $a - h$. Replace this pair by a reduction from q to $q - h$, then a split into $a - h, b$. It replaces $\int_{a-h}^{a} p$ by $\int_{q-h}^{q} p$, which is weakly smaller when p is nonincreasing since $a \leq q$. The same exchange applies to a reduction in the second entity. Repeated exchanges

move every reduction before every split. Additivity combines the reductions into one in the parent, and minimizing the subsequent division gives $\mathcal{R}_c\mathcal{S}_cT$. These exchanges preserve the reports at the end of each pair, the number of splits and feasibility of all later operations.

Both comparisons are equalities when p is constant. Finally, if T is subadditive, repeated application of its binary inequality gives $T(q) \leq \sum_j T(y_j)$ for every division. Since split costs are nonnegative and $n = 1$ is permitted, $\mathcal{S}_cT = T$. Under nonincreasing p, the second part therefore gives $V_c = \mathcal{R}_cT$.

E The sequential example

This appendix solves the firm's problem in Example 2 over all finite sequences of reductions and splits. It then certifies the joint optimum in the full schedule class and compares it with the optimum without splitting at the same net revenue.

E.1 Assumptions and unrestricted choices

The primitives are $x = 1$, $C_e(q) = (1 + e_R)q + 6q^2$, and split effort $3/4 + e_S$, with $e \in [0,1] \times [0,3]$. They are a special case of the sequential technology: $p^G = 0$, $p^E(y, \eta; e_R) = 1 + e_R + 12(y - \eta)$, and splitting effort is independent of the state. All coefficients are continuous and nonnegative. The distribution of characteristics can be concentrated on a single point; private heterogeneity is in c, uniformly distributed on $[0,1]$.

Full reduction in the parent requires effort at most 8, and every split requires baseline effort $3/4$. Lemma D.1 therefore gives the sufficient bound $m = \lfloor 1 + 8/(3/4) \rfloor = 11$. The bound gives a compact representation without changing any minimized burden.

To verify no manipulation at the top for every schedule with rate cap one and zero fee, let a plan conceal total h and have n terminal entities. Its baseline effort is at least $h + 3(n-1)/4$. If $n \geq 2$,

$$1 - \max_j r_j = h + \sum_j r_j - \max_j r_j \leq h + \frac{n-1}{n} \leq h + \frac{n-1}{2}.$$

For $n = 1$, $1 - r_1 = h$. Baseline effort thus weakly exceeds $1 - \max_j r_j$, the largest tax saving that any admissible schedule permits. Uniform costs and equal welfare weights satisfy the remaining distributional and welfare restrictions. Public cost $A(e) = (e_R^2 + e_S^2)/36$ is continuously differentiable and strictly convex. Strict feasibility is verified below.

Lemma E.1 (Unrestricted choices with a quadratic reduction potential)**.** *Suppose $x = 1$, $T(q) = tq$ with $t \geq 0$, and $C(q) = aq + bq^2/2$, where $a \geq 0$ and $b > 0$. Every split requires effort $\sigma > 0$, and every finite sequence of reductions and splits is available. Define*

$$B = \min_{n \geq 1} \left\{ a + \frac{b}{2n} + \sigma(n-1) \right\}.$$

The minimized burden is $\min\{t, cB\}$. For $c > 0$, some optimum either reports the entire base through one entity or divides it equally among a minimizing number n and reduces every report to zero. No genuinely mixed plan, with a positive amount both reported and concealed, attains the minimum. If the minimizing n and the better of the two burdens are unique, the optimal holdings and reports are unique up to labels.

Proof. We first reduce the choice of plans to a choice of entity counts. Since $p(q) = a + bq$ is increasing, Corollary 4(i) moves all splits before all reductions without raising effort. Conditional on an entity's holdings y, its burden after reductions is

$$\min_{0 \leq r \leq y} \{tr + c[C(y) - C(r)]\}.$$

For $c > 0$, the objective is strictly concave in r, so a minimizing report is 0 or y. Reporting entities can be merged. Their taxes remain t times their holdings, while merging removes positive split costs without changing the concealed entities' costs. A mixed plan therefore needs only one reporting entity and $k \geq 1$ concealed entities. If the latter hold total $h \in (0, 1)$, Jensen's inequality gives reduction effort at least $ah + bh^2/(2k)$, and the k additional entities require effort $k\sigma$. Its burden is consequently at least

$$t(1-h) + c\left(ah + \frac{bh^2}{2k} + k\sigma\right) \geq t(1-h) + ch\left(a + \sqrt{2b\sigma}\right).$$

The final inequality is the arithmetic–geometric mean inequality. Put $u = \sqrt{b/(2\sigma)}$ and $n = \lceil u \rceil$. Then

$$\frac{b}{2n} + \sigma(n-1) \leq \sigma u + \sigma(n-1) < 2\sigma u = \sqrt{2b\sigma},$$

so $B < a + \sqrt{2b\sigma}$. Every mixed plan therefore costs strictly more than

$$t(1-h) + chB \geq \min\{t, cB\}.$$

If everything is reported, one entity costs t and splitting adds effort. If everything is concealed through n entities, strict convexity of C makes equal holdings uniquely cheapest,

at effort $a + b/(2n) + \sigma(n-1)$. This expression tends to infinity with n, so its minimum exists. At $c = 0$, reporting zero through one entity attains burden zero, proving the value formula in that case as well. □

Writing the integer objective as B_n, its adjacent differences are

$$B_{n+1} - B_n = \sigma - \frac{b}{2n(n+1)}.$$

They increase strictly with n. An integer $n \geq 2$ uniquely minimizes effort when

$$\frac{b}{2n(n+1)} < \sigma < \frac{b}{2(n-1)n};$$

at either endpoint the adjacent numbers tie. One entity is uniquely optimal when $\sigma > b/4$.

In the example, $a = 1 + e_R$, $b = 12$ and $\sigma = 3/4 + e_S$. Only $n = 1, 2, 3$ can minimize: $B_4 - B_3 = \sigma - 1/2 > 0$, and subsequent differences are larger. This proves (44). Apart from ties, the preferred entity count for concealment is

$$n(e) = \begin{cases} 3, & 0 \leq e_S < 1/4, \\ 2, & 1/4 < e_S < 9/4, \\ 1, & e_S > 9/4. \end{cases} \tag{68}$$

At zero enforcement, the unique minimizing number is three, with $B = 9/2$ and holdings $1/3$ each. Two or four concealed entities each require effort $19/4$, and concealment through the parent requires effort 7. For any flat rate $0 < t \leq 1$, firms with $c < t/B(e)$ conceal using the minimizing organization and firms with $c > t/B(e)$ comply.

E.2 The joint optimum

The policy in (45) has $B = 5$ and $P(c) = \min\{1/2, 5c\}$. Integration gives $\mathbb{E}R = 9/20$, $A = 1/720$, $Q = 323/720$, $\mathbb{E}D = 1/40$ and $W = 21/40$. Section 3.2 verifies all its optimality conditions with $\mu = 9/8$. At the same enforcement, a flat rate $3/5$ raises gross receipts $66/125$; subtracting $1/720$ leaves net revenue strictly above G. This establishes strict feasibility. Theorem 1 therefore certifies the optimum over the full schedule and enforcement class. Strict convexity of public cost makes optimal enforcement unique; uniqueness of the schedule is not needed.

E.3 The optimum without splitting

The comparison retains every one-entity reduction, the same quadratic public cost, instrument bounds and requirement $G = 323/720$. Anti-splitting enforcement has no private effect, so every optimum has $e_S^0 = 0$. Under a flat rate t^0 and underreporting enforcement $u = e_R^0$, burden is $\min\{t^0, (7+u)c\}$. Let $r = t^0/(7+u)$ be the concealment cutoff. The rate and enforcement conditions give

$$\mu^0 = \frac{1-r}{1-2r}, \qquad u = \frac{9r^2}{1-r}, \qquad t^0 = (7+u)r.$$

The budget equation is

$$(7+u)r(1-r) - \frac{u^2}{36} = \frac{323}{720}, \qquad u = \frac{9r^2}{1-r}. \tag{69}$$

There is a unique root in $(0.06, 0.08)$. Indeed, after substituting for u, the left side of the budget equation is

$$H(r) = 7r(1-r) + 9r^3 - \frac{9}{4}\frac{r^4}{(1-r)^2}.$$

Its derivative exceeds 5.8 on that interval and its endpoint values straddle the requirement. The resulting policy is interior in t and e_R. It satisfies every rate score and the underreporting-enforcement score with equality. The anti-splitting score is zero at its lower boundary, and raising that instrument only adds a public cost. Together with the binding budget, these conditions give a global certificate by Theorem 1.

The formula $W^0 = 1 - t^0 + (t^0)^2/[2(7+e_R^0)]$ gives the welfare comparison reported in the main text, at the same net revenue and with taxes and enforcement optimized in both economies.

The single-operation comparison. At the joint optimum and $c = 1/11$, strict concavity of the single-entity objective in its final report makes the best single reduction an endpoint choice, with burden $\min\{1/2, 71/110\} = 1/2$. A split without subsequent reductions preserves total tax and adds positive effort. Three-entity concealment instead costs $5/11 < 1/2$, proving the comparison in the main text.

The enforcement comparison and Figure 3. For $0 < t \leq 1$, the inequality $B(e) > t$ holds throughout the enforcement rectangle. Integrating the burden formula gives

$$\mathbb{E}\sum_j r_j = 1 - \frac{t}{B(e)}, \qquad \mathbb{E}R = t\left(1 - \frac{t}{B(e)}\right), \qquad \mathbb{E}D = \frac{t^2}{2B(e)}.$$

On a region with fixed concealed-entity count n, $B_R = 1$ and $B_S = n - 1$. Increasing anti-splitting enforcement therefore raises reporting and receipts and lowers private costs while $n \geq 2$; its private effect ends at $n = 1$. At $e_S = 1/4$ and $9/4$, adjacent counts tie. Burdens, receipts and total reports remain continuous, although derivatives are one-sided and counts can depend on tie-breaking. These formulas and (68) generate both panels of Figure 3.

F Accounting and global optimality

This appendix proves the main characterization and its general-density extension under Assumption 3. Here $\mathcal{T}$ is the general schedule class of Appendix B.1. The main power model is obtained by setting $\ell_s = f_s(1) = \gamma_s$ and $I = 2$.

F.1 Burden envelopes and accounting

Proposition F.1 (Accounting and continuity)**.** *Maintain parts (i)–(iv) and (vi) of Assumption 3, and suppose $\mathbb{E}|\ell_s(c)| < \infty$. Then:*

(i) *Burden is jointly continuous in (s, c, π), concave in policy, and nondecreasing, concave and Lipschitz in c. Chosen effort weakly decreases and taxes weakly increase with c.*

(ii) *The identities (12) hold for every minimizer at almost every cost parameter, so taxes and private costs are selection-independent almost surely.*

(iii) *Welfare and net revenue are continuous in policy, and*

$$\mathbb{E}R = \mathbb{E}[(1 + \ell_s)P - f_s(1)T(x)], \qquad \mathbb{E}D = \mathbb{E}[f_s(1)T(x) - \ell_s P]. \tag{70}$$

For power densities these reduce to (13). Net revenue is concave if $1 + \ell_s(c) \geq 0$ almost surely.

Concavity of welfare, the welfare-weight inequality and strict feasibility are not required. Under the full Assumption 3, $\ell_s \geq 0$ and $\mathbb{E}\ell_s = \bar{J}$, so the integrability and concavity conditions hold automatically.

Proof. We differentiate minimized burden and integrate its derivative to recover taxes and costs. The strategy correspondence $(s, c, \pi) \mapsto \mathcal{A}(s)$ is continuous and compact-valued, and the objective is jointly continuous. Berge's maximum theorem and the measurable maximum theorem (Aliprantis and Border, 2006, Theorems 17.31 and 18.19) give continuous burden and a measurable, nonempty compact-valued minimizing correspondence with measurable selectors.

As an infimum of affine functions, P is concave separately in policy and c. Bounded nonnegative effort gives monotonicity and Lipschitz continuity in c. Adding the optimality inequalities for $c < v$ gives $K(v) \leq K(c)$. The inequality at c then gives $R(c) \leq R(v)$. Concavity gives differentiability outside a Lebesgue-null set (Aliprantis and Border, 2006, Theorem 7.22). At each differentiability point, the envelope theorem (Milgrom and Segal, 2002, Theorem 1) gives $P_c = K$ for every minimizer. Thus $R = P - cP_c$ and $D = cP_c$. A conditional density assigns probability zero to the exceptional set. Integrating over H gives the almost-sure assertion.

Local absolute continuity, the endpoint limits and integrability of $|\ell_s| f_s$ give $\int_0^1 \ell_s f_s \, \mathrm{d}c = f_s(1)$ for almost every s. Integration by parts on $[\epsilon, b] \subset (0, 1)$ yields, as $\epsilon \downarrow 0$ and $b \uparrow 1$,

$$\int_0^1 Df_s \, \mathrm{d}c = \int_0^1 cP_c f_s \, \mathrm{d}c = f_s(1)T(x) - \int_0^1 \ell_s P f_s \, \mathrm{d}c.$$

The lower boundary vanishes because P is bounded and $cf_s(c) \to 0$. At the upper boundary, $P_s(1) = T(x)$. The bound $P \leq \bar{T}$ and integrability of $|\ell_s|$ and $f_s(1)$ justify the limits and aggregation. Subtracting costs from $\mathbb{E}P$ gives the revenue identity. The preceding integral identity also gives $\mathbb{E}\ell_s = \bar{J}$.

Let $\bar{k}_i = \max_{\mathcal{A}(\mathcal{S})} k_i < \infty$. Comparing an optimizer at each policy with its cost at the other policy gives the uniform bound

$$|P(\pi) - P(\widetilde{\pi})| \leq m\|T - \widetilde{T}\|_\infty + \sum_i \bar{k}_i |e_i - \widetilde{e}_i|. \tag{71}$$

Bounded welfare weights give continuity of W. The accounting identity and dominated convergence give continuity of Q. If $1 + \ell_s \geq 0$, the identity expresses Q as a nonnegative weighted average of concave burdens, less a linear term and the convex cost A, so Q is concave. Under the full assumption, the score inequality gives $\ell_s \geq 0$; monotone convergence then gives $\mathbb{E}\ell_s = \bar{J} < \infty$. □

Proof of Proposition 1. For $f_s(c) = \gamma_s c^{\gamma_s - 1}$, both $\ell_s(c)$ and $f_s(1)$ equal $\gamma_s \in [1, \bar{\gamma}]$. Proposition F.1 therefore gives every assertion under regularity and no manipulation at

the top. Neither the welfare-weight restriction nor strict feasibility is needed. □

Proof of Corollary 1. Conditional on s, eliminate $\mathbb{E}[P \mid s]$ between its two identities to obtain (14). Its tax-gap term is nonnegative because $R \leq P \leq T(x)$. Since $\gamma_s/(1+\gamma_s) \in [1/2, \bar{\gamma}/(1+\bar{\gamma})]$, integration gives the bounds and the common-exponent equality. □

F.2 The concave reformulation and duality

The general Lagrangian is

$$\mathcal{L}_\mu = \mathbb{E}[U_s(x-P) + \mu(1+\ell_s)P - \mu f_s(1)T(x)] - \mu A(e) - \mu G. \tag{72}$$

For $\mu \geq w_-$, its burden-dependent integrand is concave and nondecreasing. Hence $\mathcal{L}_\mu$ is concave in policy.

Proposition F.2 (Existence and supporting multipliers)**.** *Under Assumption 3, an optimum exists and every optimum raises exactly G. The original problem has the same value as the concave program $\max_{Q \geq G} \mathcal{L}_{w_-}$. A policy solves the original problem exactly when it solves this program and has $Q = G$. Any auxiliary optimum can be scaled toward zero to obtain one with a binding budget. Moreover,*

$$V(G) = \min_{\mu \geq 0} \max_{\pi \in \mathcal{T} \times E} \{W(\pi) + \mu(Q(\pi) - G)\}. \tag{73}$$

The minimum is attained. Its minimizers are exactly the supporting multipliers, and all satisfy $\mu \geq w_-$. Every optimum maximizes the Lagrangian at every supporting multiplier. If some supporting multiplier exceeds w_-, every auxiliary optimum already has $Q = G$. On the strictly feasible range, V is concave and supporting multipliers are nondecreasing in G.

Proof. We establish the reformulation, then apply the convex Lagrange multiplier theorem under Slater's condition. Continuity and compactness of the nonempty feasible set give existence in both programs by the Weierstrass theorem (Aliprantis and Border, 2006, Corollary 2.35). Write $W_0 = \mathbb{E}U_s(x)$ and $\Phi(\mu) = \max_\pi \mathcal{L}_\mu(\pi)$. For $0 \leq \mu \leq w_-$, the bound $W \leq W_0 - w_-\mathbb{E}P$ gives

$$\mathcal{L}_\mu \leq W_0 - \mu G - (w_- - \mu)\mathbb{E}R - w_-\mathbb{E}D - \mu A \leq W_0 - \mu G.$$

Zero attains equality, so $\Phi(\mu) = W_0 - \mu G$ and zero maximizes $\mathcal{L}_{w_-}$.

For any feasible π, convexity of the policy set, $Q(0) = 0$ and continuity supply $a \in (0, 1]$ with $Q(a\pi) = G$. Concavity and maximality at zero imply $\mathcal{L}_{w_-}(a\pi) \geq \mathcal{L}_{w_-}(\pi)$. Thus the

concave program has a binding-budget maximizer. If its value is $\widehat{V}$, every feasible policy satisfies

$$W(\pi) = \mathcal{L}_{w_-}(\pi) - w_-(Q(\pi) - G) \leq \widehat{V},$$

strictly for $Q(\pi) > G$, whereas a binding-budget maximizer attains $\widehat{V}$.

The convex Lagrange multiplier theorem (Luenberger, 1969, §8.3, Theorem 1 and Corollary 1) applies: $-\mathcal{L}_{w_-}$ and $G - Q$ are convex on the convex policy set, the value is finite, and strict feasibility is Slater's condition. This theorem permits an arbitrary vector space of policies. It yields $\lambda \geq 0$ such that $\mu = w_- + \lambda$ satisfies $\Phi(\mu) = V$. Weak duality gives (73). The inequality $\Phi(\mu) > \Phi(w_-) \geq V$ excludes $\mu < w_-$. At any optimum π^* and supporting multiplier μ, the binding budget gives the saddle-point inequalities

$$\mathcal{L}_\mu(\pi) \leq \mathcal{L}_\mu(\pi^*) \leq \mathcal{L}_\nu(\pi^*) \qquad (\pi \in \mathcal{T} \times E,\ \nu \geq 0).$$

If $\mu > w_-$, any maximizer of the concave program satisfies $V \geq \mathcal{L}_\mu = V + (\mu - w_-)(Q - G)$, so its budget binds.

Finally, the dual formula makes V an infimum of affine functions, hence concave. Its supporting inequalities at $G_1 < G_2$ give

$$\mu_1 \leq \frac{V(G_1) - V(G_2)}{G_2 - G_1} \leq \mu_2,$$

proving monotonicity of the multipliers. □

Remark 1 (The boundary multiplier). If $\mu = w_-$ supports $V(G)$, then $V(G) = W_0 - w_- G$ and every optimum has $A(e) = \mathbb{E}D = 0$. Thus positive enforcement spending or manipulation cost at any optimum implies that every supporting multiplier exceeds w_-. With equal linear welfare weights, the converse holds: an optimum with $A = \mathbb{E}D = 0$ has supporting multiplier w.

Proof. Equality at the boundary multiplier requires equality in the welfare bound. Since $\Phi(w_-) = W_0 - w_- G$ and an optimum has a binding budget,

$$W \leq W_0 - w_- \mathbb{E}P = W_0 - w_-\{G + A(e) + \mathbb{E}D\}.$$

Thus $W = \Phi(w_-)$ forces $A = \mathbb{E}D = 0$. Conversely, with equal linear weights and zero public and private costs, $W = W_0 - wG = \Phi(w)$, so w supports the optimum. □

F.3 Coefficient representation and KKT conditions

Fix $\mu \geq w_-$. For a strategy write $\nu_\alpha = \sum_j \delta_{r_j}$, so $n_r = \nu_\alpha((r, X])$ and $n = \nu_\alpha([0, X])$. Set

$$w_s(c) = U_s'(x - P_s(c)), \quad \Omega_s(c) = \mu(1 + \ell_s(c)) - w_s(c), \quad J = \int f_s(1)\delta_{x(s)}\, \mathrm{d}H(s).$$

A selection is a measurable probability kernel supported on minimizing strategies. Expectations also integrate over this kernel. Put $\Xi_\mu = \mathbb{E}[\Omega_s \nu_\alpha] - \mu J$ and

$$\Xi_r = \mathbb{E}[\Omega_s n_r] - \mu J((r, X]), \qquad \Xi_\tau = \mathbb{E}[\Omega_s n] - \mu \bar{J}, \qquad g_i = \mathbb{E}[\Omega_s c k_i] - \mu A_i(e). \tag{74}$$

Thus Ξ_r and Ξ_τ are the tails and total mass of Ξ_μ. Randomization permits convex combinations of the derivatives of tied strategies. For a concave function F, write $\partial^+ F = -\partial(-F)$ for its superdifferential; thus $z \in \partial^+ F(\pi)$ means $F(\widetilde{\pi}) - F(\pi) \leq \langle z, \widetilde{\pi} - \pi \rangle$ for every $\widetilde{\pi}$.

Theorem F.1 (Global characterization)**.** *Under Assumption 3, π^* is optimal if and only if $Q(\pi^*) = G$ and some $\mu \geq w_-$ and one selection at π^* satisfy*

$$\int (T - T^*)\, \mathrm{d}\Xi_\mu + g \cdot (e - e^*) \leq 0 \qquad \text{for every } (T, e) \in \mathcal{T} \times E. \tag{75}$$

Every supporting multiplier admits such a selection at every optimum. In the full or fixed-bracket rate class, Theorem 1 holds with the coefficients (74).

Proof. We identify the Lagrangian's superdifferential and apply the convex optimality condition. The only model-specific step is expressing its elements through one selection of firms' minimizing strategies.

Extension to the policy space. Let $B = C([0, X]) \times \mathbb{R}^I$. Extend U_s linearly at the endpoints of $[x(s) - \bar{T}, x(s)]$, retaining concavity and $w_- \leq \widehat{U}_s' \leq w_+$. Define $\widehat{P}$ on B by the same burden minimum, and set

$$\widehat{A}(e) = \max_{v \in E} \{A(v) + \nabla A(v) \cdot (e - v)\}.$$

This convex Lipschitz extension equals A on E. At $e \in E$, the tangent at e bounds $\widehat{A}$ below and A bounds it above on a neighborhood. Differentiability of A therefore gives $\nabla \widehat{A}(e) = \nabla A(e)$, including at boundary points. Substitution of these extensions in (72) gives a globally Lipschitz concave function $\widehat{\mathcal{L}}_\mu$ on B that agrees with $\mathcal{L}_\mu$ on admissible policies. Indeed, (71) applies on B, $\mathbb{E}(1 + \ell_s) < \infty$, and the burden-dependent integrand remains concave and nondecreasing. Endpoint derivatives of U_s are measurable as limits of difference quotients, so the extension also preserves measurability.

The superdifferential. Fix an admissible policy and write $\theta = (s, c)$, with minimizing strategy set M_θ. For $d = (\Delta T, \Delta e) \in B$, let

$$b_{\theta,\alpha}(d) = \int \Delta T \, \mathrm{d}\nu_\alpha + c \sum_i \Delta e_i k_i(\alpha), \qquad h_\theta(d) = \max_{\alpha \in M_\theta} \{-\Omega_\theta b_{\theta,\alpha}(d)\}.$$

The envelope theorem (Milgrom and Segal, 2002, Corollary 4) and dominated convergence give

$$-\widehat{\mathcal{L}}'_\mu(\pi; d) = \mathbb{E} h_\theta(d) + \mu \int \Delta T \, \mathrm{d}J + \mu \nabla A(e) \cdot \Delta e.$$

The derivative is dominated by an integrable multiple of $\|d\|$: reports and effort are uniformly bounded, and $\mathbb{E}\Omega < \infty$. The measurable maximum theorem makes $h_\theta(d)$ measurable in θ. It is continuous and sublinear in d.

The directional characterization of subgradients, support-function duality and Strassen's Sublinearity Theorem (Aliprantis and Border, 2006, Theorems 7.16, 7.52 and 18.35) now identify $\partial^+ \widehat{\mathcal{L}}_\mu(\pi)$ as the Gelfand integrals of weak-* measurable selections from the closed convex hulls of $\{\Omega_\theta b_{\theta,\alpha} : \alpha \in M_\theta\}$ in the weak-* topology, less $(\mu J, \mu \nabla A(e))$. The hypotheses are exactly the separability of B, the Carathéodory property just established and the integrable bound on h_θ.

Each member of the pointwise weak-* closed convex hull is a barycenter of a probability measure on M_θ. This follows from density of finitely supported probabilities and compactness of probabilities on a compact metric space (Aliprantis and Border, 2006, Theorems 15.10–15.11). The representing probabilities can be chosen measurably by Filippov's Implicit Function Theorem (Aliprantis and Border, 2006, Theorem 18.17): impose the barycenter equalities on a countable dense set of directions in B, together with zero expected distance to M_θ. These are Carathéodory constraints on the fixed compact metric space of probabilities on $\mathcal{A}(\mathcal{S})$. Their solutions are nonempty; the distance constraint ensures support on minimizing strategies. Thus

$$\partial^+ \widehat{\mathcal{L}}_\mu(\pi) = \{(\Xi_\mu, g) : \text{one measurable randomized selection at } \pi\}.$$

Optimality. Set $C = \mathcal{T} \times E$ and $N_C(\pi) = \{z : \langle z, \widetilde{\pi} - \pi \rangle \leq 0 \text{ for all } \widetilde{\pi} \in C\}$. Fermat's rule and the Moreau–Rockafellar sum rule (Aliprantis and Border, 2006, Lemma 7.10)(Barbu and Precupanu, 2012, Theorem 3.57) give

$$\begin{aligned} \pi^* \in \arg\max_C \widehat{\mathcal{L}}_\mu \quad &\iff \quad 0 \in \partial(-\widehat{\mathcal{L}}_\mu)(\pi^*) + N_C(\pi^*) \\ &\iff \quad \partial^+ \widehat{\mathcal{L}}_\mu(\pi^*) \cap N_C(\pi^*) \neq \varnothing. \end{aligned}$$

The sum rule applies because $-\widehat{\mathcal{L}}_\mu$ is finite continuous convex on B, while the indicator of C is proper lower semicontinuous convex. Proposition F.2 and the representation above give necessity at every supporting multiplier. Conversely, a selected supergradient satisfies

$$\mathcal{L}_\mu(\pi) - \mathcal{L}_\mu(\pi^*) \le \int (T - T^*)\, \mathrm{d}\Xi_\mu + g \cdot (e - e^*). \tag{76}$$

If (75) holds and $Q(\pi^*) = G$, every feasible policy satisfies $W(\pi) \le \mathcal{L}_\mu(\pi) \le W(\pi^*)$. Hence π^* is optimal and μ supports V.

For a finite signed measure Ξ, Fubini's theorem (Aliprantis and Border, 2006, Theorem 11.27) gives

$$\int T\, \mathrm{d}\Xi = \tau \Xi([0, X]) + \int_0^X t(r) \Xi((r, X])\, \mathrm{d}r; \tag{77}$$

bounded rates and finite total variation justify the interchange. Thus (76) becomes

$$\mathcal{L}_\mu(\pi) - \mathcal{L}_\mu(\pi^*) \le (\tau - \tau^*)\Xi_\tau + \int_0^X (t - t^*)\Xi_r\, \mathrm{d}r + g \cdot (e - e^*). \tag{78}$$

The normal-cone condition separates across the factors of the policy set. Maximizing the linear functional over $0 \le t \le L$ gives (17); for a fixed bracket its coefficient is $\int_B \Xi_r\, \mathrm{d}r$. The fee has the same scalar conditions. Enforcement satisfies $g \in N_E(e^*)$, which gives the coordinate signs when E is rectangular. This proves Theorem 1 and its general-density version. □

G Policy responses and the Ramsey rules

G.1 Policy derivatives and proportional changes

Proposition G.1 (Directional policy responses)**.** *For a feasible direction $(\Delta T, \Delta e)$, suppose the objective derivative $b_\alpha = \int \Delta T\, \mathrm{d}\nu_\alpha + c \sum_i \Delta e_i k_i(\alpha)$ has the same value $b(s, c)$ at every minimizer for almost every firm. Under the hypotheses of Proposition F.1,*

$$\begin{aligned} \dot{P} &= b, & \dot{W} &= -\mathbb{E}[wb], \\ \dot{\mathbb{E}R} &= \mathbb{E}[(1 + \ell_s)b] - J(\Delta T), & \dot{\mathbb{E}D} &= J(\Delta T) - \mathbb{E}[\ell_s b], \\ \dot{Q} &= \mathbb{E}[(1 + \ell_s)b] - J(\Delta T) - \nabla A \cdot \Delta e. \end{aligned} \tag{79}$$

Without this constancy condition, right derivatives use the minimum b_α over minimizers, and feasible left derivatives use the maximum.

Proof. We differentiate the burden envelope and then the accounts. The envelope theorem for compact choice sets (Milgrom and Segal, 2002, Corollary 4) gives the minimum and maximum objective derivatives. The measurable maximum theorem (Aliprantis and Border, 2006, Theorem 18.19) supplies their selectors. Under constancy both equal b. Bound (71), bounded welfare weights and integrable $|1+\ell_s|$ justify dominated convergence in welfare and the accounts. Differentiating A gives the last formula. For an interval-rate change, $\Delta T(r) = |B \cap [0,r]|$ and $b_\alpha = \int_B n_r \, \mathrm{d}r$. Fubini gives (18) and its welfare counterpart. The derivative proof needs the accounting and integrability assumptions, but neither strict feasibility nor optimality. □

Proposition G.2 (Proportional scaling)**.** *Fix a policy and a selection under the hypotheses of Proposition F.1. Let $\widetilde{n}$ be either n_r or n, and let its compliance count $\widetilde{n}^0$ be $\mathbf{1}\{x > r\}$ or 1, respectively. At $(\lambda T, e)$, $0 < \lambda \leq 1$, select the original strategy at cost c/λ when $c < \lambda$, and compliance otherwise. This selection is optimal and its aggregate count obeys*

$$M'(1-) = \mathbb{E}[\ell_s \widetilde{n}] - \mathbb{E}[f_s(1)\widetilde{n}^0]. \tag{80}$$

Under power densities it also obeys the exact finite-change formula (19), with the corresponding counts. These statements concern this specified selection when counts differ across tied strategies.

Proof. We identify optimal choices after scaling taxes, then change variables in their aggregate counts. Dividing the objective under λT by λ produces the original objective at c/λ. This proves the selection rule for $c < \lambda$. For $c \geq \lambda$, no manipulation at the top and $K \geq 0$ give $T[\alpha] + (c/\lambda)K \geq T(x)$, so compliance minimizes. The same change of argument applies to randomized selections.

Substitution $c = \lambda v$ and $\int_0^1 f_s = 1$ give

$$M(\lambda) = \mathbb{E}_H \widetilde{n}^0 + \int \int_0^1 \lambda f_s(\lambda v)\{\widetilde{n}(s,v) - \widetilde{n}^0(s)\} \, \mathrm{d}v \, \mathrm{d}H(s).$$

Set $h(s,v) = \ell_s(v) f_s(v)$ on $(0,1)$ and zero elsewhere. This function is integrable, and local absolute continuity gives

$$\frac{f_s(v) - \lambda f_s(\lambda v)}{1-\lambda} = \frac{1}{1-\lambda} \int_\lambda^1 h(s, qv) \, \mathrm{d}q.$$

The right side converges to h in $L^1(\mathrm{d}H \, \mathrm{d}v)$. To see this, first use dominated convergence for continuous compactly supported functions, then their density in L^1 (Aliprantis and Border, 2006, Theorem 13.9) and the uniform bound $1/q \leq 2$ on the dilation operators for

$q \geq 1/2$. Bounded counts and $\int_0^1 h(s,v)\,\mathrm{d}v = f_s(1)$ now give (80). This argument allows signed ℓ_s. For power densities, $\lambda f_s(\lambda v) = \lambda^{\gamma_s} f_s(v)$; substitution gives (19) directly. □

Proof of Corollary 2. The Ramsey conditions restate the normal-cone conditions in elasticities. Formula (80) gives

$$\Xi_r = \mu M_r + \mu M_r'(1-) - \mathbb{E}[w_s n_r] = \mu M_r(1 - w_r/\mu - \varepsilon_r).$$

Dividing the rate conditions by $\mu M_r > 0$ proves (24) and its boundary inequalities. The undivided condition applies if $M_r = 0$. For a fixed bracket, integrate before dividing by $\mu \int_B M_r\,\mathrm{d}r$. Using n and compliance count one gives the fee rule, whose general-density elasticity is $-[\mathbb{E}(\ell_s n) - \bar{J}]/\mathbb{E}n$. Finally, expanding $g_i = 0$ gives

$$\mu\{\mathbb{E}[(1+\ell_s)ck_i] - A_i(e)\} = \mathbb{E}[w_s c k_i],$$

with the corresponding inequality at a bound. Setting $\ell_s = \gamma_s$ proves (25). □

Proof of Corollary 3. We substitute equal welfare weights into the rate conditions and integrate for total reports. Proposition F.2 gives $\mu \geq w$, so $\beta = 1 - w/\mu \in [0,1)$ and $\Omega_s/\mu = \gamma_s + \beta$. Thus $\Xi_r = 0$ gives (27). A common exponent gives (28). Integrating interior-rate conditions and using $\int_0^X n_r\,\mathrm{d}r = \sum_j r_j$ and $\int_0^X \mathbf{1}\{x > r\}\,\mathrm{d}r = x$ gives (29) and (30). Bounded reports and counts justify Fubini. Summing interior fixed-bracket conditions gives the same identities. □

The distribution of positive reports. Under the common-power, equal-weight and interior-rate conditions of Corollary 3, (28) equates the tails of the positive-report measure with $\gamma/(\gamma+\beta)$ times the tails of the potential-base measure. Both tails are right-continuous, so equality almost everywhere extends to every threshold in $(0, X)$. Differences of tails and the endpoint limits identify these finite measures on all Borel subsets of $(0, X]$. Their normalized distributions therefore coincide whenever they have positive mass.

G.2 Enforcement responses

Proof of Proposition 13. We use monotonicity of burden at a fixed schedule and then the accounting identities. If $e' \geq e$, every strategy has weakly higher effort, so $\Delta^e = P(T,e') - P(T,e) \geq 0$. The boundary term $\gamma_s T(x)$ is unchanged. Subtracting the accounts proves (57), and increasing U_s implies weakly lower welfare. When gross receipts rise, the

ratio of private cost savings to that rise is

$$\frac{\mathbb{E}[\gamma_s \Delta^e]}{\mathbb{E}[(1+\gamma_s)\Delta^e]},$$

a weighted average of $\gamma_s/(1+\gamma_s)$, so it lies in $[1/2, \bar{\gamma}/(1+\bar{\gamma})]$.

When ck_i is the same at every minimizer almost surely, Proposition G.1 yields

$$\partial_{e_i}\mathbb{E}R = \mathbb{E}[(1+\gamma_s)ck_i], \qquad \partial_{e_i}\mathbb{E}D = -\mathbb{E}[\gamma_s ck_i], \qquad W_{e_i} = -\mathbb{E}[w_s ck_i].$$

At an interior optimum, $g_i = 0$ gives $\mu Q_{e_i} = \mathbb{E}[w_s ck_i]$, which proves (58). This quantity is positive when $\mathbb{E}[ck_i] > 0$. With equal weights, $Q_{e_i} = (1-\beta)\mathbb{E}[ck_i]$ and $\partial_{e_i}(A + \mathbb{E}D) = \beta\mathbb{E}[ck_i]$. □

H Proofs for Section 5

Proof of Proposition 3. We identify the compliance distribution from its tails. Let $\chi_s(c)$ be the probability that the selection assigns to compliance. In a top range, only complying firms contribute entities above $r > \underline{x}$, so the interior-rate condition gives

$$\mathbb{E}\big[(\gamma + \beta(x))\chi \mathbf{1}\{x > r\}\big] = \gamma \Pr(x > r)$$

for almost every $r \in (\underline{x}, X)$. The two sides are tails of finite measures. Right-continuity extends equality to every threshold, and uniqueness of finite measures then equates their densities relative to the distribution of x. Thus $(\gamma + \beta(x))\mathbb{E}[\chi \mid x] = \gamma$ almost everywhere, including at atoms. This gives (47); since $\mathbb{E}[\chi \mid x] \le 1$, it also implies $\beta(x) \ge 0$ there.

For the second statement, fix a base at which this equality holds. Burden is nondecreasing in c, bounded above by $T^*(x)$, and equals it at $c = 1$. Compliance is therefore optimal on an interval $[\bar{c}, 1]$. Any noncompliant strategy used by the selection has all reports below $\underline{x}$. Applying no manipulation at the top to $T(r) = L(r - \underline{x})^+$ gives $k_0 \ge L(x - \underline{x}) > 0$ for that strategy. It consequently has burden strictly above $T^*(x)$ when $c > \bar{c}$. The selection thus complies exactly above $\bar{c}$, apart from the null boundary. Hence $\mathbb{E}[\chi \mid x] = 1 - \bar{c}^{\gamma}$, and the share formula gives $\bar{c} = \kappa(x)$. □

Proof of Proposition 4. For part (i), write $\check{T}(x) = \min\{\sum_{j=1}^n T^*(r_j) : r \in \mathbb{R}^n_+, \sum_j r_j = x\}$, so that $S = T^* - \check{T}$. Fix $x \in (\underline{x}, X]$ at which the conclusion of Proposition 3 holds and the marginal non-compliers split into n entities and report x in full. These firms exist because

$\beta(x) > 0$ gives $\kappa(x) > 0$. Such a firm can use any division at effort σ, and no strategy with n entities and total report x costs less, so its burden is $\check{T}(x) + c\sigma$. Firms with $c > \kappa(x)$ comply and bear $T^*(x)$. The burden is continuous in c, so $T^*(x) = \check{T}(x) + \sigma\kappa(x)$, that is, $S(x) = \sigma\kappa(x)$. This holds at almost every $x \in (\underline{x}, X]$ under the distribution of potential tax bases. We next use continuity to extend it to every such x. Every T^* in the class (2) is nondecreasing and Lipschitz with constant L. The function $\check{T}$ is nondecreasing. Lowering the reports of a minimizing division of $x' > x$ until they sum to x gives a division of x that pays no more. It is also Lipschitz with constant L. Adding $x' - x$ to one report of a minimizing division of x gives a division of x' that pays at most $L(x' - x)$ more. For $x < x'$, both tax increments lie in $[0, L(x' - x)]$, so $|S(x') - S(x)| \leq L(x' - x)$. In particular, S is Lipschitz. Since w is continuous and $\beta > 0$ on $(\underline{x}, X]$, κ is continuous. Two continuous functions that agree almost everywhere under a distribution with full support on $(\underline{x}, X]$ agree everywhere there, which proves part (i).

For parts (ii) and (iii), note that at Lebesgue-almost every $x \in (\underline{x}, X)$ both T^* and $\check{T}$ are differentiable and $T^{*\prime}(x) = t^*(x)$. Part (i) implies that $\kappa = S/\sigma$ is Lipschitz and hence differentiable almost everywhere. At a base where all three functions are differentiable, part (i) gives $t^*(x) = \check{T}'(x) + \sigma\kappa'(x)$. Let r be a minimizing division of such an x, and let $r_j > 0$ be one of its reports at which T^* is differentiable. For small $h > 0$, raising r_j by h gives a division of $x + h$, and lowering it by h gives a division of $x - h$, so

$$\frac{\check{T}(x+h) - \check{T}(x)}{h} \leq \frac{T^*(r_j + h) - T^*(r_j)}{h}, \qquad \frac{\check{T}(x) - \check{T}(x-h)}{h} \geq \frac{T^*(r_j) - T^*(r_j - h)}{h}.$$

Letting $h \downarrow 0$ gives $\check{T}'(x) \leq t^*(r_j) \leq \check{T}'(x)$, which proves part (iii). With equal weights κ is constant, and part (ii) follows. □

Proof of Corollary 5. Convexity gives $\check{T}(x) = nT^*(x/n)$. For almost every $x \in (\underline{x}, X)$, both T^* at x and at x/n are differentiable, and Proposition 4(ii) gives $T^{*\prime}(x) = T^{*\prime}(x/n)$. The derivative of a convex function is nondecreasing, so T^* is affine on $[x/n, x]$. The set of such x has full measure and is dense. The interiors of these intervals cover $(\underline{x}/n, X)$: for any y in this range there is such an x with $y < x < ny$, $\underline{x} < x < X$. Thus T^* is locally affine throughout this connected interval. Overlapping intervals have the same slope, so it is affine on the whole interval. □

Proof of Proposition 5. The argument compares the one-entity reports needed to meet the targets. At each threshold r below r_0, split firms add entities above r under T^2. The targets therefore permit fewer one-entity reports above r, which requires a higher rate. For

r with $\Pr(x > r) > 0$ and $v \in [0, L]$, let $\mathcal{Q}_r(v) = \Pr(c\, p_s > v \mid x > r)$, the share of firms with bases above r whose marginal cost of removing a unit exceeds v. No manipulation at the top requires $p_s \geq L$, because a firm with $c = 1$ must not gain from removing a unit taxed at rate L. Hence $\mathcal{Q}_r(v) = 1 - v^{\gamma}\mathbb{E}[p_s^{-\gamma} \mid x > r]$, which is strictly decreasing in v. Split firms have $x \geq 2r_0$, so $\Pr(x > r) > 0$ for every $r < r_0$. Write $\beta_k = 1 - w/\mu_k$, so that $\mu_2 \geq \mu_1$ gives $\gamma/(\gamma + \beta_2) \leq \gamma/(\gamma + \beta_1)$.

Take a representative of t^2 that is nondecreasing on $(0, r_0)$, and consider $r \in (0, r_0)$ at which both rates are interior, both targets (28) hold, and T^1 and T^2 are differentiable with derivatives $t^1(r)$ and $t^2(r)$. Almost every $r \in (0, r_0)$ qualifies. At such an r, the hypotheses give $t^2(v) \leq t^2(r)$ for almost every $v < r$ and $t^2(q) \geq t^2(r)$ for almost every $q > r$. The firms with $c\, p_s = t^k(r)$ form a null set, because cost parameters have no atoms.

A firm with one entity and report q bears $T(q) + c\, p_s(x - q)$, so it chooses q to minimize $g(q) = T(q) - c\, p_s q$ over $[0, x]$. Consider a firm with $x > r$. Under T^1, which is convex, g is convex, so the firm reports above r if $c\, p_s > t^1(r)$ and at most r if $c\, p_s < t^1(r)$. Under T^2, if $c\, p_s > t^2(r)$, then g is nonincreasing on $[0, r]$ and has a negative derivative at r, so every minimizer exceeds r. If $c\, p_s < t^2(r)$, then $g(q) - g(r) \geq (t^2(r) - c\, p_s)(q - r) > 0$ for every $q > r$, so every minimizer is at most r.

Every entity of a split firm reports at least $r_0 > r$, except on a null set, so a split firm has $n_r = n \geq 2$. Let

$$N_r = \mathbb{E}\big[\mathbf{1}\{\text{split}\}\big(n - \mathbf{1}\{c\, p_s > t^2(r)\}\big)\big],$$

the entities above r that split firms add relative to the one-entity reports they would make at T^2. Since $n \geq 2$, N_r is at least the mass of split firms, which is positive. A firm that does not split operates one entity, since there is no exit. Its report minimizes g: its strategy is optimal, and all one-entity reports at effort $p_s(x - q)$ are available. Hence

$$M_r^2 = \Pr(x > r)\, \mathcal{Q}_r(t^2(r)) + N_r, \qquad M_r^1 = \Pr(x > r)\, \mathcal{Q}_r(t^1(r)).$$

The targets give $M_r^k = \Pr(x > r)\, \gamma/(\gamma + \beta_k)$. Therefore

$$\mathcal{Q}_r(t^2(r)) = \frac{\gamma}{\gamma + \beta_2} - \frac{N_r}{\Pr(x > r)} < \frac{\gamma}{\gamma + \beta_2} \leq \frac{\gamma}{\gamma + \beta_1} = \mathcal{Q}_r(t^1(r)),$$

and $t^2(r) > t^1(r)$ because $\mathcal{Q}_r$ is strictly decreasing. This is part (i).

The fees are equal, so for every $q \in (0, r_0]$, $T^2(q) - T^1(q) = \int_0^q (t^2 - t^1)\, \mathrm{d}v > 0$ by part (i). A firm with $x \leq r_0$ does not split at T^2, except on a null set, since the entities of a split firm report at least $2r_0$ in total. Its burden under T^2 is therefore $\min_{q \leq x}\{T^2(q) + c\, p_s(x - q)\}$. Under T^1, where splitting is unavailable, its burden is the same minimum with T^1 in place

of T^2. The first minimum is at least the second, and strictly greater when it is attained at a positive report. This is part (ii). □

I Proofs for Section 6

The proof establishes the general-density version of Theorem 2 described in Appendix F. In the main model, $\ell_s = \gamma_s$ and $f_s(1) = \gamma_s$.

I.1 Proof of Theorem 2

(i) Subtract the identities (70) for the two economies. The terms $J(T)$ coincide because firms do not manipulate at the top in either economy.

(ii) We compare Lagrangians at a multiplier μ_1 supporting the smaller economy. Proposition F.2 gives $\mu_1 \geq w_-$, so $p \mapsto U_s(x-p) + \mu_1(1+\ell_s)p$ is nondecreasing. Since $P^2 \leq P^1$, we have $\mathcal{L}^2_{\mu_1} \leq \mathcal{L}^1_{\mu_1}$ at every policy. Weak duality therefore gives

$$V^2(G) \leq \max_{\pi} \mathcal{L}^2_{\mu_1}(\pi) \leq \max_{\pi} \mathcal{L}^1_{\mu_1}(\pi) = V^1(G).$$

At an optimum π_2 of the larger economy, the binding budget gives $\mathcal{L}^2_{\mu_1}(\pi_2) = V^2(G)$. Add and subtract $\mathcal{L}^1_{\mu_1}(\pi_2)$ and apply the fundamental theorem of calculus to obtain (50), with $\ell_s(c)$ replacing γ_s. Its first term is nonnegative by optimality in the smaller economy. Its integrand is at least $\mu_1(1+\ell_s) - w_+ \geq 0$, proving (51).

(iii) Suppose an optimal policy π_1 with the smaller set has $\Delta = 0$ for almost every firm. It gives the same burdens in both economies. By (70), it therefore gives the same revenue G and the same welfare $V^1(G) \geq V^2(G)$. It is optimal with the larger set, and $V^2(G) = V^1(G)$.

Conversely, suppose that $V^2(G) = V^1(G)$, let $\pi_2 = (T_2, e_2)$ be optimal with the larger set, and let μ_1 support $V^1(G)$. If $\mu_1 > w_-$, the coefficient in (51) is at least $(\mu_1 - w_-)(1+\ell_s) > 0$, so $\Delta(\pi_2) = 0$ for almost every firm. If $\mu_1 = w_-$, then $V^1(G) = W_0 - w_- G$ (Appendix F). Since $V^2(G) = W^2(\pi_2) \leq W_0 - w_- \mathbb{E}P^2(\pi_2) = W_0 - w_-[G + A(e_2) + \mathbb{E}D^2(\pi_2)]$, we obtain $A(e_2) = 0$ and $D^2(\pi_2) = 0$ for almost every firm. A firm with $c \in (0,1)$ and zero manipulation cost chooses a strategy that requires no effort, whose taxes paid are at least $T_2(x)$ by (6). Hence $P^2 \geq T_2(x) \geq P^1$, and again $\Delta(\pi_2) = 0$ for almost every firm.

In both cases π_2 has the same burdens in the two economies. By (70), it raises G and attains $V^2(G) = V^1(G)$ with the smaller set. The optimal policies with the larger set are

therefore optimal policies with the smaller set that have $\Delta = 0$, and the first paragraph gives the reverse inclusion. The last sentence of (iii) is the contrapositive. □

I.2 Proof of Proposition 9

Under the general-density assumptions of Appendix F, we show that a selection using only strategies from the smaller set would support an optimum in both economies. Suppose a selection supporting an optimum π_2 uses only those strategies for almost every firm. Those strategies minimize burden in both sets, so $P^1 = P^2$ at π_2. Accounting gives the same revenue, and the common welfare functions give the same welfare weights. The selected tax and enforcement exposures also coincide. The same multiplier and selection therefore satisfy the sufficient conditions of Theorem F.1 in the smaller economy. The policy is optimal in both economies, contradicting $V^2(G) < V^1(G)$. When splitting status is unique almost everywhere, every optimal selection has the same split share, giving the assertion about actual choices.

For the enforcement assertion, write the common welfare weight as w. If a supporting multiplier at π_2 equalled w, Remark 1 would imply zero public expenditure and zero private manipulation cost. Welfare would attain the upper bound $w(\mathbb{E}x - G)$, contradicting the strict welfare gap. Thus $\mu > w$. The equal-weight density restriction implies $\Omega = \mu(1 + \ell_s) - w > 0$, since $\ell_s \geq 0$. At $e_S = 0$ the selected net value is

$$g_S = \mathbb{E}[\Omega c k_S] - \mu A_S(e_R, 0) = \mathbb{E}[\Omega c k_S] > 0.$$

Every supporting selection uses split strategies on positive mass, and $c > 0$ almost surely, so the inequality is strict. It contradicts optimality at zero when an increase is feasible. □

I.3 The bounds of Section 6

Proof of Proposition 7. We compare the same subadditive policy in both economies. Iterated subadditivity and the merging condition give a one-entity counterpart with no greater taxes or effort for every operating strategy. Exit is shared, so $P^1 = P^2$. The accounting identities then give equal revenue and welfare. Taking suprema over feasible subadditive policies yields $V^2(G) \geq V^1_{\text{sub}}(G)$, which is (52). A linear schedule $T(r) = \tau + tr$ with $\tau \geq 0$ is subadditive. If an optimum without splitting is subadditive, its equal burdens, revenue and welfare in the two economies, together with $V^2(G) \leq V^1(G)$, prove the final assertion. □

Proof of Proposition 8. The upper bound uses weak duality at the optimum of the smaller economy. Its budget binds, so

$$V^1(G) - V^2(G) \le \mathcal{L}^1_{\mu_2}(\pi_1) - \mathcal{L}^2_{\mu_2}(\pi_1) = \mathbb{E}\int_{P^2(\pi_1)}^{P^1(\pi_1)} \{\mu_2(1+\gamma_s) - U_s'(x-p)\}\,\mathrm{d}p.$$

Bounding U_s' below by w_- proves (53). With equal weights and a common exponent, substitute $\Delta R = (1+\gamma)\mathbb{E}\Delta$ into this bound and (51) to obtain (54). □

The numerical bounds reported after Proposition 8 use exact cost integration and are rounded outward. Appendix N gives the error certificate.

J Proof of Proposition 6

Write $b_z(r) = \min\{r, z\}$, $T_a = T - ab_z$, and $B_z^a = \sum_j b_z(r_j^a)$ for a firm's chosen exposure after a cut a. We first prove the bound without splitting and concavity of receipts, then construct the counterexample required for part (ii).

Parts (i) and (iii). For $0 \le a < b$, the optimality inequalities give

$$(b-a)B_z^a \le P_a - P_b \le (b-a)B_z^b.$$

Thus exposure is nondecreasing for any choices from the minimizing sets. Without splitting, an initially exiting firm has zero taxes and exposure, so its nonnegative new taxes satisfy $R_a \ge R_0 - aB_z^0$. If an initially operating firm's exposure increases, its new report exceeds its old report q_0. Monotonicity of the new schedule then gives $R_a = T_a(q_a) \ge T_a(q_0) = R_0 - aB_z^0$. If exposure is unchanged, the new strategy is also optimal at the old policy. Almost-sure selection independence of old taxes gives the same equality, including a switch to exit. Averaging proves part (i).

Each strategy's burden is affine in a, so P_a is concave. By (13),

$$\bar{R}(a) = \mathbb{E}[(1+\gamma_s)P_a] - \mathbb{E}[\gamma_s T_a(x)]$$

is concave as well. Hence $\mathcal{X}(a) = \bar{R}(0) - \bar{R}(a) - a\mathbb{E}B_z^0$ is convex and $\mathcal{X}(0) = 0$. Its secant slopes $\mathcal{X}(a)/a$ are nondecreasing for $a > 0$. Once $\mathcal{X}(a) > 0$, it remains positive at every larger admissible cut, proving part (iii). The first argument does not require a power law; the second extends by the general-density accounting identity whenever $1 + \ell_s \ge 0$.

Part (ii): an unrestricted counterexample. Let all firms have $x = 1$, uniform $c \in [0,1]$ and welfare weight one. There is no exit. The fee is zero, the rate cap is one, $e \in [0,1]^2$, $A(e) = e_R + e_S$, and $G = 19/40$. Every finite nonnegative report vector with sum at most one is feasible, at effort

$$(10 + e_R)\Big(1 - \sum_j r_j\Big) + (\tfrac{1}{2} + e_S)(n-1).$$

Counts above 23 are dominated by full concealment through one entity, uniformly over enforcement. At $c = 0$, one zero report attains the minimum. Thus a finite compact representation preserves the unrestricted burden. With $h = 1 - \sum_j r_j$, for $n \geq 2$,

$$1 - \max_j r_j \leq h + \frac{n-1}{n}(1-h) \leq 10h + \frac{n-1}{2}.$$

For $n = 1$, $1 - r_1 = h \leq 10h$. Baseline effort therefore exceeds the largest tax saving under every admissible schedule, establishing no manipulation at the top. The other regularity and welfare conditions hold. The flat rate 0.6 at zero enforcement raises $141/250 > G$, establishing strict feasibility.

At $T(r) = r/2$ and $e = 0$, splitting is dominated for every $c > 0$. Firms report zero for $c < 1/20$ and one for $c > 1/20$. Receipts are $19/40$. Set $\mu = 19/18$ and $\Omega = 2\mu - 1 = 10/9$. At every $r \in (0,1)$, the rate score is $\Omega(19/20) - \mu = 0$. The enforcement scores are $g_R = \Omega \int_0^{1/20} c \, \mathrm{d}c - \mu < 0$ and $g_S = -\mu < 0$. The budget binds, the fee is fixed, and both enforcement choices are zero. Theorem 1 proves global optimality over the full schedule class and enforcement rectangle.

Now cut the rate below $z = 1/2$ by $a = 1/5$. The two rates become $3/10$ and $1/2$. For total report q through n entities, the minimum tax is $3q/10 + (q - n/2)^+/5$. For $n \geq 2$ this equals $3q/10$, so further entities add effort without reducing tax. Conditional on two entities, burden is affine in q; its endpoints are full reporting or zero reporting, the latter dominated by one zero report. Through one entity, burden is affine on each bracket. The unrestricted minimum is therefore the lower envelope of

reports	(0)	$(1/2)$	$(1/2, 1/2)$	(1)
burden	$10c$	$3/20 + 5c$	$3/10 + c/2$	$2/5.$

The successive cutoffs are $3/100$, $1/30$ and $1/5$. Integration gives

$$\bar{R}(1/5) = \frac{3}{20}\Big(\frac{1}{30} - \frac{3}{100}\Big) + \frac{3}{10}\Big(\frac{1}{5} - \frac{1}{30}\Big) + \frac{2}{5}\frac{4}{5} = \frac{741}{2000}.$$

The revenue loss is $209/2000$, whereas the initial mechanical estimate is $a\mathbb{E}B_z^0 = (1/5)(19/40) = 19/200$. The former is strictly larger. No firm initially splits, and all finite divisions and continuous reports remain feasible throughout. This proves part (ii). □

K Proofs for Section 7

K.1 Proof of Proposition 12

We verify the hypotheses of the projected subgradient method for $-\mathcal{L}_\mu$. In a fixed-bracket class the admissible policy set Π is a compact rectangle, and clip is its Euclidean projection. The supporting inequality (78) applies at every policy, so each g_k is a supergradient. Uniformly bounded entity counts and effort coefficients, bounded welfare weights and continuity of ∇A on E give a common bound $|g_k| \le B$. Continuity and compactness give a maximizer π°. The projection characterization gives

$$\operatorname{clip}(\pi_k + g_k) = \pi_k \quad \Longrightarrow \quad g_k \cdot (\pi' - \pi_k) \le 0 \quad (\pi' \in \Pi),$$

which proves the stopping claim by the supporting inequality.

For a nonterminating sequence, the standard projected subgradient bound (Boyd, 2014, Sections 3.2 and 6), with step sizes $\eta_j/|g_j|$, gives

$$\min_{j \le k}\{\mathcal{L}_\mu^* - \mathcal{L}_\mu(\pi_j)\} \le \frac{D^2 + \sum_{j=0}^k \eta_j^2}{2\sum_{j=0}^k \eta_j/|g_j|} \le \frac{B(D^2 + \sum_{j=0}^k \eta_j^2)}{2\sum_{j=0}^k \eta_j}.$$

The bounds $\sum_{j=0}^k \eta_j \ge 2\eta(\sqrt{k+2} - 1)$ and $\sum_{j=0}^k \eta_j^2 \le \eta^2(1 + \log(k+1))$ yield part (ii). If the rule stops, the shortfall is zero thereafter. See also Polyak (1987, Chapter 5).

The best iterates have objective values converging to $\mathcal{L}_\mu^*$. By continuity, every accumulation point in compact Π is a maximizer. Uniqueness therefore gives convergence. If μ supports the authority's optimum, strong duality makes that optimum a Lagrangian maximizer, proving part (iii). Strict feasibility guarantees supporting multipliers; convergence at a fixed μ does not use it. □

L Proofs for Section 8

Fix characteristics s and a cost parameter c. Write $T = \tau + T_0$ and let $\Phi(h, n)$ be the least burden at underreported base $h \in [0, x]$ and entity count n. We apply the monotonicity

theorem of Milgrom and Shannon (1994, Theorem 4) to $-\Phi$. Compactness and lower semicontinuity give attained minima. Increasing differences of Φ in (h, n) make $-\Phi$ supermodular in $(h, -n)$; decreasing differences make it supermodular in (h, n). For $\Phi = \Phi_0 + \theta(n-1)$, the objective has increasing differences between $(h, -n)$ and θ in the first case, and between (h, n) and $-\theta$ in the second. Increasing θ therefore raises extremal h and lowers extremal n in the first case, and lowers both in the second. For $\Phi = \Phi_0 + \rho h$ in the first case, use the order $(-h, n)$ to obtain the opposite responses to ρ. Applying the same theorem to the conditional choice of h at each n gives the comparisons with one-entity strategies.

For claims that hold under any selections, revealed preference gives the stronger own-price inequalities

$$(\theta' - \theta)(n' - n) \leq 0, \qquad (\rho' - \rho)(h' - h) \leq 0.$$

Each follows by adding the two optimality inequalities when only the corresponding parameter changes. Cross responses require extremal selections.

L.1 Proof of Proposition 14

Part (i). In the benchmark technology, a strategy with n entities and underreported base h costs at least $T_n(x-h) + c[(p_s + e_R)h + (\sigma_s + e_S)(n-1)]$, with equality for the reports that attain $T_n(x-h)$, so

$$\Phi(h, n) = T_{0,n}(x-h) + \rho h + \theta(n-1) + \tau, \qquad \rho = c(p_s + e_R), \quad \theta = \tau + c(\sigma_s + e_S).$$

The minimum defining $T_{0,n}(q)$ is attained because the feasible reports form a compact set, and $T_{0,n}$ is continuous in q. The fee enters only through θ, enforcement against splitting only through θ, and enforcement against underreporting only through ρ. By the monotonicity theorem, it remains to show that Φ has increasing differences, that is, that $T_{0,n+1}(q) - T_{0,n}(q)$ is nonincreasing in q for every feasible n, so that $\Phi(h, n+1) - \Phi(h, n)$ is nondecreasing in h.

For unrestricted divisions, Jensen's inequality gives $T_{0,n}(q) = nT_0(q/n)$. Its adjacent difference has derivative $t(q/(n+1)) - t(q/n) \leq 0$ almost everywhere. With a single division into two holdings, let $y_- = \min\{y_1, y_2\}$. Convexity makes the optimal reports as

equal as their holdings permit, so

$$T_{0,2}(q) = \begin{cases} 2T_0(q/2), & q \le 2y_-, \\ T_0(y_-) + T_0(q - y_-), & q \ge 2y_-. \end{cases}$$

Its derivative is $t(q/2)$ or $t(q - y_-)$, each at most $t(q)$. Thus $T_{0,2} - T_0$ is nonincreasing. These derivative comparisons integrate because the functions are absolutely continuous. This proves increasing differences in both cases.

Part (ii). In the convex-cost technology, a strategy with n entities and underreported base h costs at least $n\tau + t \cdot (x - h) + c[C_n(h) + (\sigma_s + e_S)(n-1)]$, where $C_n(h) = p_s h + \min\{\sum_j \psi_s(\eta_j) : \sum_j \eta_j = h,\ \eta_j \le y_j\}$ over feasible divisions. With unrestricted divisions, equal amounts $\eta_j = h/n$ can be underreported by entities with equal holdings $y_j = x/n$, since $h \le x$. Convexity of ψ_s gives a lower bound attained when the underreported amounts are equal. Thus $C_n(h) = p_s h + n\psi_s(h/n)$, and

$$\Phi(h, n) = t \cdot (x - h) + c\, p_s h + c\, n\psi_s(h/n) + \theta(n-1) + \tau, \qquad \theta = \tau + c(\sigma_s + e_S).$$

For unrestricted divisions, the adjacent cost difference has derivative $\psi_s'(h/(n+1)) - \psi_s'(h/n) \le 0$ almost everywhere. For a fixed two-entity division, the same clipped equal allocation just used for taxes applies to ψ_s: the derivative of $C_2(h) - C_1(h)$ is either $\psi_s'(h/2) - \psi_s'(h)$ or $\psi_s'(h - y_-) - \psi_s'(h)$, both nonpositive. Convexity gives absolute continuity on interior intervals, and continuity at the endpoints extends the comparison to $[0, x]$. Hence Φ has decreasing differences, and the monotonicity theorem gives the claims. The fee and enforcement against splitting enter only through θ.

□

L.2 Proof of Corollary 7

We need reporting comparisons that hold for every selection, so first establish almost-sure uniqueness of the one-entity report. Its burden is concave in c and therefore differentiable except on a Lebesgue-null set. The envelope theorem (Milgrom and Segal, 2002, Theorem 1) makes $(p_s + e_R)h$ the same at every minimizer wherever the derivative exists. Since $p_s > 0$ and costs have a density, the one-entity underreported base h^1 is unique almost surely. Proposition 14(i) then gives $h \le h^1$ for every optimal splitting strategy; a one-entity optimum has $h = h^1$. Integrating proves the reporting comparison in (i). The revenue and cost identities are Theorem 2(i).

For the comparative statics, the fee and enforcement against splitting raise θ, while enforcement against underreporting raises ρ. The revealed-preference inequalities above order entity counts and underreported bases, respectively, under any selections. Proposition 14(i) gives the cross responses under extremal selections. Integration gives the aggregate claims.

Selections are measurable by definition (Section 3). The optimal-strategy correspondence is measurable and compact-valued (Appendix F). Its largest and smallest optimal underreported bases and counts are therefore measurable functions of the firm and define selections.

For (ii), the same envelope argument applies to the one-entity effort $p_s h + \psi_s(h)$. This is strictly increasing in h, because $p_s > 0$ and a nonnegative convex ψ_s with $\psi_s(0) = 0$ is nondecreasing. Thus h^1 is again unique almost surely. Proposition 14(ii) gives $h \geq h^1$ under splitting, and its comparative statics give the remaining claims after integration. □

M Proofs for measurement

M.1 Proof of Proposition 15

(i) We integrate the envelope along the reform path. For each firm, $P(\lambda) = \min_\alpha\{\lambda T[\alpha] + cK(\alpha; e)\}$ is concave and Lipschitz, with $P(0) = 0$. Its selected tax payment $z(\lambda) = T[\alpha(\lambda)]$ is a supergradient and therefore nonincreasing. The envelope theorem and its integral formula (Milgrom and Segal, 2002, Theorem 2 and Corollary 4) give $P(1) = \int_0^1 z(\lambda)\,\mathrm{d}\lambda$, regardless of choices at kinks. Compliance gives $0 \leq z(\lambda) \leq T(x)$. If $\lambda < c$, the effective cost parameter is $c/\lambda > 1$, so no manipulation at the top forces $z(\lambda) = T(x)$. Dominated convergence gives $Z(\lambda) \to \mathbb{E}T(x)$. Fubini and $\mathbb{E}R = Z(1)$ now prove (60).

(ii) On $[\lambda_{k-1}, \lambda_k]$, monotonicity gives $Z(\lambda_k) \leq Z \leq Z(\lambda_{k-1})$, and on $(0, \lambda_1]$, $Z(\lambda_1) \leq Z \leq \mathbb{E}T(x) \leq \bar{B}$. Integrating gives $\underline{D} \leq \mathbb{E}D \leq \overline{D}$. The width is $\lambda_1[\bar{B} - Z(\lambda_1)] + \sum_{k\geq 2}(\lambda_k - \lambda_{k-1})[Z(\lambda_{k-1}) - Z(\lambda_k)]$, and the sum is at most $\max_k(\lambda_k - \lambda_{k-1})$ times the telescoping sum $Z(\lambda_1) - Z(1)$.

(iii) Under the selection rule of Section 3.3, firms with $c \geq \lambda$ comply and contribute zero to Y. For the remaining firms, substitution $c = \lambda v$ gives

$$\mathbb{E}_\lambda Y = \int\int_0^\lambda Y(s, \alpha_{s,c/\lambda})\gamma c^{\gamma-1}\,\mathrm{d}c\,\mathrm{d}H(s) = \lambda^\gamma \mathbb{E}_1 Y.$$

Apply this to $Y = T[\alpha] - T(x)$ to obtain the tax-payment formula in (61); almost-sure uniqueness of taxes makes this formula selection-independent. Integrating it in (60)

gives $\mathbb{E}D = \gamma[\mathbb{E}T(x) - Z(1)]/(1+\gamma)$ and hence the stated finite-reform formula. For an observed statistic with $\mathbb{E}_1 Y \neq 0$, $\mathbb{E}_\lambda Y / \mathbb{E}_1 Y = \lambda^\gamma$ identifies γ. Without one, the first identity at $\lambda_1 < \lambda_2 < 1$ gives the stated ratio $(1-\lambda_1^\gamma)/(1-\lambda_2^\gamma)$ whenever $Z(\lambda_2) \neq Z(1)$.

The derivative of the log ratio with respect to γ is $b(\lambda_1) - b(\lambda_2)$, where $b(\lambda) = -\lambda^\gamma \log\lambda/(1-\lambda^\gamma)$. Writing $y = -\log\lambda$ gives $b(\lambda) = y/(\exp(\gamma y) - 1)$. This expression decreases in y, so b increases in λ and the ratio strictly decreases in γ. The ratio tends to $\log\lambda_1/\log\lambda_2$ as $\gamma \downarrow 0$ and to 1 as $\gamma \to \infty$. □

M.2 Conditional exponents

When exponents differ across characteristics, the aggregate response is a mixture of the group responses in Proposition 15(iii). Its elasticity and curvature describe the composition of responding firms.

Proposition M.1 (Conditional scaling and aggregate restrictions)**.** *Maintain the strategy and policy assumptions and no manipulation at the top in Proposition 15. Let $F_s(c) = c^{\gamma_s}$ with $1 \leq \gamma_s \leq \bar{\gamma}$, and use the selection rule of Section 3.3. Scale the whole schedule by $\lambda \in (0,1]$, holding enforcement and technology fixed. For $Y = Y(s,\alpha)$ as in Proposition 15(iii),*

$$\mathbb{E}_\lambda[Y \mid s] = \lambda^{\gamma_s}\mathbb{E}_1[Y \mid s], \qquad \mathbb{E}_\lambda[D \mid s] = \lambda^{1+\gamma_s}\mathbb{E}_1[D \mid s]. \tag{81}$$

If $Y \geq 0$ and $\mathbb{E}_1 Y > 0$, write $a_s = \mathbb{E}_1[Y \mid s]$ and $Y_\lambda = \mathbb{E}_\lambda Y = \int a_s \lambda^{\gamma_s}\,\mathrm{d}H(s)$. Then

$$\lambda^{\bar{\gamma}} Y_1 \leq Y_\lambda \leq \lambda Y_1. \tag{82}$$

Define the probability measure $\mathrm{d}H_\lambda^Y(s) = a_s\lambda^{\gamma_s}\,\mathrm{d}H(s)/Y_\lambda$. The slope and curvature of the aggregate response satisfy

$$\frac{\mathrm{d}\log Y_\lambda}{\mathrm{d}\log\lambda} = \mathbb{E}_{H_\lambda^Y}\gamma_s, \qquad \frac{\mathrm{d}^2\log Y_\lambda}{\mathrm{d}(\log\lambda)^2} = \mathrm{Var}_{H_\lambda^Y}(\gamma_s) \geq 0. \tag{83}$$

The mix of splitting and underreporting is unchanged within a characteristic group. Aggregate composition changes when exponents differ. The bounds and curvature apply to nonnegative statistics such as splitting shares, exit shares, extra entity counts $(n-1)^+$ and amounts underreported. Curvature is strict when responding groups have different exponents. Realized private manipulation costs contain the private parameter c, so they have the extra power of λ in (81). With a common exponent, total private costs fall by the factor $\lambda^{1+\gamma}$.

If the data identify the relevant characteristic groups, a nonzero statistic in each group identifies its exponent from a causal reform. Proposition 15(iii) then recovers group manipulation costs, which can be aggregated. Without that group information, a single aggregate share ratio generally identifies neither a common exponent nor total manipulation costs. The full-path identity and finite-reform bounds remain available without a power law.

M.3 Proof of Proposition M.1

We apply the proportional-scaling argument conditional on s. The substitution $c = \lambda v$ proves the formula for Y in (81). For $D = cK$, it contributes one further factor of λ, proving the cost formula. Integrating over s and using $\lambda^{\bar{\gamma}} \leq \lambda^{\gamma_s} \leq \lambda$ proves the bounds.

Set $u = \log \lambda$. Bounded exponents justify differentiating $Y_{e^u} = \int a_s e^{\gamma_s u} \, \mathrm{d}H(s)$ twice. Its first logarithmic derivative is the mean of γ_s under H_λ^Y, and its second is the variance. This proves (83), with strict curvature exactly when responding firms have different exponents. □

M.4 The reform estimators in the main text

For a common exponent, (19) implies

$$\frac{\gamma}{1-\lambda^\gamma} \frac{M_r(\lambda) - M_r(1)}{M_r(1)} = \varepsilon_r.$$

At a fixed $\lambda < 1$ and positive base, consistent sample means therefore give consistency of (63) by the continuous mapping theorem. A consistent separately estimated exponent gives the same conclusion.

For the local estimator, the definition of the left derivative gives

$$\frac{M_r(\lambda) - M_r(1)}{(1-\lambda) M_r(1)} \longrightarrow \varepsilon_r.$$

When $M_r(1) > 0$, consistent sample means keep the estimated denominator away from zero in probability. Errors that are $o_p(1-\lambda)$ then leave the limit unchanged. Independent firm-level observations and bounded entity counts give errors of order $O_p(N^{-1/2})$, so $\sqrt{N}(1-\lambda) \to \infty$ is sufficient for independent samples of comparable size. Clustered assignment requires inference at the assignment level rather than treating an owner's entities as independent observations.

Nonproportional reforms. Nonproportional reforms still bound changes in manipulation costs. For a reform from T_0 to T_1 at fixed enforcement, let α_0 and α_1 be the same firm's choices before and after. Optimality under each schedule implies

$$T_0[\alpha_0] - T_0[\alpha_1] \;\le\; D_1 - D_0 \;\le\; T_1[\alpha_0] - T_1[\alpha_1], \tag{84}$$

provided the firm's costs are unchanged. The bounds also cover exit, with no entities and the same exit cost in both periods. Estimating their population averages requires a research design that identifies the reform's effect.

N Computations for the running example

N.1 Why the finite calculation is exact

For a firm of base x, let q denote its total report. The sequential technology with constant marginal reduction effort and fixed effort per split gives

$$\sum_j T(r_j) + c\big[(p_x + e_R)(x - q) + (\sigma + e_S)(n - 1)\big]$$

for every report vector with $q \le x$. Conversely, splitting the initial holdings and then reducing reports implements every such vector. Costs depend only on the total reduction and number of splits, so this representation preserves all feasible burdens.

Consider a general nonnegative two-bracket tax with threshold b. If $t_2 \ge t_1$, then

$$\min_{\substack{r_j \ge 0 \\ \sum_j r_j = q}} \sum_j T(r_j) = n\tau + t_1 q + (t_2 - t_1)(q - nb)^+.$$

Indeed, $\sum_j \min\{r_j, b\} \le \min\{q, nb\}$, and a division attains this upper bound. Adding the linear reduction cost makes the objective piecewise affine in q. A minimum occurs at 0, $\min\{nb, x\}$ or x. At $q = 0$, extra entities are dominated. If $n > \lceil x/b \rceil$, fewer entities can put all of q below the bracket threshold, preserving the smallest possible variable tax while lowering fees and split costs. Hence the finite candidate set is

$$\begin{aligned} &n = 1: && q \in \{0, \min\{b, x\}, x\}, \\ &2 \le n \le \lceil x/b \rceil: && q = nb \text{ if } nb < x, \quad \text{or } q = x. \end{aligned}$$

When $t_2 \le t_1$, the variable tax is concave and vanishes at zero. Merging all reports into one

entity weakly lowers tax and lowers splitting effort, so the one-entity choices suffice. Thus the same finite set attains the unrestricted minimum over the entire policy box, including decreasing rates. At $c = 0$, extra entities can tie when the fee is zero, but one of the listed strategies still attains the minimum.

For each listed n, q, the calculation uses $(b, \dots, b, q - (n-1)b)$ as one representative division. All minimum-tax divisions have the same bracket exposures when $t_1 \neq t_2$, and the displayed optimal policies have unequal rates. The calculations of total reports, count shares and burdens do not require a unique partition.

N.2 Assumptions and compact representation

The inequalities $p_x \geq L$ and $\sigma \geq Lx/2$ verify no manipulation at the top by the bound in Appendix B.1. Constant welfare weights and uniform costs satisfy the welfare restriction.

For a compact representation valid beyond two brackets, compare any plan to one entity reporting zero. Nonnegative marginal rates give a tax payment of at least $n\tau$, and

$$\text{burden(plan)} - \text{burden(zero)} \geq (n-1)\tau + c\big[\sigma(n-1) - (p_x + e_R)x\big].$$

Counts greater than $1 + (p_x + \bar{e}_R)x/\sigma$ are therefore dominated for $c > 0$; at $c = 0$ one zero-report entity is sufficient. The example needs at most 27 entities for this uniform representation across its entire enforcement rectangle and full capped schedule class. The much sharper bounds $1, 2, 4$ follow from the two-bracket tax geometry. Neither is a primitive constraint on the firm's feasible plans.

The characteristic distribution can be supported on three points of an interval. Interpolate x and p_x continuously between them. Each bounded-count report set is compact and varies continuously with the base. Public enforcement cost is convex and smooth. A flat rate 0.4 at zero enforcement raises $58/375 > 3/20$ with fee zero, establishing strict feasibility in each policy comparison. Since a flat schedule gives no tax saving from splitting in this technology, this check is common to the economies with and without splitting.

N.3 Exact integration and the global error bound

For any fixed policy, each candidate's burden is affine in c, with tax as its intercept and effort as its slope. Intersecting the inequalities against every other candidate gives its interval of optimal cost types. Integrating on these intervals is exact under the uniform

distribution. In particular,

$$\mathbb{E}R = 2\mathbb{E}P - \mathbb{E}T(x), \qquad \mathbb{E}D = \mathbb{E}T(x) - \mathbb{E}P.$$

The candidate set is policy-independent, so expected burden is concave in the policy vector z. For $\mu \geq 1/2$, the Lagrangian is

$$\mathcal{L}_\mu(z) = \mathbb{E}x + (2\mu - 1)\mathbb{E}P(z) - \mu[\mathbb{E}T_z(x) + A(z) + G].$$

Let H_μ be an upper bound on $\mathcal{L}_\mu$ over the policy box. Weak duality gives, for every revenue-feasible z,

$$0 \leq V(G) - W(z) \leq \max_{z'} H_\mu(z') - \mathcal{L}_\mu(z) + \mu[Q(z) - G]. \tag{85}$$

In this example, a common selection of minimum-burden strategies supplies a supergradient g. Concavity permits $H_\mu(z') = \mathcal{L}_\mu(z) + g \cdot (z' - z)$. Its maximum is explicit: choose each coordinate of z' at its upper bound when $g_j > 0$ and at zero otherwise.

Without splitting, e_S only raises public cost, so its optimum is zero. Eliminating it preserves the comparison with both instruments available. The independent checker treats each saved decimal policy and multiplier as an exact rational number. It verifies $Q(z) \geq G$ and calculates the bound without floating-point arithmetic. The five policies in Table 1 have error bounds below 9.46×10^{-12}. There is no cost-type discretization error. The certificate covers both rates, the fee and the permitted enforcement instruments at the fixed bracket threshold. It does not cover changes in the threshold or arbitrary nonlinear schedules.

N.4 A many-base version with freely chosen rates

Figure 1 uses the same benchmark manipulation technology with a richer distribution of potential bases. For $i = 1, \ldots, 40$, let

$$x_i = \frac{i}{40}, \qquad m_i = \frac{1/i}{\sum_{j=1}^{40} 1/j}, \qquad p(x_i) = 0.6(1 + 4x_i), \qquad \sigma = 0.3.$$

Cost parameters are uniform, welfare weights equal one, and there is no exit. The fee and enforcement are zero. The authority chooses any continuous tax schedule with marginal rates in $[0, 0.6]$. Progressivity is not required. The revenue requirement is $G \simeq 0.06$, with its exact rational value recorded as `G_exact` in

`code/running_examples/plateau_certificates.json`. Firms may use any finite sequence of reductions and splits, with continuous reports and arbitrary divisions. Every plan has effort $p(x)(x - \sum_j r_j) + \sigma(n-1)$, and every nonnegative report vector whose sum is at most x can be implemented.

Continuous reports and arbitrary entity counts. Let $h = 1/40$ and first consider a schedule that is affine between consecutive mesh points. For a fixed entity count, partition the report space into cells according to the mesh interval containing each report. Burden is affine on each cell intersected with $\sum_j r_j \le x$. At a minimizing vertex, all reports are cell endpoints, except possibly one when the sum constraint binds. That report equals x minus a sum of mesh points and is therefore itself on the mesh. Thus a minimizing report vector can be chosen on the mesh, even though reports are continuous.

A zero-report entity is removable if another entity remains. Positive reports on the mesh are at least h, so at most x/h entities are needed, with one entity reporting zero as the remaining possibility. Independently of the mesh, a plan with $n - 1 > p(x)x/\sigma$ costs more than full concealment through one entity for every $c > 0$: its splitting effort alone exceeds $p(x)x$ and its tax is nonnegative. At $c = 0$, reporting zero through one entity is optimal. Hence an optimal plan exists with

$$n \le \max\{1, \min\{x/h, \lfloor p(x)x/\sigma \rfloor + 1\}\}.$$

The resulting bound is 11 in this calibration. It follows from dominance.

For n positive-report entities and total report qh, compute the smallest tax payment recursively:

$$B_{n,q} = \min_{1 \le k \le q} \{T(kh) + B_{n-1,q-k}\}, \qquad B_{0,0} = 0,$$

with impossible states assigned $+\infty$. Add the single entity reporting zero separately. The firm's burden is the lower envelope of the finitely many lines

$$B_{n,q} + c\{p(x)(x - qh) + \sigma(n-1)\}.$$

The recursion permits every division and does not impose equal splits or a two-entity limit. If $B + cK$ minimizes burden on (a, b), its contributions to expected burden and receipts are $B(b-a) + K(b^2 - a^2)/2$ and $B(b-a)$, respectively. All reported outcomes integrate over cost types exactly.

Why the calculation covers the full schedule class. For any admissible schedule T, let S interpolate its values on the mesh. Each slope of S is an average slope of T and belongs to $[0, 0.6]$. Every potential base is a mesh point, so $S(x) = T(x)$. By the vertex argument,

$$P(c;T) \leq \min_{\text{mesh reports}} \{T[\alpha] + cK(\alpha)\} = \min_{\text{mesh reports}} \{S[\alpha] + cK(\alpha)\} = P(c;S).$$

The conditions $p(x) \geq L$ and $\sigma \geq LX/2$ hold here. As shown earlier in this appendix, they imply compliance at $c = 1$ for every schedule in the full class, including nonprogressive schedules. The accounting identity therefore holds throughout that class. For $\mu \geq 1$, and omitting the common constant $\mathbb{E}x - \mu G$,

$$\mathcal{L}_\mu(T) = (2\mu - 1)\mathbb{E}P(c;T) - \mu\mathbb{E}T(x) \leq (2\mu - 1)\mathbb{E}P(c;S) - \mu\mathbb{E}S(x) = \mathcal{L}_\mu(S).$$

A global bound on the mesh therefore bounds the entire capped nonlinear class. This argument uses a Lagrangian comparison. It does not assert that interpolation raises welfare and receipts separately, nor does it restrict feasible schedules or manipulations.

Independent global certificates. The independent checker integrates the saved feasible plans with exact rational arithmetic and forms convex combinations of their affine burden bounds. These remain upper bounds even when the proposed cost intervals are approximate. Substitution into the Lagrangian gives an affine upper bound H_μ, maximized explicitly over the rate box. Equation (85) then certifies welfare; the preceding interpolation argument extends the certificate to the full schedule class.

With $\mu = 1.3$, the bound for the displayed splitting policy is below 5×10^{-11}. The requirement is rounded down from that policy's receipts, and the certificate includes the resulting revenue slack. The policy without splitting meets the same requirement, with a positive rounding allowance for feasibility, and its independent welfare bound is below 10^{-12}. The harmonic-mean Ramsey formula is used to construct that candidate; the affine certificate verifies it directly. Both bounds cover the full nonlinear schedule class.