Online Appendix to Optimal Taxation When Firms Can Underreport and Split

Faheem Gilani

October 25, 2026

This appendix contains supplementary results, extensions and computational details. The paper's attached appendix provides the proofs and supporting arguments for its main results. References without an S prefix refer to the paper; results and equations numbered S belong to this appendix.

S.1 Further model and policy implications

S.1.1 Firms restricted to compliance

Firms restricted to compliance are already included: give them a separate characteristic type with $\mathcal{A}(s) = \{a^0(s)\}$ and any continuous cost distribution satisfying Assumption 3. Their cost parameter is irrelevant, their burden is $T(x)$, and the accounting and optimality formulas give their ordinary tax and welfare contributions. Atoms in the cost distribution of firms that can manipulate are different: they can place positive mass on ties with different tax payments and require a separate argument.

S.1.2 The policy class

The common schedule is without loss relative to a deterministic payment rule $\phi(r, z)$ with costless messages z that affect only payments. If messages are independently available to each entity at every report and minima are attained, $\widetilde{T}(r) = \min_z \phi(r, z)$ gives

$$\min_{z_1,\ldots,z_n} \sum_j \phi(r_j, z_j) = \sum_j \widetilde{T}(r_j).$$

Strategies and manipulation costs are unchanged, so firms face the same burden problem. This is the taxation principle (Hammond, 1979; Guesnerie, 1995). It does not establish the nonnegativity, fee cap, rate cap or enforcement restrictions of the paper's policy class. Ownership-based taxes, report-dependent audits and penalties, and costly messages are additional instruments outside that class.

Unbounded enforcement and coercivity. With the maintained compact set E, enforcement cannot diverge, even if A vanishes along a direction. If instead $e \in \mathbb{R}^I_+$, impose coercivity: $A(e) \to \infty$ as $\|e\| \to \infty$ within the nonnegative orthant. Continuity then makes the sublevel sets of A compact. Every revenue-feasible policy satisfies

$$A(e) \leq \mathbb{E}R - G \leq \bar{T} - G.$$

Choose a box whose upper faces lie strictly beyond the compact sublevel set $\{e : A(e) \leq \bar{T} - G\}$. Restricting enforcement to this box changes no feasible policy and leaves its artificial upper bounds inactive. The compact-policy results therefore apply. Convexity alone does not supply this bound. For example, $A(e) = (e_R - e_S)^2$ is zero along $e_R = e_S \to \infty$ and is not coercive. With unbounded enforcement this technology can provide arbitrarily strong enforcement for free and need not admit a finite optimizer.

Other compact policy classes. The class

$$\{T \in AC[0, X] : T \geq 0,\ 0 \leq T(0) \leq F,\ L_- \leq T' \leq L \text{ almost everywhere}\}, \tag{S.1}$$

where $L_- \leq 0$, is convex, compact in the uniform norm and contains zero. If the remaining general assumptions hold for it, Theorem F.1, Proposition F.2, Theorem 2, Proposition S.12.3, Proposition 7 and Proposition 15 apply unchanged. The approximation bounds in Theorem S.13.1 apply with $L - L_-$ replacing L in the general bound. The continuity bounds (S.64)–(S.68) and Proposition S.13.2 also apply; the condition $T \geq 0$ is closed, so the class with at most N brackets remains compact. The pointwise rate conditions need not hold: nonnegativity links rates across thresholds. The normal-cone condition remains valid.

S.1.3 Sufficient conditions for a strict multiplier

By Remark 1, positive manipulation cost at any optimum implies that every supporting multiplier exceeds w_-. Positive manipulation costs arise if (a) exit is available to almost every firm, (b) a one-entity zero report is feasible for almost every firm and the optimal

fee is zero, or (c) the optimal schedule is nondecreasing, the fee is positive, and exit is available on a set of characteristics of positive mass.

To verify these criteria, if $c > 0$ and $D = 0$, then $K = k_0 = 0$. No manipulation at the top then gives $P = T[\alpha] \geq T(x)$. Consequently $P < T(x)$ implies $D > 0$. Positive receipts and $R \leq T(x)$ imply $T(x) > 0$ on a positive mass of characteristics. In (a) or (b), their exit or zero-report strategy has zero tax and finite effort, so $P < T(x)$ for sufficiently small positive c. Positive conditional densities then give $\mathbb{E}D > 0$. In (c), $T(x) \geq \tau > 0$ on the positive mass with exit, and the same argument applies.

S.1.4 Report distributions, fees and exit

Proposition S.1.1 (Report distributions and participation)**.** *Maintain Assumption 3 except strict feasibility. Let π^* be optimal in the full schedule class $\mathcal{T}$, with a supporting multiplier $\mu \geq w_-$ and selection satisfying (75).*

(i) *If $0 < t^* < L$ almost everywhere on an open interval $O \subset (0, X)$, then, for every Borel $B \subset O$,*

$$\mathbb{E}[\Omega \#\{j : r_j \in B\}] = \mu \mathbb{E}[f_s(1)\mathbf{1}\{x \in B\}].$$

If no firm has potential base $q \in O$ with positive probability, the mass of firms with an entity reporting exactly q is zero.

(ii) *The fee score satisfies*

$$\Xi_\tau = \mu - \mathbb{E}w_s(c) + \mathbb{E}[\Omega(n-1)^+] - \mathbb{E}[\Omega\mathbf{1}\{n = 0\}]. \tag{S.2}$$

Thus no exit and $\mu > \mathbb{E}w$ imply $\tau^* = F$. This conclusion also holds for fixed brackets.

(iii) *Suppose $U_s(y) = wy$ and the exponent is a common $\gamma > 0$. Write $\beta = 1 - w/\mu$. If rates are interior almost everywhere, then, for every Borel $B \subset (0, X]$ and every $p > 0$,*

$$\mathbb{E}\sum_j \mathbf{1}\{r_j \in B\} = \frac{\gamma}{\gamma + \beta}\Pr(x \in B), \tag{S.3}$$

$$\mathbb{E}\sum_j r_j^p = \frac{\gamma}{\gamma + \beta}\mathbb{E}x^p. \tag{S.4}$$

If $x > 0$ almost surely and $Z = \mathbb{E}\sum_j \mathbf{1}\{r_j = 0\}$, then

$$\Pr(n = 0) + Z = \mathbb{E}(n-1)^+ + \frac{\beta}{\gamma + \beta}. \tag{S.5}$$

If also $F > 0$ and $\tau^* < F$, then $Z = 0$. For either schedule class, an interior fee requires

$\mathbb{E}n = \gamma/(\gamma+\beta)$.

(iv) *Suppose* $L > 0$, $U_s(y) = wy$, *the exponent is a common* $\gamma > 0$, $\beta > 0$ *and* $x > 0$ *almost surely. If firms never exit and every entity reports a strictly positive amount under the supporting selection, then* $t^* = L$ *almost everywhere on an interval immediately above zero.*

The entity and total-report targets of Corollary 3 also hold for the power distributions in part (iii), including $0 < \gamma < 1$, *with their stated interiority conditions.*

Proof. We recover the report measure from the rate scores, then use the fee score to count entry and exit. For part (i), the rate conditions give $\Xi_\mu((r, X]) = 0$ almost everywhere on O. Right-continuity extends this equality to every $r \in O$. Differences of tails identify the signed measure on half-open intervals, hence on all Borel subsets of O.

For $q \in O$ with $J(\{q\}) = 0$, the weighted mass of reports at q is zero. If $\mu > w_-$, the score restriction gives $\Omega \geq (\mu - w_-)(1 + \ell_s) > 0$, so the unweighted mass is zero as well. If $\mu = w_-$, Remark 1 gives $D = 0$ and hence $k_0 = 0$ almost surely. The full-class top condition in Appendix B then requires $\max_j r_j = x$ whenever $x > 0$. Since reports sum to at most x, every other report is zero. Thus a positive report q requires $x = q$, which has probability zero.

For part (ii), substitute $n = 1 + (n-1)^+ - \mathbf{1}\{n = 0\}$ and $\mathbb{E}\Omega = \mu(1 + \bar{J}) - \mathbb{E}w$ into the fee score. Without exit, $\Xi_\tau \geq \mu - \mathbb{E}w > 0$, so its optimality condition puts the fee at its cap.

For part (iii), $\Omega = \mu(\gamma + \beta)$ and $f_s(1) = \gamma$. The same tail argument on $(0, X)$, including the limit as $r \downarrow 0$, proves (S.3). Integrating r^p gives (S.4). The positive-report mass is $\mathbb{E}n - Z = \gamma/(\gamma+\beta)$ when $x > 0$ almost surely. The identity $\mathbb{E}n = 1 + \mathbb{E}(n-1)^+ - \Pr(n = 0)$ gives (S.5). The fee score is $\mu\{(\gamma+\beta)\mathbb{E}n - \gamma\}$, so an interior fee gives the stated count target. Under interior rates it equals $\mu(\gamma+\beta)Z$; if $F > 0$ and $\tau^* < F$, the fee condition requires this to be nonpositive, giving $Z = 0$. The proof of Corollary 3 uses only these score identities and Fubini, so it applies for every $\gamma > 0$.

Finally, in part (iv), $n_r \to n$ almost surely as $r \downarrow 0$. Bounded counts and dominated convergence give

$$\lim_{r \downarrow 0} \Xi_r = \mu\{(\gamma+\beta)\mathbb{E}n - \gamma\} \geq \mu\beta > 0, \tag{S.6}$$

because $n \geq 1$. The rate conditions therefore require the cap throughout an interval above zero. □

With positive total mass, (S.3) says that a randomly selected positive entity report has the distribution of potential bases conditional on being positive. These are aggregate restrictions. They use the common supporting selection and become independent of tie-breaking when almost every firm's report measure is unique. Fixed brackets impose only

integrated targets and can therefore induce bunching that part (i) rules out under freely chosen interior rates.

For heterogeneous power exponents and equal weights, an interior fee gives the participation balance

$$\mathbb{E}[(\gamma_s + \beta)\mathbf{1}\{n = 0\}] = \mathbb{E}[(\gamma_s + \beta)(n-1)^+] + \beta.$$

Under the interior-rate conditions of part (iii), a fee below its cap makes $Z = 0$, so this balance and (S.5) coincide for a common exponent. Part (iv) explains why fully interior rates require exit or zero reports when $\beta > 0$: without either, the optimum reaches the rate cap near zero.

S.1.5 One-sided enforcement responses

At a kink, let u_i^+ and u_i^- be the minimum and maximum of ck_i over minimizers. The directional-response proposition gives right and left burden derivatives u_i^+, u_i^-, and gross-receipt derivatives $\mathbb{E}[(1+\gamma_s)u_i^+]$, $\mathbb{E}[(1+\gamma_s)u_i^-]$. At an interior optimum the one-sided Lagrangian conditions are

$$\mathbb{E}[\Omega u_i^+] \leq \mu A_i \leq \mathbb{E}[\Omega u_i^-].$$

A supporting kernel can produce equality between these extremes without describing the response to an increase. Gross receipts remain nondecreasing in enforcement. The right derivative of net receipts need not be positive.

S.2 Further labor-supply examples

S.2.1 A quasilinear family

The broader quasilinear family $u(C, q) = C + (1 - c\eta)x(1-q) - c\kappa x(1-q)^2/2$ gives $K_x(h) = \eta h + \kappa h^2/(2x)$. Conditions $\eta > 0$, $\kappa \geq 0$ and $\eta + \kappa \leq 1$ make leisure desirable with weakly diminishing marginal utility. The restriction $L \leq \eta$ ensures no manipulation at the top. Wages in a compact subset of $(0, \infty)$ preserve continuity.

S.2.2 A general-density example matching aggregate earnings

An exact welfare representation differs from matching a tax response. Consider the logarithmic preferences used in the example of Mirrlees (1971, p. 193), with fixed wage w

and nonlabor income $B \in (0, w)$:

$$u(C, q) = \log C + \log(1 - q), \qquad C = (1 - t)wq + B.$$

The interior first-order condition yields

$$r_M(t) = \frac{w}{2} - \frac{B}{2(1-t)}, \qquad x = r_M(0) = \frac{w - B}{2}. \tag{S.7}$$

Put $\rho = 1 - B/w$. A one-entity burden model with no exit or enforcement and

$$K(h) = \rho h, \qquad F(c) = \frac{(1-\rho)c}{1-\rho c}, \qquad c \in [0, 1], \tag{S.8}$$

has firms report zero if $c < t/\rho$ and x if $c > t/\rho$. For $0 \le t \le L \le \rho$,

$$\mathbb{E}r(t) = x\{1 - F(t/\rho)\} = r_M(t), \qquad \mathbb{E}T(r(t)) = t\, r_M(t). \tag{S.9}$$

The density is not a power law, but

$$f(c) = \frac{1-\rho}{(1-\rho c)^2}, \qquad \ell(c) = \frac{1+\rho c}{1-\rho c} \ge 1,$$

so it satisfies Assumption 3 with equal welfare weights. Also $K(h) \ge Lh$ verifies the upper cost boundary. For $w = 3$, $B = 1$, $L = 0.5$ and $G = 0.1$, the policy $t = 0.2$ raises $7/40 > G$. The remaining regularity conditions follow from the compact one-entity strategy set.

This construction matches aggregate earnings and receipts along proportional reforms with B fixed. It does not identify welfare. The reduced-form economy has heterogeneous discrete choices; the logarithmic workers have an interior labor choice and income effects.

S.3 Further sequential technologies

The operations and effort notation are those of Appendix D. The results below allow state-dependent splitting costs beyond the report-dependent continuation costs in the main recursive characterization.

Example S.3.1 (Order of operations)**.** A firm with $x = 1$ ends with two entities that report 0.2 each. A split requires effort 0.1. Underreporting one unit requires effort equal to the entity's holdings: 1 in the parent and 0.5 in each entity after a split into halves. Splitting into halves and then underreporting 0.3 in each requires $0.1 + 2(0.5)(0.3) = 0.4$.

Underreporting 0.6 in the parent and then splitting requires $0.6 + 0.1 = 0.7$. The reports are identical, but the effort differs. The cheaper order can also reverse. If a split requires effort 0.5 times the group's report at the time of the split and underreporting requires 0.5 per unit, splitting first requires 0.8, whereas underreporting first requires 0.5.

Without positive effort per split, a finite bound preserving unrestricted burdens need not exist, even though every fixed bound gives a compact strategy set. For example, set the split effort to zero and let reduction effort be generated by $C(q) = q + 6q^2$, as in Example 2 at zero enforcement. Dividing a unit base equally among n entities and concealing everything requires effort $1 + 6/n$, which approaches 1 without reaching it. Under $T(r) = r$ and $0 < c < 1$, the burden has infimum c but no finite plan attains it: any reported base costs more per unit, and any positive concealed amount has a strictly positive quadratic cost. Positive split effort prevents this failure of attainment.

Two-step plans. For some technologies, a firm loses nothing by doing all of its splitting before all of its underreporting, or the reverse.

Proposition S.3.1 (Two-step plans)**.** *In the sequential manipulation technology, suppose that every coefficient of $\sigma_s(x, q_G, q, z; e)$ is nonincreasing in the report q of the splitting entity and nondecreasing in the transfer z. Take any plan with n entities, let y_j, η_j and $q_j = y_j - \eta_j$ be the final holdings, accumulated underreporting and reports of entity $j = 1, \dots, n$, and let $\eta_G = \sum_j \eta_j$.*

(i) *Suppose also that every coefficient of σ_s is nonincreasing in the group's total report q_G, and that every coefficient of $p_s^E(y, \eta; e)$ is nondecreasing in holdings y. For some ordering $j_1, \dots, j_n$, the following plan preserves the original reports without raising any effort coefficient. The parent first creates entities with holdings $y_{j_2}, \dots, y_{j_n}$, in that order, and keeps y_{j_1}. Each entity j then underreports η_j. Its effort is*

$$\sum_{k=2}^{n} \sigma_s\Big(x, x, x - \sum_{l=2}^{k-1} y_{j_l}, y_{j_k}; e\Big) + \sum_{j=1}^{n} \int_0^{\eta_j} p_s^E(y_j, v; e)\, \mathrm{d}v + \int_0^{\eta_G} p_s^G(v; e)\, \mathrm{d}v. \tag{S.10}$$

(ii) *Suppose instead that σ_s does not depend on q, written $\sigma_s(x, q_G, z; e)$, that every coefficient of σ_s is nondecreasing in q_G, and that $p_s^E \equiv 0$. The following plan preserves the original reports without raising any effort coefficient. The parent first underreports η_G and then creates entities with reports $q_2, \dots, q_n$. Its effort is*

$$\int_0^{\eta_G} p_s^G(v; e)\, \mathrm{d}v + \sum_{j=2}^{n} \sigma_s(x, x - \eta_G, q_j; e). \tag{S.11}$$

(Proof in Appendix S.3.1.) In case (i), a firm can split first and then underreport. Split costs increase with the amount transferred and decrease with the splitting entity's report. These conditions can arise when a new entity must take over a real line of business. Underreporting is also costlier for larger entities. Moving underreporting after the splits therefore weakly lowers its cost and leaves larger reports at which to make the splits.

In case (ii), a firm can underreport first and then split. Underreporting is a group-wide operation, such as keeping a single set of false accounts. A new entity attracts more scrutiny when the group's reported base is larger, as under rules for assessing related companies together. Underreporting first lowers the cost of later splits without changing underreporting costs. A constant splitting cost $\sigma_s \equiv \bar{\sigma}_s > 0$ with $p_s^E \equiv 0$, that is, a fixed cost per additional entity together with any group-level underreporting cost, satisfies both sets of conditions.

The two cases of Example S.3.1 satisfy parts (i) and (ii), respectively.

Without these conditions, neither order need be optimal. Let $\sigma_s = 0.1 + 0.5q_G$, $p_s^E = y$ and $p_s^G = 0$. Splitting is then costlier when the group reports more, and underreporting is costlier in larger entities. A firm with $x = 1$ ends with two entities reporting 0.3 and 0.2. If the firm splits first, the split requires effort 0.6, and the best division of holdings, $21/40$ and $19/40$, brings the total effort to $679/800$. Underreporting 0.5 first and then splitting requires $0.5 + 0.35 = 0.85$. The cheapest plan underreports 0.35 in the parent, creates an entity with holdings 0.35 at an effort of $17/40$, and lets that entity underreport 0.15 at an effort of $21/400$, for a total of $331/400$. The parent underreports before the split to lower the scrutiny of the split, and the new entity underreports after it because it is small.

Proposition S.3.1 simplifies the firm's problem in examples such as Figure S.1. It also verifies the merging condition of Proposition 7.

A two-instrument example. Take no exit, a zero fee, rates $0 \leq t \leq L$, and the effort

$$K = (p + e_R)\,\varphi\Big(x - \sum_j r_j\Big) + (\sigma + e_S)(n-1),$$

where e_R and e_S are enforcement against underreporting and against splitting and $p, \sigma > 0$ are the corresponding efforts at zero enforcement. Underreporting requires effort proportional to $\varphi(h)$, where h is the amount underreported, and each additional entity incurs a fixed effort. Let φ be continuous and increasing on $[0, X]$ with $\varphi(0) = 0$. Example 1 has this form with linear φ, a positive entity fee and a firm-specific marginal effort of underreporting. Its divisions are unrestricted; Appendix N derives a count bound valid uniformly over policies. Compare any strategy with the one-entity strategy that reports a

minimizer of T on $[0, x]$. The latter pays no more tax, because taxes are nonnegative, and its manipulation cost is at most $c(p + e_R)\varphi(X)$. For $c > 0$, every optimal strategy therefore satisfies

$$n - 1 \leq \frac{(p + e_R)\varphi(X)}{\sigma + e_S},$$

which is uniform when enforcement is bounded. At $c = 0$, a one-entity strategy attains the same minimum tax. Restricting to this bound therefore preserves every burden. Reports now range over a finite union over n of the simplices $\{r \in \mathbb{R}^n_+ : \sum_j r_j \leq x\}$. The effort coefficients are continuous in reports, so the strategy set is compact. If φ is convex, $p\varphi'(0) \geq L$ and $\sigma \geq LX/2$, then (6) follows from

$$Lh + \frac{LX}{2}(n-1) \;\geq\; L\Big(x - \max_j r_j\Big), \qquad h = x - \sum_j r_j,$$

because $\varphi(h) \geq \varphi'(0)h$. For $n \geq 2$, use $\sum_j r_j - \max_j r_j \leq x(n-1)/n \leq X(n-1)/2$, and for $n = 1$ the two sides are equal. Convexity is used only in this sufficient condition. A unit of e_R raises a firm's manipulation cost by $c\varphi(h)$ and a unit of e_S raises it by $c(n-1)$, so e_R raises the cost of underreporting and e_S the cost of creating entities.

S.3.1 Proof of Proposition S.3.1

We construct plans that preserve terminal reports and weakly lower each effort coefficient. Number entities by creation, with the parent first. Holdings fall only when an entity splits; its accumulated underreporting rises only when it reduces its own report. Costs can be compared coefficient by coefficient because the constructions do not depend on enforcement.

Part (i). The first step moves all underreports to the end. Fix a plan ω. Let ω' perform the splits of ω in the same order, by the same entities and with the same transfers z_k, and then let each entity j underreport η_j. Holdings do not depend on underreports, so at each split the splitting entity has the same holdings in both plans. In ω' its report equals its holdings. This is at least its report in ω and hence at least z_k. Thus ω' is feasible and preserves final holdings and reports. At each split the group's total report is x in ω' and at most x in ω, and the splitting entity's report is weakly larger in ω'. Since σ_s is nonincreasing in q_G and in q, every split is weakly cheaper. The group-level cost of underreporting is $\int_0^{\eta_G} p^G_s$ in both plans. Entity j reaches accumulated underreporting v in ω with holdings at least y_j. In ω' its holdings are y_j. Since p^E_s is nondecreasing in holdings, its entity-level cost is

weakly lower in ω'.

The second step makes the parent the only entity that splits. Suppose a splitting entity i creates another splitting entity i'. Write $y_i(k), y_{i'}(k)$ for their holdings after split k, and let $y_\wedge$ be the smaller of their final holdings. Replace their creation split by a transfer of $y_\wedge$. The recipient never splits; the other entity performs all subsequent splits originally assigned to either one, with the original transfers. After each such split its holdings are

$$y_i(k) + y_{i'}(k) - y_\wedge \;\geq\; \max\{y_i(k), y_{i'}(k)\},$$

since both original holdings remain at least their final values. Every later transfer is therefore feasible, and its splitting entity has a weakly larger report. The creation transfer is weakly smaller. Monotonicity of σ_s in reports and transfers makes every split weakly cheaper. Final holdings are preserved up to labels, so assigning each holding its original underreport preserves reduction costs and final reports. This removes one splitting entity. Repeating leaves only the parent, giving (S.10) in the order in which it creates the final entities.

Part (ii). Let ω'' underreport η_G in the parent and then create, for each entity $j \geq 2$ of ω, an entity reporting q_j. The parent keeps $x - \eta_G - \sum_{j\geq 2} q_j = q_1$, so the reports are unchanged. Underreporting costs are $\int_0^{\eta_G} p_s^G$ in both plans because $p_s^E \equiv 0$. In ω, entity j was created with a transfer $z_j \geq q_j$, since reports never increase. At that time the group reported at least its final total, $x - \eta_G$. Since σ_s does not depend on q and is nondecreasing in q_G and in z, the cost $\sigma_s(x, x - \eta_G, q_j; e)$ of creating j in ω'' is no larger than its cost in ω. □

S.4 Further implications of splitting

S.4.1 Split-rate restrictions with discrete potential bases

With finitely many potential tax bases, the counterpart of Proposition 4(ii)–(iii) bounds average rates between adjacent bases.

Lemma S.4.1 (Split rates with finitely many potential tax bases)**.** *Maintain the assumptions of Proposition 4 except full support. Let $x_1 < x_2$ be potential tax bases in $(\underline{x}, X]$ at which part (i) holds, and put $\delta = x_2 - x_1$. Let $r_{\max}$ be the largest report of a minimizing division of x_1, and $r'_{\max}$*

the largest report of a minimizing division of x_2. If $\delta \leq r'_{\max}$, then

$$\begin{aligned}\frac{T^*(r'_{\max}) - T^*(r'_{\max} - \delta)}{\delta} &\leq \frac{T^*(x_2) - T^*(x_1)}{\delta} - \sigma \frac{\kappa(x_2) - \kappa(x_1)}{\delta} \\ &\leq \frac{T^*(r_{\max} + \delta) - T^*(r_{\max})}{\delta}.\end{aligned}$$

Proof. Part (i) at x_1 and x_2 gives $T^*(x_2) - T^*(x_1) - \sigma(\kappa(x_2) - \kappa(x_1)) = \check{T}(x_2) - \check{T}(x_1)$. Raising the largest report of the division of x_1 by δ gives a division of x_2, so $\check{T}(x_2) \leq \check{T}(x_1) + T^*(r_{\max} + \delta) - T^*(r_{\max})$. Lowering the largest report of the division of x_2 by δ gives a division of x_1, which is feasible by $r'_{\max} \geq \delta$, so $\check{T}(x_1) \leq \check{T}(x_2) - T^*(r'_{\max}) + T^*(r'_{\max} - \delta)$. Dividing by δ gives the two inequalities. □

The condition $\delta \leq x_1/(n-1)$ is sufficient because $r'_{\max} \geq x_2/n$. Each side of the lemma is an average marginal rate over an interval of length δ. With equal weights the terms in κ vanish, so the average rate on $(x_1, x_2]$ lies between the average rates on $(r'_{\max} - \delta, r'_{\max}]$ and $(r_{\max}, r_{\max} + \delta]$, and it equals that rate when the rate takes the same value on both intervals. With equal weights, if the equal divisions are minimizing at both bases, then $\check{T}(x) = nT^*(x/n)$ at both, and the average rate on $(x_1, x_2]$ equals the average rate on $(x_1/n, x_2/n]$.

S.4.2 Relation to the running example

The three-base economy of Example 1 illustrates rate compression and incidence in a two-bracket policy class. Appendix N proves that the finite calculation covers arbitrary finite manipulation plans and certifies the optimum. The example lacks the continuous support and independently movable rates required for pointwise equality in Proposition 4. Lemma S.4.1 gives the corresponding restriction for discrete bases.

S.4.3 The tax saving retained by splitters

Corollary S.4.1 (A retained tax saving)**.** *Under the maintained assumptions, consider an optimum with $e_S > 0$. Suppose $k_S = 0$ on one-entity and exit strategies, $k_S > 0$ on split strategies, and every split strategy has a feasible one-entity counterpart with the same total report and no greater baseline or other enforcement effort coefficient. Then almost every selected splitter has*

$$B_{\text{split}} = T\left(\sum_j r_j\right) - \sum_j T(r_j) > 0. \tag{S.12}$$

If a positive mass splits, the optimal schedule is not subadditive. Set $B_{\text{split}} = 0$ for firms that do not split. If both small increases and decreases in e_S are feasible, then, using the same supporting selection as in its optimality condition,

$$\mathbb{E}\left[\frac{\Omega}{\mu} B_{\text{split}}\right] \geq e_S A_S(e). \tag{S.13}$$

With equal linear welfare weights in the main model, $\Omega/\mu = \gamma_s + \beta$, where $\beta = 1 - w/\mu$. With a common exponent this becomes $\mathbb{E}B_{\text{split}} \geq e_S A_S/(\gamma + \beta)$.

Proof. We compare a split with its one-entity counterpart. Merging eliminates $k_S > 0$ and weakly reduces the other coefficients, so it saves at least $ce_S k_S$ in private cost. Optimality of the selected split requires $B_{\text{split}} \geq ce_S k_S > 0$ for $c > 0$. On non-split strategies $k_S = 0$, so the weak inequality holds for every selected strategy almost surely. Multiply it by $\Omega/\mu \geq 0$, average and use the interior enforcement condition $\mathbb{E}[\Omega c k_S]/\mu = A_S$. □

Proposition 9 supplies the condition $e_S > 0$ under a strict welfare loss and zero initial marginal public enforcement cost. The present result itself does not need those additional assumptions. The premium holds a firm's total report fixed. It is not the revenue that the authority could recover by preventing splitting, since the firm could then change that report. If the public cost of anti-splitting enforcement is a separable quadratic component $A^S(e_S)$, then $e_S A_S = 2A^S(e_S)$. The result also holds under the general-density assumptions, with $\Omega/\mu = \ell_s(c) + \beta$ under equal linear weights in (S.13).

S.5 Further results on lower-bracket relief

This appendix develops the marginal and finite revenue effects of a lower-bracket cut beyond Proposition 6. The arguments allow exit, heterogeneous welfare weights and costs that depend on the order of manipulation choices.

S.5.1 The reform and firms' exposure to it

Fix $z \in (0, X)$, put $b_z(r) = \min\{r, z\}$, and consider

$$T_a(r) = T(r) - a b_z(r), \qquad 0 \leq a \leq \bar{a}, \tag{S.14}$$

where every schedule along the path is admissible. The cut lowers marginal rates below z and leaves higher rates, the fee and enforcement unchanged. Progressivity is not required.

For a chosen strategy, define the total report $q = \sum_j r_j$ and its *exposure* to the concession,

$$B_z = \sum_j b_z(r_j) = \int_0^z n_r \, \mathrm{d}r.$$

Holding this strategy fixed, a unit cut costs B_z in revenue. Two nonnegative quantities compare its exposure with full reporting through one entity:

$$\delta_z = B_z - b_z(q), \qquad \eta_z = b_z(x) - b_z(q). \tag{S.15}$$

Splitting adds δ_z to exposure at a given total report; underreporting removes η_z. Exit gives $q = B_z = \delta_z = 0$ and $\eta_z = b_z(x)$. Splitting adds no exposure if the total report is below z.

Write P_a and $\bar{R}(a) = \mathbb{E}R(T_a, e)$ for burden and gross receipts, and use superscript a for choices under T_a. A policy is *regular for this reform* if almost every firm has the same B_z at all its optimal strategies. This condition permits an envelope derivative without requiring unique strategies or continuous reports.

S.5.2 Marginal and finite revenue effects

The initial mechanical estimate of a cut's cost is $a\mathbb{E}B_z^0$. A marginal comparison at a later policy instead uses its current exposure $\mathbb{E}B_z^a$. The distinction matters because firms can change their exposure along the reform path.

Proposition S.5.1 (Revenue effects along the reform path)**.** *Maintain the regularity, power-law and no-manipulation-at-the-top assumptions. For every admissible path* (S.14)*:*

(i) *Each firm's exposure B_z^a is nondecreasing in a. At every regular point,*

$$-\bar{R}'(a) = \mathbb{E}B_z^a + \mathcal{F}_z(a), \qquad \frac{\mathrm{d}\mathbb{E}D_a}{\mathrm{d}a} = \mathcal{F}_z(a), \tag{S.16}$$

where

$$\mathcal{F}_z(a) = \mathbb{E}[\gamma_s(B_z^a - b_z(x))] = \mathbb{E}[\gamma_s(\delta_z^a - \eta_z^a)] \tag{S.17}$$

is nondecreasing. The marginal behavioral effect can therefore change from offsetting the current mechanical loss to increasing it at most once, allowing an interval of equality.

(ii) *The additional private gain from changing strategy is*

$$\Delta_a = P_0 - P_a - aB_z^0 \geq 0. \tag{S.18}$$

For any measurable initial optimal selection,

$$\bar{R}(0) - \bar{R}(a) - a\mathbb{E}B_z^0 = a\mathcal{F}_z(0) + \mathbb{E}[(1+\gamma_s)\Delta_a]. \tag{S.19}$$

If firms can operate at most one entity, including the option to exit, then

$$\bar{R}(0) - \bar{R}(a) \leq a\mathbb{E}B_z^0. \tag{S.20}$$

With splitting, this bound can fail even when the initial policy is globally optimal and no firm initially splits.

(iii) *Receipts $\bar{R}(a)$ are concave. Hence the excess loss*

$$\mathcal{X}(a) = \bar{R}(0) - \bar{R}(a) - a\mathbb{E}B_z^0 \tag{S.21}$$

is convex, vanishes at zero, and has nondecreasing secant slopes $\mathcal{X}(a)/a$ for $a > 0$. Once a cut exceeds its initial mechanical estimate, every larger admissible cut does too.

The local formula separates the extra exposure from splitting from the exposure lost through underreporting and exit. Both terms use firms' own cost exponents. The finite formula also includes gains from switching strategies. It remains valid across reporting jumps and organizational thresholds. Neither comparison requires optimality or the welfare-weight restriction. Since enforcement is fixed, the same comparisons apply to net receipts.

Cuts from an optimum. Suppose the initial policy is regular and optimal, and the rates below z can move together in both directions. Integrating their supporting conditions gives

$$\mathbb{E}[\Omega_s^0(c)B_z^0] = \mu\mathbb{E}[\gamma_s b_z(x)]. \tag{S.22}$$

Thus, with $w_s^0(c) = U_s'(x - P_0)$,

$$-\bar{R}'(0) = \frac{\mathbb{E}[w_s^0(c)B_z^0]}{\mu}. \tag{S.23}$$

If $\mathbb{E}B_z^0 > 0$, a marginal cut costs less than its mechanical estimate precisely when the exposure-weighted welfare weight is below μ. For equal weights w, put $\beta = 1 - w/\mu$. The finite identity becomes

$$\mathcal{X}(a) = -a\beta\mathbb{E}B_z^0 + \mathbb{E}[(1+\gamma_s)\Delta_a]. \tag{S.24}$$

Switching gains must outweigh the initial fiscal offset before the finite cut exceeds its mechanical estimate.

More generally, absolute continuity gives

$$\Delta_a = \int_0^a (B_z^u - B_z^0)\,\mathrm{d}u. \tag{S.25}$$

Consequently $\mathcal{X}(a) > 0$ exactly when

$$\frac{1}{a}\int_0^a \mathbb{E}[(1+\gamma_s)(B_z^u - B_z^0)]\,\mathrm{d}u > \mathbb{E}\left[\left(1 - \frac{w_s^0(c)}{\mu}\right) B_z^0\right]. \tag{S.26}$$

If the right side is strictly positive, sufficiently small cuts cost less than their mechanical estimate. With equal weights, a common exponent γ and $\mathbb{E}B_z^0 > 0$, the finite and marginal reversal conditions are, respectively,

$$\frac{a^{-1}\int_0^a \mathbb{E}B_z^u\,\mathrm{d}u}{\mathbb{E}B_z^0} > 1 + \frac{\beta}{1+\gamma}, \tag{S.27}$$

$$\frac{\mathbb{E}B_z^a}{\mathbb{E}B_z^0} > 1 + \frac{\beta}{\gamma}. \tag{S.28}$$

The first compares the full cut with its initial mechanical estimate. The second, at a regular point, asks whether a further small cut costs more than the current mechanical estimate. With heterogeneous exponents, the second condition is $\mathbb{E}[\gamma_s(B_z^a - B_z^0)] > \beta\mathbb{E}B_z^0$.

Proof of Proposition S.5.1 and the optimal-policy comparisons. We establish monotonicity by revealed preference, then apply the burden envelope and accounting identities. For $a < b$, the two 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. The infimum of the affine strategy burdens is concave in a, and $0 \le B_z \le mz$ makes it uniformly Lipschitz. The envelope theorem gives $P_a' = -B_z^a$ at regular points and almost everywhere along each firm's reform path. Differentiating

$$\bar{R}(a) = \mathbb{E}[(1+\gamma_s)P_a] - \mathbb{E}[\gamma_s T_a(x)] \tag{S.29}$$

and the private-cost identity proves part (i). Bounded exposures and exponents justify

differentiation under the integral. Integrating $P_a' = -B_z^a$ proves (S.25) and gives

$$\bar{R}(0) - \bar{R}(a) = \int_0^a \{\mathbb{E}B_z^u + \mathcal{F}_z(u)\}\,\mathrm{d}u. \tag{S.30}$$

Fubini's theorem permits either order of integration. The integral uses current exposure at each successive policy.

Keeping the initial strategy gives $\Delta_a \geq 0$. Subtracting the accounting identities proves (S.19). For the one-entity bound, the displayed revealed-preference inequalities give $B_z^a \geq B_z^0$. An initially exiting firm has zero receipts and exposure, so its nonnegative new receipts satisfy the bound. If an initially operating firm's exposure increases, its new report exceeds its old report q_0. The new nondecreasing schedule then gives

$$R_a = T_a(q_a) \geq T_a(q_0) = R_0 - aB_z^0.$$

If exposure is unchanged, the new strategy is also optimal at the old policy. Almost-sure selection independence of old taxes gives the same inequality with equality, including a switch to exit. Averaging proves (S.20). This argument needs neither the power law nor restrictions on welfare weights. The counterexample below establishes failure with splitting.

Equation (S.29) makes receipts concave. Convexity of $\mathcal{X}$ and monotonicity of its secant slopes prove part (iii). At a regular optimal start, integrating the rate conditions gives (S.22), so

$$\mathcal{F}_z(0) = -\mathbb{E}\left[\left(1 - \frac{w_s^0(c)}{\mu}\right) B_z^0\right].$$

Substitution proves (S.23)–(S.24). Combining it with (S.25) gives (S.26). A strictly negative $\mathcal{X}_+'(0) = \mathcal{F}_z(0)$ gives the stated initial interval. Under equal weights, $\mathcal{F}_z(a) = \mathbb{E}[\gamma_s(B_z^a - B_z^0)] - \beta\mathbb{E}B_z^0$. Division by the relevant positive exposure gives the two common-exponent thresholds. □

Ties and general conditional densities. At a nonregular policy, the right and left burden derivatives use $-\max B_z$ and $-\min B_z$ over optimal strategies. The corresponding one-sided receipt derivatives follow from the accounting identity. A supporting kernel can average these exposures, so its Ramsey condition need not describe either actual directional response. The finite identities remain valid for any measurable initial optimal selection, including a randomized selection.

Under Assumption 3, all the results above extend with $\ell_s(c)$ replacing γ_s and $f_s(1)T_a(x)$ as the accounting boundary term. Indeed, $\mathbb{E}[\ell_s(c) \mid s] = f_s(1)$, so the feedback is

$\mathbb{E}[\ell_s(c)(B_z^a - b_z(x))]$. Its monotonicity follows from $\ell_s(c) \geq 0$. Integrability of ℓ_s and bounded exposures justify the same differentiation and integration steps. Only the ratio formulas (S.27)–(S.28) require a common power exponent.

An isolated organizational switch. Suppose two strategies determine the lower envelope near an interior cutoff $c_*(a)$, with no third-strategy tie there. The split strategy has additional effort $\Delta K > 0$ and exposure $\Delta B > 0$. Indifference gives $c_* = (R_U - R_S)/\Delta K$ and $c_*' = \Delta B/\Delta K$. Its moving-boundary contribution to the revenue loss is therefore

$$-(R_S - R_U) f_s(c_*) c_*' = c_* f_s(c_*) \Delta B = \gamma_s c_*^{\gamma_s} \Delta B. \tag{S.31}$$

This loss can occur without a fall in total reports. Direct tax changes enter the mechanical term, and other switches may offset this contribution. The formula concerns switches that raise effort; splitting can instead replace costly underreporting.

S.5.3 Relief from an optimum with no splitting

This auxiliary counterexample establishes the stronger claim that relief can have excess fiscal cost even when no firm splits at the initial optimum. All firms have $x = 1$, c is uniform on $[0, 1]$, and welfare weights equal one. The schedule class has $F = 0$ and rate cap $L = 1$. Every finite report vector with $\sum_j r_j \leq 1$ is feasible, at effort

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

There is no exit. Let $e \in [0, 1]^2$ and $A(e) = e_R + e_S$. The revenue requirement is $G = 19/40$. Every division and operation order is feasible. Cost depends only on total reduction and entity count.

Assumptions and global optimality. Complete reduction through one entity dominates counts above 23 uniformly over the policy class. A compact representation therefore preserves the unrestricted minimum. No manipulation at the top holds through one entity because $10(1 - q) \geq 1 - q$. For any strategy with $n \geq 2$ and underreported base $h = 1 - \sum_j r_j$,

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

Thus baseline effort exceeds the largest tax saving under the full capped class. Uniform costs and equal weights satisfy the restriction on welfare weights. The flat schedule $T(r) = 0.6r$ with zero enforcement raises $141/250 > G = 19/40$, establishing strict feasibility.

At $T(r) = r/2$ and $e = 0$, splitting is strictly dominated for $c > 0$. One-entity reporting is zero below $c = 1/20$ and one above it. Receipts are $19/40$. At every threshold $r \in (0, 1)$, $\mathbb{E}n_r = 19/20$ and $\Pr(x > r) = 1$. The supporting multiplier and the net values of the marginal rates are

$$\mu = \frac{19}{18}, \qquad \beta = \frac{1}{19}, \qquad \Omega = \frac{10}{9}, \qquad \Xi_r = \frac{10}{9}\frac{19}{20} - \frac{19}{18} = 0. \tag{S.32}$$

The net values of enforcement are

$$g_R = \Omega \int_0^{1/20} c \, \mathrm{d}c - \mu < 0, \qquad g_S = -\mu < 0.$$

The fee cap is zero, so the fee condition does not apply. The budget binds and all net values satisfy Theorem 1. The policy is therefore globally optimal over the full rate class and enforcement choices.

Choices after the cut. Set $z = 1/2$ and $0 \le a \le 1/2$. The new rates are $t_1 = 1/2 - a$ and $t_2 = 1/2$. Through one entity the objective is linear on each bracket, so an optimum is among reports $0, 1/2, 1$. For any number of entities, the finite-reduction argument of Appendix N, with threshold $1/2$, shows that at most two entities are needed. Two entities can put the entire base in the lower bracket, so further splits add effort without reducing tax. Partial two-entity reports are dominated by a one-entity report or by the full division $(1/2, 1/2)$. The relevant strategies are therefore

Reports	$R(a)$	K	B_z
(0)	0	10	0
$(1/2)$	$1/4 - a/2$	5	$1/2$
(1)	$1/2 - a/2$	0	$1/2$
$(1/2, 1/2)$	$1/2 - a$	$1/2$	1

For $0 \le a \le 1/20$, there is no splitting. The active cutoffs are $(1/2 - a)/10$ and $1/20$. For $a > 1/20$, the successive strategies in the table are (0), $(1/2)$, $(1/2, 1/2)$, and (1), with

cutoffs

$$c_0 = \frac{1/2 - a}{10}, \qquad c_1 = \frac{1/2 - a}{9}, \qquad c_2 = a. \tag{S.33}$$

The splitting interval opens precisely when $c_1 < c_2$, or $a > 1/20$. Its mass is

$$S(a) = \begin{cases} 0, & a \le 1/20, \\ (10a - 1/2)/9, & a > 1/20. \end{cases} \tag{S.34}$$

Conditional on s, no atom crosses an organizational boundary, so this share and aggregate receipts are continuous at entry.

Integrating taxes over the active intervals gives

$$\bar{R}(a) = \begin{cases} \frac{19}{40} - \frac{9}{20}a - \frac{1}{20}a^2, & 0 \le a \le 1/20, \\ \frac{341}{720} - \frac{71}{180}a - \frac{109}{180}a^2, & 1/20 < a \le 1/2. \end{cases} \tag{S.35}$$

The no-splitting counterfactual follows the first expression throughout $[0, 1/2]$. Initial mechanical exposure is $\mathbb{E}B_z^0 = 19/40$, so the mechanical loss is $19a/40$. After splitting begins, the difference between the actual and mechanical losses is

$$\frac{1 - 58a + 436a^2}{720}.$$

It becomes positive at

$$a > \frac{29 + 9\sqrt{5}}{436} \simeq 0.11. \tag{S.36}$$

At $a = 1/5$, the cutoffs are $3/100$, $1/30$, and $1/5$. The split share is $1/6$, receipts are $741/2000$, and the loss is $209/2000$. The mechanical loss is $19/200$, so their ratio is $11/10$. Total reports rise from $19/20$ to $581/600 \simeq 0.97$. Without splitting, the loss is $23/250$.

Independent calculation. The code in `code/reform/` integrates the lower envelope of the four strategies and checks the revenue formula, reform identities and global optimality conditions. The dominance argument above establishes that these calculations solve the unrestricted manipulation problem.

S.5.4 Large fiscal overruns

Proposition S.5.2 (Large fiscal overruns with repeated concessions)**.** *For each integer $m \ge 2$, there is an economy satisfying the maintained assumptions, with at most m entities per firm, in*

which a flat tax is globally optimal and no firm initially splits, but an admissible lower-bracket cut gives

$$\frac{\bar{R}(0) - \bar{R}(a)}{a\mathbb{E}B_z^0} = \frac{9m + 10}{19}. \tag{S.37}$$

Total reporting rises after the cut. Thus the ratio of actual to mechanical revenue loss has no uniform bound across economies with increasing entity caps.

The construction keeps the initial tax rate, cost distribution and revenue requirement fixed. It sets the relief threshold at $1/m$ and lets half the firms claim the concession m times. The mechanical estimate shrinks as m grows, while the actual loss stays bounded away from zero. Absolute revenue loss remains bounded. The construction uses declining incremental costs of creating entities. It does not give an unbounded ratio at a fixed entity cap or threshold, or under a common positive lower bound on incremental splitting costs across economies.

Proof of Proposition S.5.2. Fix $m \geq 2$. All firms have potential base one, costs are uniform on $[0, 1]$, and welfare weights equal one. The fee and enforcement are unavailable, marginal rates lie in $[0, 1]$, and $G = 19/40$. A firm can operate any number $n \in \{1, \ldots, m\}$ of entities, each reporting in $[0, 1/n]$, at effort

$$10\left(1 - \sum_{j=1}^{n} r_j\right) + \kappa_n, \qquad \kappa_n = 1 - \frac{1}{n}.$$

All intermediate entity counts are available. For each m, the strategy set is a finite union of compact sets and effort is bounded. Reporting one through a single entity is feasible at zero effort. To check no manipulation at the top, let $h = 1 - \sum_j r_j$. The largest possible tax saving satisfies

$$1 - \max_j r_j \leq 1 - \frac{1-h}{n} = \kappa_n + \frac{h}{n} \leq \kappa_n + 10h.$$

The cost and welfare restrictions therefore hold. The flat rate 0.6 raises $141/250 > G$, giving strict feasibility.

At the initial schedule $T(r) = r/2$, using several entities adds effort without changing taxes on a given total report. It is therefore strictly dominated for $c > 0$. Firms report zero below $c = 1/20$ and one above it. Receipts are $19/40$, and at every $r \in (0, 1)$,

$$\mathbb{E}n_r = \frac{19}{20}, \qquad \Pr(x > r) = 1.$$

With $\mu = 19/18$, the rate condition is

$$(2\mu - 1)\mathbb{E}n_r - \mu \Pr(x > r) = \frac{10}{9}\frac{19}{20} - \frac{19}{18} = 0.$$

The budget binds and all marginal rates are interior. Theorem 1 proves global optimality over the full admissible schedule class.

Set $z = 1/m$ and cut the lower rate to zero, so $a = 1/2$ and

$$T_a(r) = \frac{1}{2}(r - 1/m)^+.$$

This is admissible. Reporting fully through m entities pays no tax and costs $c\kappa_m$. For $n < m$ and $0 < c < 1/20$, the best report through each entity is $1/m$: burden falls below this report and rises above it. The resulting effort exceeds κ_m by

$$10(1 - n/m) + \kappa_n - \kappa_m = \frac{m - n}{m}\left(10 - \frac{1}{n}\right) > 0.$$

Thus these firms use m entities and report fully.

For $c > 1/20$, a best strategy at each entity count reports the full base. Its tax is $(1 - n/m)/2$ and its effort is κ_n. If $c \leq 1/2$, its burden minus the burden with m entities is

$$\frac{m - n}{m}\left(\frac{1}{2} - \frac{c}{n}\right) \geq 0.$$

If $c \geq 1/2$, its burden minus the burden with one entity is

$$(n - 1)\left(\frac{c}{n} - \frac{1}{2m}\right) \geq 0.$$

Consequently, firms with $c < 1/2$ use m entities and those with $c > 1/2$ use one. The boundary cost parameters have zero mass. Every firm reports its full base. Receipts after the cut are $\bar{R}(a) = \kappa_m/4$. The actual and mechanical losses are

$$\bar{R}(0) - \bar{R}(a) = \frac{9}{40} + \frac{1}{4m}, \qquad a\mathbb{E}B_z^0 = \frac{1}{2}\frac{1}{m}\frac{19}{20} = \frac{19}{40m}.$$

Their ratio is $(9m + 10)/19$, and total reporting rises from $19/20$ to one. □

S.6 Extensions of the policy-learning rule

The full rate class. In the class (2), measure the fee and enforcement as before and the marginal rate as an element of the space of square-integrable functions on $[0, X]$. Projection clips the rate at each threshold. The net value of the rates is the function $r \mapsto \Xi_r$. The diameter of the admissible set and the bound on net values remain finite. The argument for parts (i) and (ii) then goes through unchanged. The full class is compact in the uniform norm on schedules. By (71), the Lagrangian is continuous in that norm and attains its maximum at some π°. The same compactness gives part (iii), with the best schedules converging uniformly to the unique maximizer. Convergence of the marginal rates in the norm of square-integrable functions needs further conditions.

S.6.1 Net revenue and the value of public funds

For $w_- \le \mu < \mu'$, let π maximize $\mathcal{L}_\mu$ and π' maximize $\mathcal{L}_{\mu'}$. Then $\mathcal{L}_\mu(\pi) \ge \mathcal{L}_\mu(\pi')$ and $\mathcal{L}_{\mu'}(\pi') \ge \mathcal{L}_{\mu'}(\pi)$. Adding the two inequalities, the welfare terms cancel and

$$(\mu' - \mu)\big(Q(\pi') - Q(\pi)\big) \ge 0,$$

so net revenue at the maximizers is nondecreasing in μ. By Proposition F.2, the supporting multipliers minimize the function $\mu \mapsto \max_\pi \mathcal{L}_\mu(\pi)$, which is convex because it is a maximum of functions affine in μ. Its subdifferential at μ contains $Q(\pi_\mu) - G$ for every maximizer π_μ. A projected subgradient update increases μ when the selected exact maximizer raises less than G and decreases it when it raises more. A finite step can overshoot, and approximate policy maximization introduces error. At a supporting multiplier, several maximizers may have different revenues. The proof of Proposition F.2 then selects a point in their convex set with $Q = G$, using continuity of revenue. Monotonicity alone does not guarantee convergence of an arbitrary coupled policy–multiplier iteration. Primal–dual methods such as those of Arrow et al. (1958) and Nedić and Ozdaglar (2009) require their own step-size, boundedness and approximation conditions.

S.7 Further enforcement results

S.7.1 Resource shares of marginal enforcement

Under equal linear welfare weights $U_s(y) = wy$, put $\beta = 1 - w/\mu$ and $H_i = \mathbb{E}[ck_i]$. At a regular interior optimum,

$$A_i = \mathbb{E}[(\gamma_s + \beta)ck_i], \qquad Q_{e_i} = (1-\beta)H_i, \qquad \partial_{e_i}\{A + \mathbb{E}D\} = \beta H_i. \tag{S.38}$$

For $H_i > 0$, define $\gamma_i = \mathbb{E}[\gamma_s ck_i]/H_i$. The shares of the marginal gross receipts spent on public enforcement and retained as net receipts are

$$\frac{A_i}{\partial_{e_i}\mathbb{E}R} = \frac{\gamma_i + \beta}{1 + \gamma_i}, \qquad \frac{Q_{e_i}}{\partial_{e_i}\mathbb{E}R} = \frac{1-\beta}{1+\gamma_i}. \tag{S.39}$$

The relevant exponent weights each firm's cost distribution by its exposure to the instrument. With a common exponent, $\gamma_i = \gamma$.

Online Appendix S.10.3 derives (S.39), and Proposition S.10.1 relates total resource costs to variable taxes and entity fees.

S.7.2 Both motives

The comparison of the two motives can be made without assuming that tax savings and private costs are minimized by the same strategy. Fix s and $c > 0$, and the variable schedule $T_0 = T - \tau$. For an underreported base h and entity count n, define the joint minimization over strategies

$$B(h,n) = \min_{\substack{\alpha \in \mathcal{A}(s) \\ x - \sum_j r_j = h,\ n(\alpha) = n}} \left\{ \sum_j T_0(r_j) + ck_0(\alpha) \right\}. \tag{S.40}$$

This retains differences in costs between strategies with the same final reports.

Proposition S.7.1 (A mixed-motive criterion)**.** *Exclude exit, suppose $k_R = h$ and $k_S = n - 1$, and suppose every $(h,n) \in [0,x] \times \{1,\dots,m\}$ is feasible. Assume the minimum in (S.40) is attained and B is lower semicontinuous in h. Hold T_0 and e_R fixed. If $B(h,n+1) - B(h,n)$ is nondecreasing in h for each $n < m$, raising e_S or τ weakly raises the largest and smallest optimal underreported bases and weakly lowers the largest and smallest optimal entity counts. If these differences are nonincreasing in h, both the extremal underreported bases and entity counts weakly fall.*

The condition compares the two motives directly. Where variable taxes and baseline effort can be minimized jointly, write $B(h,n) = T_{0,n}(x-h) + cC_n^0(h)$, with $T_{0,n} = T_n - n\tau$ and C_n^0 the minimum baseline effort. If these functions are absolutely continuous, the substitution condition is equivalent to

$$T'_{0,n}(x-h) - T'_{0,n+1}(x-h) \;\geq\; c\{(C_n^0)'(h) - (C_{n+1}^0)'(h)\} \tag{S.41}$$

for almost every h and every $n < m$. The marginal tax saving from an extra entity then dominates its marginal saving in underreporting costs. Reversing the inequality gives complementarity. The decomposition requires the tax and cost minima to be attained by the same strategy. With an arbitrary coefficient k_S in place of $n-1$, own-price monotonicity orders that coefficient, but need not order entity counts.

Corollary S.7.1 (Conditional joint enforcement)**.** *Suppose the conditions of Proposition S.7.1 hold for almost every firm, with the same sign of cross differences. Let $U_s(y) = wy$, fix T and $\mu \geq w$, and put $\beta = 1 - w/\mu$. Public cost is $A^R(e_R) + aA^S(e_S)$, with continuous components, A^S nondecreasing and $a \geq 0$. Over the fixed rectangular enforcement set, maximize*

$$\Psi_a(e) = \mathbb{E}[(\gamma_s + \beta)P(T,e)] - A^R(e_R) - aA^S(e_S).$$

If the responses are substitutes, lowering a weakly raises both maximizing instruments. If they are complements, it weakly raises e_S and lowers e_R. The comparison holds for unique maximizers, or for the least and greatest maximizers in the order (e_R, e_S) in the first case and $(-e_R, e_S)$ in the second.

This comparison fixes the tax schedule and the value of public funds. It does not sign enforcement changes after taxes and the multiplier are reoptimized at a fixed revenue requirement.

S.7.3 Proof of Proposition S.7.1

At a given (h,n), reporting and splitting enforcement add the same costs to every strategy in the fiber. The firm's minimized burden conditional on (h,n) is therefore

$$\Phi(h,n) = B(h,n) + ce_R h + (\tau + ce_S)(n-1) + \tau.$$

Lower semicontinuity and the compact rectangle give attained minima and compact sets of minimizers, so the largest and smallest coordinates exist. Adding $ce_R h$ preserves the cross differences of B. A nondecreasing adjacent difference in n gives increasing differences for

every pair of counts by summation; a nonincreasing difference gives decreasing differences. The monotonicity theorem, with $\theta = \tau + ce_S$ and the orders used in Appendix L, proves both conclusions.

For the decomposition stated after the proposition, the adjacent difference is $T_{0,n+1}(x-h) - T_{0,n}(x-h) + c\{C^0_{n+1}(h) - C^0_n(h)\}$. It is absolutely continuous, with derivative $T'_{0,n}(x-h) - T'_{0,n+1}(x-h) - c\{(C^0_n)'(h) - (C^0_{n+1})'(h)\}$ almost everywhere. It is nondecreasing exactly when this derivative is nonnegative almost everywhere, giving (S.41). The reverse sign gives decreasing differences. Finally, revealed preference for an arbitrary coefficient gives $(e'_S - e_S)(k'_S - k_S) \leq 0$ when only e_S rises. This orders k_S, not n, unless the technology links them. □

S.7.4 Proof of Corollary S.7.1

Burden is Lipschitz in enforcement. Wherever its derivative in e_R exists, the envelope theorem gives $P_{e_R} = ch$. Proposition S.7.1 orders the extremal optimal values of h as e_S rises. Integrating these derivatives over an interval of e_R shows that P has increasing differences in (e_R, e_S) under substitution and decreasing differences under complementarity. This argument permits jumps in optimal strategies.

The weights $\gamma_s + \beta$ are positive and independent of enforcement. Expectations preserve the corresponding differences, and separable public costs do not change them. Thus Ψ_a is supermodular in (e_R, e_S) under substitution and in $(-e_R, e_S)$ under complementarity. It has increasing differences between $-a$ and e_S because A^S is nondecreasing. The monotonicity theorem (Milgrom and Shannon, 1994, Theorem 4) on the compact rectangle gives the stated ordering of extremal maximizers. Under uniqueness, this orders the single maximizer. □

S.8 Measurement: illustrations and implementation

The identification proof and the conditional-scaling results are in Appendix M of the paper. This appendix provides the numerical illustration and additional estimation details.

S.8.1 The tax-payment curve in the running example

In Figure S.1, firms face the optimal schedule with splitting and zero enforcement. Write $J = \mathbb{E}T(x)$ and $D_1 = \mathbb{E}_1 D$ for its original no-manipulation tax and private costs. Since

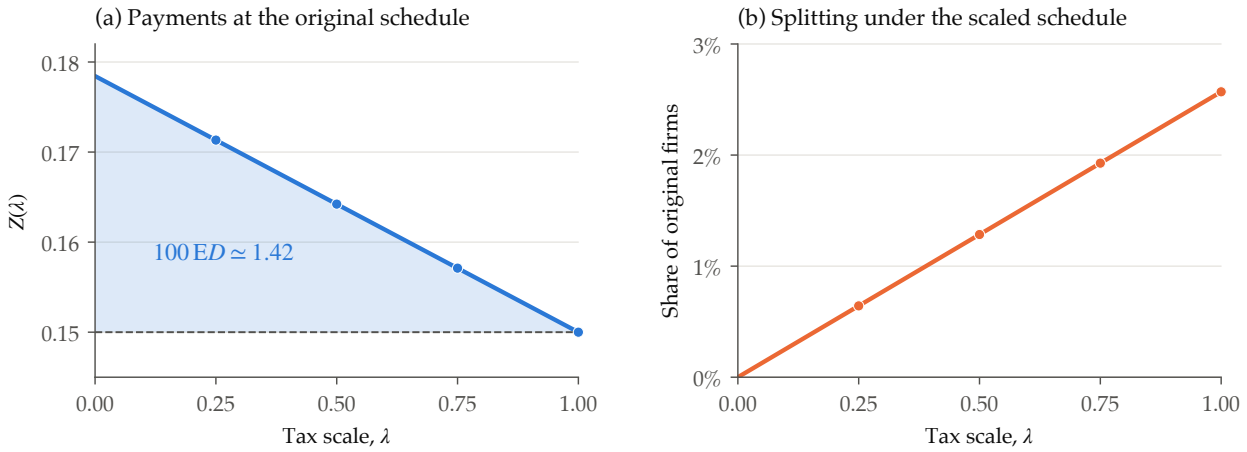


Figure S.1: Private manipulation costs from proportional reforms in the running example. Start at the optimum with splitting and zero enforcement. Left: $Z(\lambda)$ evaluates the reports chosen under λT at the original schedule. Its intercept is the no-manipulation tax $J = \mathbb{E}T(x) \simeq 0.18$, and $Z(1) = 0.15$. The shaded area between the curve and $Z(1)$ is private manipulation cost $\mathbb{E}D \simeq 1.42 \times 10^{-2}$. Right: the split share scales with λ, since cost parameters are uniform. Markers evaluate the firms' choices at the displayed reforms; lines give the exact formulas.

$\gamma = 1$, the exact tax-payment curve and split share are

$$Z(\lambda) = J - 2\lambda D_1, \qquad \mathbb{E}R(\lambda) = \lambda J - 2\lambda^2 D_1, \qquad S(\lambda) = \lambda S(1),$$

where $J \simeq 0.18$, $D_1 \simeq 1.42 \times 10^{-2}$ and $S(1) \simeq 2.57 \times 10^{-2}$. The cost integral in Proposition 15(i) is the area of the triangle above the original tax payments. Uniform costs also imply $\mathbb{E}_\lambda D = \lambda^2 D_1$.

Reforms at $\lambda = 1/4, 1/2, 3/4, 1$, together with the known upper bound $\bar{B} = J$, give outward-rounded bounds on private manipulation cost of 1.06×10^{-2} and 1.78×10^{-2} by Proposition 15(ii). The split-share ratio at $\lambda = 1/2$ identifies $\gamma = 1$, and part (iii) then recovers D_1 exactly from that reform. The calculations integrate cost types analytically and leave reports unrestricted.

S.8.2 Entity-count estimators

Heterogeneous exponents. With heterogeneous exponents, write $M_{r,s}(\lambda) = \mathbb{E}_\lambda[n_r \mid s]$. Given identified characteristic groups, the finite-reform estimator is

$$\hat{\varepsilon}^{\text{PL,het}}_{r,\lambda} = \frac{1}{\widehat{M}_r(1)} \int \frac{\widehat{\gamma}_s}{1 - \lambda\widehat{\gamma}_s} \left[\widehat{M}_{r,s}(\lambda) - \widehat{M}_{r,s}(1)\right] \mathrm{d}\widehat{H}(s), \tag{S.42}$$

where $\widehat{H}$ estimates the original population's distribution of characteristics. Its population counterpart equals ε_r exactly. For fixed $\lambda < 1$, consistency requires convergence of the estimated conditional counts, exponents and population weights under the integral. Arbitrary observable bins need not be homogeneous in the model's characteristics, and an average exponent cannot replace the group-specific corrections.

Apply conditional scaling to $n_r - \mathbf{1}\{x(s) > r\}$, which is bounded and zero at the no-manipulation strategy. This gives

$$M_{r,s}(\lambda) = \lambda^{\gamma_s} M_{r,s}(1) + (1 - \lambda^{\gamma_s})\mathbf{1}\{x(s) > r\}.$$

Bounded entity counts and exponents permit differentiation under the integral at $\lambda = 1$. Thus

$$-M_r'(1-) = \int \gamma_s \left[\mathbf{1}\{x(s) > r\} - M_{r,s}(1)\right] \mathrm{d}H(s).$$

Eliminating the bracketed term with the finite-change identity proves that the population counterpart of (S.42) equals ε_r. With a common exponent it gives (63). For fixed $\lambda < 1$, the correction factor is continuous and bounded on $[1, \bar{\gamma}]$. The stated convergence under the integral therefore gives consistency of the plug-in estimator. With finitely many identified groups, consistency of these estimates in every group suffices.

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.

Estimation. With random assignment of scale factors, independent across firms, sample means are consistent and asymptotically normal. The bounds in Proposition 15(ii) are linear in these means. Confidence intervals for $\mathbb{E}D$ that allow for partial identification follow from Imbens and Manski (2004) and Stoye (2009).

If scale factors could be drawn from the uniform distribution on $(0, 1)$, the average of $T[\alpha(\lambda_{s,c})]$ over a sample of n_0 independently drawn firms, each evaluated at its own scale factor $\lambda_{s,c}$, would be an unbiased estimator of $\mathbb{E}P$. Taxes paid at the chosen strategies

lie in $[0, \bar{T}]$, so by the two-sided form of Hoeffding's inequality its error would exceed $\bar{T}\sqrt{\log(2/\delta)/(2n_0)}$ with probability at most δ (Hoeffding, 1963, Theorem 2). Subtracting the mean taxes paid by a separate group at $\lambda = 1$ then estimates $\mathbb{E}D$. An interval for $\mathbb{E}D$ must also allow for that group's sampling error.

S.9 Scope: the power law and aggregate uncertainty

Table S.1 separates distribution-free results from those requiring a power law. Exact identities for positive reports require equal weights, a common exponent, and freely chosen interior rates (Proposition S.1.1). The local weighted restriction, including its bunching implication, extends to general densities and unequal weights (Proposition S.1.1). Fixed brackets impose integrated restrictions only.

S.9.1 Aggregate uncertainty

The characterization also allows one policy chosen before an aggregate state z is realized, with firms observing z before choosing reports and the budget required to hold in expectation. Treat (z, s) as the characteristics index and impose the regularity, uniform bounds and strict feasibility assumptions on this compact joint space. Theorem 1 then applies directly: expectations include states and firms, and there is one supporting multiplier μ. The conditional exponent $\gamma_{z,s}$ stays inside expectations. No independence across firms is needed because welfare and receipts are additive and choices create no payoff externalities. State-by-state budgets would instead introduce separate constraints and shadow values.

S.10 Resource accounting and one-entity tax schedules

The first result relates collections to resource costs at an optimum. The second solves the rate condition when firms can neither split nor exit.

S.10.1 Resources, variable taxes and entity fees

A proportional change in marginal rates holds the entity fee fixed. A proportional change in the whole schedule changes both. The distinction matters for relating tax collections to resource costs.

Proposition S.10.1 (Resource accounting at an optimum)**.** *Maintain the main conditional-power assumptions, let $U_s(y) = wy$, and use a multiplier and selection from Theorem 1. Write*

Result	Distributional scope and additional restrictions
Global characterization and Ramsey rule	Extend to the general conditional densities of Appendix F, with its restriction on welfare weights and no manipulation at the top. The elasticity formula (22) and the targets (28) and (30) use a common exponent, and all report targets use equal welfare weights.
Sequential continuation and ordering	Distribution-free. The recursive characterization and ordering results require the report-dependent reduction costs and constant split effort specified in Section 4.
Private costs and the tax gap	The accounting identities extend with $\ell_s(c)$ and $f_s(1)$ in place of the exponent. The fixed fraction of each group's tax gap uses the power law (Corollary 1).
Enforcement, receipts and private costs	The aggregate signs extend. The explicit shares and their bounds use the conditional power law (Proposition 13).
Lower-bracket relief	The bound without splitting does not require a power law. The decompositions, monotone feedback and convexity of excess loss extend under the restriction on $\ell_s(c)$, and conditional powers give the displayed weights. The large relative-loss construction uses uniform costs and varies the entity limit and relief threshold (Proposition 6, Appendix J).
The threat of splitting	Section 5 uses a common exponent and linear welfare weights that depend only on the potential tax base. Propositions 3 and 4 also use the properties of the optimum stated there, and Proposition 5 uses equal weights, no exit and a constant marginal effort of underreporting.
Coverage ratios	Section 7 uses a common exponent, and the ranking of rates in Corollary 6 uses equal weights. The learning rule of Proposition 12 rests on the concavity of the Lagrangian under Assumptions 1 and 2.
Welfare and residual splitting	The welfare comparison, exact decomposition and supporting-selection result extend to general densities. Positive enforcement additionally uses equal weights and zero initial marginal public cost (Section 6.3).
Splitting term in optimal rates	The displayed weighted formula uses conditional power laws, equal weights, no exit, a progressive schedule and the specified one-entity technology. The closed-form rates use a common exponent. Signed compression also requires the tax-capacity technology (Section 5.2 and Appendix S.11).
Comparisons of optimal rate ratios	Use a common exponent and the additional technology, feasibility and regularity conditions in Appendix S.11.
Measurement from tax reforms	For any distribution, the behavioral revenue response measures the change in private costs. Without mass at a zero cost parameter, full-path identification and finite-reform bounds measure their level. Conditional power laws give exact scaling within characteristic groups, and a common exponent gives aggregate scaling and one-reform identification (Section 9).

Table S.1: Which results depend on the power law. The shared cost parameter and the stated policy and technology restrictions remain in force.

$\beta = 1 - w/\mu$,

$$v = \sum_j \int_0^{r_j} t(r)\,\mathrm{d}r, \qquad R_{\text{var}} = \mathbb{E}v, \qquad N = n - 1, \qquad D_0 = ck_0, \qquad D_e = c\sum_i e_i k_i,$$

so $R = \tau n + v$ *and* $D = D_0 + D_e$. *Then:*

(i) *The always-feasible proportional cut in the whole schedule gives*

$$\mathbb{E}[(1+\gamma_s)D] \le \beta \mathbb{E}R. \tag{S.43}$$

If $\mathbb{E}R > 0$, *then* $\mathbb{E}D/\mathbb{E}R \le \beta/2$. *With a common exponent* γ, *the sharper expression is* $\mathbb{E}D/\mathbb{E}R \le \beta/(1+\gamma)$.

(ii) *If small proportional changes in all marginal rates are feasible in both directions at fixed fee and enforcement, then*

$$\beta R_{\text{var}} = \mathbb{E}[(1+\gamma_s)D] + \tau\mathbb{E}[\gamma_s N]. \tag{S.44}$$

For a common exponent this is $\beta R_{\text{var}} = (1+\gamma)\mathbb{E}D + \gamma\tau\mathbb{E}N$.

(iii) *If, in addition, small proportional changes in enforcement are feasible in both directions, then*

$$\beta R_{\text{var}} = e \cdot \nabla A + \mathbb{E}[(\gamma_s + \beta)D_0] + (1-\beta)\mathbb{E}D + \tau\mathbb{E}[\gamma_s N]. \tag{S.45}$$

If A *is homogeneous of degree* ν, *then* $e \cdot \nabla A = \nu A$.

(Proof in Appendix S.10.3.) These identities retain the full path technology and the common-scale exit option. Without exit, $N \ge 0$, so a positive fee and positive extra entity counts make the fee term positive. With exit, $N = -1$ for exiting firms, and that interpretation no longer applies. Fees are transfers, not resource losses: the term records the additional collections that the authority must account for when it changes marginal rates at a fixed fee.

Under Assumption 3, the same identities hold with the score $\ell_s(c)$ in place of γ_s. The bound $\beta/2$ uses $\gamma_s \ge 1$ and need not hold for a general density. The common β still requires equal linear welfare weights.

S.10.2 Marginal rates when firms can neither split nor exit

Fix enforcement and suppose that reducing a firm's report by $h \in [0, x]$ requires effort $\int_0^h p_s(v)\,\mathrm{d}v$, where $p_s > 0$ is continuous. A schedule is *single-crossing* if it has a marginal-rate representative for which $q \mapsto t(q)/p_s(x-q)$ is nondecreasing for every s. This

condition makes each firm's burden first decrease and then increase in its report. It holds, for example, when the schedule is convex and each p_s is nondecreasing.

Proposition S.10.2 (Marginal rates as power means)**.** *Maintain Assumption 3 except strict feasibility, with equal linear welfare weights w and a common power exponent $\gamma > 0$. Suppose firms can neither split nor exit, and their reduction effort is as above. Fix enforcement and $\mu \geq w$, and put $\beta = 1 - w/\mu$. For a single-crossing schedule in $\mathcal{T}$, the rate conditions (17) hold if and only if, at almost every r with $\Pr(x > r) > 0$,*

$$t(r) = \min\left\{L,\ \left(\frac{\beta}{\gamma+\beta}\right)^{1/\gamma}\left(\mathbb{E}\left[p_s(x-r)^{-\gamma} \mid x > r\right]\right)^{-1/\gamma}\right\}. \tag{S.46}$$

For $\beta > 0$ and $L > 0$, this equivalence requires the conditional moment to be finite. When $\beta = 0$, the condition is $t(r) = 0$. No manipulation at the top requires $t(r) \leq p_s(x-r)$ for almost every firm with $x > r$.

(Proof in Appendix S.10.3.) An uncapped rate is proportional to a power mean of marginal efforts among firms above the threshold. The order is $-\gamma$, giving the harmonic mean when $\gamma = 1$. The calculation uses single crossing. Without it, rates elsewhere in the schedule affect whether a firm reports above r.

In Example 1, marginal efforts at zero enforcement are 0.6, 2.4 and 6, with masses 0.5, 0.3 and 0.2. Their conditional harmonic means are $120/119$ below 0.3, $60/19$ between 0.3 and 0.6, and 6 above 0.6. A full-class no-splitting optimum multiplies these by $\beta/(1+\beta)$ where the cap does not bind, with β determined in that policy problem. It generally needs thresholds at both 0.3 and 0.6. The main numerical family has only the first threshold, so the last two intervals share one rate.

S.10.3 Proofs

S.10.3.1 Waste per dollar of net revenue

We apply the interior optimality condition to the resource identity. Under equal linear weights, consider any rate, fee or enforcement direction that can move both ways, and use the common supporting selection. Let b be its selected burden exposure, q its net-revenue score and d its total-resource-cost score, including public enforcement spending. Since $\mathbb{E}P = Q + A + \mathbb{E}D$, the accounting formulas in Proposition G.1 give

$$\mathbb{E}b = q + d, \qquad -w\mathbb{E}b + \mu q = 0.$$

Consequently $(1-\beta)d = \beta q$, and the resource cost per net dollar is $\beta/(1-\beta)$ whenever $q > 0$. For a marginal rate, $b = n_r$ and $d = J_r - \mathbb{E}[\ell_s n_r]$; integrate these scores for a fixed bracket. For enforcement i, $b = ck_i$ and $d = A_i - \mathbb{E}[\ell_s ck_i]$.

These are scores from one common supporting selection. They equal policy derivatives when the relevant exposures agree at all minimizing strategies almost surely, as in Proposition G.1. For a marginal rate the assertion holds almost everywhere as the density of interval effects. No extra differentiability of firms' choices is required.

S.10.3.2 Proof of Proposition S.10.1

Put $\Gamma_s = \gamma_s + \beta$. The supporting value of a proportional increase in the whole schedule, divided by μ, is

$$\mathbb{E}[\Gamma_s R] - \mathbb{E}[\gamma_s T(x)].$$

A proportional decrease is always feasible, so this value is nonnegative. The accounting identity gives $\mathbb{E}[\gamma_s T(x)] = \mathbb{E}[\gamma_s P] + \mathbb{E}D$. Using $P = R + D$, the supporting inequality becomes $\beta \mathbb{E}R - \mathbb{E}[(1+\gamma_s)D] \geq 0$. This proves (S.43) and both stated ratio bounds.

For the variable-rate change, write $v^0 = \int_0^x t(r)\,\mathrm{d}r$. Feasibility in both directions gives $\mathbb{E}[\Gamma_s v] = \mathbb{E}[\gamma_s v^0]$. Since $T(x) = \tau + v^0$ and $P = \tau n + v + D$, accounting also gives

$$\mathbb{E}D = \mathbb{E}[\gamma_s\{v^0 - v - D - \tau(n-1)\}].$$

Substituting the variable-rate condition proves (S.44).

If proportional enforcement changes are feasible in both directions, Theorem 1(iii) gives $e \cdot \nabla A = \mathbb{E}[\Gamma_s D_e]$. Decompose $(1+\gamma_s)D = \Gamma_s D + (1-\beta)D$ and use $D = D_0 + D_e$ in (S.44). This proves (S.45). Euler's identity gives the homogeneous-cost case.

The general-density version uses $\Omega/\mu = \ell_s + \beta$, $\mathbb{E}[\ell_s T(x)] = J(T)$ and $\mathbb{E}D = J(T) - \mathbb{E}[\ell_s P]$ in the same steps. The bound is then $\mathbb{E}[(1+\ell_s)D] \leq \beta \mathbb{E}R$. The equal-weight shape restriction gives $\ell_s \geq 0$, but the factor $1/2$ requires the stronger lower bound $\ell_s \geq 1$. All these calculations use the common supporting selection and therefore do not require differentiable reporting choices.

For (S.39), regularity gives $\partial_{e_i}\mathbb{E}R = \mathbb{E}[(1+\gamma_s)ck_i]$ and $\partial_{e_i}\mathbb{E}D = -\mathbb{E}[\gamma_s ck_i]$. Use the interior condition $A_i = \mathbb{E}[(\gamma_s + \beta)ck_i]$ and divide by $H_i = \mathbb{E}[ck_i] > 0$. □

S.10.3.3 Proof of Proposition S.10.2

We identify the firms reporting above each threshold and substitute their mass into the rate condition. A firm minimizes $T(q) + c\int_0^{x-q} p_s(v)\,\mathrm{d}v$ on $[0, x]$. Its derivative is

$$p_s(x-q)\left[\frac{t(q)}{p_s(x-q)} - c\right]$$

almost everywhere. Single crossing makes the bracketed expression nondecreasing. For almost every $r < x$, the minimizing report therefore exceeds r exactly when $c\,p_s(x-r) > t(r)$, apart from zero-probability ties. To justify the qualification, each monotone ratio has at most countably many discontinuities; Fubini and the continuous conditional cost distribution remove these thresholds and ties.

At $c = 1$, compliance minimizes. If the ratio exceeded one on a set of positive measure, monotonicity would make burden strictly increase on a terminal interval, contradicting compliance. Hence $t(r) \leq p_s(x-r)$ almost everywhere. The power law now gives, for $t(r) > 0$,

$$M_r = \Pr(x > r)\left\{1 - t(r)^{\gamma} H_r\right\}, \qquad H_r = \mathbb{E}[p_s(x-r)^{-\gamma} \mid x > r] < \infty.$$

The rate score divided by $\mu \Pr(x > r)$ is consequently $\beta - (\gamma + \beta)t(r)^{\gamma} H_r$. It vanishes at an interior rate and is nonnegative at the cap. These conditions give (S.46). At a zero rate, $M_r = \Pr(x > r)$ and the score is β; thus zero is optimal only if $\beta = 0$ or the cap is zero. If $\beta > 0$ and $L > 0$, every admissible solution has positive rates and therefore finite H_r. Conversely, the formula with finite H_r gives the required score signs. When $\beta = 0$, every positive rate has strictly negative score, so the rate condition requires zero. □

S.11 Splitting and optimal tax differentiation

When splitting expands access to lower brackets, it raises optimal rates at low reports and lowers them at high ones. The comparison is with the rates justified by one-entity responses at the same enforcement and value of public funds. With equal weights, an interior optimum matches the entities above each threshold, weighted by $\gamma_s + \beta$, with the firms whose potential tax base exceeds it, weighted by γ_s (Corollary 3). With a common exponent, the entities above each threshold are the fraction $\gamma/(\gamma + \beta)$ of those firms. Splitting adds entities at low reports and removes them from high ones. Rate adjustments restore the target by changing underreporting. Comparing actual reports with the reports

firms would choose through one entity gives an exact additional term in each rate condition. This decomposition holds for any splitting technology. When splitting only expands access to lower brackets, it also determines the direction of adjustment and bounds its size.

Throughout this appendix, impose equal linear welfare weights, $U_s(y) = wy$, and exclude exit. Retain the conditional cost distributions $F_s(c) = c^{\gamma_s}$ of the main model. The first two propositions allow heterogeneous conditional exponents. The tax-shape comparison in Appendix S.11.5 imposes a common exponent and states its further restrictions.

S.11.1 The additional term from splitting

Consider fixed brackets I_j, with independently chosen rates $t_j \in [0, L]$, and write

$$b_j(q) = \int_{I_j} \mathbf{1}\{q > r\}\, \mathrm{d}r, \qquad B_j = \mathbb{E}b_j(x), \qquad T(q) = \tau + \sum_j t_j b_j(q).$$

Suppose the minimum effort of making a one-entity report $q \in [0, x]$ is $C_s(x - q; e_R)$, with $C_s(0; e_R) = 0$. The function C_s is differentiable, increasing and convex in the amount underreported, with positive marginal effort. Every such report is feasible, and its cost remains affine in enforcement. The splitting plans can otherwise retain the sequence dependence of the sequential manipulation technology of Section 2.1.

At a progressive optimal schedule, let q^N be the report the firm would choose if restricted to one entity at the same taxes and enforcement. Assume this report is unique for almost every firm and that $t_j / C_s'(x - r; e_R) \leq 1$ for almost every $r \in I_j$ with $x > r$. The sufficient condition $C_s'(0; e_R) \geq L$ ensures the latter restriction. Define

$$\Delta b_j = \sum_a b_j(r_a) - b_j(q^N), \qquad \widetilde{\Delta}_j = \mathbb{E}[(\gamma_s + \beta)\Delta b_j], \qquad \beta = 1 - w/\mu, \tag{S.47}$$

using a selection that supports the optimum. Thus Δb_j compares a firm's bracket exposure with and without splitting at the actual policy. It does not compare separately optimized policies.

For a candidate rate v whose cutoffs lie in $[0, 1]$, put

$$\Phi_j(v) = \mathbb{E}_s \int_{I_j} \mathbf{1}\{x > r\}(\gamma_s + \beta) \left(\frac{v}{C_s'(x - r; e_R)} \right)^{\gamma_s} \mathrm{d}r. \tag{S.48}$$

The expectation here is over characteristics. The function measures the weighted exposure that one-entity underreporting would remove from the bracket at rate v.

Proposition S.11.1 (The splitting term in optimal rates)**.** *Maintain the main assumptions and the restrictions of this subsection. At an independently movable interior bracket rate,*

$$\boxed{\Phi_j(t_j^*) = \beta B_j + \widetilde{\Delta}_j.} \tag{S.49}$$

At a rate equal to L, the left side is at most the right side. At a zero rate the inequality is reversed. All quantities use the actual optimal enforcement, fee and value of public funds.

(Proof in Appendix S.11.3.) Without splitting, the additional term is zero. Compare the optimal rate with a one-entity reference at the same enforcement and value of public funds. If that reference lies in the admissible rate interval, added exposure raises the required rate and lost exposure lowers it. Both the costs of underreporting and their conditional distribution enter this reference. An average exponent cannot replace γ_s in either term.

With a common exponent γ, the condition has a closed form. Define the *marginal-effort index* $m_j(e_R)$ of bracket j by

$$m_j(e_R)^{-\gamma} = \frac{1}{B_j}\mathbb{E}_s \int_{I_j} \mathbf{1}\{x > r\}[C_s'(x-r;e_R)]^{-\gamma}\,\mathrm{d}r,$$

a power mean of order $-\gamma$ of marginal underreporting efforts. At each report level in the bracket, it includes firms whose potential base exceeds that level. Then $\Phi_j(v)/B_j = (\gamma+\beta)(v/m_j)^\gamma$, and at an interior rate (S.49) becomes

$$t_j^* = m_j(e_R^*)\left(\frac{\beta + \widetilde{\Delta}_j/B_j}{\gamma+\beta}\right)^{1/\gamma}. \tag{S.50}$$

Without splitting, $\widetilde{\Delta}_j = 0$, and each rate is $(\beta/(\gamma+\beta))^{1/\gamma}$ times the bracket's marginal-effort index. Rates are then higher in brackets whose firms are harder to move, as when larger firms face higher marginal efforts of underreporting. Appendix S.10.2 gives the corresponding formula for the full class of schedules. When $\beta > 0$, splitting multiplies the one-entity rate at the same enforcement and value of public funds by $(1 + \widetilde{\Delta}_j/(\beta B_j))^{1/\gamma}$. This factor exceeds one where splitting adds exposure and falls below one where it removes exposure.

There is an equivalent elasticity form. Suppose small changes in t_j preserve progressivity and the cutoff restrictions. Let $M_j^N = \mathbb{E}b_j(q^N) > 0$ and define its elasticity at fixed

enforcement by $Z_j^N = -\partial \log M_j^N / \partial \log t_j$. At an interior rate,

$$Z_j^N - \beta = \frac{\widetilde{\Delta}_j}{M_j^N}. \tag{S.51}$$

This measures how far the one-entity response at the actual policy departs from its own Ramsey condition because splitting is available.

S.11.2 Splitting to obtain lower rates

To sign the exposure changes, specialize to two brackets with threshold b. Suppose an n-entity firm can attain the minimum variable tax

$$t_1 \min\{q, \ell_n(s)\} + t_2(q - \ell_n(s))^+$$

at total report q, where lower-bracket capacity $\ell_n(s) \in [0, x]$ is nondecreasing in n and $\ell_1(s) = \min\{x, b\}$. Unrestricted division gives $\ell_n(s) = \min\{x, nb\}$. Its effort is

$$C_s(x - q; e_R) + \sigma_s(n) + e_S(n-1), \qquad \sigma_s(1) = 0, \quad \sigma_s(n) \geq 0. \tag{S.52}$$

The minimum tax and this cost must be attainable together. Splitting then expands access to lower rates without changing the cost of underreporting a given total amount. Proposition S.11.1 does not require these restrictions.

Let q_L and q_H be the reports a firm would choose if every reported unit faced t_1 and t_2, respectively. For a progressive schedule, $q_H \leq q_L$. Conditional on n, its lower- and upper-bracket exposures are

$$L_n = \min\{q_L, \ell_n\}, \qquad U_n = (q_H - \ell_n)^+.$$

An additional entity weakly raises the first and lowers the second. It follows that $\Delta b_1 \geq 0$ and $\Delta b_2 \leq 0$ for each firm. Define the difference between the normalized rate conditions by

$$\mathcal{C} = \frac{\Phi_1(t_1^*)}{B_1} - \frac{\Phi_2(t_2^*)}{B_2}, \qquad B_1, B_2 > 0. \tag{S.53}$$

Proposition S.11.2 (Rate differentiation, fees and enforcement)**.** *Maintain the preceding restrictions and let the optimal schedule be progressive, with $\beta > 0$ and interior bracket rates. Then*

$$\mathcal{C} = \frac{\widetilde{\Delta}_1}{B_1} - \frac{\widetilde{\Delta}_2}{B_2} \geq 0. \tag{S.54}$$

The fee is $\tau^* = F$, and

$$\boxed{B_1(t_2^* - t_1^*)\mathcal{C} \;\geq\; F\mathbb{E}[(\gamma_s + \beta)(n-1)] + \mathbb{E}[(\gamma_s + \beta)c\sigma_s(n)] + e_S^* A_S(e^*).} \tag{S.55}$$

If the upper rate is at L and the lower rate is interior, (S.54) holds with $\geq$ in place of its equality, and (S.55) continues to hold. The bound also holds when anti-splitting enforcement is at either boundary of its feasible interval.

(Proof in Appendix S.11.4.) A firm splits only if the extra access to the lower rate saves enough to cover its additional fees and private costs. The enforcement condition prices part of those private costs through marginal public expenditure. Positive terms on the right of (S.55) therefore require a strict departure from the one-entity rate conditions, even when both enforcement instruments are chosen jointly. The firm must cover the fees in this bound, but these payments are transfers rather than social costs.

With a common exponent γ, the direction has a simple rate-ratio interpretation. Since $\widetilde{\Delta}_1 \geq 0 \geq \widetilde{\Delta}_2$, the closed form (S.50) gives

$$\frac{t_2^*}{t_1^*} \leq \frac{m_2(e_R^*)}{m_1(e_R^*)} \tag{S.56}$$

when the lower rate is positive, with a strict inequality if the bound's right side is positive. The right side is the ratio of the one-entity rates at the same enforcement and value of public funds. The comparison therefore uses marginal underreporting costs after enforcement has been chosen. Appendix S.11.5 gives conditions under which splitting also compresses the ratio relative to the optimum without splitting. The comparison holds enforcement effects on one-entity costs fixed.

S.11.3 Proof of Proposition S.11.1

We first calculate one-entity exposure at the policy under consideration. The firm minimizes

$$\tau + \sum_j t_j b_j(q) + cC_s(x - q; e_R).$$

The objective is convex in q. Except at bracket endpoints and a zero-measure set of cost parameters, its minimizer exceeds $r < x$ if and only if $cC_s'(x - r; e_R) > t(r)$. The assumed cutoff restriction and the conditional power law therefore give

$$\mathbb{E}[b_j(q^N) \mid s] = \int_{I_j} \mathbf{1}\{x > r\} \left[1 - \left(\frac{t_j}{C_s'(x - r; e_R)}\right)^{\gamma_s}\right] \mathrm{d}r.$$

Let $\Gamma_s = \gamma_s + \beta$. Multiplication by Γ_s and integration over characteristics give

$$\mathbb{E}[\Gamma_s b_j(q^N)] = \mathbb{E}[\Gamma_s b_j(x)] - \Phi_j(t_j).$$

Under equal weights, $\Omega_s(c) = \mu\Gamma_s$. The bracket score of Theorem 1 divided by μ is consequently

$$\begin{aligned}\Xi_j/\mu &= \mathbb{E}\left[\Gamma_s \sum_a b_j(r_a)\right] - \mathbb{E}[\gamma_s b_j(x)] \\ &= \beta B_j - \Phi_j(t_j) + \widetilde{\Delta}_j.\end{aligned}$$

An interior rate has score zero, a capped rate has nonnegative score, and a zero rate has nonpositive score. These give the stated equality and inequalities. The calculation uses the actual optimal reports in Δb_j and places no further restriction on split paths. □

For (S.51), differentiate the one-entity exposure at fixed enforcement. This gives $Z_j^N M_j^N = \mathbb{E}_s \int_{I_j} \mathbf{1}\{x > r\}\gamma_s[t_j/C_s'(x-r;e_R)]^{\gamma_s}\,\mathrm{d}r$, hence $(Z_j^N - \beta)M_j^N = \Phi_j(t_j) - \beta B_j = \widetilde{\Delta}_j$.

S.11.4 Proof of Proposition S.11.2

We first sign the exposure changes. Fix s, c, e and the two rates. Conditional on lower-bracket capacity ℓ, the reporting objective is convex. Its minimizer is

$$q(\ell) = \text{median}\{q_H, \ell, q_L\},$$

because its marginal tax is t_1 below ℓ and t_2 above it. Hence lower exposure is $\min\{q_L, \ell\}$ and upper exposure is $(q_H - \ell)^+$. Increasing capacity weakly raises the former and lowers the latter. Relative to $n = 1$, this proves $\Delta b_1 \geq 0$ and $\Delta b_2 \leq 0$. Since $\Gamma_s > 0$, their weighted expectations have the same signs. Applying Proposition S.11.1 to both brackets proves (S.54), including the indicated inequality at a capped upper rate.

For the bound, define minimized variable burden at capacity ℓ by

$$g(\ell) = \min_{0 \leq q \leq x} \{t_1 \min(q, \ell) + t_2(q - \ell)^+ + cC_s(x - q; e_R)\}.$$

For $\ell < q_H$, its derivative is $-(t_2 - t_1)$; for $\ell > q_L$, it is zero. Between q_H and q_L, it is

$t_1 - cC_s'(x - \ell; e_R)$, which lies between these two values. Thus

$$g(\ell_1) - g(\ell_n) \leq (t_2 - t_1)[\min(q_L, \ell_n) - \min(q_L, \ell_1)] = (t_2 - t_1)\Delta b_1.$$

A chosen strategy with n entities must weakly improve on the best one-entity strategy. Thus the left side is at least $\tau(n-1) + c\sigma_s(n) + ce_S(n-1)$. This inequality also holds, with both sides zero, at $n = 1$. Multiply by Γ_s and average. With no exit and $\beta > 0$, the fee condition gives $\tau = F$. At an interior anti-splitting choice, $A_S = \mathbb{E}[\Gamma_s c(n-1)]$. At its upper bound, the KKT condition gives $A_S \leq \mathbb{E}[\Gamma_s c(n-1)]$. At zero both products with e_S vanish. Consequently, at every optimal enforcement choice,

$$(t_2 - t_1)\widetilde{\Delta}_1 \geq F\mathbb{E}[\Gamma_s(n-1)] + \mathbb{E}[\Gamma_s c\sigma_s(n)] + e_S A_S.$$

Finally $B_1\mathcal{C} \geq \widetilde{\Delta}_1$, using the sign of $\widetilde{\Delta}_2$, proves (S.55). These steps use one common supporting selection, including at policy kinks. If $\gamma_s \equiv \gamma$, then (S.48) gives $\Phi_j(t_j)/B_j = (\gamma + \beta)(t_j/m_j)^\gamma$, which proves (S.56) and its strict version. □

Costly splitting in this benchmark implies an interior lower rate when $\beta > 0$. At zero its score is positive because $\Phi_1(0) = 0$ and $\widetilde{\Delta}_1 \geq 0$. The firm's comparison above requires a strictly positive variable-tax saving to cover positive splitting costs. With $t_1 = t_2$ that saving is zero. Thus $t_1 < t_2 \leq L$, and the lower rate cannot be capped either. A cap can add to measured compression. The equality in (S.54) attributes all of it to splitting only when both rates are interior.

S.11.5 Separating the tax level from its allocation across brackets

For the comparisons in this subsection, impose a common exponent $\gamma > 0$, equal linear weights, no exit and a common fixed fee F. Consider finitely many fixed brackets, put $B = (B_1, \ldots, B_J)$ with positive components, and write $T_t(r) = F + \sum_j t_j b_j(r)$. The economies with and without splitting share the one-entity strategies and all policy instruments. Enforcement in this comparison changes only split strategies; one-entity costs are fixed. Maintain the regularity assumptions, no manipulation at the top and strict feasibility in each economy. Let

$$H^k(t, e) = \mathbb{E}[T_t(x) - P^k(t, e)], \qquad k \in \{N, S\}.$$

These functions are convex in policy. Common-power accounting gives

$$\mathcal{L}_\mu^k = w\mathbb{E}x - \mu G + (\mu - w)(F + B \cdot t) - [\mu(\gamma + 1) - w]H^k(t, e) - \mu A(e). \tag{S.57}$$

Proposition S.11.3 (The allocation of optimal rates)**.** *Under the restrictions of this subsection, let t^0 be an optimal no-splitting rate vector with $s^0 = B \cdot t^0$ positive. Suppose H^N is differentiable and the interior shape conditions $\nabla H^N(t^0) = \lambda_0 B$ hold. Let (t^*, e^*) be an optimum with splitting and positive $s^* = B \cdot t^*$. Suppose $\bar{t} = (s^*/s^0)t^0$ is admissible, including no manipulation at the top, and aggregate burdens are differentiable in rates there. Put*

$$\Delta M = M^S(\bar{t}, e^*) - M^N(\bar{t}), \qquad M_j^k = \mathbb{E}\sum_a b_j(r_a^k).$$

Then

$$\Delta M \cdot (t^* - \bar{t}) \;\geq\; H^S(\bar{t}, e^*) - H^S(t^*, e^*) \;\geq\; 0. \tag{S.58}$$

Proof. We compare tax shapes at the common variable-tax level $B \cdot t = s^*$. In (S.57), only the strictly negative coefficient times H^S varies along this constraint. Thus t^* minimizes $H^S(t, e^*)$ there, giving the second inequality.

The one-entity gain relative to compliance satisfies $\phi_s(at, c) = a\phi_s(t, c/a)$: the fee cancels and variable taxes scale by a. No manipulation at the top makes this gain zero for $c \geq 1$. Extending the cost integral to $(0, \infty)$ and substituting $c = av$ therefore gives

$$H^N(at) = a^{\gamma+1} H^N(t) \tag{S.59}$$

whenever both schedules satisfy the top condition. Consequently $\nabla H^N(\bar{t}) = \lambda B$ for some λ. The envelope theorem gives $\nabla H^k = B - M^k$, so $\nabla H^S(\bar{t}, e^*) = \lambda B - \Delta M$. The convex tangent inequality and $B \cdot (t^* - \bar{t}) = 0$ now give

$$H^S(t^*, e^*) - H^S(\bar{t}, e^*) \geq (\lambda B - \Delta M) \cdot (t^* - \bar{t}) = -\Delta M \cdot (t^* - \bar{t}),$$

which is the first inequality. □

Corollary S.11.1 (Compression relative to the no-splitting optimum)**.** *In Proposition S.11.3, take two brackets and positive lower rates. Suppose a small change from $\bar{t}$ in the direction $(B_2, -B_1)$ is feasible. If*

$$B_2 \Delta M_1 - B_1 \Delta M_2 > 0,$$

then $t_2^*/t_1^* < t_2^0/t_1^0$.

Proof. On the line $B \cdot t = s^*$ write $t = \bar{t} + z(B_2, -B_1)$. Its convex objective $H^S(t, e^*)$ has a strictly negative derivative at $z = 0$. Every minimizer must therefore have $z > 0$. The rate ratio decreases strictly with z, since its derivative is $-s^*/t_1^2$. The reference $\bar{t}$ has the same rate ratio as t^0. □

Evaluate the exposure condition at the old tax shape rescaled to the new tax level, using the new optimal enforcement. Its sign at the original policy alone does not suffice. The result allows a positive fee and arbitrary split technologies satisfying the exposure condition. It orders the rate ratio, not the two rates separately.

S.12 Additional splitting results and examples

These results give an exact splitting test for an additive technology, describe the change in policy scores when splitting becomes available, and characterize subadditive schedules. Results written with ℓ_s or $f_s(1)$ use Assumption 3; in the main model both quantities equal γ_s.

The running examples are introduced in the paper. Example 1 permits every finite division and continuous report, and Appendix N gives its exact finite representation. Example 2 isolates the cost advantage of splitting before concealment.

S.12.1 The splitting test

Theorem 2(iii) asks whether some firm strictly gains from splitting at an optimum without splitting. When reductions have constant marginal effort and each new entity has fixed effort, the test has a closed-form expression for any schedule. It compares the two uses of effort.

The benchmark technology. Fix characteristics s with potential tax base x. Here the benchmark technology can restrict how the base is divided. The firm divides its potential tax base among $n \geq 1$ entities with holdings $y = (y_1, \ldots, y_n)$ in a feasible set $Y_s(n)$ of vectors with nonnegative coordinates summing to x, with $Y_s(1) = \{x\}$ and $Y_s(n) = \varnothing$ for $n > m$. Divisions are *unrestricted* if $Y_s(n)$ contains every such vector for $n \leq m$. Each entity reports at most its holdings. The group's *removal* is $h = x - \sum_j r_j$, where $r_1, \ldots, r_n$ are the reports. It requires effort $p_s h$ with $p_s > 0$. Each additional entity requires effort $\sigma_s > 0$. Enforcement $e = (e_R, e_S)$ raises these efforts to $p_s + e_R$ and $\sigma_s + e_S$, so that

$$K(\alpha; e) = (p_s + e_R)\Big(x - \sum_j r_j\Big) + (\sigma_s + e_S)(n-1). \tag{S.60}$$

There is no exit. Each feasible set $Y_s(n)$ is compact and varies continuously with s. The cost of reduction depends only on the firm. The effort of a split does not depend on when it is made. Thus the order of splits and reductions is immaterial. With unrestricted

divisions, this is the benchmark technology of Section 2.1, which satisfies the conditions of Proposition S.3.1(ii). In Example 1, p_s takes the values 0.6, 2.4 and 6, $\sigma_s = 0.3$, and all divisions are available. The finite entity bound is derived by dominance (Appendix N); it does not constrain the underlying plans. No manipulation at the top holds in the class (2) if $p_s \geq L$ and, for every $n \geq 2$ and $y \in Y_s(n)$,

$$\sigma_s(n-1) \;\geq\; L\Big(x - \max_j y_j\Big), \tag{S.61}$$

because the largest saving from a division is the rate cap times the holdings outside the largest entity (Appendix S.12.5).

The reduction equivalent of a split. Let $\delta_s = \sigma_s / p_s$. With the effort of one split, a firm could instead remove δ_s from its reports. A strategy with n entities, reports r_j and removal h requires the same effort as the one-entity strategy that reports $\sum_j r_j - (n-1)\delta_s$, if that number is nonnegative. The two strategies differ only in taxes paid, whatever the firm's cost parameter. A firm with characteristics s therefore prefers the split strategy only if its entities pay less tax in total than the one-entity report with the same effort. For a schedule T define

$$\Pi_s(T) = \max\Big\{T\big(x-(n-1)\delta_s\big) - \sum_{j=1}^{n} T(y_j) :\; n \geq 2,\; y \in Y_s(n),\; (n-1)\delta_s \leq x\Big\}, \tag{S.62}$$

with $\Pi_s(T) = -\infty$ if no such n and y exist. The first term is the tax the firm would pay if it spent the effort of its splits on reduction instead. The sum is the tax on its entities' holdings without further reduction, including a fee for each entity. A schedule *passes the splitting test* for s if $\Pi_s(T) \leq 0$.

Proposition S.12.1 (The splitting test)**.** *Let characteristics s have the benchmark technology and satisfy no manipulation at the top. Let T belong to the class (2), with nondecreasing marginal rate, and fix enforcement at zero. Compare this economy with one in which only its one-entity strategies are available.*

(i) *The largest gain from splitting over cost parameters is $\sup_{c\in(0,1)} \Delta_{s,c} = \max\{\Pi_s(T), 0\}$.*

(ii) *The set of cost parameters c with $\Delta_{s,c} > 0$, at which firms with characteristics s gain strictly from splitting, has positive measure if and only if $\Pi_s(T) > 0$. When positive, the gain $\Pi_s(T)$ is attained by the firms that, without splitting, report $x-(n-1)\delta_s$ for a maximizing n in (S.62).*

(iii) *If $\Pi_s(T) \leq 0$, no firm with characteristics s gains from splitting, $\Delta_{s,c} = 0$ for every c. For any schedule in the class (2), the sufficient test maximizes $T(\sum_j r_j - (n-1)\delta_s) - \sum_j T(r_j)$*

over feasible reports $r_j \le y_j$ with $\sum_j r_j \ge (n-1)\delta_s$, and requires that maximum to be nonpositive.

With positive enforcement, the same statements hold with $p_s + e_R$ and $\sigma_s + e_S$ in place of p_s and σ_s.

Under the assumptions of Theorem 2, with schedule class $\mathcal{T}$ or $\mathcal{T}_{\mathcal{P}^N}$, compare an economy with only one-entity strategies with one that also permits the benchmark splitting strategies. Let $V^1_\circ(G)$ be the supremum of welfare among policies in the first economy that raise at least G and pass the test in part (iii) for almost every s, using efforts at the chosen enforcement. Then

$$V^1_\circ(G) \le V^2(G) \le V^1(G).$$

Every such policy gives the same revenue and welfare in both economies. In particular, an optimum without splitting that passes the test remains optimal with splitting.

(Proof in Appendix S.12.5.) Under a schedule with nondecreasing rates, every one-entity report is chosen by some firm, and the report $x - (n-1)\delta_s$ is chosen by the firms whose marginal cost of removing a unit equals the rate there. For these firms, the split strategy costs the same effort and pays $\Pi_s(T)$ less in tax. For any other firm, the one-entity report with the same effort as any split strategy is available and pays at most $\Pi_s(T)$ more. Since the two burdens are continuous in the cost parameter, a strict gain at one cost parameter gives a strict gain on an interval.

Dividing a report into n pieces can never save more than $n-1$ times the splitting premium $\Psi(t)$ of Section S.12.3 before fees, so $\Pi_s(T) \le (n-1)(\Psi(t) - \tau)$ for the maximizing n, and a subadditive schedule passes the test. The test is sharper because it charges each split its effort, converted into forgone reduction. A linear schedule with a positive fee or a positive rate fails to reward splitting at all: $\Pi_s(T) = \max_n\{-(n-1)(\tau + t\delta_s)\} < 0$.

In the running example, a large firm has $\delta = 0.3/6 = 0.05$. At the optimum without splitting and zero enforcement, the largest gain in (S.62) is approximately 0.16, attained by four entities. Every entity count is considered: counts with $(n-1)\delta > 1$ cannot beat complete reduction, and for each remaining count convexity makes equal division tax-minimizing. The test therefore detects profitable splitting without assuming an organizational form.

S.12.2 The response of policy

The following comparison holds taxes, enforcement and the value of public funds at their optimum without splitting. It signs the changes in supporting scores when splitting becomes available. Appendix S.11 gives the additional conditions needed to compare

separately optimized rate ratios.

Proposition S.12.2 (The first-order response to splitting)**.** *Suppose that the assumptions of Theorem 2 hold, that every strategy of the smaller set has at most one entity, that the schedule class is $\mathcal{T}$ or $\mathcal{T}_{\mathcal{P}^N}$, and that $\pi_1 = (T_1, e_1)$ is optimal with the smaller set, with a supporting multiplier μ_1 and a selection satisfying the conditions of Theorem 1. With the larger set at π_1, call a firm with $\Delta(\pi_1) > 0$ a* splitter*. Let each splitter choose an optimal strategy and every other firm keep its selected strategy. Compute the net values (16) with μ_1 and this selection. Suppose that splitters have positive mass and that every splitter's optimal strategies have at least two entities, and let $\underline{q}$ and $\bar{q}$ be the essential infimum and supremum of the reports of splitters' entities.*

(i) *For $r < \underline{q}$, the net value of raising the marginal rate rises by at least $\mathbb{E}[\Omega^1 \mathbf{1}\{\Delta > 0\}]$, which is positive when $\mu_1 > w_-$; for $r \geq \bar{q}$ it does not rise. The net value of raising the fee rises by at least the same amount.*

(ii) *Net revenue at π_1 falls by $\mathbb{E}[(1 + \ell_s)\Delta(\pi_1)] > 0$.*

Here Ω^1 is the net value of a dollar of burden at π_1 with the smaller set. These are comparisons of supporting scores. If $\mu_1 > w_-$ and the rate conditions (17) held with equality at π_1, the selected score is positive for a rise in rates below $\underline{q}$ and for a rise in the fee, and nonpositive for a rise in rates above $\bar{q}$. It equals the actual directional derivative for a proposed reform δT if $\sum_j \delta T(r_j)$ is the same at all optimal strategies for almost every firm in the larger economy. At fixed strategies, a higher value of public funds raises the selected score of every rate with a positive revenue score.

(Proof in Appendix S.12.5.) Extra entities expose splitters to low rates more than once, while their smaller reports reduce exposure to high rates. Reoptimization can also change the value of public funds, so these score comparisons alone do not order it or the final rates. In the running example at the optimum without splitting, admitting unrestricted splitting raises the bottom bracket's supporting score and lowers the top bracket's score, while gross receipts fall by about 1.27×10^{-2}. Reoptimization lowers the split share from 13.15 to 2.57 percent and compresses the rate difference from 38.45 to 21.35 percentage points. Table 1 gives the comparison with the same enforcement instruments available in both economies.

S.12.3 Subadditive schedules

A schedule is *subadditive* if $T(r_1 + r_2) \leq T(r_1) + T(r_2)$ whenever $r_1, r_2 \geq 0$ and $r_1 + r_2 \leq X$. Dividing a report among entities then weakly increases total taxes. For a schedule in (2),

define the *splitting premium*

$$\Psi(t) = \sup_{r_1, r_2 \geq 0,\, r_1 + r_2 \leq X} \int_0^{r_2} \big[t(r_1 + v) - t(v)\big]\, \mathrm{d}v.$$

This is the largest saving from dividing a report before paying the new entity's fee. Subadditivity requires the fee to cover that saving. If merging never raises manipulation costs, it also preserves an optimum when splitting becomes available.

Proposition S.12.3 (Subadditive schedules)**.** *A schedule in the class* (2) *is subadditive if and only if* $\Psi(t) \leq \tau$*. In particular, every schedule with nonincreasing marginal rates is subadditive, and a schedule with nondecreasing marginal rates is subadditive if and only if* $T(X) \leq 2T(X/2)$*.*

(Proof in Appendix S.12.5.) Under rising marginal rates, dividing a report among entities lowers the rates applied to part of the total reported base. A fee can offset that reward, so even strictly rising rates can be subadditive. When firms can exit, however, the same fee discourages participation.

The merging condition in Proposition 7 holds under Proposition S.3.1(ii): perform the reduction through the parent and omit the costly splits. It also holds in the benchmark technology. Without this condition, firms may gain from splitting even under a subadditive tax, as when splitting lowers later reduction costs (Example 2).

Merging never raises costs in Example 1: it preserves total reporting and avoids all split effort. For its threshold 0.3, the splitting premium is $0.3(t_2 - t_1)$ when rates increase. With the fee capped at 0.01, subadditivity permits a rate difference of at most $1/30$. The optimal difference with splitting and zero enforcement is about 21.35 percentage points, so the authority accommodates some splitting. In Example 2, by contrast, merging raises the effort of concealment. Its optimal flat tax has no tax premium, yet ten percent of firms split.

S.12.4 Examples

Example 2 and Appendix E illustrate the cost motive for splitting. The following example isolates the participation margin.

Example S.12.1 (A fee and participation)**.** Let $x = 1$, cost parameters uniformly distributed on $(0, 1)$, no enforcement, the no-manipulation strategy as the only operating strategy and an exit cost $c\kappa$. Consider constant schedules $T \equiv \tau$ with $0 \leq \tau \leq F < \kappa/2$. Since $F < \kappa$, firms with $c = 1$ do not manipulate under any such schedule. A firm exits when $c < \tau/\kappa$, so

$$\mathbb{E}R = \tau\Big(1 - \frac{\tau}{\kappa}\Big), \qquad \mathbb{E}D = \frac{\tau^2}{2\kappa}, \qquad \mathbb{E}P = \tau - \frac{\tau^2}{2\kappa}.$$

With equal welfare weights, a requirement $0 < G < F(1 - F/\kappa)$ is met at least cost by

$$\tau^* = \frac{\kappa}{2}\Big(1 - \sqrt{1 - 4G/\kappa}\Big), \qquad \mu = w\,\frac{1 - \tau^*/\kappa}{1 - 2\tau^*/\kappa}.$$

Revenue and the burden both rise with the fee on the feasible interval, so the smallest fee that raises G is optimal. Its base, the number of entities per firm, has elasticity $(\tau^*/\kappa)/(1 - \tau^*/\kappa) = 1 - w/\mu$, as Corollary 2 requires.

S.12.5 Proofs

Where a proof uses ℓ_s or $f_s(1)$, it establishes the general-density version described in Appendix F. In the main model both quantities equal γ_s. Additional common-exponent, welfare or technology restrictions are imposed in the statements being proved.

Fix characteristics s with the benchmark technology. Write x, p, σ, $\delta = \sigma/p$ and $Y(n)$ for its primitives. Let $T = \tau + T_0$ belong to the class (2), with $T_0(r) = \int_0^r t$. A strategy α with n entities, whose reports are at most the holdings of their entities under some division $y \in Y(n)$, has total report $q_\alpha = \sum_j r_j$, taxes paid $T[\alpha] = \sum_j T(r_j)$ and, at zero enforcement, effort $K_\alpha = p(x - q_\alpha) + \sigma(n - 1)$. The one-entity strategy that reports q has effort $p(x - q)$ and is written $\alpha(q)$. Positive enforcement replaces p and σ by $p + e_R$ and $\sigma + e_S$ throughout, so all statements extend to it.

No manipulation at the top. The largest tax saving is at most $L(x - \max_j r_j)$, as shown in Appendix B. If $h = x - \sum_j r_j$, then $\max_j r_j \geq \max_j y_j - h$. Hence (S.61) and $p \geq L$ give

$$T(x) - T[\alpha] \leq L\Big(x - \max_j y_j\Big) + Lh \leq \sigma(n - 1) + ph = K_\alpha,$$

which is (6).

S.12.5.1 Proof of Proposition S.12.1

We compare each split strategy with a one-entity strategy of the same effort, then show that the largest tax advantage is attainable. Write P^1 and P^2 for burdens without and with splitting. For a split strategy with total report $q_\alpha \geq (n - 1)\delta$, put $\varrho(\alpha) = T(q_\alpha - (n - 1)\delta) - T[\alpha]$ and let $\widetilde{\Pi}$ be its supremum over such strategies.

The one-entity report $q_\alpha - (n - 1)\delta$ has the same effort as α. Therefore

$$T[\alpha] + cK_\alpha \geq P^1(c) - \varrho(\alpha) \geq P^1(c) - \widetilde{\Pi}. \tag{S.63}$$

If $q_\alpha < (n-1)\delta$, reporting zero through one entity has weakly lower taxes and strictly less effort. Taking minima gives $0 \leq \Delta(c) \leq \max\{\widetilde{\Pi}, 0\}$, proving (iii) for an arbitrary admissible schedule.

Suppose now that T is convex. For reports $r \leq y$, let $q = \sum_j r_j \geq (n-1)\delta$. If $\sum_{i \neq j} r_i < (n-1)\delta$ for some j, then $T(q-(n-1)\delta) \leq T(r_j)$ and $\varrho(r) \leq 0$. Otherwise $q-(n-1)\delta \geq r_j$ for every j. Increasing any report weakly raises ϱ, by increasing increments of a convex function, and preserves these inequalities. Raising the reports to their holdings gives $\varrho(r) \leq \varrho(y) \leq \Pi_s(T)$. Consequently $\max\{\widetilde{\Pi}, 0\} = \max\{\Pi_s(T), 0\}$, which gives the required upper bound and the no-gain claims.

If $\Pi_s(T) > 0$, choose maximizing n, y and put $\tilde{q} = x-(n-1)\delta$. Positivity of the gain and $\delta > 0$ imply $0 < \tilde{q} < x$. The one-entity report $\tilde{q}$ minimizes burden exactly when $cp \in \partial T(\tilde{q})$. This subdifferential contains a value in $(0, p)$: its upper endpoint cannot be zero when the gain is positive, and its lower endpoint cannot equal p. Indeed, $t \leq L \leq p$ would then force $t = p$ on $[\tilde{q}, x]$, making the split strategy's burden at $c = 1$ equal to $T(x) - \Pi_s(T) < T(x)$, contrary to no manipulation at the top.

Choose the resulting $\tilde{c} \in (0,1)$. The reports y cost the same effort as $\tilde{q}$ and pay $\Pi_s(T)$ less tax. Hence $\Delta(\tilde{c}) \geq \Pi_s(T)$, with equality by the upper bound. Continuity of both burdens gives a positive gain on a neighborhood of $\tilde{c}$. This proves (i) and (ii).

For the welfare bounds, Theorem 2(ii) gives $V^2(G) \leq V^1(G)$. At every policy passing the test, part (iii) gives equal burdens. The accounting identities then give equal revenue and welfare, so taking suprema over those policies yields $V^1_\circ(G) \leq V^2(G)$. Equality follows if an optimum without splitting passes the test. □

The bound by the splitting premium. For holdings $y \in Y(n)$ with partial sums $S_k = y_1 + \cdots + y_k$, the definition of $\Psi(t)$ gives $T_0(S_k) - T_0(S_{k-1}) - T_0(y_k) \leq \Psi(t)$ for each $k \geq 2$. Adding over k and using $T(x-(n-1)\delta) \leq T(x)$,

$$T\big(x-(n-1)\delta\big) - \sum_j T(y_j) \;\leq\; T_0(x) - \sum_j T_0(y_j) - (n-1)\tau \;\leq\; (n-1)\big(\Psi(t)-\tau\big),$$

so $\Pi_s(T) \leq (n-1)(\Psi(t)-\tau)$ for the maximizing n, and a subadditive schedule passes the test. Under a linear schedule, $T_0(x-(n-1)\delta) - \sum_j T_0(y_j) = -t(n-1)\delta$, and $\Pi_s(T) = \max_n\{-(n-1)(\tau+t\delta)\}$, which is negative when $\tau + t\delta > 0$.

S.12.5.2 Proof of Proposition S.12.2

With the smaller set, write ν^1, n^1 and Ω^1 for the report measure, entity count and net burden value of a firm's selected strategy at π_1, and ν^2, n^2 and Ω^2 for those with the larger set under the described selection; by hypothesis $n^1 \leq 1$ and, for splitters, $n^2 \geq 2$. For a firm with $\Delta(\pi_1) = 0$ the strategy and the burden are the same, so $\Omega^2 = \Omega^1$ and $\nu^2 = \nu^1$. For a splitter, $P^2 < P^1$, so $U_s'(x - P^2) \leq U_s'(x - P^1)$ and $\Omega^2 \geq \Omega^1 \geq 0$, the last inequality by (65) and $\mu_1 \geq w_-$. The boundary terms $\mu_1 \mathbb{E}[f_s(1)\mathbf{1}\{x > r\}]$ and $\mu_1 \mathbb{E} f_s(1)$ in (74) are the same in both economies. Hence

$$\Xi_r^2 - \Xi_r^1 = \mathbb{E}\big[(\Omega^2 n_r^2 - \Omega^1 n_r^1)\mathbf{1}\{\Delta > 0\}\big], \qquad \Xi_\tau^2 - \Xi_\tau^1 = \mathbb{E}\big[(\Omega^2 n^2 - \Omega^1 n^1)\mathbf{1}\{\Delta > 0\}\big].$$

For $r < \underline{q}$, every entity of almost every splitter reports more than r, so $n_r^2 = n^2 \geq 2$, while $n_r^1 \leq 1$; thus $\Omega^2 n_r^2 - \Omega^1 n_r^1 \geq \Omega^1(n_r^2 - n_r^1) \geq \Omega^1$. For $r \geq \bar{q}$, $n_r^2 = 0$ and the integrand is $-\Omega^1 n_r^1 \leq 0$. For the fee, $n^2 \geq 2$ and $n^1 \leq 1$, so the integrand is at least Ω^1. When $\mu_1 > w_-$, (65) gives $\Omega^1 \geq (\mu_1 - w_-)(1 + \ell_s) > 0$, so $\mathbb{E}[\Omega^1 \mathbf{1}\{\Delta > 0\}] > 0$. This proves (i). For a proposed tax reform whose exposure is the same at all optimal strategies for almost every firm in the larger economy, the directional envelope theorem identifies its derivative with the selected score integrated against that reform. Without this regularity, the right derivative uses the minimum exposure among tied strategies, so the score comparison alone does not determine it.

Part (ii) is Theorem 2(i), with $\mathbb{E}[(1 + \ell_s)\Delta] > 0$ because $1 + \ell_s \geq 1$ and $\Delta > 0$ on a set of positive measure. For the last claim, at fixed strategies $\Xi_r = \mu \mathbb{E}[(1 + \ell_s)n_r] - \mathbb{E}[w_s(c)n_r] - \mu \mathbb{E}[f_s(1)\mathbf{1}\{x > r\}]$ is affine in μ with slope R_r, the revenue response of Appendix S.10.3.1. □

S.12.5.3 Proof of Proposition S.12.3

For $T = \tau + \int_0^{\cdot} t$ and $r_1, r_2 \geq 0$ with $r_1 + r_2 \leq X$,

$$T(r_1 + r_2) - T(r_1) - T(r_2) = \int_{r_1}^{r_1 + r_2} t - \int_0^{r_2} t - \tau = \int_0^{r_2} \big[t(r_1 + v) - t(v)\big]\, \mathrm{d}v - \tau,$$

so T is subadditive exactly when $\Psi(t) \leq \tau$. If t is nonincreasing, the integrand is nonpositive and $\Psi(t) = 0 \leq \tau$. If t is nondecreasing, the integral is nondecreasing in r_1, so the supremum over r_1 is attained at $r_1 = X - r_2$, where it equals $T(X) + \tau - T(X - r_2) - T(r_2)$. Since T is then convex, $T(X - r_2) + T(r_2)$ is convex and symmetric about $X/2$ and is smallest at $r_2 = X/2$, so $\Psi(t) = T(X) + \tau - 2T(X/2)$, and $\Psi(t) \leq \tau$ exactly when

$T(X) \leq 2T(X/2)$. □

S.13 Tax schedules with few brackets

The unrestricted optimum need not be a schedule with a few brackets. When the optimal marginal rate has bounded variation and an affine schedule can raise more than the required revenue, the welfare loss from using N brackets is at most a constant times $1/N^2$. Under additional conditions, that order is also a lower bound.

The upper bound holds because a piecewise-linear tax schedule can closely approximate the optimal schedule, and small uniform changes in the schedule have bounded effects on burdens, revenue and welfare. Firms may switch strategies abruptly without invalidating either property.

Throughout this section, policies use the full schedule class $\mathcal{T}$ in (2). An *N-bracket schedule* belongs to $\mathcal{T}$ and is affine on each of at most N intervals partitioning $[0, X]$. Let Π_N be the set of policies with an N-bracket schedule. Then $V_N(G) = \sup\{W(\pi) : \pi \in \Pi_N,\ Q(\pi) \geq G\}$ is the best welfare attainable with N brackets. For $T \in \mathcal{T}$ let

$$d_N(T) = \min\{\|T - S\|_\infty :\ S \in \mathcal{T} \text{ is an } N\text{-bracket schedule with } S(0) = T(0)\}$$

be the uniform distance from T to the nearest N-bracket schedule with the same fee.[1]

The approximations consist of chords between chosen thresholds (Appendix S.13.3). Chords preserve endpoint taxes, and their slopes are averages of admissible rates, so they remain in $\mathcal{T}$. Concentrating thresholds where rates vary most gives the $1/N^2$ bound even when a bounded-variation rate has jumps, as a piecewise-monotone rate may. Equally spaced thresholds generally give only order $1/N$ at a jump.

Small schedule errors change burdens and the Lagrangian at most proportionally. If $\|S - T\|_\infty \leq d$, then at any common enforcement level

$$|P(S, e) - P(T, e)| \leq md, \qquad |\mathcal{L}_\mu(S, e) - \mathcal{L}_\mu(T, e)| \leq C_\mu\, d, \tag{S.64}$$

where the first bound holds for every firm and the second for $\mu \geq w_-$. Analogous bounds hold for welfare and net revenue, with C_μ given in Appendix S.13.3. A strategy has at most m reports, so its tax payment changes by at most md. Minimizing over strategies preserves this bound. The accounting identity and the bounds on welfare weights then

[1] The minimum is attained because N-bracket schedules in $\mathcal{T}$ form a compact set. Their thresholds can be described by points of a simplex, with coincident thresholds representing empty brackets, and their rates and fees lie in bounded intervals.

control aggregate outcomes.

Theorem S.13.1 (The cost of few brackets)**.** *Maintain regularity, no manipulation at the top and the welfare-weight restriction. Use the full schedule class $\mathcal{T}$, and let $\pi^* = (T^*, e^*)$ be optimal with a supporting multiplier $\mu \geq w_-$.*

(i) *For every $N \geq 1$,*

$$0 \leq V(G) - \sup_{\pi \in \Pi_N} \mathcal{L}_\mu(\pi) \leq C_\mu \, d_N(T^*).$$

(ii) *Suppose that some policy with an affine schedule raises net revenue strictly above G. Then policies in Π_N can raise G for every $N \geq 1$, and there is a constant C, given in Appendix S.13.3, such that*

$$0 \leq V(G) - V_N(G) \leq C \, d_N(T^*) \qquad \text{for every } N \geq 1. \tag{S.65}$$

For every $T \in \mathcal{T}$, the approximation error satisfies

$$\begin{aligned}
&d_N(T) \leq LX/(4N), && N \geq 1, \\
&d_N(T) \leq \mathrm{Lip}(t)X^2/(8N^2), && N \geq 1 \quad \text{if } t \text{ is Lipschitz}, \\
&d_N(T) \leq X\,\mathrm{TV}(t)/[4(N-1)^2], && N \geq 2 \quad \text{if } t \text{ has bounded variation}.
\end{aligned}$$

Here $\mathrm{TV}(t)$ is the least variation among representatives of the marginal rate. Thus both loss bounds are of order $1/N$ in general and $1/N^2$ when t^* has bounded variation.

(Proof in Appendix S.13.3.) Part (i) bounds loss in the Lagrangian, which is a lower bound on loss in welfare at the same revenue requirement, because every feasible policy satisfies $\mathcal{L}_\mu \geq W$. Part (ii) directly bounds the welfare loss by adjusting the approximation to restore funding.

The approximation bounds use only the schedule class. The policy-loss bounds take a supporting multiplier as given and do not otherwise require strict feasibility. Strict feasibility with an affine schedule makes the funding adjustment available for every N. Without it, the N-bracket class may fail to raise G for some N.

For fixed thresholds the policy class is convex, so Theorems F.1 and 1 characterize optimal rates whenever that class is strictly feasible. Choosing thresholds as well is not a concave problem, because the set of all N-bracket schedules is not convex. The approximation bounds concern optimal values and do not require this convexity.

Refining a fixed grid preserves every old schedule. Best attainable welfare therefore weakly increases, although individual firms may lose. Welfare converges to the unrestricted optimum as the grid becomes arbitrarily fine. This does not require bounded variation of

the optimal rate or an affine strictly feasible policy.

Proposition S.13.1 (The value of additional brackets)**.** *Maintain the assumptions of Proposition 1 and strict feasibility for the full class (2). Let $(\mathcal{P}_k)$ be nested finite partitions of $[0, X]$ whose maximum interval length tends to zero. Let $V_{\mathcal{P}_k}(G)$ be optimal welfare when rates are constant on each interval of $\mathcal{P}_k$, with the same fee and enforcement bounds. These problems are strictly feasible for all sufficiently large k. Their values are nondecreasing wherever feasible, and*

$$V_{\mathcal{P}_k}(G) \longrightarrow V(G). \tag{S.66}$$

Every accumulation point of their optimal policies, with uniform convergence of schedules and convergence of enforcement, is optimal in the full class.

Proof. We approximate a strictly feasible policy and then the unrestricted optimum on the refining grids. Let $I_k T$ interpolate T at the thresholds of $\mathcal{P}_k$, and let h_k be the maximum interval length. Interpolation preserves the fee and keeps rates in $[0, L]$. By (S.67), $\|I_k T - T\|_\infty \le L h_k/4 \to 0$, so (S.68) gives convergence of welfare and revenue at unchanged enforcement. Interpolating a strictly feasible policy π^+ therefore gives strict feasibility on every sufficiently fine grid. Nested grids give nested feasible policy sets, so their values are nondecreasing and bounded above by $V(G)$.

Let π^* be a full-class optimum. For $\eta \in (0, 1)$, concavity of net revenue gives

$$Q\big((1-\eta)\pi^* + \eta\pi^+\big) \ge G + \eta\{Q(\pi^+) - G\} > G.$$

For fixed η, interpolating the schedule of this policy gives feasible grid policies for all large k, whose welfare converges to $W((1-\eta)\pi^* + \eta\pi^+)$. Hence $\liminf_k V_{\mathcal{P}_k}(G)$ is at least this welfare. Letting $\eta \downarrow 0$ proves (S.66). Compactness of the full policy set and continuity of W and Q then make every accumulation point of grid optima feasible with welfare $V(G)$. □

The same proof applies under the general-density restrictions of Appendix F. It also gives $V_N(G) \to V(G)$ when thresholds are freely chosen, since each fixed-grid policy is available once N is large enough. The stronger conditions in Theorem S.13.1(ii) give a quantitative bound for every N.

The policy calculations in Example 1 use a fixed threshold at 0.3. Their global certificates apply within that two-bracket class and do not measure the gain from additional thresholds. By contrast, the flat schedule in Example 2 is optimal over the full class, so no additional brackets improve welfare there.

S.13.1 A lower bound

The $1/N^2$ upper bound is sharp in order for some economies. A constant optimal rate, as in Example S.12.1, needs only one bracket. The lower bound requires a varying optimal rate and a positive density of entities that can adjust their reports. Under the conditions below, restricting policy to N brackets loses at least a positive constant times $1/N^2$.

Proposition S.13.2 (A lower bound)**.** *Maintain the assumptions of Theorem S.13.1, and fix an optimum $\pi^* = (T^*, e^*)$ with a supporting multiplier μ and selection satisfying (75). Suppose that there are a closed interval $\mathcal{I} \subset (0, X)$, constants $\kappa, L_p, m_0, \xi_{\min} > 0$, and a set $\Theta_{\mathcal{I}}$ of firms with the following three properties.*

(a) *The optimal marginal rate t^* is continuous and monotone on $\mathcal{I}$, with $|t^*(q) - t^*(r)| \geq \kappa|q - r|$ for all $q, r \in \mathcal{I}$.*

(b) *The selection assigns to each firm $(s, c) \in \Theta_{\mathcal{I}}$ a single strategy $\alpha_{s,c}$ with an entity whose report $q_{s,c}$ lies in $\mathcal{I}$, and feasible strategies $\alpha_{s,c}(q)$, $q \in \mathcal{I}$, change only this report to q, with $\alpha_{s,c}(q_{s,c}) = \alpha_{s,c}$. The effort of $\alpha_{s,c}(q)$ is differentiable in q with $\partial_q K(\alpha_{s,c}(q); e) = -\xi(e)\, p_{s,c}(q)$ for every $e \in E$. Here $p_{s,c}$ is L_p-Lipschitz on $\mathcal{I}$, and ξ is a function of enforcement alone, the same for all firms, with $\xi(e^*) = 1$ and $\xi(e) \geq \xi_{\min}$ for every $e \in E$.*

(c) *The measure $\mathcal{O} \mapsto \mathbb{E}\big[\Omega_s(c)\mathbf{1}\{(s, c) \in \Theta_{\mathcal{I}},\ q_{s,c} \in \mathcal{O}\}\big]$ has a density of at least m_0 on $\mathcal{I}$.*

Then, for every $N \geq 1$,

$$V(G) - V_N(G) \;\geq\; V(G) - \sup_{\pi \in \Pi_N} \mathcal{L}_\mu(\pi) \;\geq\; \frac{\xi_{\min}\, m_0\, |\mathcal{I}|^3}{64\, N^2} \min\Big\{\frac{\kappa^2}{L_p},\, \kappa\Big\},$$

and the constant 64 can be replaced by 24 if t^* is decreasing on $\mathcal{I}$.

(Proof in Appendix S.13.3.) At the optimum, each affected entity's marginal reduction cost equals the rate it faces. Replacing a varying rate by a constant creates a gap and allows a gain from moving the report that is quadratic in that gap. On a fixed fraction of the interval, a step function with at most N steps must differ from a rate whose slope is bounded away from zero by order $1/N$.

The optimality conditions make the effect at fixed strategies nonpositive, and each dollar gained by adjusting a report reduces the Lagrangian by its net burden value. The resulting loss has order $1/N^2$. Reoptimizing enforcement rescales marginal costs by the factor in (b); its positive lower bound limits how far this can reduce the loss. Thus, when t^* has bounded variation, an affine policy is strictly feasible, and (a)–(c) hold on an interval, the best N-bracket welfare loss lies between two positive constants times $1/N^2$.

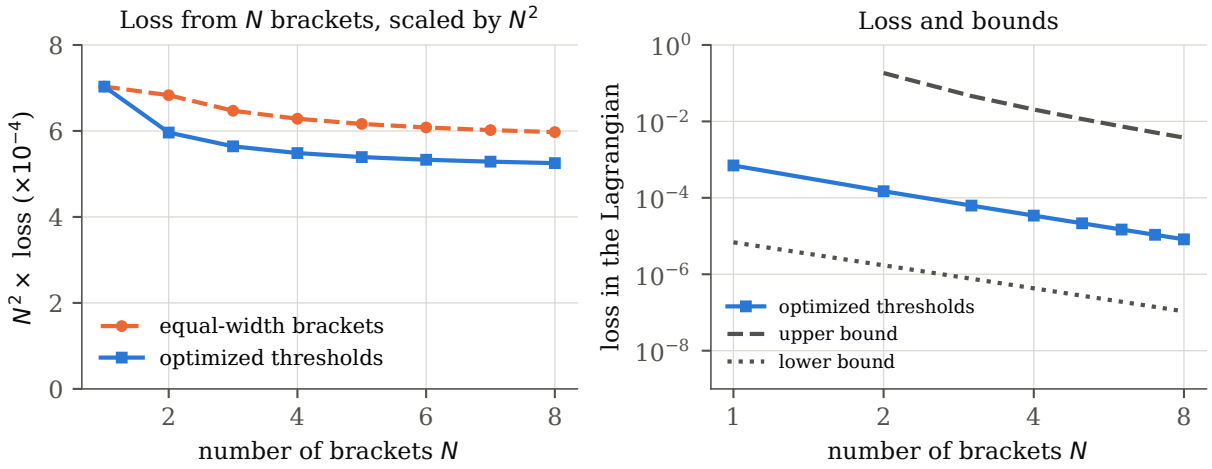


Figure S.2: Computed bracket losses fall at approximately $1/N^2$. Firms can neither split nor exit; potential tax bases and cost parameters are independent and uniformly distributed. The optimal rate falls from about 0.48 to 1/3. Losses are in the Lagrangian at $\mu = 2$. Left: N^2 times the loss, with rates optimized for equal-width brackets and with rates and thresholds optimized jointly. Right: optimized-threshold losses and theoretical bounds on logarithmic scales. The upper bound is $C_\mu X \,\mathrm{TV}(t^*)/[4(N-1)^2]$, with $X = 1$ and C_μ from (S.64). The lower bound uses $[0.1, 0.5]$ in Proposition S.13.2; since the optimal rate falls there, its constant 64 becomes 24. The bounds concern the smallest loss over all N-bracket schedules, which the computed schedules need not attain.

S.13.2 An example

In a falling-rate example, computed losses fall like $1/N^2$ and lie far from both general bounds (Figure S.2).

Let potential tax bases and cost parameters be independent and uniformly distributed. Exclude splitting and exit, and let reducing by h cost $c(h + h^2/2)$. Take no fee or enforcement, $L = 1$, $w = 1$ and $\beta = 1/2$. Proposition S.10.2 gives $t^*(r) = (1-r)/[3\log(2-r)]$, falling from about 0.48 to 1/3.

The threshold problem is not concave. The joint search uses equal-width brackets, quantiles of $|t^{*\prime}|^{2/3}$ and six random partitions; the best deterministic and random results differ by less than 0.03 percent. The plotted schedules need not minimize Lagrangian loss over all N-bracket schedules, so their losses alone do not certify a welfare bound.

S.13.3 Proofs and approximation bounds

Where a proof uses ℓ_s or $f_s(1)$, it establishes the general-density version described in Appendix F. In the main model both quantities equal γ_s. Additional common-exponent, welfare or technology restrictions are imposed in the statements being proved.

S.13.3.1 Proof of Theorem S.13.1: approximation errors

We first bound the error on each interval, then choose a partition. On a subinterval $[a, b]$ let $T^{a,b}(r) = [(b-r)T(a) + (r-a)T(b)]/(b-a)$ be the chord of T. Subtracting it gives

$$T(r) - T^{a,b}(r) = \frac{(r-a)(b-r)}{b-a}\Big[\frac{1}{r-a}\int_a^r t - \frac{1}{b-r}\int_r^b t\Big],$$

a multiple of the difference between the average rates on (a, r) and on (r, b). Hence

$$|T(r) - T^{a,b}(r)| \le \frac{(r-a)(b-r)}{b-a} \operatorname*{osc}_{(a,b)} t \le \frac{b-a}{4} \operatorname*{osc}_{(a,b)} t, \tag{S.67}$$

where $\operatorname{osc}_{(a,b)} t$ is the essential supremum minus the essential infimum of t on (a, b). Use these chords on a partition of $[0, X]$ into at most N cells. The resulting piecewise affine schedule keeps the fee and has nonnegative values at the thresholds. Its slopes are averages of admissible rates, so it is an N-bracket schedule in $\mathcal{T}$.

(a) The oscillation is at most L, and equal cells of length X/N give the bound. In the class (S.1) with negative rates, the same argument gives $(L - L_-)X/(4N)$.

(b) Let $q = T - T^{a,b}$. Then $q(a) = q(b) = 0$ and $|q''| \le \mathrm{Lip}(t)$ almost everywhere. The functions $\pm q - \mathrm{Lip}(t)(r-a)(b-r)/2$ are convex and vanish at a and b. They are therefore nonpositive, giving $|q| \le \mathrm{Lip}(t)(r-a)(b-r)/2 \le \mathrm{Lip}(t)(b-a)^2/8$. Equal cells give the bound.

(c) The claim is immediate if $X\,\mathrm{TV}(t) = 0$. Otherwise take a right-continuous representative with least variation, and let $|\,\mathrm{d}t|$ be its variation measure. The Lebesgue–Stieltjes construction gives $|\,\mathrm{d}t|((0, X]) \le \mathrm{TV}(t)$ and bounds the oscillation on each interval by its variation (Folland, 1999, Section 1.5). Let $\varrho_N = X\,\mathrm{TV}(t)/(N-1)^2$. Starting from $a_0 = 0$, choose

$$a_i = \sup\Big\{z \in [a_{i-1}, X] : \ (z - a_{i-1}) \operatorname*{osc}_{(a_{i-1},z)} t \le \varrho_N\Big\}.$$

The product is nondecreasing and left-continuous in z, so the supremum is attained. Every cell has positive length and length times oscillation at most ϱ_N. If $a_i < X$, letting $z \downarrow a_i$ gives

$$(a_i - a_{i-1})\, O_i \ge \varrho_N, \qquad O_i = \lim_{z \downarrow a_i} \operatorname*{osc}_{(a_{i-1},z)} t \le |\,\mathrm{d}t|\big((a_{i-1}, a_i]\big),$$

where a jump at the right end of a cell is assigned to that cell and so is not counted twice. For any k cells that end before X, the Cauchy–Schwarz inequality and the disjointness of

the intervals $(a_{i-1}, a_i]$ give

$$k\sqrt{\varrho_N} \le \sum_{i=1}^{k} \sqrt{(a_i - a_{i-1})O_i} \le \Big(\sum_{i=1}^{k}(a_i - a_{i-1})\Big)^{1/2}\Big(\sum_{i=1}^{k} O_i\Big)^{1/2} \le \sqrt{X\,\mathrm{TV}(t)}.$$

Since $\sqrt{X\,\mathrm{TV}(t)} = (N-1)\sqrt{\varrho_N}$, at most $N-1$ cells end before X, and the construction reaches X in at most N cells. By (S.67), each chord error is at most $\varrho_N/4$, which proves (c), for negative as well as nonnegative rates. □

S.13.3.2 Proof of the continuity bounds

The burden bound in (71), bounded welfare weights and the accounting identity (70) give, with $C_W = mw_+$ and $C_Q = m(1+\bar{J}) + \bar{J}$,

$$|W(S,e) - W(T,e)| \le C_W d, \qquad |Q(S,e) - Q(T,e)| \le C_Q d. \tag{S.68}$$

Define $C_\mu = m\big[\mu(1+\bar{J}) - w_-\big] + \mu\bar{J}$. For $\mu \ge w_-$, the function $P \mapsto U_s(x-P) + \mu(1+\ell_s)P$ has slope between 0 and $\mu(1+\ell_s) - w_-$, so by (72) the Lagrangian changes by at most $\mathbb{E}[(\mu(1+\ell_s) - w_-)md] + \mu\bar{J}d = C_\mu d$. □

S.13.3.3 Proof of Theorem S.13.1: policy losses

(i) The first inequality holds because π^* maximizes $\mathcal{L}_\mu$. Let S be an N-bracket schedule with $\|S - T^*\|_\infty = d_N(T^*)$ and $S(0) = T^*(0)$. By (S.64), $\mathcal{L}_\mu(S, e^*) \ge \mathcal{L}_\mu(\pi^*) - C_\mu d_N(T^*) = V(G) - C_\mu d_N(T^*)$.

(ii) We restore the revenue requirement by mixing the approximation with a strictly feasible affine policy. Let $\pi^\circ = (T^\circ, e^\circ)$ have an affine schedule and $Q(\pi^\circ) \ge G + \vartheta$ for some $\vartheta > 0$. Put $\bar{k}_i = \sup\{k_i(\alpha) : \alpha \in \mathcal{A}(s),\ s \in \mathcal{S}\}$ and $C_0 = w_+\big[m\bar{T} + \sum_i \bar{k}_i |e_i^\circ - e_i^*|\big]$. The policy $\hat{\pi}_N = (S, e^*)$ raises at least $G - \delta_N$ with $\delta_N = C_Q d_N(T^*)$ and has welfare at least $V(G) - C_W d_N(T^*)$, by (S.68). Take the convex combination that puts weight $\chi_N = \delta_N/(\vartheta + \delta_N)$ on π°. Because Q is concave,

$$Q\big((1-\chi_N)\hat{\pi}_N + \chi_N\pi^\circ\big) \ge (1-\chi_N)(G - \delta_N) + \chi_N(G + \vartheta) = G.$$

Adding an affine schedule creates no new thresholds, and the combination stays in $\mathcal{T}$ because the constraints of (2) are convex. If each firm keeps a strategy that is optimal at $\hat{\pi}_N$, its burden rises by at most $\chi_N[m\bar{T} + \sum_i \bar{k}_i|e_i^\circ - e_i^*|]$, so welfare falls by at most $\chi_N C_0$.

Since $V_N(G)$ is at least the welfare of this combination,

$$0 \le V(G) - V_N(G) \le C_W\, d_N(T^*) + C_0\, \frac{C_Q\, d_N(T^*)}{\vartheta + C_Q\, d_N(T^*)} \le \Big(C_W + \frac{C_0 C_Q}{\vartheta}\Big) d_N(T^*),$$

which is (S.65) with $C = C_W + C_0 C_Q/\vartheta$. The middle expression is a sharper bound, and it never exceeds $C_W\, d_N(T^*) + C_0$. □

S.13.3.4 Proof of Proposition S.13.2

Weak duality gives the first inequality, since $W \le \mathcal{L}_\mu$ at every policy that raises G. For the second, we bound the gains firms can obtain by moving their reports within the proposed brackets. Let $\pi_N = (T_N, e_N) \in \Pi_N$ with marginal rate t_N, and write $\xi_N = \xi(e_N) \ge \xi_{\min}$. For a firm in $\Theta_{\mathcal{I}}$ write $\alpha = \alpha_{s,c}$ and $q_0 = q_{s,c}$. By (c), q_0 lies in the interior of $\mathcal{I}$ for Ω-weighted almost every firm in $\Theta_{\mathcal{I}}$. Since $\alpha(q)$ is feasible for $q \in \mathcal{I}$ and α is optimal at π^*, q_0 minimizes $T^*(q) + cK(\alpha(q); e^*)$ over $\mathcal{I}$. The first-order condition gives $t^*(q_0) = c\, p_{s,c}(q_0)$, using $\xi(e^*) = 1$.

A refined supporting inequality. Let ν_α and Ξ_μ be the report measure and the measure of Appendix F, and let $\mathcal{I}_1, \dots, \mathcal{I}_k$, $k \le N$, be the intervals into which the thresholds of T_N divide $\mathcal{I}$, with lengths $l_1, \dots, l_k$ that add up to $|\mathcal{I}|$. On $\mathcal{I}_j$ the rate t_N is a constant $\hat{t}_j$. For firms in $\Theta_{\mathcal{I}}$ let $\varpi_{s,c} = T_N[\alpha] + cK(\alpha; e_N) - P_s(c; \pi_N) \ge 0$ be the gain from reoptimizing at the proposed policy; set it to zero for other firms. This is measurable, and

$$P_s(c; \pi_N) - P_s(c; \pi^*) \le \int (T_N - T^*)\, \mathrm{d}\nu_\alpha + c \sum_i (e_{N,i} - e_i^*) k_i(\alpha) - \varpi_{s,c}.$$

Repeating the argument for the supporting inequality (76) with this inequality, and using $\Omega \ge 0$ and (75), gives

$$\mathcal{L}_\mu(\pi_N) - \mathcal{L}_\mu(\pi^*) \le \int (T_N - T^*)\, \mathrm{d}\Xi_\mu + g \cdot (e_N - e^*) - \mathbb{E}[\Omega\, \varpi_{s,c}] \le -\mathbb{E}\big[\Omega\, \varpi_{s,c}\big].$$

The gain from moving one report. Put $\widetilde{P}(q) = T_N(q) + cK(\alpha(q); e_N)$. Feasible changes of this report give $\varpi_{s,c} \ge \widetilde{P}(q_0) - \min_{q \in \mathcal{I}_j} \widetilde{P}(q)$. On $\mathcal{I}_j$, $\widetilde{P}'(q) = \hat{t}_j - \xi_N c\, p_{s,c}(q)$, and $\xi_N c\, p_{s,c}$ is $\xi_N L_p$-Lipschitz because $c \le 1$. At q_0 the slope has absolute value $\delta_0 = |\hat{t}_j - \xi_N t^*(q_0)|$. Moving the report by h in the direction in which $\widetilde{P}$ decreases, while staying in $\mathcal{I}_j$, lowers $\widetilde{P}$

by at least $\int_0^h (\delta_0 - \xi_N L_p v)\,\mathrm{d}v$. If $\bar{h}$ is the distance from q_0 to the end of $\mathcal{I}_j$ in that direction,

$$\varpi_{s,c} \geq \max_{0 \leq h \leq \bar{h}} \Big(\delta_0 h - \frac{\xi_N L_p h^2}{2} \Big) \geq \frac{1}{2} \min\Big\{ \frac{\delta_0^2}{\xi_N L_p},\ \delta_0 \bar{h} \Big\}.$$

The function $\xi_N t^*$ is continuous and monotone on $\mathcal{I}_j$, with secant slopes of absolute value at least $\xi_N \kappa$. Let v_0 be the point of the closure of $\mathcal{I}_j$ at which $\xi_N t^* - \hat{t}_j$ changes sign, or the endpoint at which $|\xi_N t^* - \hat{t}_j|$ is smallest if it does not change sign, and let $z = |q_0 - v_0|$. Then $\delta_0 \geq \xi_N \kappa z$, so with $\varrho_* = \min\{\kappa^2 / L_p, \kappa\}$,

$$\varpi_{s,c} \geq \frac{\xi_N}{2}\, \varrho_*\, z \min\{z, \bar{h}\}.$$

If t^* is increasing, the slope of $\widetilde{P}$ at q_0 is positive to the left of v_0 and negative to the right. Thus its initial descent direction points away from v_0. The distance $\bar{h}$ runs from q_0 to the end of $\mathcal{I}_j$ on the far side of v_0. If t^* is decreasing, its initial descent direction points toward v_0, so $\bar{h} \geq z$ and the bound becomes $\frac{\xi_N}{2} \varrho_* z^2$, which is larger. Write $\mathcal{I}_j$ as $[v_0 - b_1, v_0 + b_2]$ with $b_1 + b_2 = l_j$. Since $\int_0^y z \min\{z, y - z\}\,\mathrm{d}z = y^3/8$, in both cases

$$\int_{\mathcal{I}_j} z \min\{z, \bar{h}\}\,\mathrm{d}q_0 \ \geq\ \frac{b_1^3 + b_2^3}{8} \ \geq\ \frac{l_j^3}{32},$$

and when t^* is decreasing the integral is at least $(b_1^3 + b_2^3)/3 \geq l_j^3/12$.

Integrating the gains. The lower bound on $\varpi_{s,c}$ depends on the firm only through q_0, so by (c),

$$\mathbb{E}\big[\Omega\, \varpi_{s,c}\big] \geq \sum_{j=1}^{k} m_0\, \frac{\xi_{\min}}{2}\, \varrho_*\, \frac{l_j^3}{32} \geq \frac{\xi_{\min} m_0 \varrho_*}{64} \cdot \frac{|\mathcal{I}|^3}{k^2} \geq \frac{\xi_{\min} m_0 \varrho_* |\mathcal{I}|^3}{64 N^2},$$

using $\sum_j l_j^3 \geq |\mathcal{I}|^3/k^2$ for lengths that add up to $|\mathcal{I}|$. With $l_j^3/12$ in place of $l_j^3/32$, the constant 64 becomes 24 when t^* is decreasing. Since $\mathcal{L}_\mu(\pi^*) = V(G)$ and the bound holds for every $\pi_N \in \Pi_N$, the result follows. □

S.14 Transfers and outside opportunities

The maintained enforcement instruments raise real private costs. This appendix considers payments to the authority and exit losses with a different dependence on private information. The general density score $\ell_s(c)$ is defined in Appendix F. Under the main specification

it equals γ_s.

S.14.1 Public receipts from manipulation costs

When firms cannot exit, the results extend to a common public share $\iota \in [0,1]$ of every manipulation cost. Receipts then satisfy

$$\mathbb{E}(R + \iota D) = \mathbb{E}\big[(1 + (1-\iota)\ell_s)P - (1-\iota)f_s(1)T(x)\big].$$

Use $1 + (1-\iota)\ell_s$ in place of $1 + \ell_s$ and multiply the boundary term by $1 - \iota$. Impose the corresponding shape restriction. With exit, this expression exceeds receipts by $\iota\mathbb{E}[\mathbf{1}_{\text{exit}}D]$, because exiting firms' lost surplus is not a payment to the authority. The extension therefore does not cover different public shares across strategies or general audit and penalty policies.

S.14.2 General exit losses

If exit instead entails a general loss $\omega_s(c)$, the accounting identity is

$$R = P - cP_c - \mathbf{1}_{\text{exit}}\{\omega_s(c) - c\omega_s'(c)\}. \tag{S.69}$$

Integration adds the exit correction and leaves the boundary term $\mathbb{E}[f_s(1)P_s(1;\pi)]$, which need not be linear in policy. With uniformly distributed cost parameters and no manipulation at the top, the Lagrangian can still be nonconcave for every value of public funds (Appendix S.15).

S.15 General outside opportunities

Arbitrary exit losses can destroy concavity of the Lagrangian even when cost parameters are uniformly distributed and firms with the highest cost parameter do not manipulate.

Let the no-manipulation strategy be the only operating strategy. The tax schedule is constant, $T \equiv \tau$ with $\tau \in [0,1]$, and cost parameters are uniform on $(0,1)$. For $\psi > 1$ let the loss from exit be

$$\omega(c) = -\frac{1}{\psi}\log\{1 - c(1 - \exp(-\psi))\}.$$

The loss $\omega(c)$ has a truncated exponential distribution on $[0,1]$. Since $\omega(1) = 1 \geq \tau$, the no-manipulation strategy is optimal at $c = 1$. The burden is $\min\{\tau, \omega(c)\}$ and revenue is

$\tau \Pr(\omega(c) \geq \tau)$. With equal welfare weights,

$$\frac{\mathrm{d}^2}{\mathrm{d}\tau^2}\big[W(\tau) + \mu\,\mathbb{E}R(\tau)\big] = \frac{\psi \exp(-\psi\tau)}{1-\exp(-\psi)}\big[w - 2\mu + \mu\psi\tau\big].$$

At $\mu = w$, this is positive for $\tau > 1/\psi$. For $\psi \geq 2$, it is positive at $\tau = 1$ for every $\mu \geq 0$: the factor $w - 2\mu + \mu\psi\tau$ then equals $w + \mu(\psi - 2) > 0$. With $\psi \geq 2$, no multiplier makes the Lagrangian concave.

Equation (S.69) follows by checking operating and exiting firms separately. Under absolute continuity and integrability, it integrates to

$$\mathbb{E}R = \mathbb{E}\big[(1+\ell_s)P\big] - \mathbb{E}\big[f_s(1)P_s(1)\big] - \mathbb{E}\big[\mathbf{1}_{\text{exit}}\big(\omega_s(c) - c\omega_s'(c)\big)\big],$$

and both corrections must be kept.

S.16 Numerical implementation and additional calculations

S.16.1 Replication of the three-base example

Replication. The files in `code/running_examples/` record the full-precision policies, exact integration and independent rational certificates. The checker also compares the finite reduction with a separate partition dynamic program in 132 policy/type cases, allowing every grid partition and up to 27 entities. All values agree exactly. The grid check supplements the continuous-choice proof above. The reform and proportional-tax calculations are evaluated by the same exact integration, without discretizing cost types.

S.16.2 Numerical search and display of the many-base example

The calibration, unrestricted-choice reduction and full-class certificates are in Appendix N.4 of the paper.

Without splitting, marginal rates rise from 17.70 to 56.02 percent. With splitting, the bottom rate rises to 17.92 percent and the top rate falls to 41.72 percent. The latter rate applies both on $(17/40, 11/20)$ and on $(7/8, 1)$, with a hump between these ranges. Rates below 0.3 rise by between about 0.22 and 1.20 percentage points. The share of firms splitting falls from 0.59 percent at the old schedule to 0.16 percent after reoptimization. Every firm that splits uses two entities at the displayed optimum, although all finite counts

are feasible.

For each potential base, a selected feasible plan on each cost interval gives an affine upper bound on expected burden as a function of the marginal rates. The numerical search successively adds these bounds and maximizes their minimum. The search treats all 40 rates as independent variables in $[0, 0.6]$. A final refinement uses the adjacent-rate equalities found by the search. The certificate still compares the candidate with every admissible schedule.

For display, Figure 1 uses shape-preserving cubic interpolation through the bracket rates at their midpoints, with constant values over the outer half-brackets. The interpolation preserves every computed rate and does not overshoot adjacent values. All calculations and certificates use the original step schedules.

S.16.3 Additional calculations for the sequential example

Appendix E of the paper establishes unrestricted optimal choices and certifies both policy optima.

The solution and welfare comparison are

$$\begin{aligned} r &\simeq 0.07, & e_R^0 &\simeq 0.05, \\ t^0 &\simeq 48.16\%, & \mu^0 &\simeq 1.08, \\ W^0 &\simeq 0.53, & W^0 - W &\simeq 9.87 \times 10^{-3}. \end{aligned}$$

Here $W^0 = 1 - t^0 + (t^0)^2/[2(7 + e_R^0)]$. The comparison holds net revenue fixed and optimizes taxes and enforcement in each economy.

S.16.3.1 Operation order and single-operation tests

At zero enforcement, $C(q) = q + 6q^2$. Splitting into thirds and reducing each report to $1/6$ requires effort

$$2\left(\frac{3}{4}\right) + 3[C(1/3) - C(1/6)] = \frac{3}{2} + 3\left(1 - \frac{1}{3}\right) = \frac{7}{2}.$$

Reducing the parent to $1/2$ first and then dividing that report into thirds requires

$$C(1) - C(1/2) + 2\left(\frac{3}{4}\right) = 7 - 2 + \frac{3}{2} = \frac{13}{2}.$$

Both plans end at reports $(1/6, 1/6, 1/6)$, while every reduction remains additive along its path.

At the joint optimum, $C_e(1) = 71/10$. For $c = 1/11$, the burden from a single reduction is minimized at an endpoint by strict concavity in its final report, and is at least

$$\min\{T(1), cC_e(1)\} = \min\{1/2, 71/110\} = 1/2.$$

A single split leaves total tax at $1/2$ and adds positive effort. Yet three-entity concealment costs $5/11 < 1/2$. Initial one-operation tests therefore fail to detect a profitable sequence.

S.16.3.2 Enforcement responses and the figure

At a flat rate $0 < t \leq 1$, $B(e) > t$ throughout the enforcement rectangle. Hence

$$\begin{aligned} \Pr(\text{concealment}) &= \frac{t}{B(e)}, & \mathbb{E}\sum_j r_j &= 1 - \frac{t}{B(e)}, \\ \mathbb{E}R &= t\left(1 - \frac{t}{B(e)}\right), & \mathbb{E}D &= \frac{t^2}{2B(e)}, \\ \mathbb{E}P &= t - \frac{t^2}{2B(e)}, & W &= 1 - t + \frac{t^2}{2B(e)}. \end{aligned} \tag{S.70}$$

Public expenditure must be subtracted from gross receipts to obtain net revenue. Away from the two organizational ties, the split share is $t/B(e)$ when $n(e) \geq 2$ and zero when $n(e) = 1$, and $\mathbb{E}n = 1 + (n(e) - 1)t/B(e)$.

Within a region with fixed n, $B_R = 1$ and $B_S = n - 1$. For either instrument i,

$$\frac{\partial \mathbb{E}R}{\partial e_i} = \frac{t^2 B_i}{B^2}, \qquad \frac{\partial \mathbb{E}\sum_j r_j}{\partial e_i} = \frac{tB_i}{B^2}, \qquad \frac{\partial \mathbb{E}D}{\partial e_i} = -\frac{t^2 B_i}{2B^2}. \tag{S.71}$$

Anti-splitting enforcement makes concealment more costly and therefore raises reporting. This effect ends when firms that conceal use only one entity. Both instruments reduce aggregate private manipulation costs even though they weakly raise the effort of every fixed plan.

At $e_S = 1/4$, two and three concealed entities tie for every manipulating firm; at $e_S = 9/4$, one and two tie. Burdens, receipts and total reports are continuous at these boundaries, though their derivatives are one-sided. Entity counts and split shares can depend on tie-breaking.

Figure 3(a) compares $\min\{1/2, 5c\}$ with the reductions-only burden $\min\{1/2, 7.1c\}$ at the same policy. Panel (b) fixes $t = 1/2$ and $e_R = 1/10$, varies e_S, and plots $1 - t/B(e)$.

This experiment holds the tax policy fixed and lets net revenue vary. The calculations are analytic. An independent dynamic program on grids of reports, permitting every reduction to a lower grid point and every binary division, reproduces the burden formula in 2,805 checks without imposing the entity count, equal divisions or operation order.

References

Arrow, K. J., Hurwicz, L., and Uzawa, H. (1958). *Studies in Linear and Non-Linear Programming*. Stanford University Press, Stanford, CA.

Folland, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*. Wiley, New York, 2nd edition.

Guesnerie, R. (1995). *A Contribution to the Pure Theory of Taxation*. Econometric Society Monographs 25. Cambridge University Press, Cambridge.

Hammond, P. J. (1979). Straightforward individual incentive compatibility in large economies. *Review of Economic Studies*, 46(2):263–282.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857.

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.

Nedić, A. and Ozdaglar, A. (2009). Subgradient methods for saddle-point problems. *Journal of Optimization Theory and Applications*, 142(1):205–228.

Stoye, J. (2009). More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315.