The Division of Surplus and the Burden of Proof

Deniz Kattwinkel (University College London) **Justus Preusser** (Bocconi)

Many contractual relationships rely on **evidence** to resolve asymmetric info.

- Tax enforcement: agent's hidden wealth / revenue.
- Investor-entrepreneur: profit from investment.
- Procurement: cost of public good.

Problem: acquiring evidence is **costly**.

- Who bears burden of acquiring evidence?
- How is underlying decision distorted?

Fundamental tension: who bears cost of evidence?

Uncertain Tax Treatment policy in UK.

- Firms must report uncertain tax positions earlier.
- Benefit: reduce compliance discovery, avoid later litigation (HMRC).
- Concern: firms' cost disproportionate (House of Lords Economic Affairs Committee).

German bureaucracy reduction act (*Bürokratieentlastungsgesetz IV*).

- Reduce record-keeping requirements.
- Benefit: reduce citizens' and firms' costs.
- Concern: insufficient time to uncover complex tax cases (CumEx & CumCum).

Canonical model: **surplus division**, Townsend (1979).

Principal and agent divide surplus.

Only agent knows true amount of surplus & has limited liability.

Townsend: principal can acquire costly evidence ("verify").

This paper: principal & agent can acquire costly evidence.

Full solution: division of surplus & burden of proof.

Key challenge: multiple instruments.

Transfer contingent on who provides evidence & what it shows.

New solution technique for multi-instrumental mechanism design.

Both principal and agent acquire evidence, but for fundamentally different reasons.

- Principal checks for misreports.
- If agent shows evidence, principal can save costs of checking.
- Incentivizing agent to provide evidence also incentivizes agent to report truthfully.

In all optimal mechanisms:

- Principal's effort **decreasing** in surplus.
- Agent's effort **non-monotone**, characterized by five intervals: $\searrow, \nearrow, \searrow, \nearrow, \searrow$.

Roadmap.

1. Model.
2. Literature.
3. Basic economics of evidence acquisition.
4. Main characterization.
5. Solution technique.
6. Takeaways and conclusion.

Model

Agent has privately known **surplus** $x \in [\underline{x}, \bar{x}] \subset [0, \infty)$, CDF F.

Principal can only seize **revealed surplus**.

Two ways of revealing surplus:

- Agent makes **advance payment** $y \in [\underline{x}, x] \Rightarrow$ principal can seize up to y.
- **Evidence** reveals $x - y \Rightarrow$ principal can seize up to x.

Principal: evidence with probability e_P ("effort") has cost $c_P(e_P)$, effort is contractible.

Agent: evidence with probability e_A ("effort") has cost $c_A(e_A)$, effort is hidden.

Principal has private funds $\tau > 0$.

Payoff (for both) = allocated portion of $(x + \tau)$ − private effort costs.

Mechanisms $m = (e_A, e_P, r_A, r_P, r_\emptyset)$ have following timing:

1. Agent advances $y \in [\underline{x}, x]$.
2. Mechanism recommends effort level $e_A(y)$ & agent exerts effort covertly.
3. If agent provided evidence, principal acquires no evidence.
   Else, principal exerts effort $e_P(y)$.
4. The principal seizes
   - x if there is evidence and $y \neq x$,
   - $x - r_A(x)$ if agent provided evidence and $y = x$,
   - $x - r_P(x)$ if principal provided evidence and $y = x$,
   - $y - r_\emptyset(y)$ if there is no evidence.

Feasibility: $r_A(y)$, $r_P(y)$, $r_\emptyset(y)$ are in $[0, y + \tau]$.

Def. Mechanism m is **incentive-compatible (IC)** if agent finds it optimal to

make full advance payment ($y = x$),

exert recommended effort $e_A(x)$,

disclose evidence whenever obtained.

Lemma. Revelation Principle & basic optimality: IC mechanisms are without loss.

No additional communication.

No alternative order of moves.

Rest of talk: Mechanism means feasible IC mechanism.

Assumptions.

c_A, c_P smooth, str. incr., str. cvx., $c_A(0) = c_P(0) = c_A'(0) = c_P'(0) = 0$.

$e_A c_A'(e_A)$ str. convex.

$c_A'(1) > c_P(1)$.

$\tau \geq B(c_A, c_P)$ for a bound $B(c_A, c_P) < \infty$.

Example: $c_A(e) = c_P(e) = e^2/2$ and $\tau \geq 1$.

Surplus division with verification. Townsend (1979), Border and Sobel (1987), Mookherjee and Png (1989), Chander and Wilde (1998), Kattwinkel and Preusser (2025).

Principal acquires evidence. Gale and Hellwig (1985), Ben-Porath et al. (2014), Halac and Yared (2020), Kattwinkel and Knoepfle (2023), Ahmadzadeh and Waizmann (2024), etc.

Agents acquire/disclose evidence. Jovanovic (1982), Verrecchia (1983), Ben-Porath et al. (2026), etc.

Counterparties provide verifiable info. Stahl and Strausz (2017), Ben-Porath et al. (2019), Bester et al. (2021), Lichtig and Mass (2024).

Empirical work on tax evasion. Allingham and Sandmo (1972), Andreoni et al. (1998), etc.

Incentives and profit

Agent's expected utility when advancing full surplus and exerting effort $\tilde{e}_A$:

$$(1-\tilde{e}_A)\underbrace{[e_P(x)r_P(x)+(1-e_P(x))r_\emptyset(x)]}_{R_{NA}(x)}+\tilde{e}_A r_A(x)-c_A(\tilde{e}_A)$$
$$=R_{NA}(x)+\tilde{e}_A\left(r_A(x)-R_{NA}(x)\right)-c_A(\tilde{e}_A). \qquad (*)$$

Obedience: recommended effort $e_A(x)$ maximizes $(*)$.

evidence rent $u_A(e_A(x))=e_A(x)\left(r_A(x)-R_{NA}(x)\right)-c_A(e_A(x))$
$$=e_A(x)c_A'(e_A(x))-c_A(e_A(x)).$$

$u_A(e_A(x))$ strictly increasing in $e_A(x)$.

Type x's expected utility from advancing the surplus and exerting $e_A(x)$:

$$U_m(x) = R_{NA}(x) + u_A(e_A(x)).$$

Type x's expected utility from advancing smaller amount $y \in [\underline{x}, x)$:

$$x - [e_P(y)x + (1 - e_P(y))(y - r_\emptyset(y))]\,.$$

Lowest loss from deviating:

$$\lambda_m(x) = \inf_{y \leq x} [e_P(y)x + (1 - e_P(y))(y - r_\emptyset(y))].$$

IC for type $x \Leftrightarrow U_m(x) \geq x - \lambda_m(x) =$ **info rent of type** x.

$$
\begin{aligned}
U_m(x) &= R_{NA}(x) + u_A(e_A(x)) \\
&\geq x - \lambda_m(x) \\
&= x - \inf_{y \leq x} [e_P(y)x + (1 - e_P(y))(y - r_{\emptyset}(y))].
\end{aligned}
$$

$$
\Pi_m(x) = x - U_m(x) - c_A(e_A(x)) - (1 - e_A(x))c_P(e_P(x)).
$$

Benefits of **principal's effort** $e_P(x)$:

Relaxes IC by raising λ_m.

Potentially relaxes IC by raising $U_m(x)$.

$$
\begin{aligned}
U_m(x) &= R_{NA}(x) + u_A(e_A(x)) \\
&\geq x - \lambda_m(x) \\
&= x - \inf_{y \leq x} [e_P(y)x + (1 - e_P(y))(y - r_\emptyset(y))].
\end{aligned}
$$

$$
\Pi_m(x) = x - U_m(x) - c_A(e_A(x)) - (1 - e_A(x))c_P(e_P(x)).
$$

Benefits of **agent's effort** $e_A(x)$:

Raises $\Pi_m(x)$ by lowering audit costs $(1 - e_A(x))c_P(e_P(x))$.

Relaxes IC by raising evidence rent $u_A(e_A(x))$.

Counterfactual: ignore evidence rent $u_A(e_A(x))$.

$$
e_A^{\text{eff}}(x) = \text{argmin}_{\tilde{e}_A \in [0,1]} c_A(\tilde{e}_A) + (1 - \tilde{e}_A)c_P(e_P(x)).
$$

Tight mechanisms

Recall: $\lambda_m(x) =$ type x's lowest loss when deviating in mechanism m

Def. m is **tight** if there does **not** exist m^* such that

$$\forall x, \quad \Pi_m(x) \leq \Pi_{m^*}(x) \quad \text{and} \quad \lambda_m(x) \leq \lambda_{m^*}(x),$$

with at least one inequality strict for at least one type.

Lemma. For every mechanism m there exists a tight mechanism m^* such that

$$\forall x, \quad \Pi_m(x) \leq \Pi_{m^*}(x) \quad \text{and} \quad \lambda_m(x) \leq \lambda_{m^*}(x).$$

Theorem. Let m be a tight mechanism such that e_P is interior on $[\underline{x}, \bar{x})$.

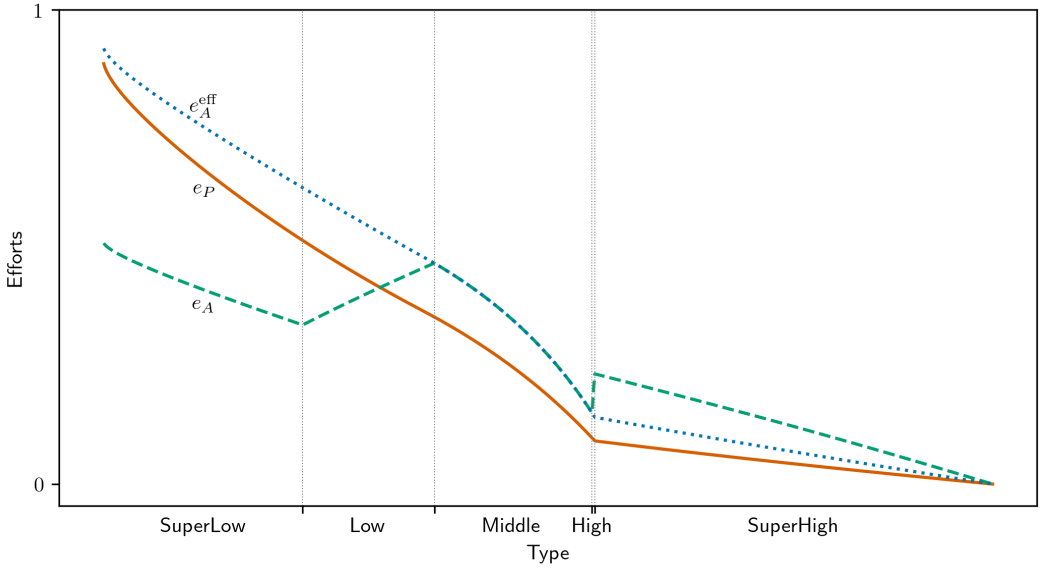
Then, $[\underline{x}, \bar{x})$ is partitioned into five consecutive intervals,

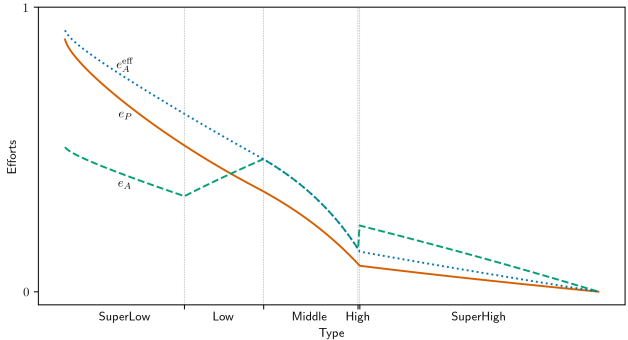
$$\text{SuperLow}, \text{Low}, \text{Middle}, \text{High}, \text{SuperHigh},$$

each with non-empty interior, such that

- e_P is decreasing on $[\underline{x}, \bar{x}]$, strictly so on $\text{SuperLow} \cup \ldots \cup \text{High}$.
- e_A is strictly decreasing on each of SuperLow, Middle, and SuperHigh, but strictly increasing on Low and High.
- e_P and e_A are continuous and approach 0 at $\bar{x}$.

($\rightarrow$ General characterization)

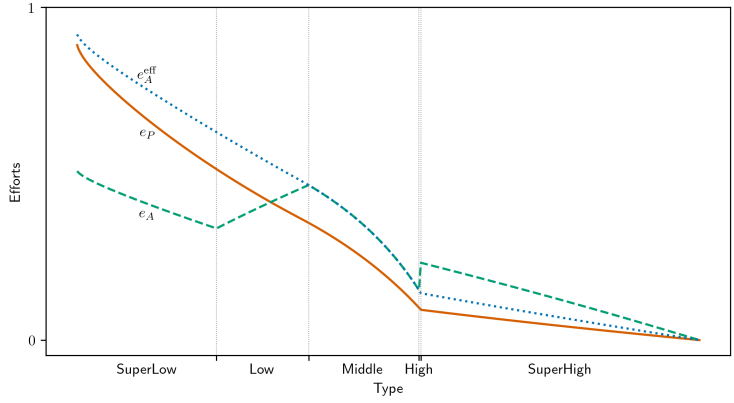

1
0
Efforts
e_A^{eff}
e_P
e_A
SuperLow
Low
Middle
High
SuperHigh
Type



Facts:

$e_P(x)$ and $e_A^{\text{eff}}(x)$ decrease in x.

$x - \lambda_m(x)$ increases in x (remember IC: $U_m(x) \geq x - \lambda_m(x)$).

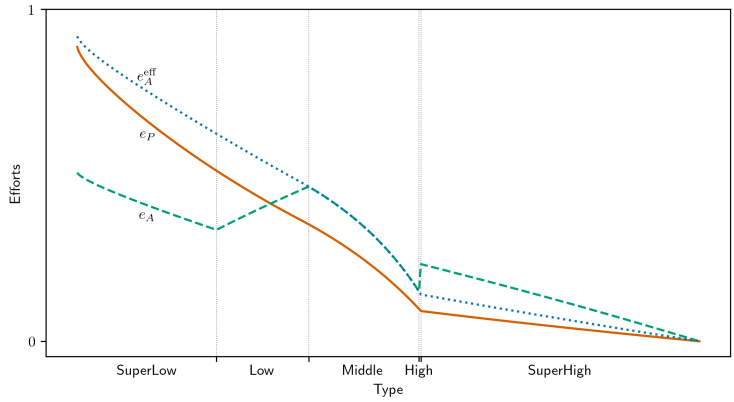


Types $x \in \underline{\text{SuperLow}}$: IC slack, $U_m(x) > x - \lambda_m(x)$.

For $\underline{x}$, advancing truthfully + evidence rent $\succ$ losing everything.

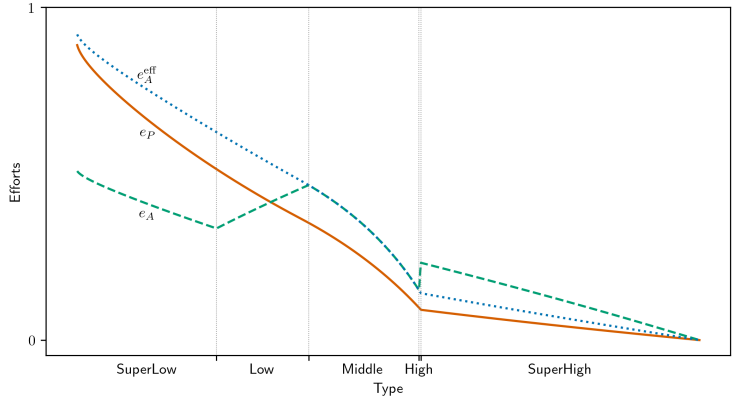
$e_P(x) > 0 \Rightarrow e_A(x) > 0 \Rightarrow u_A(e_A(x)) > 0$.

Since IC slack: $e_A(x) < e_A^{\text{eff}}(x)$ and $r_P(x) = r_\emptyset(x) = 0$.



From Low onward, IC binds: $U_m(x) = x - \lambda_m(x)$.

How to make info rent $x - \lambda_m(x)$?

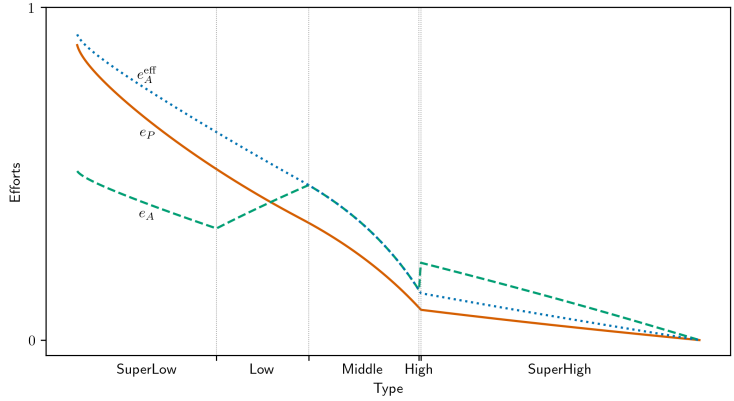


Types $x \in \underline{\text{Low}}$:

Evidence rent pays info rent: $u_A(e_A(x)) = x - \lambda_m(x)$.

$r_P(x) = r_\emptyset(x) = 0$.

$e_A(x)$ increases, reaches $e_A^{\text{eff}}(x)$ at $x = \sup \text{Low}$.



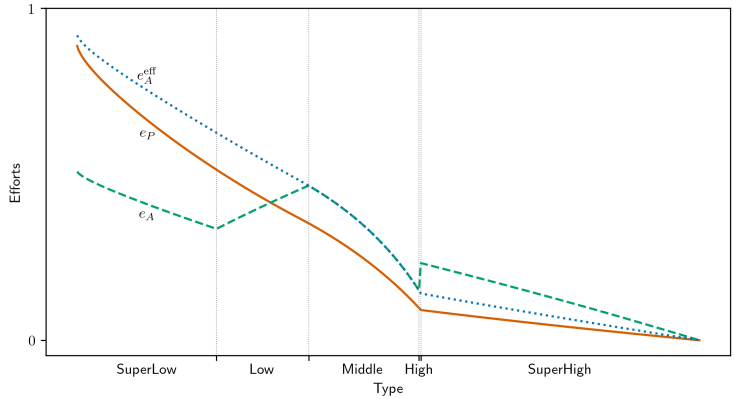
Types $\underline{x \in \text{Middle}}$: $e_A(x) = e_A^{\text{eff}}(x)$ and decreasing.

Provide missing info rent using $\underline{r_P(x)}$ or $e_A(x)$ or $r_\emptyset(x)$?

$r_P(x)$ = refund after principal gets evidence.

$r_\emptyset(x)$ = refund if no evidence.

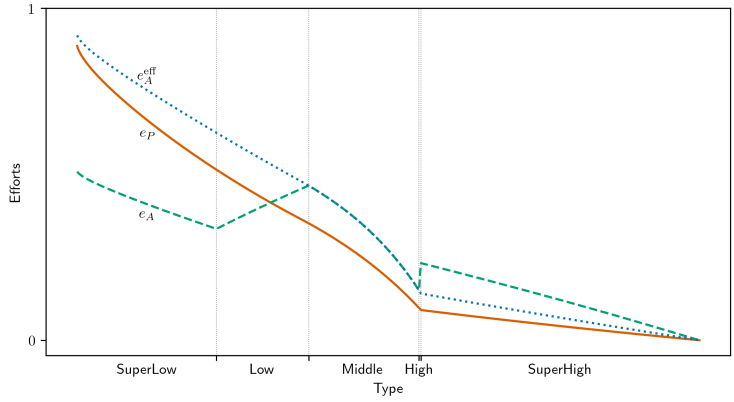
$r_P(x)$ increases, $r_\emptyset(x) = 0$.



Types $x \in$ High: $r_P(x) = x + \tau$.

Provide missing info rent using $e_A(x)$ or $r_\emptyset(x)$?

e_A increases: $u_A(e_A(x)) = x - \lambda_m(x) - e_P(x)(x + \tau)$, $r_\emptyset(x) = 0$.



Types $x \in \underline{\text{SuperHigh}}$:

$e_A(x)$ so high that principal uses $r_\emptyset(x)$.

$e_A(x)$ & $e_P(x)$ decrease in x.

Solution technique: Incentive programming

Recall loss function:

$$\lambda_m(x) = \inf_{y \le x} \; e_P(y)\, x + (1 - e_P(y))(y - r_\emptyset(y)).$$

$$m \xrightarrow{\text{calculate loss}} \lambda_m \xrightarrow{\text{re-optimize } m \text{ type-wise}} T(x, \lambda_m)$$

Type-wise maximization given x and $\tilde{\lambda}$:

$$
\begin{aligned}
T(x, \tilde{\lambda}) = \operatorname*{argmax}_{m(x)} \Pi(x, m(x)) \\
\text{s.t. } x - \tilde{\lambda}(x) \leq x\text{'s on-path utility}, \\
x \text{ obeys } e_A(x), \\
\forall w \geq x \quad \tilde{\lambda}(w) \leq w\text{'s loss when deviating to } x.
\end{aligned}
$$

$\tilde{\lambda}$ summarizes cross-type ICs (cf. value function in dynamic programming).

Let $\Lambda =$ set of incr. concave $\tilde{\lambda}$ such that $\tilde{\lambda} \leq \text{id}$ and $\tilde{\lambda}(\underline{x}) = \underline{x}$.

Theorem. If m is tight, then $\lambda_m \in \Lambda$ and $m(x) \in T(x, \lambda_m)$ for all x.

Tightness characterization: comparative statics of $T(x, \lambda_m)$.

($\rightarrow$ Detailed steps)

Optimal mechanisms

Mechanism is **optimal** if it maximizes expected profit

$$\int \Pi_m(x)\,\mathrm{d}F(x)$$

across all mechanisms m.

Lemma. A tight optimal mechanism exists. Every optimal mechanism coincides with some tight mechanism F-almost everywhere.

Theorem. Suppose F has no mass point at $\bar{x}$. If m is tight and optimal, then:

Agent gets a rent: $\inf_x U_m(x) > 0$.

No extreme audits: $\sup_x e_P(x) < 1$.

$e_P(x) = 0$ if and only if $x = \bar{x}$.

Recall: $c_A'(0) = c_P'(0) = 0$.

No mass point at $\bar{x} \Rightarrow U_m(\bar{x}) > 0$.

For $x \in [\underline{x}, \bar{x})$:

$c_P'(0) = 0 \Rightarrow e_P(x) > 0$ to extract more from higher types.

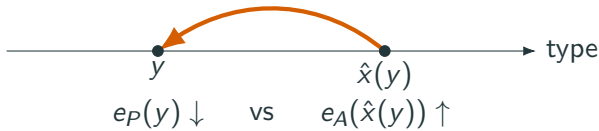
$c_A'(0) = 0 \Rightarrow e_A(x) > 0$ to reduce costs.

$\Rightarrow u_A(e_A(x)) > 0 \Rightarrow U_m(x) > 0$.

Since $\inf_x U_m(x) > 0$, auditing with prob. 1 is excessive $\Rightarrow \sup_x e_P(x) < 1$.

Trade-offs across types: audit-the-poor-or-burden-the-rich.

Audit cost at y vs. surplus seized from deviating type $\hat{x}(y)$.



Formally: $\lambda_m(\hat{x}(y)) = e_P(y)\hat{x}(y) + (1 - e_P(y))(y - r_\emptyset(y))$.

In tight mechanisms:

- $\hat{x}(y) > y$ whenever $e_P(y) > 0$ ("lie big!").
- $\hat{x}$ is increasing since e_P is decreasing (selection).

For a moment, assume **double uniqueness (DU)**:

unique binding IC $\hat{x}$ is **str. incr.** before reaching $\bar{x}$.

For mechanism with DU, perturbation gives necessary FOCs for optimality.

Under regularity, these FOCs are sufficient (even without DU).

Every optimal tight m approximated by tight $(m_n)_{n\in\mathbb{N}}$ satisfying DU:

$m_n \to m$ and $\Pi_{m_n} \to \Pi_m$ point-wise on $(\underline{x}, \bar{x})$.

Takeaways

Both e_P and e_A are used, but for different reasons:

- e_P deters deviations,
- e_A reduces principal's cost & provides info rent.

Agent is optimally left rent from showing evidence (cf. Allingham and Sandmo (1972)).

Tightness:

- IC imposes little structure, too many instruments.
- Refine IC by type-wise optimality of instruments.
- Comparative statics of type-wise problem.
- Fundamental question: how to make info rent?

Appendix

So far: assumed e_P interior on $[\underline{x}, \bar{x})$.

This assumption is met by all optimal tight mechanisms.

General characterization of tight mechanisms:

- Initial interval on which $e_P = 1$ and $e_A > 0$ (possibly empty).
- Interval on which $e_P \in (0, 1)$ is partitioned into SuperLow, ..., SuperHigh.
- Final interval on which $e_P = e_A = 0$. Always includes $\bar{x}$.

If e_P deterministic $\Rightarrow$ debt contract with relief clause

($\rightarrow$ Back to theorem)

How to provide info rent $x - \tilde{\lambda}(x)$ most cheaply?

Virtual costs:

$$\kappa_\bullet(x) = \text{cost of giving agent one util} = -\frac{\partial_\bullet \Pi(x, m(x))}{\partial_\bullet U(x, m(x))}.$$

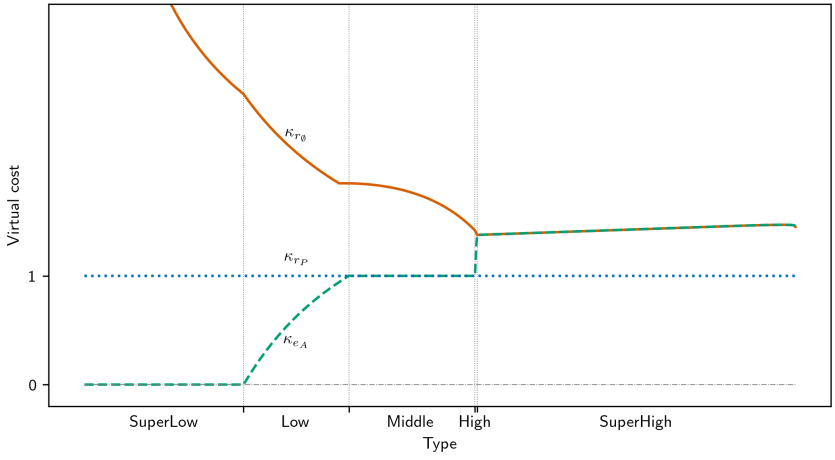
For $\bullet = e_A, e_P, r_P$, direct computation.

For $\bullet = r_\emptyset$, also adjust e_P:

$$\kappa_{r_\emptyset}(x) = -\frac{\partial_{\tilde{r}_\emptyset} \Pi(x, m(x)) + \partial_{\tilde{e}_P} \Pi(x, m(x)) \partial_{\tilde{r}_\emptyset} e_P^{\min}(x, r_\emptyset(x))}{\partial_{\tilde{r}_\emptyset} U(x, m(x)) + \partial_{\tilde{e}_P} U(x, m(x)) \partial_{\tilde{r}_\emptyset} e_P^{\min}(x, r_\emptyset(x))}.$$

Solution at each x: compare virtual costs.

(→ Back to technique)


Virtual cost
1
0
κ_{r_θ}
κ_{r_P}
κ_{e_A}
SuperLow
Low
Middle
High
SuperHigh
Type

Fix $(x, \tilde{e}_A)$.

Optimize $(\tilde{e}_P(x), \tilde{r}_A(x), \tilde{r}_P(x), \tilde{r}_\emptyset(x))$ as per virtual-cost comparison.

Three possible regimes:

1. x's IC slack $\Rightarrow$ profit $\pi_1(x, \tilde{e}_A)$.
2. x's IC binds and $\tilde{r}_\emptyset(x) = 0 \Rightarrow$ profit $\pi_2(x, \tilde{e}_A)$.
3. x's IC binds and $\tilde{r}_\emptyset(x) > 0 \Rightarrow$ profit $\pi_3(x, \tilde{e}_A)$.

Def:

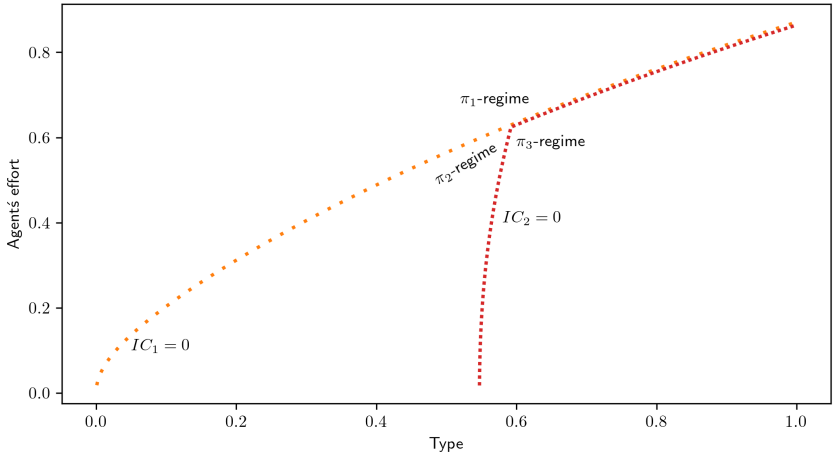
$$IC_1 = \{(x, \tilde{e}_A) \colon \pi_1(x, \tilde{e}_A) = \pi_2(x, \tilde{e}_A)\},$$
$$IC_2 = \{(x, \tilde{e}_A) \colon \pi_2(x, \tilde{e}_A) = \pi_3(x, \tilde{e}_A)\}.$$

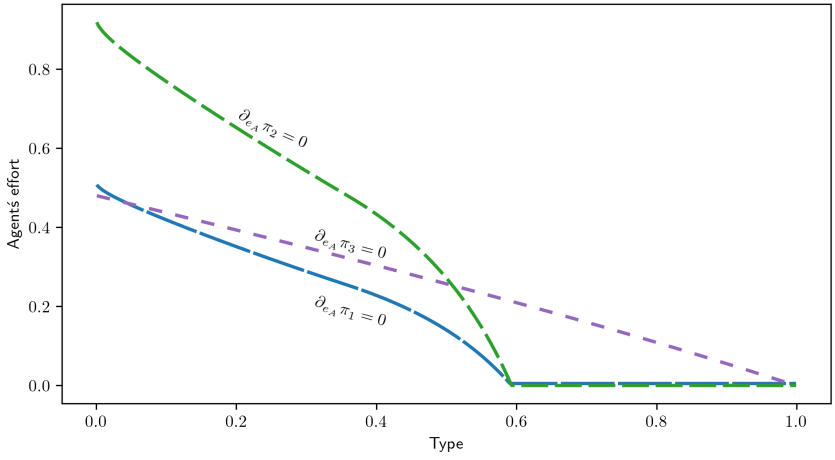
Recall:

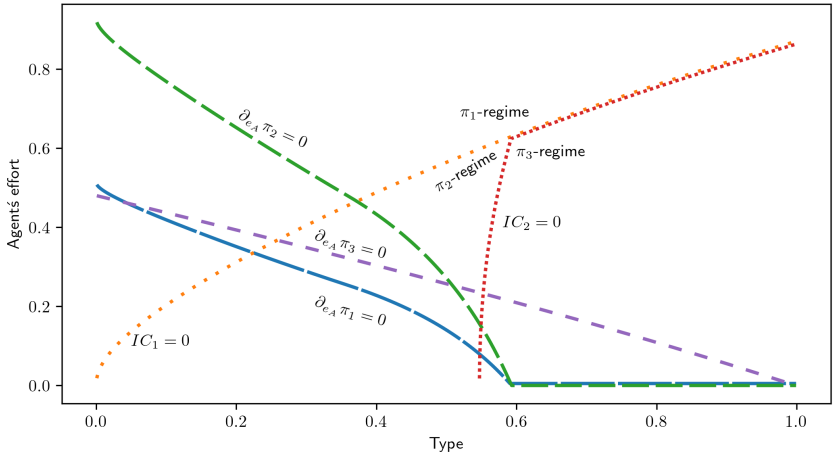
$$\begin{aligned}\lambda_m(x) &= \text{lowest deviation loss} \\ &= \inf_{y \le x} e_P(y)x + (1 - e_P(y))(y - r_\emptyset(y)).\end{aligned}$$

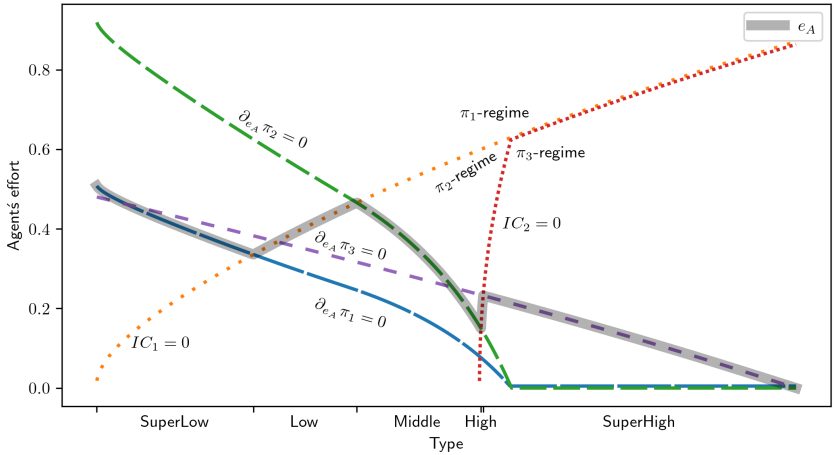
Let $\Lambda =$ set of incr. concave $\tilde{\lambda}$ s.t. $\tilde{\lambda} \le \text{id}$ and $\tilde{\lambda}(\underline{x}) = \underline{x}$.

Lemma. If m is tight, then $\lambda_m \in \Lambda$.


0.8
0.6
0.4
0.2
0.0
Agents effort
0.0
0.2
0.4
0.6
0.8
1.0
Type
π_1-regime
π_2-regime
π_3-regime
$IC_2 = 0$
$IC_1 = 0$


0.8
0.6
0.4
0.2
0.0
Agents effort
$\partial_{e_A} \pi_2 = 0$
$\partial_{e_A} \pi_3 = 0$
$\partial_{e_A} \pi_1 = 0$
0.0
0.2
0.4
0.6
0.8
1.0
Type


0.8
0.6
0.4
0.2
0.0
Agents effort
0.0
0.2
0.4
0.6
0.8
1.0
Type
$\partial_{e_A}\pi_2 = 0$
π_1-regime
π_3-regime
π_2-regime
$IC_2 = 0$
$\partial_{e_A}\pi_3 = 0$
$\partial_{e_A}\pi_1 = 0$
$IC_1 = 0$


e_A
0.8
0.6
0.4
0.2
0.0
Agents effort
$\partial_{e_A}\pi_2 = 0$
π_1-regime
π_2-regime
π_3-regime
$IC_2 = 0$
$\partial_{e_A}\pi_3 = 0$
$\partial_{e_A}\pi_1 = 0$
$IC_1 = 0$
SuperLow
Low
Middle
High
SuperHigh
Type

References

- Allingham, Michael G. and Agnar Sandmo (1972). **"Income tax evasion: A theoretical analysis".** In: Journal of Public Economics 1.3-4, pp. 323–338.
- Townsend, Robert M. (1979). **"Optimal contracts and competitive markets with costly state verification".** In: Journal of Economic Theory 21.2, pp. 265–293.
- Jovanovic, Boyan (1982). **"Truthful disclosure of information".** In: Bell Journal of Economics, pp. 36–44.
- Verrecchia, Robert E. (1983). **"Discretionary disclosure".** In: Journal of Accounting and Economics 5, pp. 179–194.
- Gale, Douglas and Martin Hellwig (1985). **"Incentive-compatible debt contracts: The one-period problem".** In: Review of Economic Studies 52.4, pp. 647–663.

- Border, Kim C and Joel Sobel (1987). **"Samurai accountant: A theory of auditing and plunder".** In: Review of Economic Studies 54.4, pp. 525–540.
- Mookherjee, Dilip and Ivan Png (1989). **"Optimal auditing, insurance, and redistribution".** In: Quarterly Journal of Economics 104.2, pp. 399–415.
- Andreoni, James, Brian Erard, and Jonathan Feinstein (1998). **"Tax compliance".** In: Journal of Economic Literature 36.2, pp. 818–860.
- Chander, Parkash and Louis L. Wilde (1998). **"A general characterization of optimal income tax enforcement".** In: Review of Economic Studies 65.1, pp. 165–183.
- Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2014). **"Optimal allocation with costly verification".** In: American Economic Review 104.12, pp. 3779–3813.
- Stahl, Konrad and Roland Strausz (2017). **"Certification and market transparency".** In: Review of Economic Studies 84.4, pp. 1842–1868.

- Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2019). **"Mechanisms with evidence: Commitment and robustness".** In: Econometrica 87.2, pp. 529–566.
- Halac, Marina and Pierre Yared (2020). **"Commitment versus flexibility with costly verification".** In: Journal of Political Economy 128.12, pp. 4523–4573.
- Bester, Helmut, Matthias Lang, and Jianpei Li (2021). **"Signaling versus auditing".** In: RAND Journal of Economics 52.4, pp. 859–883.
- Kattwinkel, Deniz and Jan Knoepfle (2023). **"Costless information and costly verification: A case for transparency".** In: Journal of Political Economy 131.2, pp. 504–548.
- Ahmadzadeh, Amirreza and Stephan Waizmann (2024). **"Mechanism Design with Costly Inspection".** In: Working paper.
- Lichtig, Avi and Helene Mass (2024). **"Optimal Testing in Disclosure Games".** In: Working paper.

Kattwinkel, Deniz and Justus Preusser (2025). **"Tight Samurai Accountant".** In: Working paper.

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman (2026). **"Sequential Mechanisms for Evidence Acquisition".** In: Working paper.