Optimal Audit Policy with Prediction Uncertainty

Aviv Caspi, Stanford University
Jacob Goldin, University of Chicago
Daniel E. Ho, Stanford University
Daniel Reck, University of Maryland

NTA 2023

Nov 2 2023

Motivation: How to Use Machine Learning Predictions?

- Tax authorities use machine learning and other tools to predict audit outcomes
- How can machine learning predictions inform optimal audits? How do we account for the fact that predictions are uncertain?

Motivation: How to Use Machine Learning Predictions?

- Tax authorities use machine learning and other tools to predict audit outcomes
- How can machine learning predictions inform optimal audits? How do we account for the fact that predictions are uncertain?
- Uncertainty interacts with distributional concerns: e.g. targeting audits toward those with income $> \$400$k w/o observing true income
- Feasibility constraint: audits must be selected based on information observable to government

Motivation: How to Use Machine Learning Predictions?

- Tax authorities use machine learning and other tools to predict audit outcomes
- How can machine learning predictions inform optimal audits? How do we account for the fact that predictions are uncertain?
- Uncertainty interacts with distributional concerns: e.g. targeting audits toward those with income $>$ \$400k w/o observing true income
- Feasibility constraint: audits must be selected based on information observable to government
- We develop a sufficient statistics approach to optimal audit selection based on predicted audit outcomes to address these questions.

Outline

1. Setup: Taxpayer and Government Information
2. Derive Suff Stat Characterization for Baseline Static Model
3. Extension: New Information
4. Extension: Dynamic Information Effects

Related Literature

- **Theories of Optimal Audit Selection:** Reinganum & Wilde, 1985; Sanchez & Sobel, 1993; Cremer & Gahvari 1996; Mookherjee & Png 1989; Graetz, Reinganum & Wilde 1986
  - **Our Contribution:** richer (high-dimensional) information environment, sufficient statistics (implicit) characterization of optimum
- **Optimal Tax Systems and Enforcement:** Mayshar 1991, Slemrod & Yitzhaki 1996, 2001; Kleven & Kreiner 2006; Hendren 2016; Keen & Slemrod 2017; Hendren & Sprung-Keyser 2020; **Boning et al 2023**
  - **Our contribution:** focus on return-level audit selection, characterize optimal audit rate (c.f. Saez 2001 for optimal tax rates)
- **Machine learning and policy problems:** Kleinberg et al 2015; Black et al., 2022; Henderson et al., 2023; Elzayn et al., 2023
  - **Our contribution:** welfarist objective, connect to optimal tax theory

Setup and Order of Events

1. Individuals, endowed with private information (their type $\theta \in \mathbb{R}^N$) file a return reporting information ($\hat{\theta} \in \mathbb{R}^N$) to the government and remit taxes
   - True tax liability $T(\theta)$, reported liability $T(\hat{\theta})$
   - Reported liability maximizes expected utility given risk of an audit
   - Risk of audit depends on (unobserved) gov't information and self-report
   - Penalties, tax schedule, audit procedures, true incomes all held fixed

Setup and Order of Events

1. Individuals, endowed with private information (their type $\theta \in \mathbb{R}^N$) file a return reporting information ($\hat{\theta} \in \mathbb{R}^N$) to the government and remit taxes
2. The government observes taxpayer reports $\hat{\theta}$ and additional private signal $\sigma \in \mathbb{R}^M$ (e.g. third-party/whistleblower info), and implements an audit selection rule $A(\hat{\theta}, \sigma)$.
    - Individuals do not observe government's signal σ

Setup and Order of Events

1. Individuals, endowed with private information (their type $\theta \in \mathbb{R}^N$) file a return reporting information ($\hat{\theta} \in \mathbb{R}^N$) to the government and remit taxes
2. The government observes taxpayer reports $\hat{\theta}$ and additional private signal $\sigma \in \mathbb{R}^M$ (e.g. third-party/whistleblower info), and implements an audit selection rule $A(\hat{\theta}, \sigma)$.
3. Revenues are rebated lump-sum, ex post payoffs are realized

Setup and Order of Events

1. Individuals, endowed with private information (their type $\theta \in \mathbb{R}^N$) file a return reporting information ($\hat{\theta} \in \mathbb{R}^N$) to the government and remit taxes
2. The government observes taxpayer reports $\hat{\theta}$ and additional private signal $\sigma \in \mathbb{R}^M$ (e.g. third-party/whistleblower info), and implements an audit selection rule $A(\hat{\theta}, \sigma)$.
3. Revenues are rebated lump-sum, ex post payoffs are realized

We characterize the socially optimal audit selection rule in a rational expectations (Bayesian) equilibrium of this game

- $\implies$ Distribution $f(\theta, \sigma)$ common knowledge, *aggregate learning* deferred to future work
- Agents anticipate others' actions correctly given their information

Information and Beliefs

- Individuals know their type but are uncertain about what the government knows $\implies$ make decisions given beliefs $f(\sigma|\theta)$
- Government makes decisions based on $f(\theta|\hat{\theta}, \sigma)$
  - Non-degenerate when there is pooling of types in reporting behavior, e.g. when non-compliant types attempt to appear compliant

Information and Beliefs

- Individuals know their type but are uncertain about what the government knows $\implies$ make decisions given beliefs $f(\sigma|\theta)$
- Government makes decisions based on $f(\theta|\hat{\theta}, \sigma)$
- Audit selection rule $A(\hat{\theta}, \sigma)$ maps gov't info to an audit rate in $[0, 1]$
  - $\implies$ where $A(\hat{\theta}, \sigma) \in (0, 1)$, some randomness in audit selection

Information and Beliefs

- Individuals know their type but are uncertain about what the government knows $\Longrightarrow$ make decisions given beliefs $f(\sigma|\theta)$
- Government makes decisions based on $f(\theta|\hat{\theta},\sigma)$
- Audit selection rule $A(\hat{\theta},\sigma)$ maps gov't info to an audit rate in $[0,1]$
- Individuals form beliefs about their audit probability given their self-report:

$$p_\theta(\hat{\theta}, A) = \int_\sigma A(\hat{\theta},\sigma) dF(\sigma|\theta)$$

Information and Beliefs

- Individuals know their type but are uncertain about what the government knows $\implies$ make decisions given beliefs $f(\sigma|\theta)$
- Government makes decisions based on $f(\theta|\hat{\theta},\sigma)$
- Audit selection rule $A(\hat{\theta},\sigma)$ maps gov't info to an audit rate in $[0,1]$
- Individuals form beliefs about their audit probability given their self-report:

$$p_\theta(\hat{\theta}, A) = \int_\sigma A(\hat{\theta},\sigma) dF(\sigma|\theta)$$

- A perturbation to audit selection rule $dA(\hat{\theta},\sigma)$ affects individual θ's audit risk according to

$$dp_\theta(\hat{\theta},\sigma) = \int_\sigma dA(\hat{\theta},\sigma) dF(\sigma|\theta)$$

Illustration: Audit Rule with Simpler Setup

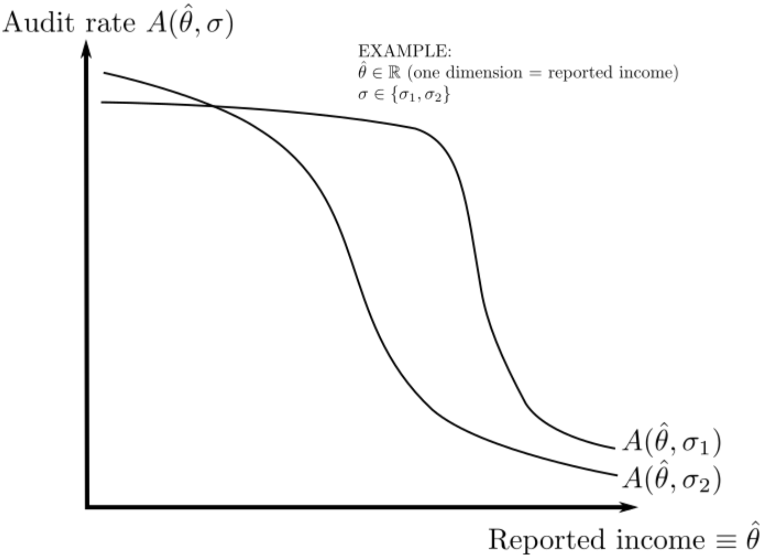
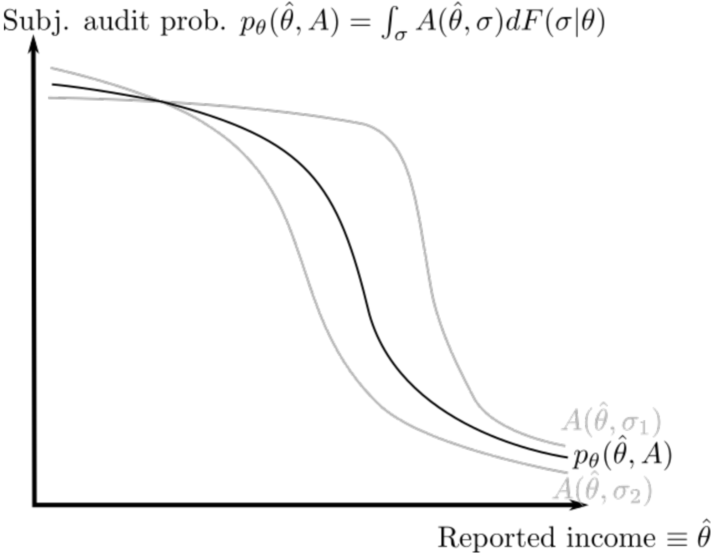


Illustration: Individual Beliefs

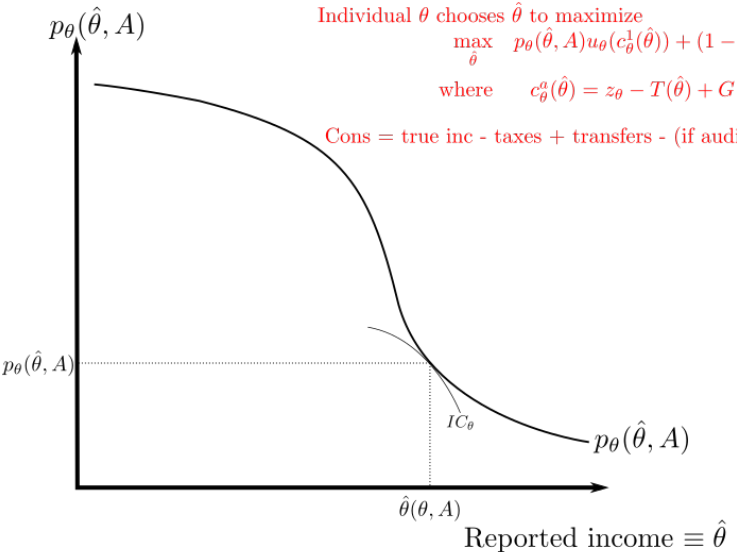


Individual θ chooses $\hat{\theta}$ to maximize

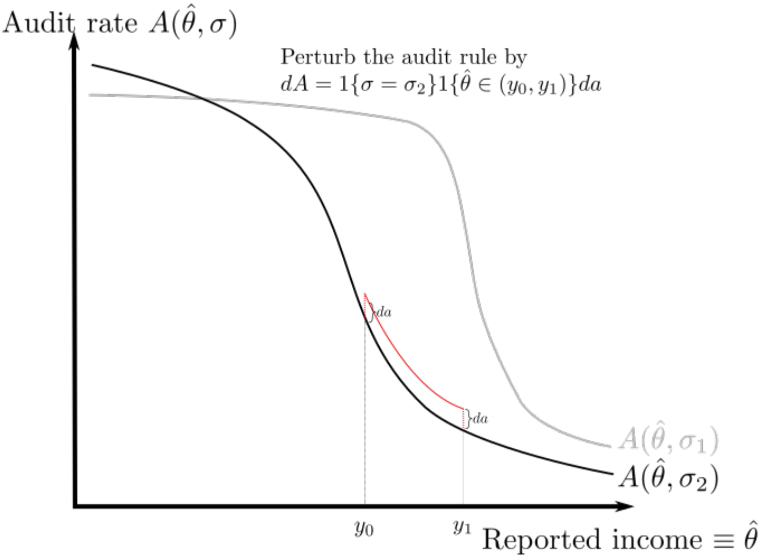
$$\max_{\hat{\theta}} \quad p_\theta(\hat{\theta}, A)u_\theta(c_\theta^1(\hat{\theta})) + (1 - p_\theta(\hat{\theta}, A))u_\theta(c_\theta^0(\hat{\theta}))$$

$$\text{where} \qquad c_\theta^a(\hat{\theta}) = z_\theta - T(\hat{\theta}) + G - a[R_\theta(\hat{\theta}) + H_\theta]$$

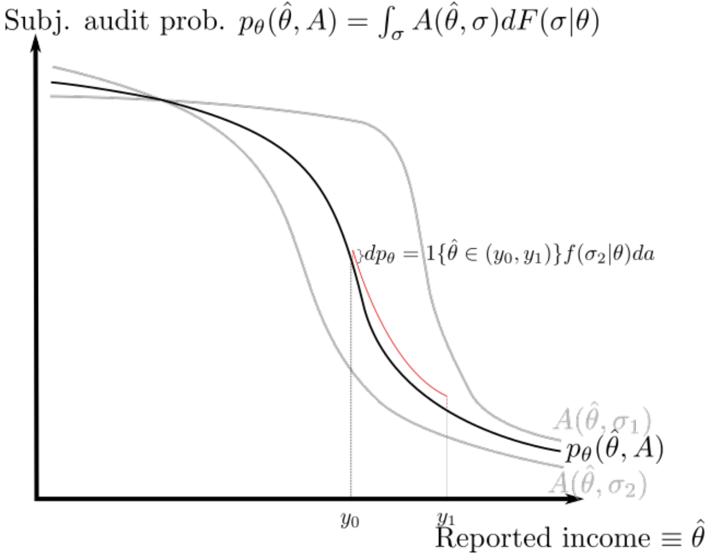
Cons = true inc - taxes + transfers - (if audited) audit revenues & costs



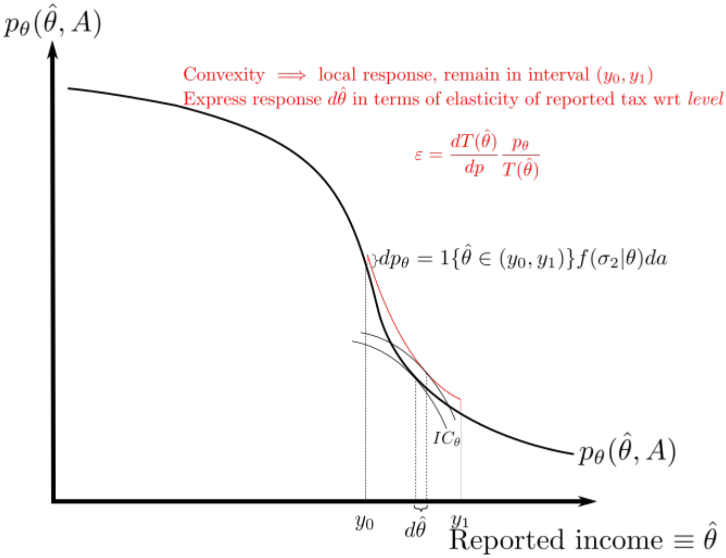
Local Perturbation of Audit Selection Rule



Resulting Perturbation of Audit Probability



Behavioral Responses



Individual Behavior/Welfare

For an arbitrary perturbation $dA(\hat{\theta}, \sigma)$

$$dp_\theta(\hat{\theta}) = \int_\sigma dA(\hat{\theta}, \sigma) f(\sigma|\theta)$$

The effect on individual welfare $v_\theta(p_\theta)$ is

$$dv_\theta = [u_\theta(c_\theta^1) - u_\theta(c_\theta^0)] dp_\theta(\hat{\theta}) \approx -EMU_\theta[R_\theta + H_\theta] dp_\theta(\hat{\theta})$$

- Behavioral response $d\hat{\theta}$ is second-order for private welfare (envelope theorem)

Individual Behavior/Welfare

For an arbitrary perturbation $dA(\hat{\theta}, \sigma)$

$$dp_\theta(\hat{\theta}) = \int_\sigma dA(\hat{\theta}, \sigma) f(\sigma|\theta)$$

The effect on individual welfare $v_\theta(p_\theta)$ is

$$dv_\theta = [u_\theta(c_\theta^1) - u_\theta(c_\theta^0)] dp_\theta(\hat{\theta}) \approx -EMU_\theta[R_\theta + H_\theta] dp_\theta(\hat{\theta})$$

- Behavioral response $d\hat{\theta}$ is second-order for private welfare (envelope theorem)
- *Direct Private Welfare Loss*: audit revenues (R_θ) + private (compliance) cost of audit H_θ
- Individual values these losses at expected marginal utility of consumption EMU_θ

Social Welfare

The government aims to maximize generalized utilitarian welfare

$$W(A) = \int_\theta \psi_\theta v(p_\theta) dF(\theta)$$

subject to Government Budget Constraint

$$GBC \equiv \int_\theta \int_\sigma T(\hat{\theta}_\theta) + A(\hat{\theta}_\theta, \sigma)(R_\theta - C_\theta) dF(\sigma|\theta) dF(\theta) \geq G$$

where C_θ is the admin cost of an audit.

Social Welfare

The government aims to maximize generalized utilitarian welfare

$$W(A) = \int_\theta \psi_\theta v(p_\theta) dF(\theta)$$

subject to Government Budget Constraint

$$GBC \equiv \int_\theta \int_\sigma T(\hat{\theta}_\theta) + A(\hat{\theta}_\theta, \sigma)(R_\theta - C_\theta) dF(\sigma|\theta) dF(\theta) \geq G$$

where C_θ is the admin cost of an audit. For a marginal perturbation dA

$$dGBC = \underbrace{(R_\theta - C_\theta) dp_\theta}_{\text{Direct Revenue Effect}} + \underbrace{\left(\frac{dT_\theta}{dp_\theta} + p_\theta \frac{dR_\theta}{dp_\theta} \right) dp_\theta}_{\text{Behavioral Revenue Effect}}$$

Social Welfare

The government aims to maximize generalized utilitarian welfare

$$W(A) = \int_\theta \psi_\theta v(p_\theta) dF(\theta)$$

subject to Government Budget Constraint

$$GBC \equiv \int_\theta \int_\sigma T(\hat{\theta}_\theta) + A(\hat{\theta}_\theta, \sigma)(R_\theta - C_\theta) dF(\sigma|\theta) dF(\theta) \geq G$$

where C_θ is the admin cost of an audit. For a marginal perturbation dA

$$dGBC = \underbrace{(R_\theta - C_\theta) dp_\theta}_{\text{Direct Revenue Effect}} + \underbrace{\left(\frac{dT_\theta}{dp_\theta} + p_\theta \frac{dR_\theta}{dp_\theta}\right) dp_\theta}_{\text{Behavioral Revenue Effect}}$$

Assuming a linear penalty, we can just track the change in T:

$$R_\theta = (1+\rho)[T(\hat{\theta}^*) - T(\hat{\theta})] \implies dR_\theta = -(1+\rho) dT_\theta$$

.

Welfare Effect of a Local Perturbation

With social welfare weights $g_\theta = \frac{\psi_\theta EMU_\theta}{\lambda}$, i.e. normalizing by $\lambda = E_\theta[EMU_\theta]$, the social welfare effect of a perturbation is

$$dW = \int_\theta dp_\theta \left[R_\theta - C_\theta - g_\theta (R_\theta + H_\theta) + \frac{dT_\theta}{dp_\theta}(1 - p_\theta(1+\rho)) \right] dF(\theta)$$

Welfare Effect of a Local Perturbation

With social welfare weights $g_\theta = \frac{\psi_\theta EMU_\theta}{\lambda}$, i.e. normalizing by $\lambda = E_\theta[EMU_\theta]$, the social welfare effect of a perturbation is

$$dW = \int_\theta dp_\theta \left[R_\theta - C_\theta - g_\theta(R_\theta + H_\theta) + \frac{dT_\theta}{dp_\theta}(1 - p_\theta(1+\rho)) \right] dF(\theta)$$

Express this in terms of conditional means, covariances, and the *elasticity of reported tax due wrt p*, ε_θ

$$\begin{aligned} dW = \int_\sigma \int_{\hat\theta} \quad & dA(\hat\theta, \sigma)[\overline{R}(1-\overline{g}) - \overline{g}\overline{H} - \overline{C} \\ & - Cov(g_\theta, R_\theta + H_\theta | \hat\theta, \sigma) + T(\hat\theta)\bar\varepsilon \tfrac{1-A(1+\rho)}{A}] dF(\hat\theta|\sigma) dF(\sigma) \end{aligned}$$

where $\overline{R}(\hat\theta, \sigma) = \int_\theta R_\theta dF(\theta|\hat\theta, \sigma)$ is the conditional mean of audit revenue given the government's information and other terms are similar conditional means.

Unpacking the Covariance Term

$$dW = \int_\sigma \int_{\hat{\theta}} \quad dA(\hat{\theta}, \sigma)[\overline{R}(1 - \overline{g}) - \overline{g}\overline{H} - \overline{C} \\ - Cov(g_\theta, R_\theta + H_\theta | \hat{\theta}, \sigma) + T(\hat{\theta})\overline{\varepsilon}\frac{1 - A(1+\rho)}{A}] dF(\hat{\theta}|\sigma) dF(\sigma)$$

Covariance between audit outcomes and welfare weight is probably negative $\implies$ higher audit rates under more uncertainty

- Audit recovers more revenue than expected $\implies$ taxpayer has higher income $\implies$ lower welfare weight than expected

Unpacking the Covariance Term

$$dW = \int_\sigma \int_{\hat\theta} \quad dA(\hat\theta,\sigma)[\overline{R}(1-\overline{g}) - \overline{g}\overline{H} - \overline{C} \\ -Cov(g_\theta, R_\theta + H_\theta|\hat\theta,\sigma) + T(\hat\theta)\overline{\varepsilon}\frac{1-A(1+\rho)}{A}]dF(\hat\theta|\sigma)dF(\sigma)$$

Covariance between audit outcomes and welfare weight is probably negative $\implies$ higher audit rates under more uncertainty

- Audit recovers more revenue than expected $\implies$ taxpayer has higher income $\implies$ lower welfare weight than expected
- (OR compliance-based welfare weight imposed ex ante? see paper)
- Expressing welfare weight as a function of audit revenues $g^*(R_\theta)$, we have

$$dW \approx \int_\sigma \int_{\hat\theta} \quad dA(\hat\theta,\sigma)[\overline{R}(1-g^*(\overline{R})) - g^*(\overline{R})\overline{H} - \overline{C} \\ -\frac{dg^*(\overline{R})}{dR} Var(R_\theta|\hat\theta,\sigma) + T(\hat\theta)\overline{\varepsilon}\frac{1-A(1+\rho)}{A}]dF(\hat\theta|\sigma)dF(\sigma)$$

Unpacking the Behavioral Response

$$dW = \int_\sigma \int_{\hat{\theta}} \quad dA(\hat{\theta}, \sigma)[\overline{R}(1 - \overline{g}) - \overline{g}\overline{H} - \overline{C} \\ - Cov(g_\theta, R_\theta + H_\theta | \hat{\theta}, \sigma) + T(\hat{\theta})\overline{\varepsilon}\frac{1 - A(1+\rho)}{A}]dF(\hat{\theta}|\sigma)dF(\sigma)$$

- ε_θ is the effect of a marginal level increase in the probability of audit
- $\overline{\varepsilon}$ is a conditional mean of ε_θ, weighted by exposure to the perturbation $\frac{dp_\theta}{E[dA]}$.
- Note we assume local incentive compatibility is sufficient to ensure global incentive compatibility $\Longleftarrow$ convexity, as in optimal tax theory

Corner Solutions

- Optimality: for any feasible perturbation dA, $dW \leq 0$.
- Simpler notation: all direct effects denoted $\overline{D}(\hat{\theta}, \sigma)$

$$\overline{D} \equiv \overline{R}(1 - \overline{g}) - \overline{g}\overline{H} - \overline{C} - Cov(g_\theta, R_\theta + H_\theta | \hat{\theta}, \sigma)$$

Corner Solutions

- Optimality: for any feasible perturbation dA, $dW \leq 0$.
- Simpler notation: all direct effects denoted $\overline{D}(\hat{\theta}, \sigma)$

$$\overline{D} \equiv \overline{R}(1 - \overline{g}) - \overline{g}\overline{H} - \overline{C} - Cov(g_\theta, R_\theta + H_\theta | \hat{\theta}, \sigma)$$

- Corner solution 1: $dW \geq 0$ at 100% audit rate

$$A^*(\hat{\theta}, \sigma) = 1 \iff \overline{D} > T(\hat{\theta})\rho\bar{\varepsilon}$$

  - Positive direct effect is not sufficient for 100% audit rate because deterrence reduces penalty collections at a high audit rate

Corner Solutions

- Optimality: for any feasible perturbation dA, $dW \leq 0$.
- Simpler notation: all direct effects denoted $\overline{D}(\hat{\theta}, \sigma)$

$$\overline{D} \equiv \overline{R}(1 - \overline{g}) - \overline{g}\overline{H} - \overline{C} - Cov(g_\theta, R_\theta + H_\theta | \hat{\theta}, \sigma)$$

- Corner solution 1: $dW \geq 0$ at 100% audit rate

$$A^*(\hat{\theta}, \sigma) = 1 \iff \overline{D} > T(\hat{\theta})\rho\overline{\varepsilon}$$

  - Positive direct effect is not sufficient for 100% audit rate because deterrence reduces penalty collections at a high audit rate
- Corner solution 0: $dW \leq 0$ at 0% audit rate (re-express behavioral resp using semi-elasticity $\eta \equiv \frac{dT}{dp}\frac{1}{T}$):

$$A^*(\hat{\theta}, \sigma) = 0 \iff \overline{D} < -T(\hat{\theta})\overline{\eta}$$

  - Could arise w/high audit costs vs revenue, high welfare weights, and/or weak deterrence effects

Optimal Audit Rate at an Interior Optimum

If neither corner condition is met, we must have

$$dW = 0 \implies A^*(\hat{\theta}, \sigma) = \frac{T(\hat{\theta})\overline{\varepsilon}}{T(\hat{\theta})\overline{\varepsilon}(1+\rho) - \overline{\overline{D}}}.$$

Our sufficient statistics for evaluating optimality of audit selection for *any group that is distinguishable with gov't information*:

- Predicted (mean) audit revenues and admin/private costs
- Welfare weight at predicted audit outcome etc.
- Uncertainty: sensitivity of welfare weight to audit outcomes, variance of audit revenues/private costs
- Reported tax due
- Deterrence elasticity of reported tax due

Optimal Audit Rate at an Interior Optimum

If neither corner condition is met, we must have

$$dW = 0 \implies A^*(\hat{\theta}, \sigma) = \frac{T(\hat{\theta})\overline{\varepsilon}}{T(\hat{\theta})\overline{\varepsilon}(1+\rho) - \overline{D}}.$$

Our sufficient statistics for evaluating optimality of audit selection for *any group that is distinguishable with gov't information*:

- Predicted (mean) audit revenues and admin/private costs
- Welfare weight at predicted audit outcome etc.
- Uncertainty: sensitivity of welfare weight to audit outcomes, variance of audit revenues/private costs
- Reported tax due
- Deterrence elasticity of reported tax due

Note we have derived *implicit* characterizations of *whether status quo is optimal*, *holding all other policies fixed* (c.f. Saez 2001).

Sketch of Extension 1: New Information

- Now suppose the government can observe some additional information in σ
- Result 1: How optimal audit rates change
  - Where new information is discriminating, increase audits where gains are high and decrease them elsewhere
  - Express this in terms of how mean predictions (e.g. $\overline{R}$) and covariance change.

Sketch of Extension 1: New Information

- Now suppose the government can observe some additional information in σ
- Result 1: How optimal audit rates change
  - Where new information is discriminating, increase audits where gains are high and decrease them elsewhere
  - Express this in terms of how mean predictions (e.g. $\overline{R}$) and covariance change.
- Result 2: Effect on social welfare of new information
  - Proportional to amount of variance in individual-specific welfare effect of a marginal audit that is *explained by new information* – related to partial R^2
  - Quantifies the social value of new info under optimal selection, could be traded off against the costs of collecting/using information.

Sketch of Extension 2: Dynamic Information Effects

- Audits reveal information about future periods $\implies$ an audit in t modifies σ_{t+1}. How does this modify optimal audits in t?
- We model the case where information is revealed *exclusively about the audited individual* (and individuals and gov't know what is revealed)
  - Private direct effect now includes effect on NPV of future consumption
  - Fiscal direct effect includes effect on future gov't budget

Sketch of Extension 2: Dynamic Information Effects

- Audits reveal information about future periods $\implies$ an audit in t modifies σ_{t+1}. How does this modify optimal audits in t?
- We model the case where information is revealed *exclusively about the audited individual* (and individuals and gov't know what is revealed)
  - Private direct effect now includes effect on NPV of future consumption
  - Fiscal direct effect includes effect on future gov't budget
  - Both of these include the effect of info on future tax payments $\leftarrow$ subject of empirical literature (e.g. DeBacker et al 2018)
  - Also incorporated in Boning et al (2023) (labelled a deterrence effect)

Sketch of Extension 2: Dynamic Information Effects

- Audits reveal information about future periods $\implies$ an audit in t modifies σ_{t+1}. How does this modify optimal audits in t?
- We model the case where information is revealed *exclusively about the audited individual* (and individuals and gov't know what is revealed)
  - Private direct effect now includes effect on NPV of future consumption
  - Fiscal direct effect includes effect on future gov't budget
  - Both of these include the effect of info on future tax payments $\leftarrow$ subject of empirical literature (e.g. DeBacker et al 2018)
  - Also incorporated in Boning et al (2023) (labelled a deterrence effect)
- Broader insight: how we value information revealed by period-t audits depends on for whom that information is relevant
  - e.g. information spillovers through preparer networks, business ownership networks
  - More research needed to understand which information effects matter.

Conclusion & Next Steps

- We derive a sufficient statistics characterization of the optimal audit selection rule with a welfarist objective
- Admits a test of the optimality of increasing/decreasing audit rates for *any group of individuals that is observable to the government.*

Conclusion & Next Steps

- We derive a sufficient statistics characterization of the optimal audit selection rule with a welfarist objective
- Admits a test of the optimality of increasing/decreasing audit rates for *any group of individuals that is observable to the government.*
- We used this to quantify the value of new information.
- We extended it to account for dynamic effects of audits.

Conclusion & Next Steps

- We derive a sufficient statistics characterization of the optimal audit selection rule with a welfarist objective
- Admits a test of the optimality of increasing/decreasing audit rates for *any group of individuals that is observable to the government.*
- We used this to quantify the value of new information.
- We extended it to account for dynamic effects of audits.
- As with other sufficient statistics characterizations, we found an *implicit* characterization *holding all other policies fixed*
- Next step: implement our sufficient statistics characterization with machine-learning predictions trained on real audit data
  - Requires circumspection around welfare weights, deterrence elasticity
  - Imposing modest structure on deterrence could be useful here too (e.g. elasticity must be zero for compliant types...)