Research, Applied Analytics, and Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

12th Annual IRS/TPC Joint Research Conference on Tax Administration

June 16, 2022

Housekeeping

- Event is being recorded and the recording will be posted online afterward.
- Hide captions or adjust settings with the Live Transcript button.
- The slides, speaker biographies, and abstracts are available online.
- All participants are muted.
- Type your questions or comments into the Q&A box at any time.
- Please complete the survey at the end of the event.

Session 1:
Balancing Audits: Enforcement vs. Measuring Non-Compliance

Combining Operational and NRP data with Statistical Controls for Bias

IRS-TPC Research Conference June 16, 2022

IRS—Research, Applied Analytics, and Statistics

Ishani Roy, Brett Collins, Alex Turk, and Mark Payne

Disclaimer: The views in this paper are those of the authors only and do not necessarily reflect the positions of the IRS

Combining Operational and NRP data with Statistical Controls for Bias

National Research Program (NRP) data

- Representative (stratified) Sample of Individual Income tax returns
- Replacement for Taxpayer Compliance Measurement Program (2001)
  - Annual samples starting in 2006
  - Reductions in numbers of audits due to various resource constraints
- Audits primary conducted with **Field (Revenue Agents)** and **Office (Tax Compliance Officers)** audits
- An NRP redesign working group recommended exploring methods to supplementing NRP audits with operational (OP) audits
  - Non-random selection
  - Measurement differences
  - Erard and Ho (2002)

Focus of this paper

- NRP data provide a unique opportunity to explore bias controls for models using data from non-probability samples
    - Out of sample “ground truth” data
    - Evaluate an “active learning” approach to sampling
- Focus on comparing model ability to estimate the tax misreporting (predicted tax change) and probability audit determining the taxpayer fully reporting the tax liability (no change) when non-random data are added to model parameter estimation (for returns claiming EITC)
  - Across the distribution of compliance behavior
  - Expected audit adjustment and the percentage with no adjustment are key features of the distribution of observed reporting compliance behavior (Clotfelter 1983)
    - Full reporting of liability
    - Measurement error in audit data

Combining Operational and NRP data with Statistical Controls for Bias

General Approach: Pool operational and NRP audit data for model development

- Develop probit models of the probability that a return is audited
  - A roughly equal number of non-audited returns in the same activity codes are randomly selected for this model
  - **Use only revenue Field and office audits, which are the most similar to NRP audits**
- Use the audit probability model to calculate the Heckman selection bias control
- Pool 2010-2014 NRP audit data with 2010-2014 Field and Office operational audit data (TY 2015 used for validation)
  - Weights on the NRP cases are normalized by the average weight, operational audits weights are equal to 1
  - On average an NRP and an operational case are weighted equally

Combining Operational and NRP data with Statistical Controls for Bias

General Approach: Pool operational and NRP audit data for model development

- Estimate a two-step parametric switching model
  - Multinomial logit for categorical outcome (negative, low, or high tax change), OLS for conditional audit adjustment category
    - Include the Heckman correction for selection bias in operational audits
    - Include a dummy for NRP vs Operational audits to control for measurement differences
- Use the model to estimate an "NRP measured" predicted tax change and probability of a "no change"

Combining Operational and NRP data with Statistical Controls for Bias - Models

$p = P(Audit \mid X) = \Phi(X\beta)$	Audit Probability Model
$\lambda = \phi\left((X\beta)\right)/p$	Heckman Selection Bias control
$log\left(\frac{P(C_j)}{P(C_1)}\right) = X\gamma_j + \gamma_{dj}D + \tau_j\lambda, \quad j = 2,3$	Multinomial logit for negative, low and high tax change
$T_j = X\alpha_j + \alpha_{dj}D + \theta_j\lambda, \quad j = 1,2,3$	OLS models for audit adjustments
$PTC = p_1X\alpha_1 + p_2X\alpha_2 + p_3X\alpha_3$	Predicted tax change setting *D* and λ equal to 0

Combining Operational and NRP data with Statistical Controls for Bias

Focus on activity codes 270 and 271, which both involve income tax returns with EITC claims

- Activity code 270 cases have EITC claims, less than $200,000 in total positive income (TPI), and either no attached Schedule C or F, or that include one or both schedules with total gross receipts under $25,000
- Activity code 271 cases have EITC claims, less than $200,000 in total positive income (TPI), and at least one Schedule C or F with total gross receipts over $24,999
  - AC 271 cases have a particularly low number of NRP audits, so they may benefit more from supplementing with operational audits
- EITC claims were capped at $6,728 for TY2021, which helps to limit the variability in audit adjustment

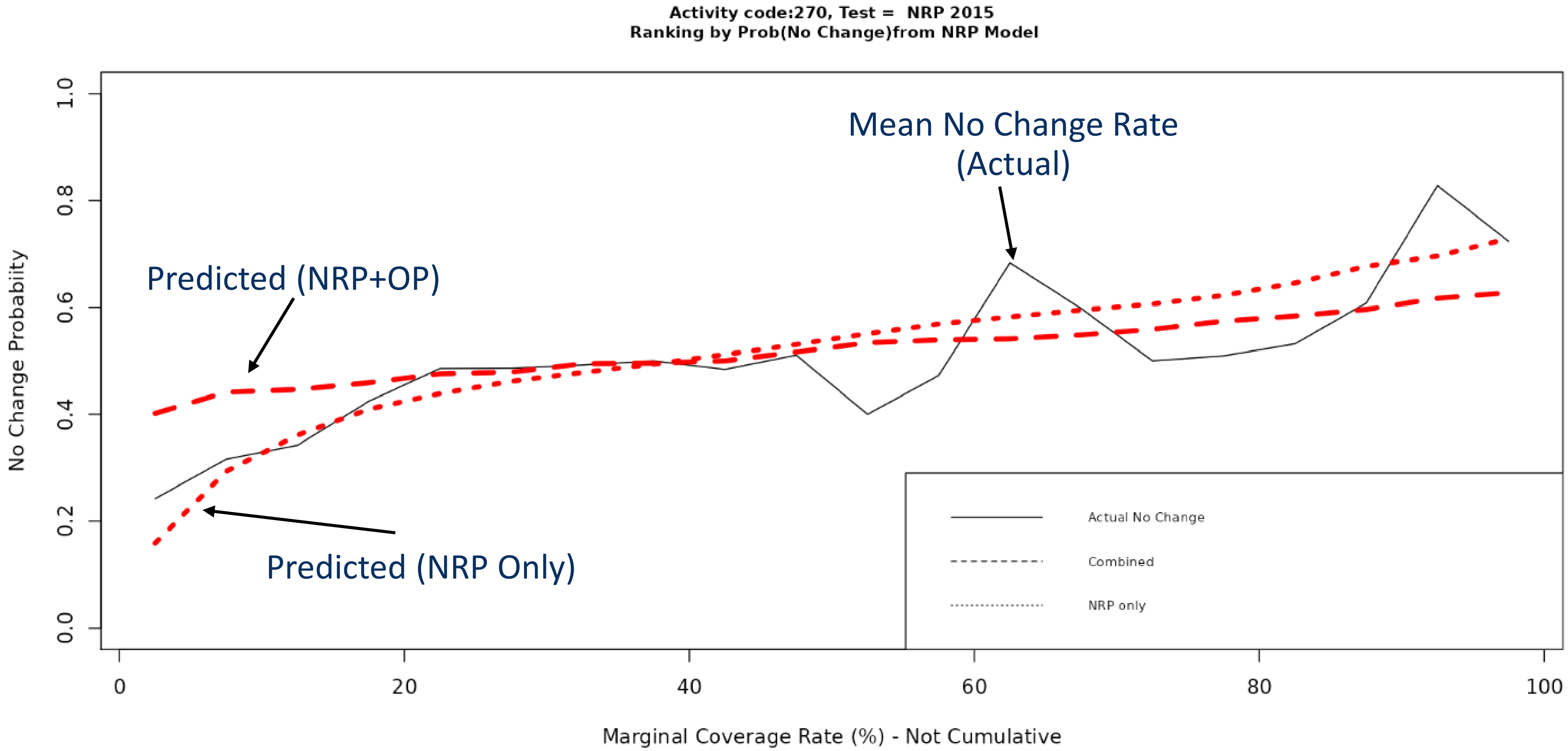
Combining Operational and NRP data with Statistical Controls for Bias - Results

- Pooling training data across five years provides a sizeable sample, even when restricting the OP cases to Field and Office audits
  - NRP training data for activity code 271 is the exception to this

Activity Code	TY 2019 Population (Cycle 202108)	NRP Training N (TY2010-TY2014)	OP Training N (TY2010-TY2014)
270	23,494,000	12,060	112,872
271	2,035,000	943	52,293

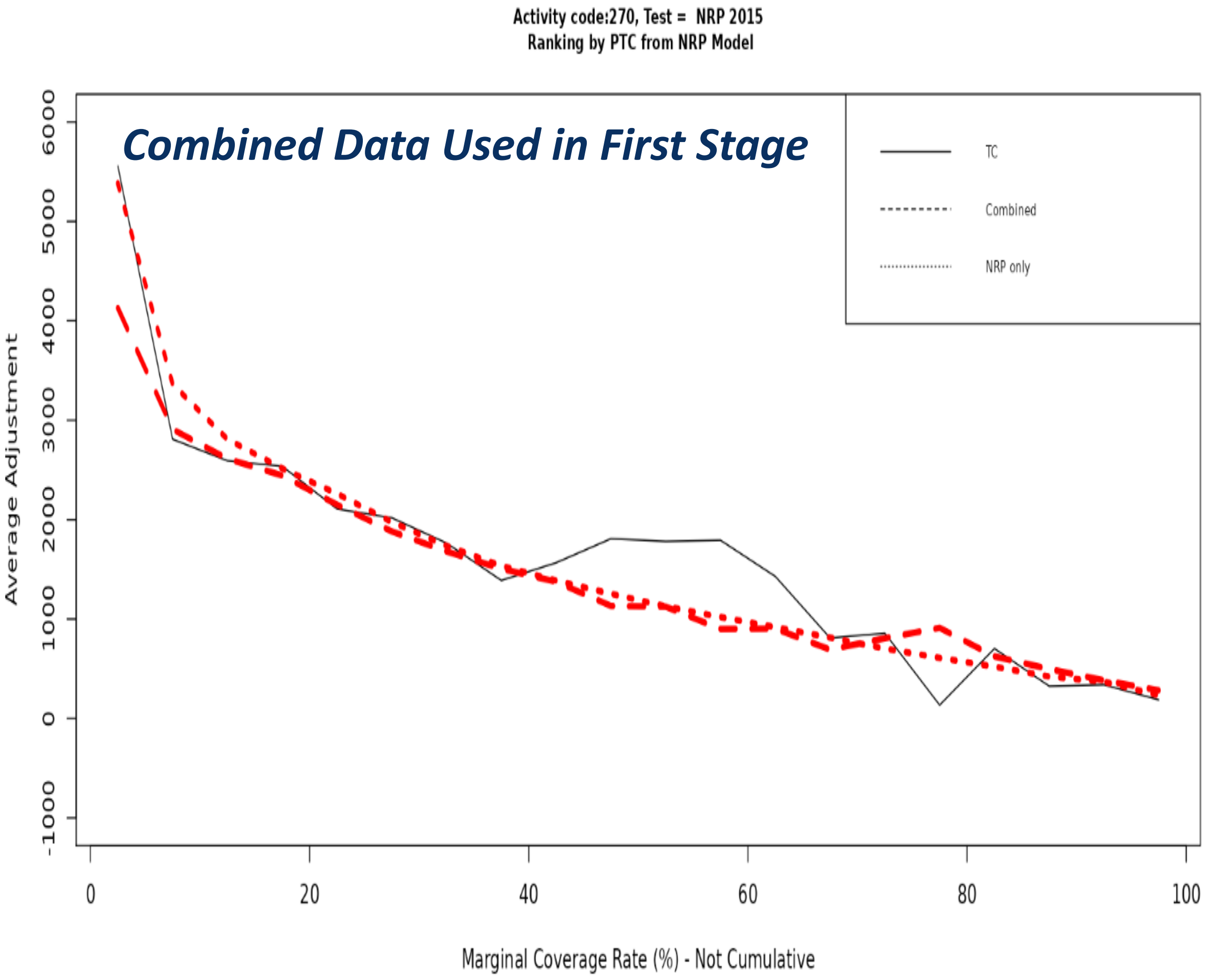
Test: NRP	Average Tax Change TY2015 (models using full data)			Probability of No Change TY2015 (models using full data)		
Activity Code	Observed (actual)	Predicted (NRP+OP)	Predicted (NRP only)	Observed (actual)	Predicted (NRP+OP)	Predicted (NRP only)
270	1,743	1,658	1,765	0.48	0.49	0.49
271	6,195	6,854	7,617	0.21	0.19	0.19

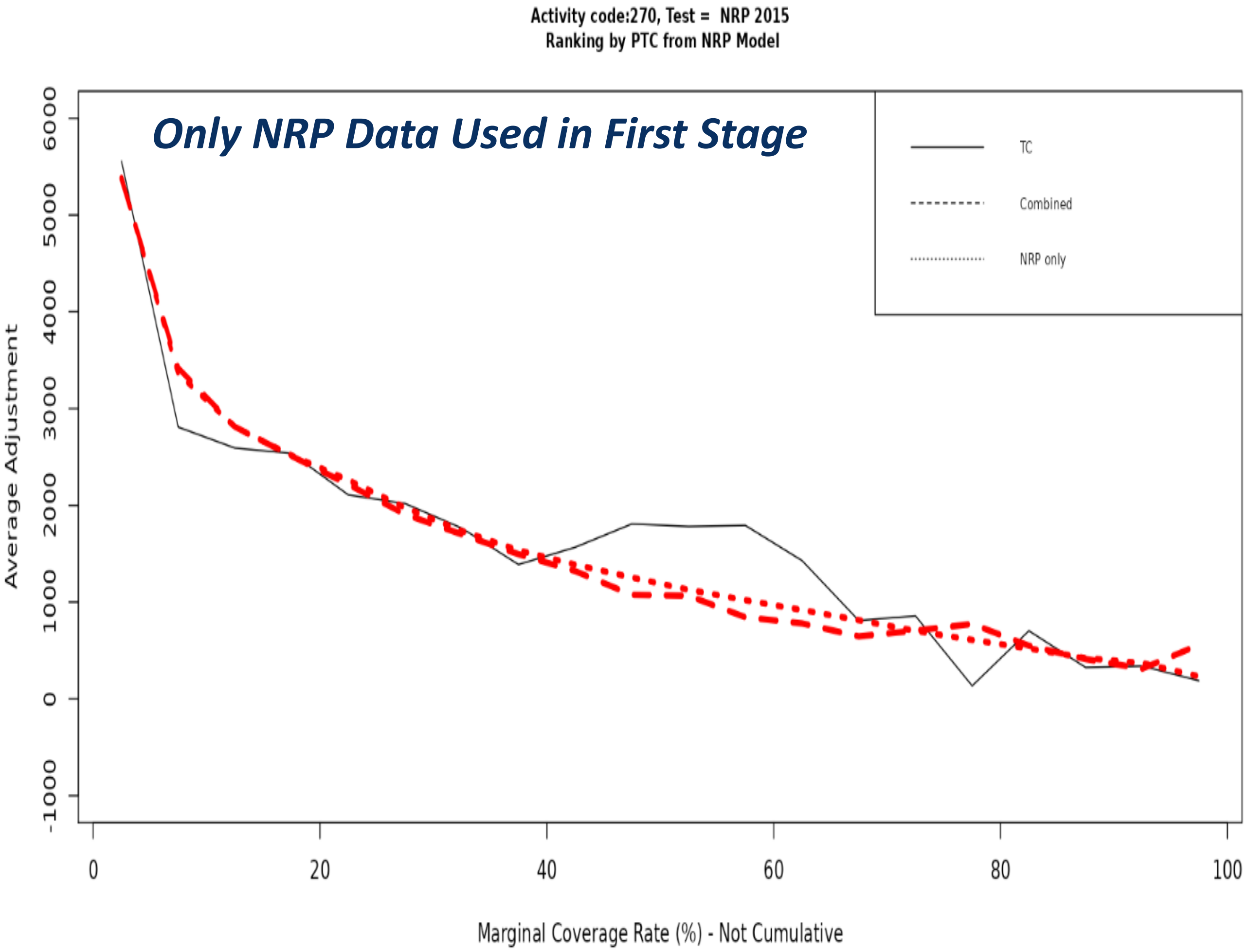
AC 270 Probability of No Change on NRP Data



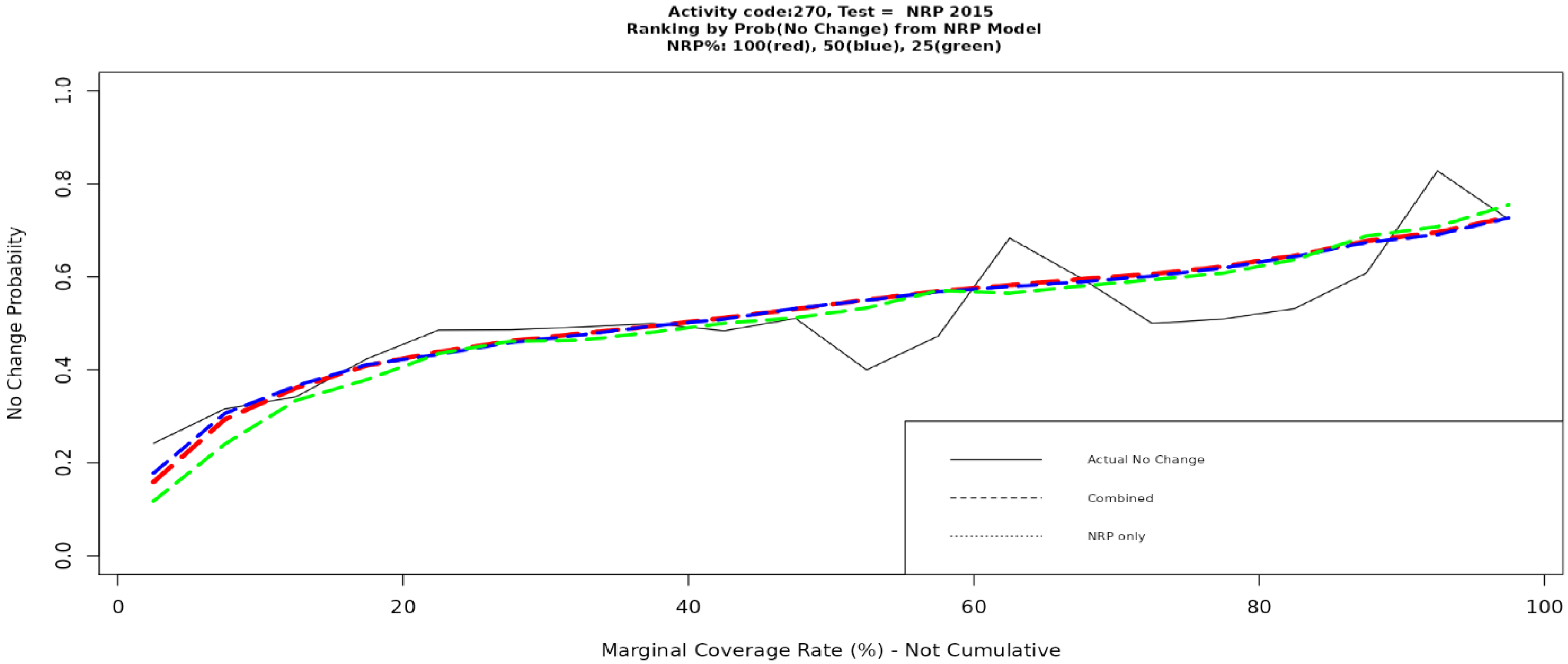
Combining Operational and NRP data with Statistical Controls for Bias - Results

AC 270 Tax Change Predictions on NRP Data, Ranked by Predicted Tax Change based on NRP-Only Model



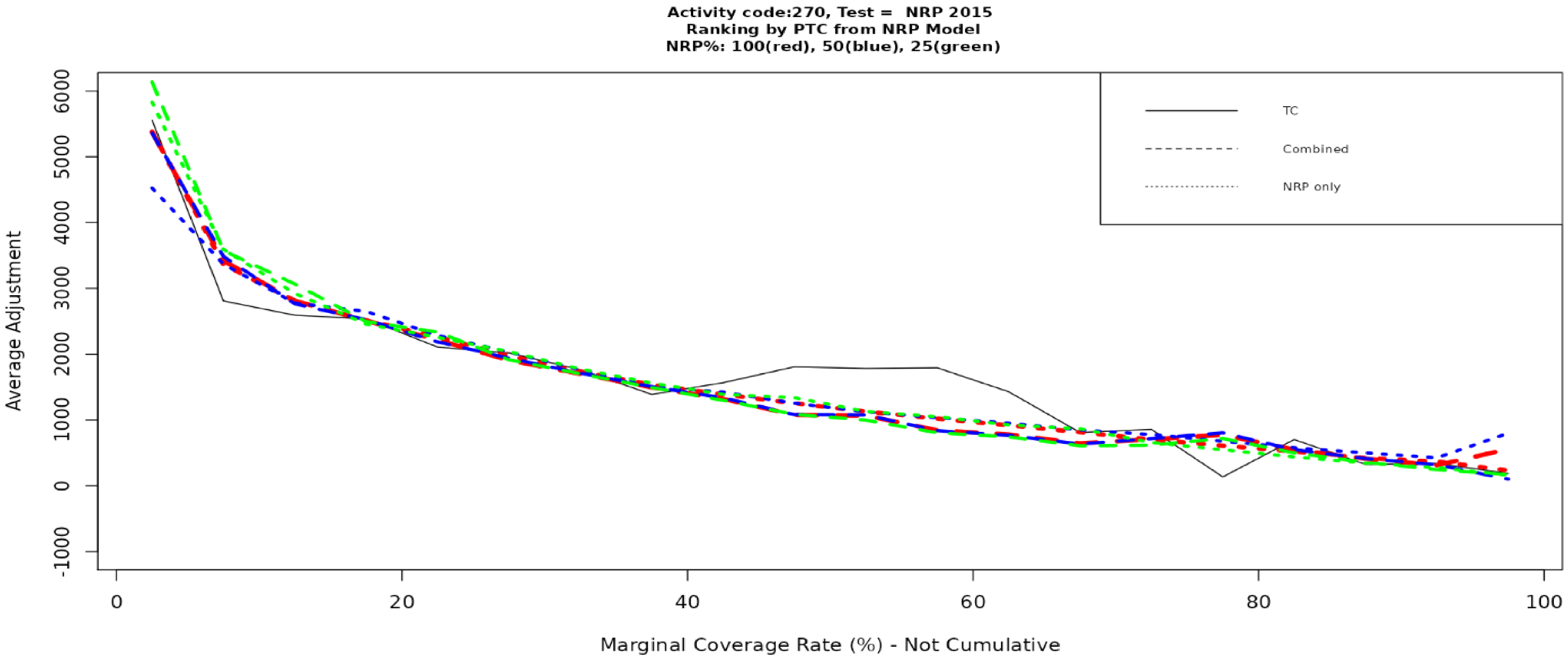


AC 270 Probability of No Change on NRP Data, Ranked by Predicted Probability of No Change Based on the 100% NRP-Only Model

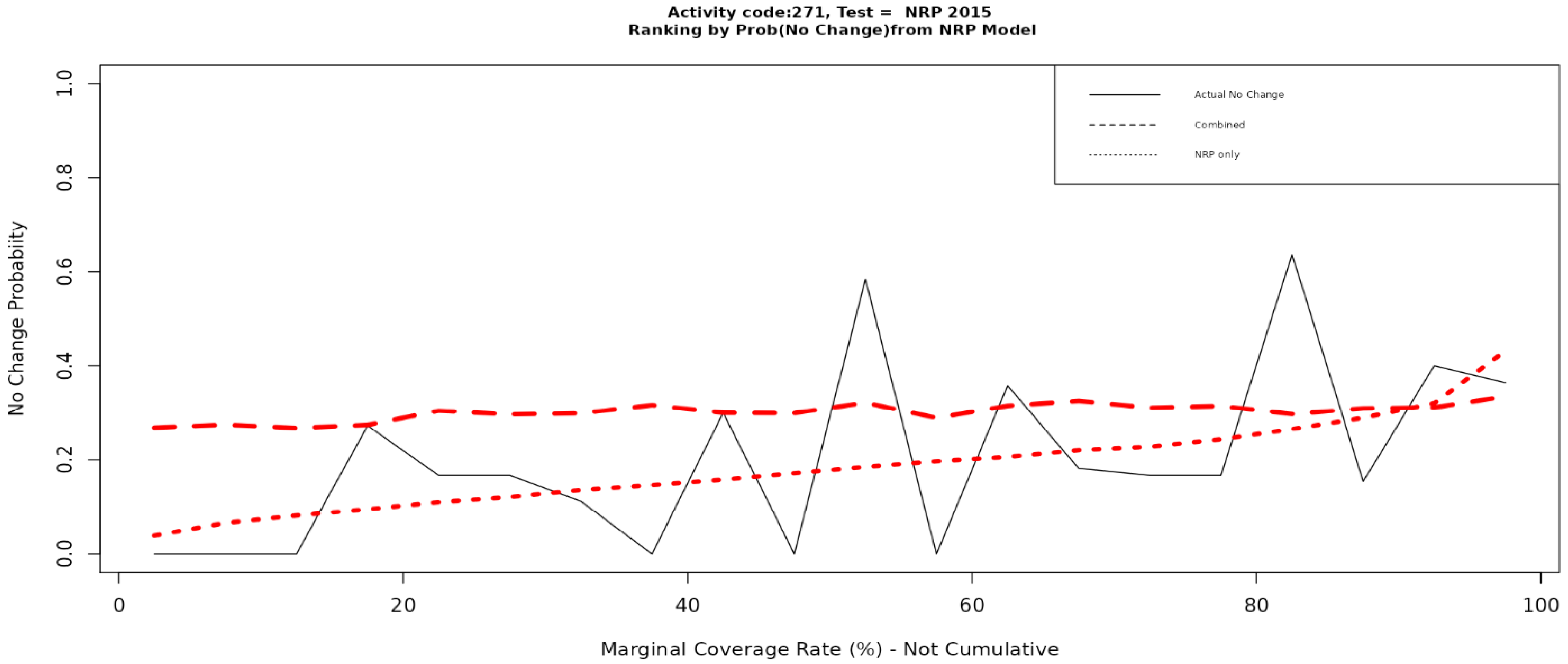


Combining Operational and NRP data with Statistical Controls for Bias - Results

AC 270 Tax Change Predictions on NRP Data, Ranked by Predicted Tax Change Based on the 100% NRP-Only Model

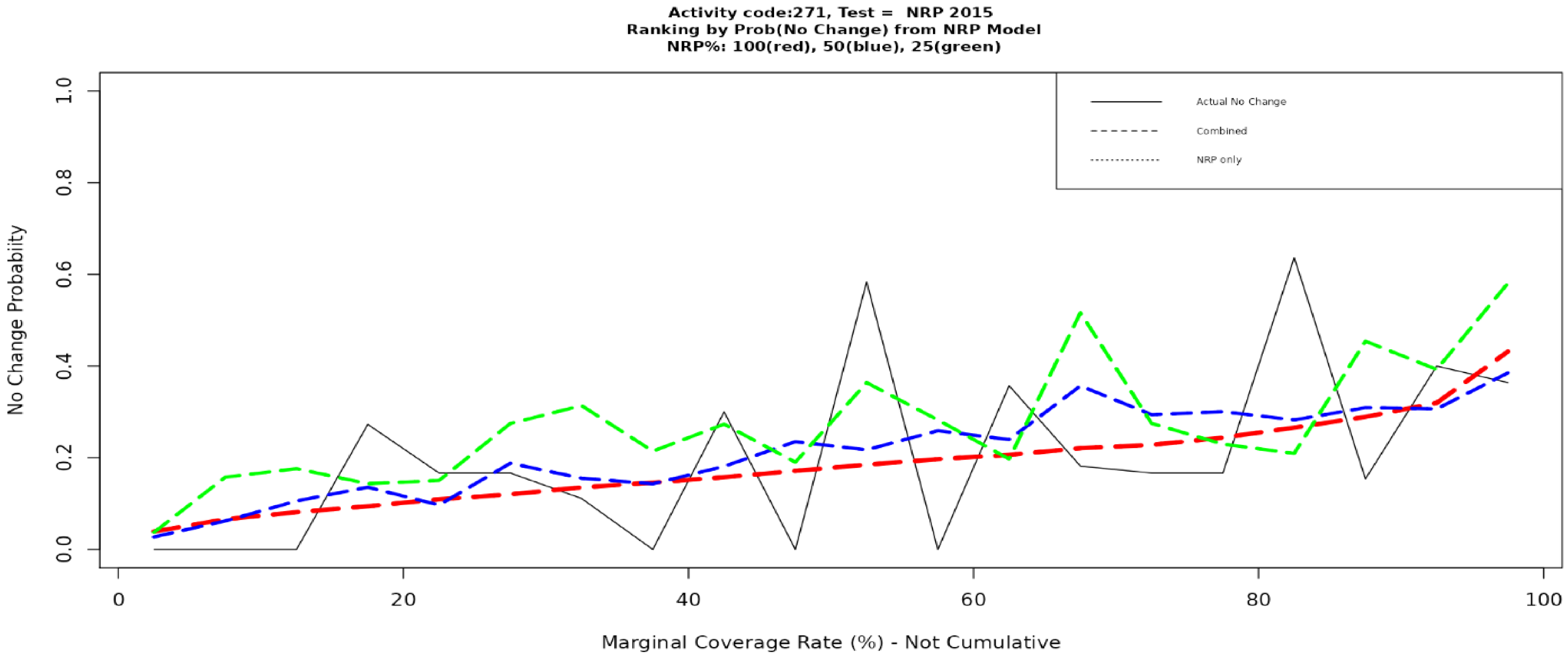


AC 271 Probability of No Change on NRP Data



Combining Operational and NRP data with Statistical Controls for Bias - Results

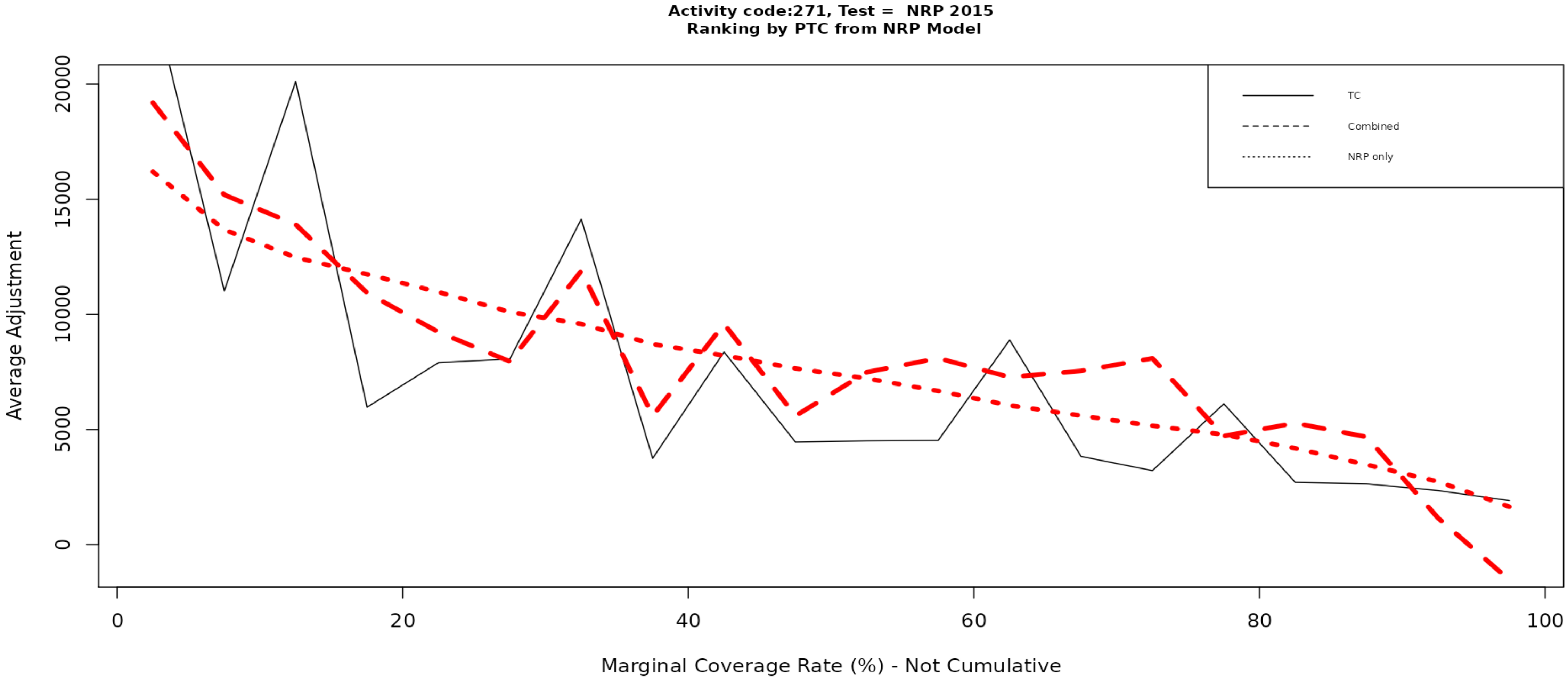
AC 271 Probability of No Change on NRP Data. (The combined data are used only in the second stage of the two-stage risk model)



Combining Operational and NRP data with Statistical Controls for Bias - Results

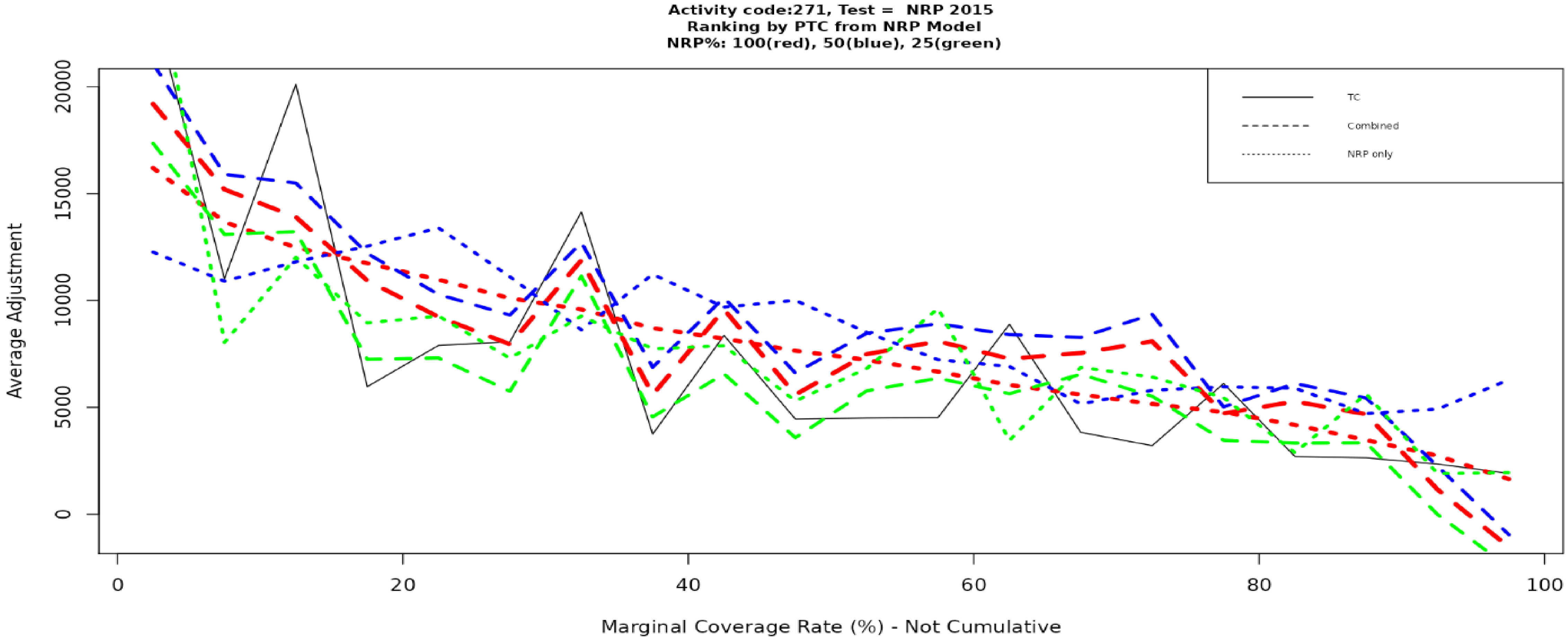
AC 271 Tax Change Predictions on NRP Data, Ranked by PTC From the Model With 100% NRP Data

(The combined data are used only in the second stage of the two-stage risk model)



AC 271 Tax Change Predictions on NRP Data, Ranked by PTC From the Model With 100% NRP Data

(The combined data are used only in the second stage of the two-stage risk model)



Conclusions

- Models trained using operational cases may have bias in estimating the categorical probabilities of audit adjustment. E.g. “no change” percentage (extensive margin)
  - Simple bias controls may not be sufficient
  - Reinforcing the need for the representative, consistently labeled data produced by NRP
- Restricting categorical models to use only NRP data can correct for this bias, leading to results close to those achieved by a representative sample
- Simple bias controls seem to work for modeling conditional audit adjustment amounts (intensive margin)
- Potential Enhancements—more rigorous approaches that
  - Weight operational cases based on the quality of the label
  - Better account for compliant cases ‘screened-out’ of the operational audit pipeline prior to audit

Westat

Improving Lives Through Research®

Augmenting National Research Program Tax Change Estimates by Incorporating Operational Audit Information: A New RAAS Research Initiative

Louis Rizzo, John Riddles, Xiaoshu Zhu, Richard Valliant (Westat), Kimberly Henry (IRS, RAAS)

National Research Program (NRP)

- The NRP is a stratified probability sample of tax returns that are then audited by uniform audit procedures.
- The difference between the audited return and the submitted return is called the ‘tax change’.
- The NRP sample design allows for a reliable representation of the full universe of tax returns in the US, and an unbiased estimator of the US national tax change.

2015 National Research Program (NRP)

Major Stratum	TY15 Sample Size	TY15 Mean Tax Change	Std Error of Mean	CV of Mean (%)	Lower Bound 95% CL for Mean	Upper Bound 95% CL for Mean	2015 Sum of Weights	2015 Total $ value (in billions $)
270	2,172	1,771	57	3.2	1,659	1,884	23,553,072	41.7
271	169	7,958	849	10.7	6,294	9,623	1,320,484	10.5
272	2,086	291	23	7.9	246	336	76,085,502	22.1
273	1,041	1,467	117	8.0	1,238	1,697	14,356,577	21.1
274	934	1,288	80	6.2	1,132	1,444	9,743,186	12.5
275	879	7,057	3,278	46.5	632	13,482	2,664,310	18.8
276	727	6,160	479	7.8	5,220	7,099	731,809	4.5
277	636	15,511	2,817	18.2	9,989	21,034	565,964	8.8
278	855	1,883	162	8.6	1,566	2,201	1,178,086	2.2
279	1,339	1,661	207	12.5	1,256	2,067	4,708,198	7.8
280	876	4,139	353	8.5	3,447	4,831	1,937,794	8
281	1,193	6,568	1,581	24.1	3,469	9,667	495,136	3.3
Other	835	1,992	216	10.8	1,569	2,415	7,436,308	14.8
All	13,742	1,217	67	5.5	1,087	1,348	144,776,427	176.2

Operational (OP) Audits

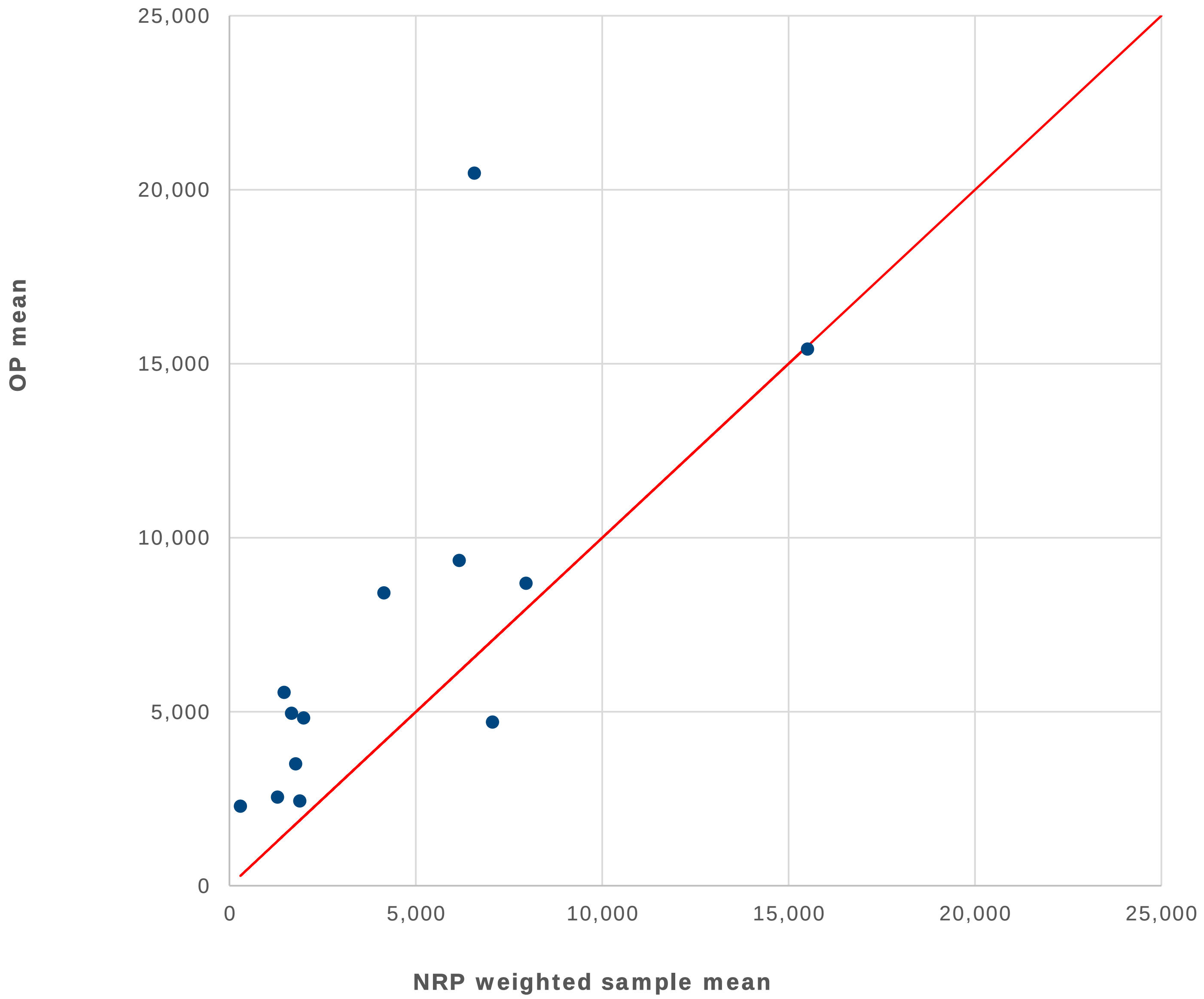
- In recent years, budget constraints for IRS have resulted in a reduction of NRP sample sizes.
  - Tax Year 2015 sample size: 13,742
- A large source of auxiliary information to supplement the declining NRP sample sizes are the operational audits (OP), the ‘ordinary’ audits IRS performs to ensure compliance. The counts for OP are very large, many times that of the NRP.
- From the viewpoint of the NRP these audits are ‘free’: they are available (and paid for) from routine compliance work, and only need to be processed for NRP research.
- But they suffer from selection bias: they are not chosen randomly.

Comparison of NRP and Operational Audits

- For 2015, the overall NRP tax-change mean was $1,217. The OP tax-change mean was $3,315. This difference is highly significant.

2015 Returns	Percentage
NRP Weighted Negative Tax Change	8.3%
NRP Weighted Zero Tax Change	43.4%
NRP Weighted Positive Tax Change	48.3%
NRP 2015 Weighted Total	100.0%
OP Negative Tax Change	5.3%
OP Zero Tax Change	39.8%
OP Positive Tax Change	54.9%
OP 2015 Total	100.0%

Comparison of NRP and Operational Audits



Propensities to be an OP Audit

- Using the 2015 tax return data, and later the 2013-2015 Tax Years together, we estimated “propensities to be OP”.
- The predictor matrix was all the self-reported information on the tax return itself, which is present for all tax returns.
- We used CFOREST, a ML propensity estimation package in R. CFOREST implements a version of random forests.
- If we assume propensity is completely determined by known tax-return information, then these propensities can be treated as if they were probabilities, and weighting with them adjusts for selection.

Estimated Propensities of a Tax Return Being OP Audited.

Major Stratum	NRP Return Total Count	Mean of Pred Prop	Min. of Pred Prop	1st Qrt of Pred Prop	Median of Pred Prop	3rd Qrt of Pred Prop	95th Perc of Pred Prop	Max. of Pred Prop
270	23,898,238	1.35%	0.04%	0.23%	0.73%	1.85%	4.27%	22.2%
271	1,334,259	1.49%	0.24%	0.44%	0.78%	2.45%	4.09%	16.7%
272	76,187,820	0.10%	0.004%	0.02%	0.03%	0.06%	0.36%	6.8%
273	14,461,352	0.63%	0.02%	0.06%	0.16%	0.59%	2.74%	14.7%
274	9,827,635	0.68%	0.01%	0.05%	0.11%	0.38%	3.10%	18.6%
275	2,680,147	0.51%	0.02%	0.14%	0.29%	0.64%	1.71%	7.8%
276	739,791	0.89%	0.11%	0.41%	0.69%	1.25%	2.11%	6.8%
277	574,019	1.47%	0.29%	0.85%	1.27%	1.90%	3.02%	8.3%
278	1,181,166	0.33%	0.06%	0.16%	0.25%	0.41%	0.82%	3.7%
279	4,724,458	0.38%	0.03%	0.13%	0.23%	0.43%	1.18%	4.9%
280	1,957,898	1.11%	0.07%	0.48%	0.78%	1.34%	3.26%	16.5%
281	505,216	1.96%	0.19%	0.80%	1.29%	2.28%	5.82%	23.1%
Other	7,511,262	0.74%	0.06%	0.24%	0.41%	0.84%	2.70%	21.8%

NRP Augmented by Self-Representing OP I

- A conservative estimator we developed is an NRP estimator augmented by treating the OP as self-representing.
- By self-representing we mean the OP audits are brought in with a weight of 1, representing themselves only.
- The NRP returns are brought in with their NRP weight multiplied to 1 minus the estimated propensity of the NRP return being OP.
- $\hat{T}_h = \sum_{\mathcal{S}_h(NRP)} \left\{ w_{hi}^{(NRP)} \left(1 - \hat{\pi}_{hi}^{(OP)}\right) y_{hi} \right\} + \sum_{\mathcal{S}_h(OP)} y_{hi}$

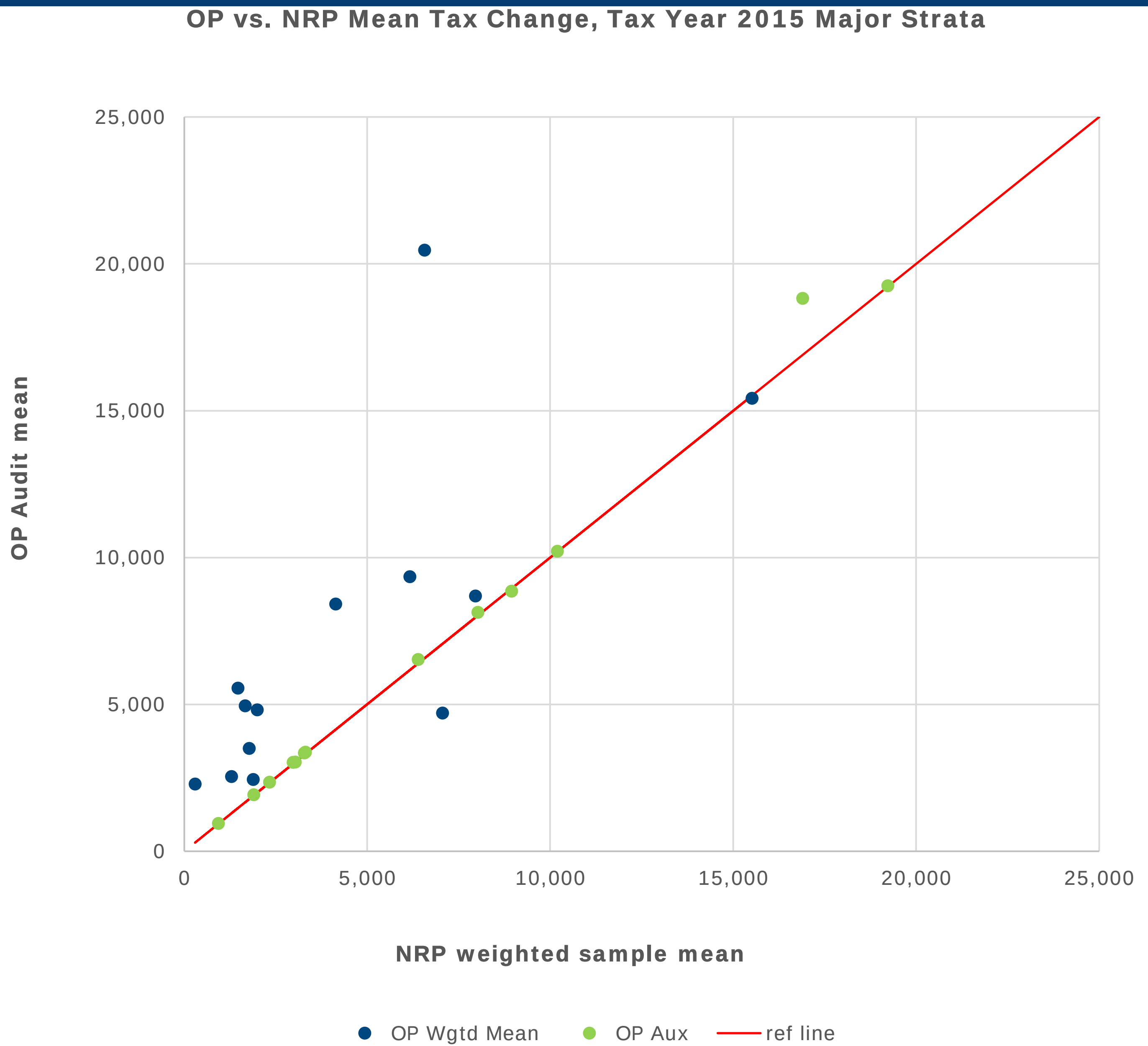
NRP Augmented by Self-Representing OP II

- The final estimator is the (unweighted total OP tax change) + (NRP total tax change reweighted to represent the population excluding OPs)
- Usually, the share of the OP is very small (800,000 returns come in as 1%-2% of the weighted estimator, so it is the equivalent of 100-200 NRP units with their full weights), with only a small reduction in SE. But assumptions are very limited, so risk of bias is also very low

NRP Augmented by Self-Representing OP III

Stratum	NRP Weight (Pos Only)	Simple NRP mean (Pos only)	Simple NRP Mean Std Err	OP Aux NRP Mean (Pos only)	Std Error OP Aux NRP Mean	Ratio OP Aux NRP Mean to NRP Mean	Ratio OP Aux NRP Std Err to NRP Std Err
270	12,933,039	$3,283	$1,558	$3,347	$1,610	1.019	1.033
271	1,047,502	$10,193	$5,451	$10,212	$5,391	1.002	0.989
272	26,372,078	$929	$597	$941	$616	1.013	1.032
273	9,470,435	$2,331	$349	$2,354	$344	1.01	0.987
274	6,776,473	$1,897	$1,006	$1,920	$1,033	1.012	1.027
275	2,118,279	$8,940	$13,915	$8,859	$13,622	0.991	0.979
276	572,643	$8,023	$3,171	$8,132	$3,219	1.014	1.015
277	460,844	$19,222	$12,277	$19,252	$12,103	1.002	0.986
278	760,881	$3,030	$2,033	$3,031	$2,029	1.001	0.998
279	2,690,371	$3,312	$2,528	$3,368	$2,564	1.017	1.014
280	1,359,468	$6,390	$3,978	$6,530	$4,116	1.022	1.035
281	265,850	$16,900	$8,424	$18,824	$8,826	1.114	1.048
Other	5,079,449	$2,975	$3,033	$3,021	$2,995	1.015	0.988
Total	**69,907,312**	**$2,639**	**$3,102**	**$2,675**	**$3,070**	**1.014**	**0.990**

NRP Augmented by Self-Representing OP III



Composite Estimators

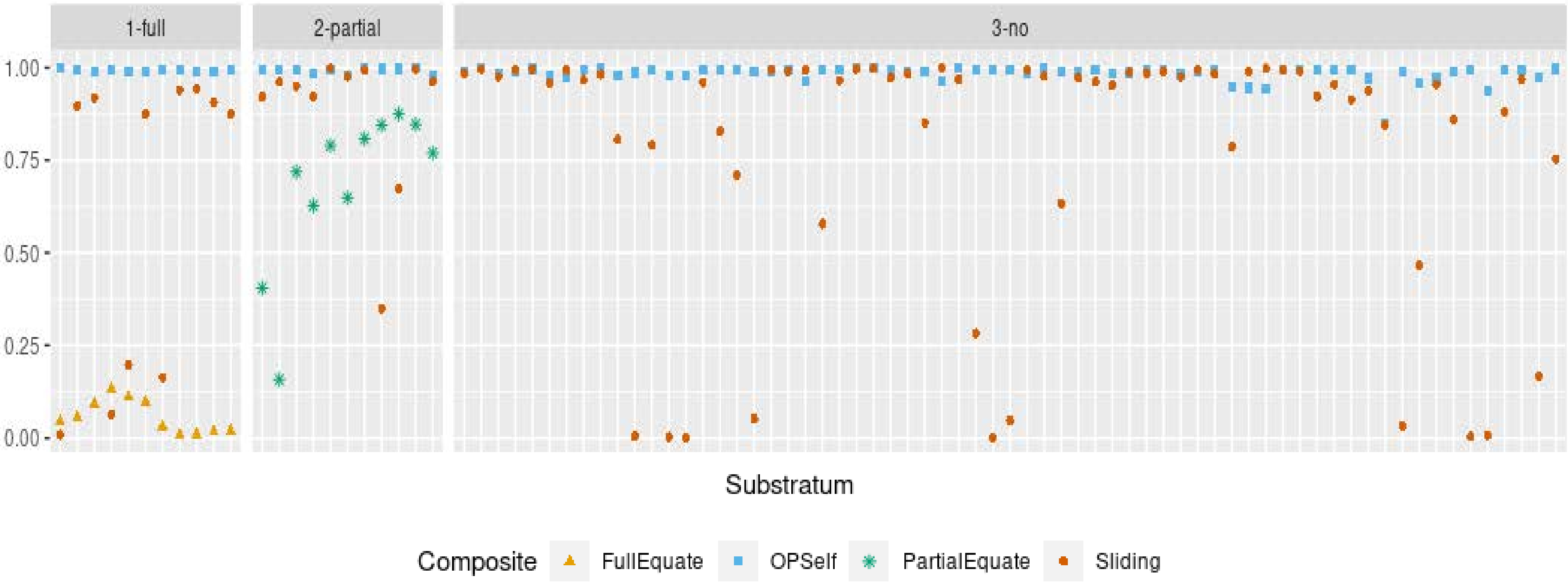
- If we can assure ourselves that the OP estimator is unbiased for NRP stratum h (tax-return domain U_h), then we can use a composite estimator:
- $\hat{T}_h = \alpha_h \, \hat{t}_{NRP}(U_h) \; + \; (1-\alpha_h)\hat{t}_{OP}(U_h)$
- α_h is based on relative size of the variances of the two estimators (it is computed to be inversely proportional to relative sizes of variances).
- If OP is not unbiased for the full stratum domain U_h, but OP units are unbiased over a partial stratum domain A_h, then we can composite over the equated domain A_h:
- $\hat{T}_h = \hat{t}_{NRP}(U_h - A_h) \; + \; \{\alpha_h \, \hat{t}_{NRP}(A_h) \; + \; (1-\alpha_h)\hat{t}_{OP}(A_h)\}$

Sliding-Scale Estimator I

- Another much less conservative estimator (as it explicitly allows a bias) is the sliding-scale estimator.
- The form of the estimator is the same as the simple composite:
- $\hat{T}_h = \alpha_h \, \hat{t}_{NRP}(U_h) \; + \; (1-\alpha_h)\hat{t}_{OP}(U_h)$
- In this case, we do not require the operational estimator in stratum h to be unbiased (i.e., to equate to the NRP estimator).
- α_h is inversely proportional to relative size of estimated mean-squared-error.
- MSE of NRP estimator is the variance (as it is unbiased). MSE of OP estimator is computed with squared difference between NRP and OP as the primary squared-bias component. Unbiased MSE is generated.

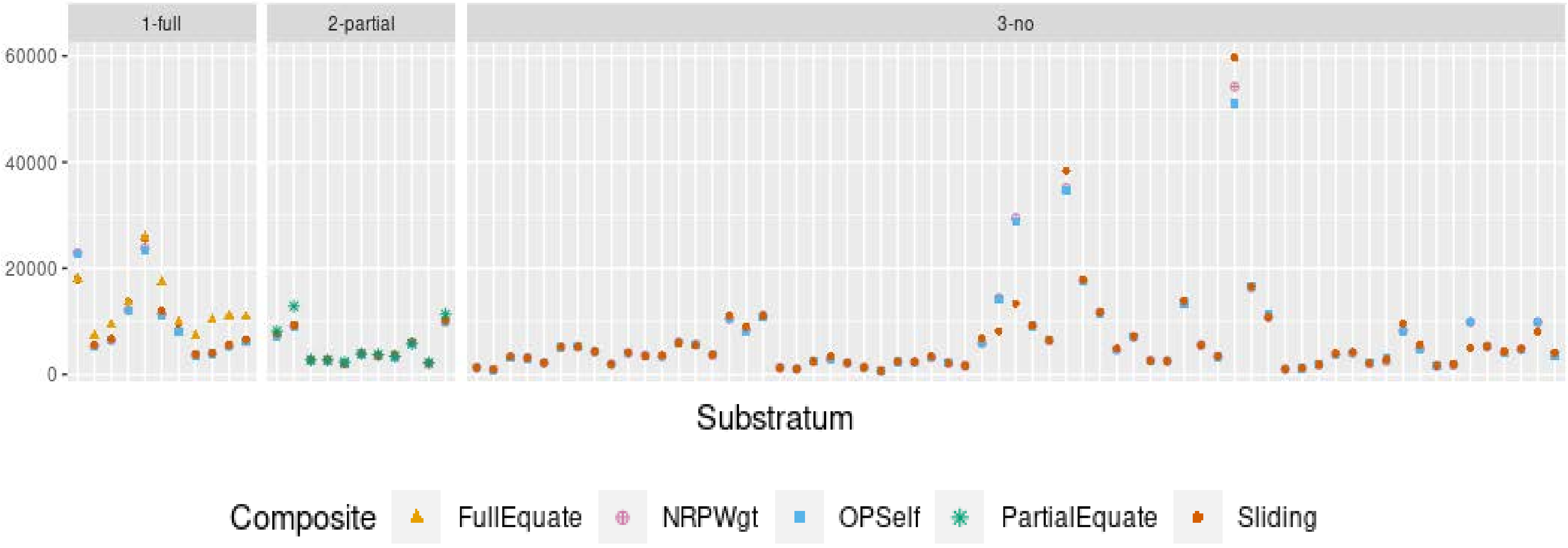
Proportion of Estimated Mean from NRP Estimator for All Types of Estimators

› Proportion of Mean Due to NRP, by equated-type of substratum:



Final Estimates of Means for All Types of Estimators

› Average tax change estimates, by equated-type of substratum:

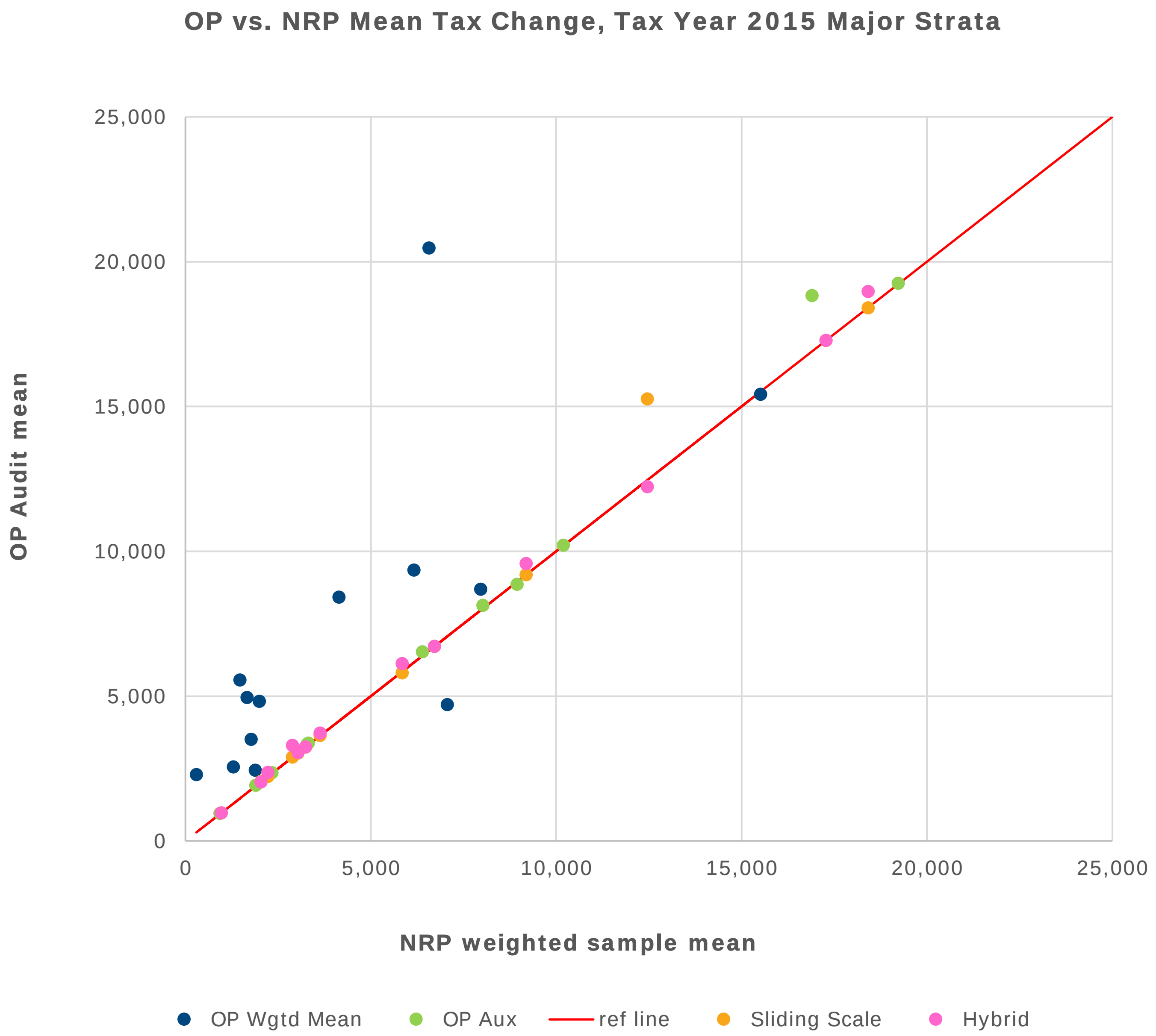


Summary of Estimators

› Hybrid estimators, by equated-type of substratum:

Substratum	NRP Means Estimate	NRP Means Std Err	OP Self-rep Means Estimate	OP Self-Rep Means Std Err	Sliding scale Means Estimate	Sliding scale Means Std Err	Hybrid Means Estimate	Hybrid Means Std Err
270	$3,235	$49	$3,169	$47	$3,237	$49	$3,235	$49
271	$12,456	$3,716	$12,309	$3,684	$15,257	$1,021	$12,224	$624
272	$967	$47	$968	$47	$967	$47	$968	$47
273	$2,218	$108	$2,205	$107	$2,220	$108	$2,362	$132
274	$2,034	$243	$2,022	$242	$2,051	$242	$2,035	$243
275	$5,845	$1,351	$5,796	$1,320	$5,801	$407	$6,116	$1,353
276	$9,185	$957	$9,132	$933	$9,184	$945	$9,573	$938
277	$18,409	$1,870	$18,196	$1,824	$18,401	$1,834	$18,966	$1,861
278	$3,030	$136	$3,022	$136	$3,030	$136	$3,043	$156
279	$3,628	$352	$3,636	$349	$3,632	$248	$3,720	$343
280	$6,717	$503	$6,642	$468	$6,716	$502	$6,717	$503
281	$17,278	$2,058	$17,283	$1,780	$17,278	$2,057	$17,278	$2,058
Other	$2,887	$204	$2,887	$200	$2,887	$204	$3,294	$202

Sliding-Scale and Hybrid Estimators I



Conclusions I

- The NRP is a high-quality probability sample of tax-returns that provide unbiased estimators of tax-change.
- Sample sizes for the NRP have been declining in recent years due to budget constraints.
- This work is part of an IRS-RAAS research initiative to consider incorporation of operational audits into the NRP estimators.
- The primary issue in this is ‘selection’: operational audits are not selected randomly and are not necessarily representative of the larger tax-return universe.

Conclusions II

- We have developed using CFOREST a methodology for estimating 'propensity to be an operational audit'. Roughly 0.6% of returns are OP. In our research we found propensities ranging from 0.004% to 23.1%.
- We have developed a relatively conservative estimator by bringing in the OP audits as self-representing (a weight of 1), among the NRP units with their NRP weights.
- We have studied equating of NRP and OP audits and found in some substrata the two distributions closely match. We use composite estimation in these cases.
- In some cases, equating is possible in a subdomain, in which case composite estimation is done in the subdomain only (NRP for the complement).

Conclusions III

- We have developed a ‘sliding scale’ estimator which does full composition but allows for a bias in the OP estimator (reducing its share accordingly).
- These alternatives show that there is a way forward for IRS to incorporate operational audits judiciously into the NRP program.

Thank you

- Thank you for participating in our presentation. Our emails for questions or comments are lourizzo@Westat.com and Kimberly.A.Henry@irs.com.

Integrating Reward Maximization and Population Estimation

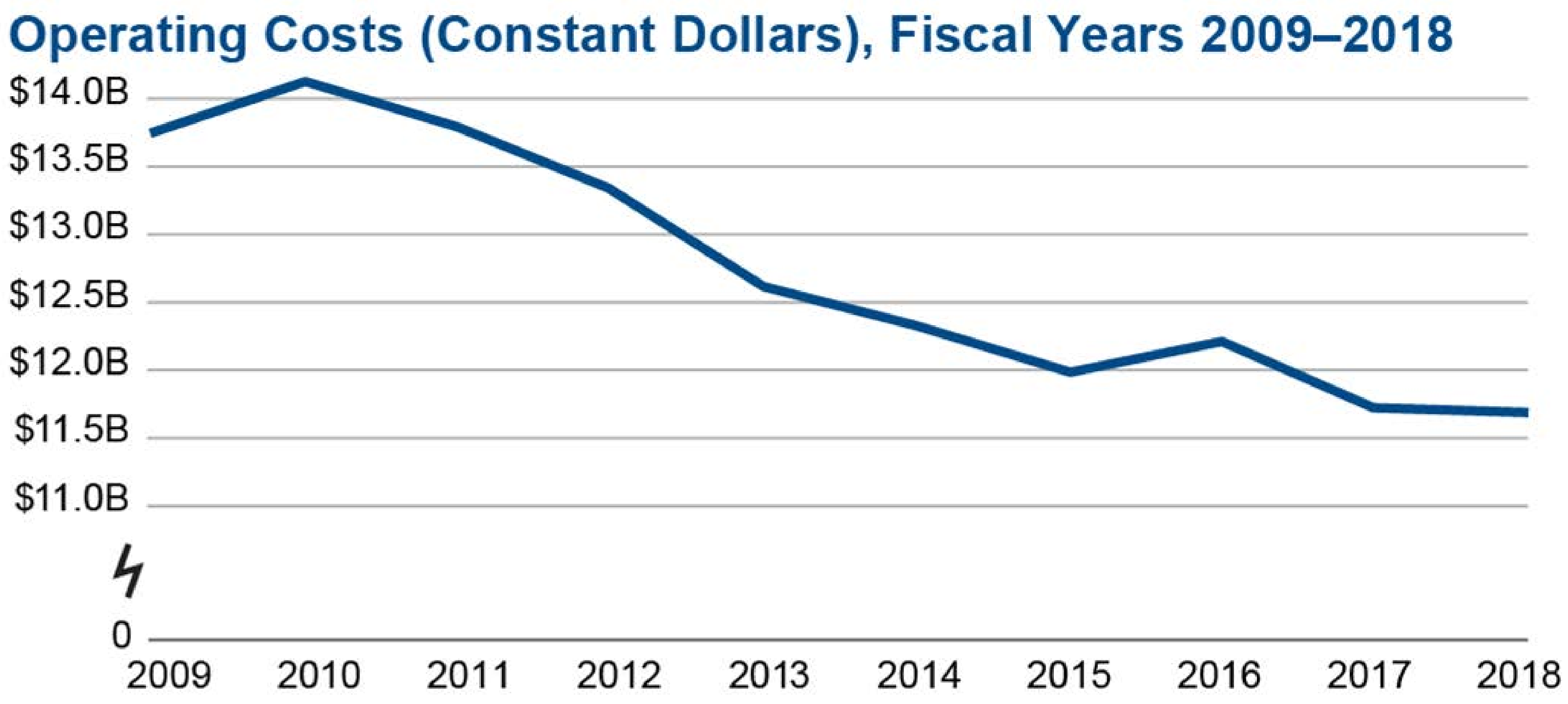
Sequential Decision-Making for Internal Revenue Service Audit Selection

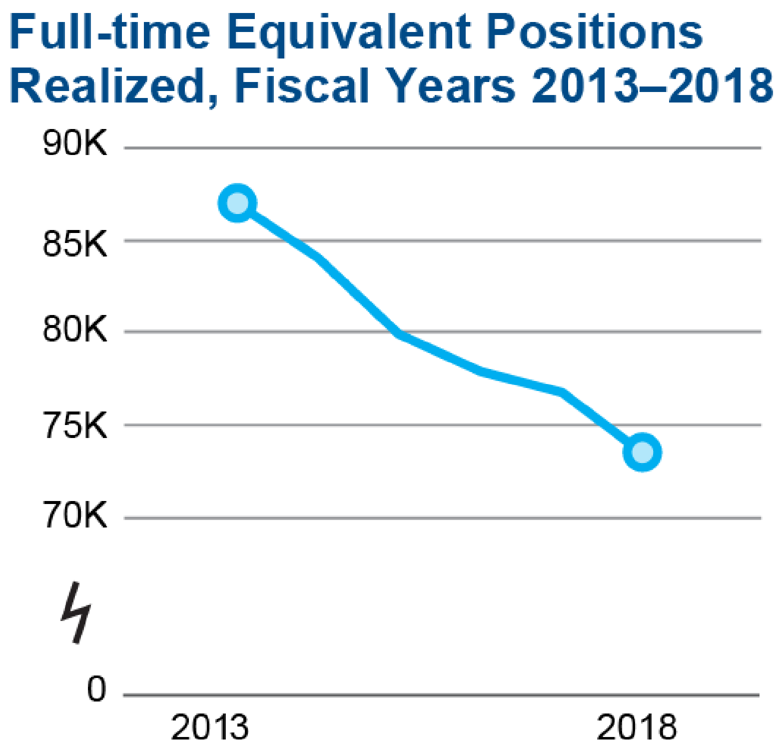
Peter Henderson[1], Ben Chugg[1], Brandon Anderson[12], Kristen Altenburger[1], Alex Turk[2], John Guyton[2], Jacob Goldin[1], Daniel E. Ho[1]

[1]Stanford University [2]IRS RAAS

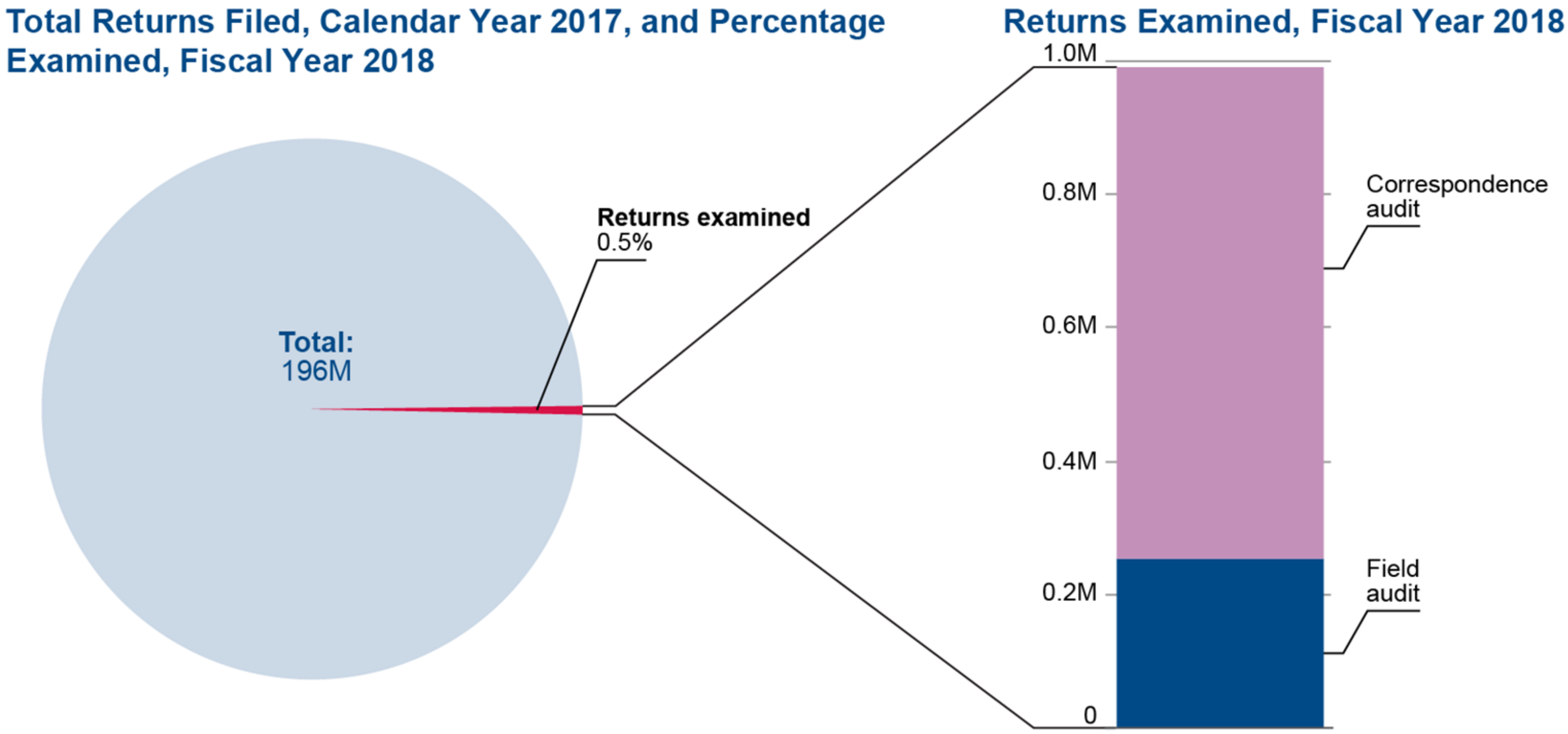
All views and opinions expressed in this presentation are our own and not of any of our co-authors, nor of the Internal Revenue Service or any other company or government entity.

Institutional Context





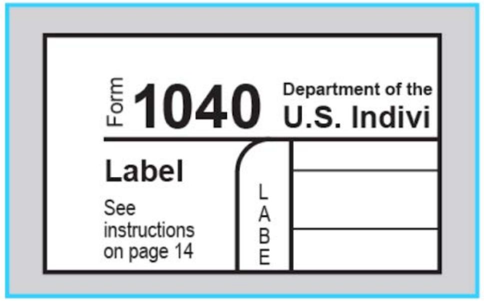
Institutional Context



SOURCE: 2018 *IRS Data Book* Table 9a

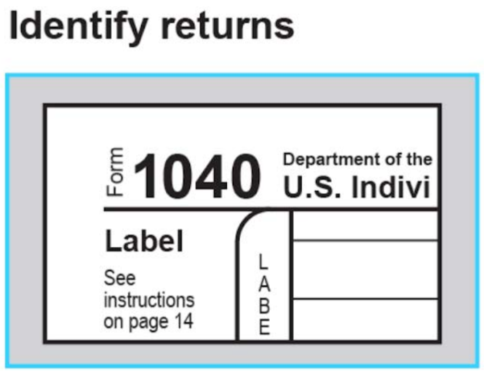
Stylized Program

Identify returns

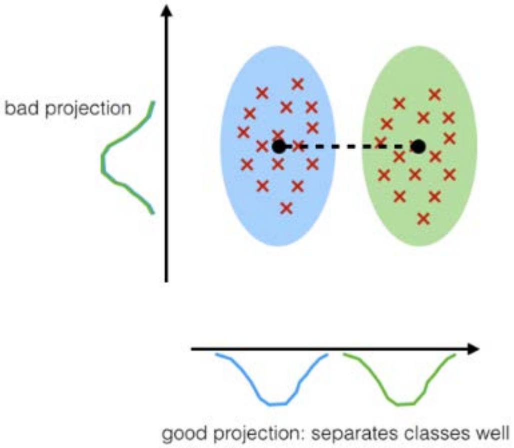


Random sample
(~15k / year, 2006-14)

Stylized Program

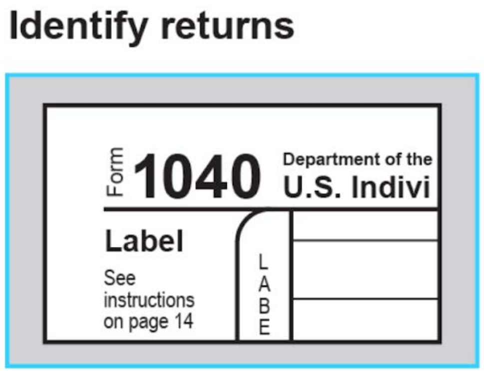


Random sample
(~15k / year, 2006-14)

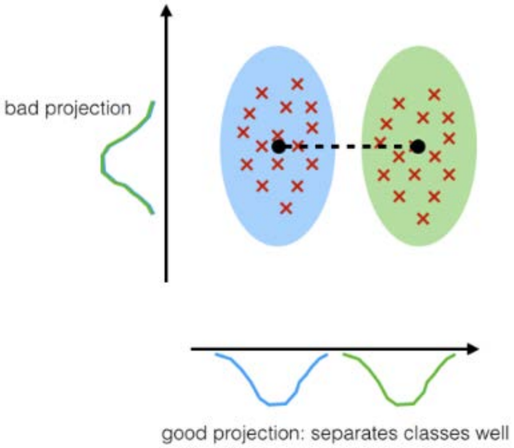


Risk model

Stylized Program



Random sample
(~15k / year, 2006-14)

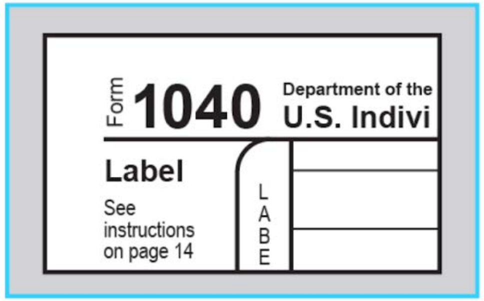


Risk model

Risk selected Op Audits
(>500k / year)

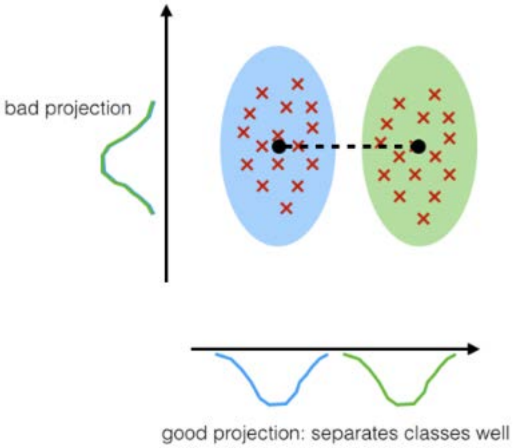
Stylized Program

Identify returns



Random sample
(~15k / year, 2006-14)

Tax Gap Estimate

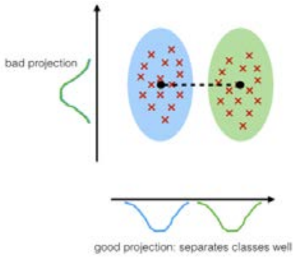


Risk model

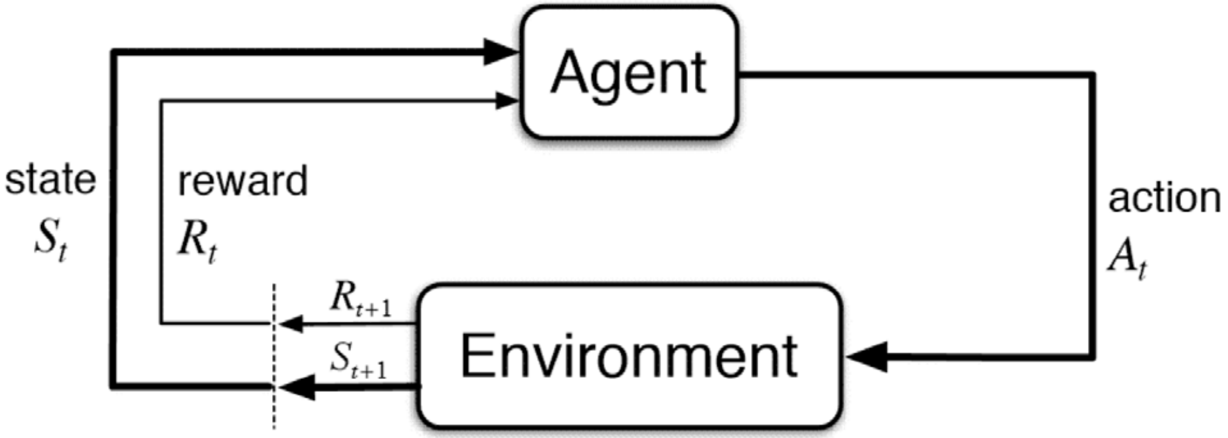
Audit returns

Risk selected Op Audits
(>500k / year)

Sequential Decision-Making (Machine Learning)

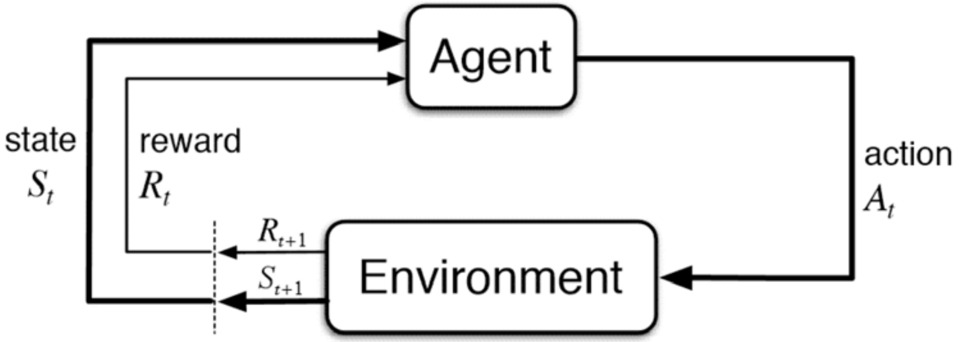


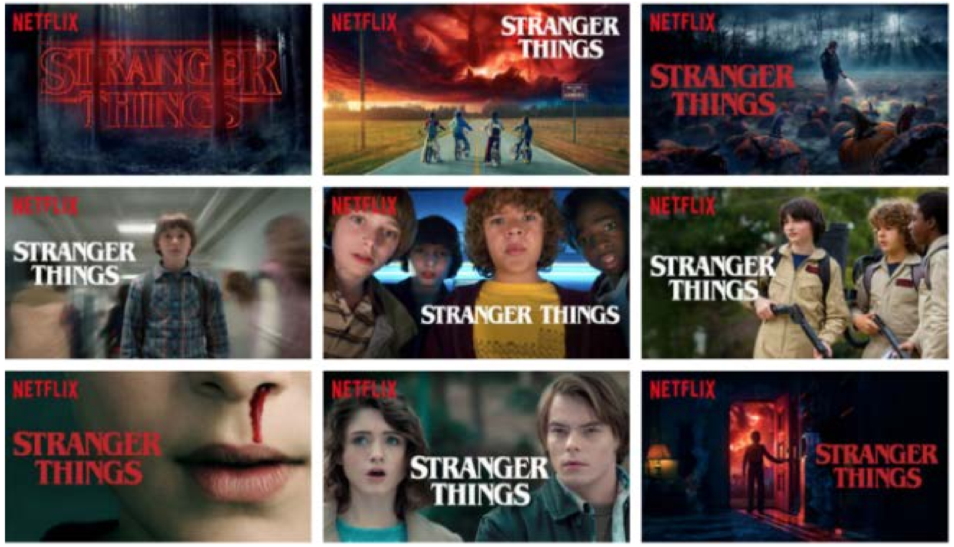
Reward-prediction
Model



Sequential Decision-Making In the Real World

Example: NETFLIX



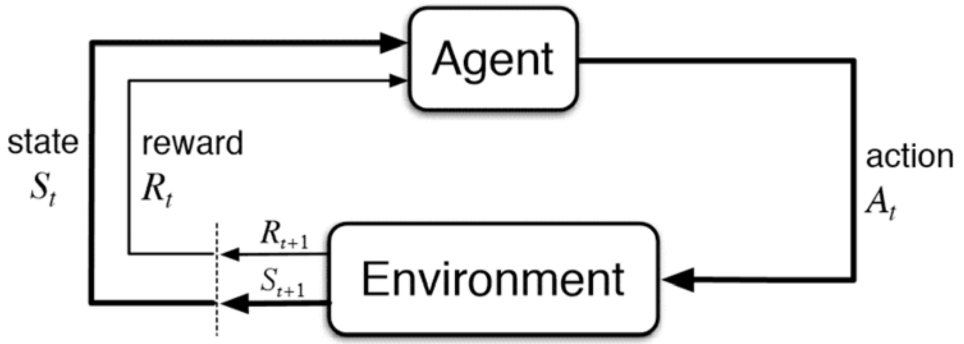


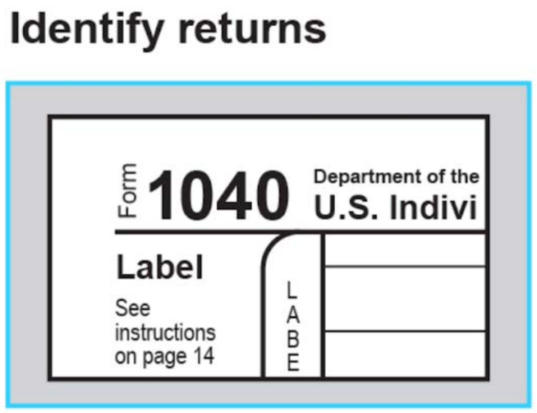
Context	User information on device (environment)
Actions	Set of movie banners to show
Reward	User engagement (click-through, minutes)
Learner	Identify policy to maximize cumulative reward

Explore new movies / preferences vs. Exploit known preferences

Sequential Decision-Making In the Real World

Example: IRS





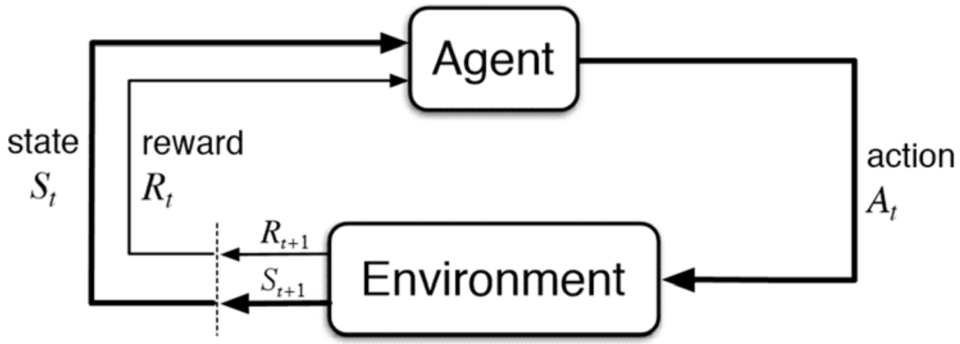
Context	Tax return information (taxpayer, stratum, etc.)
Actions	Selecting returns to audit
Reward	Under-reporting
Learner	Identify policy to maximize cumulative reward

Explore forms of underreporting vs. Exploit known underreporting

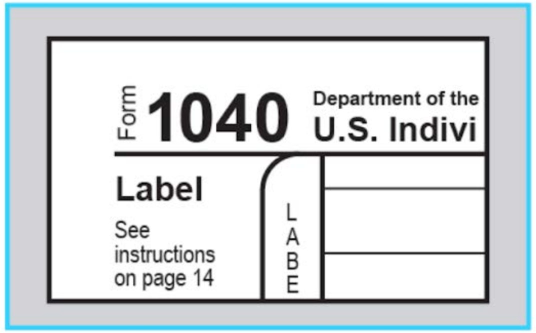
Stanford

Sequential Decision-Making In the Real World

Example: IRS



Identify returns

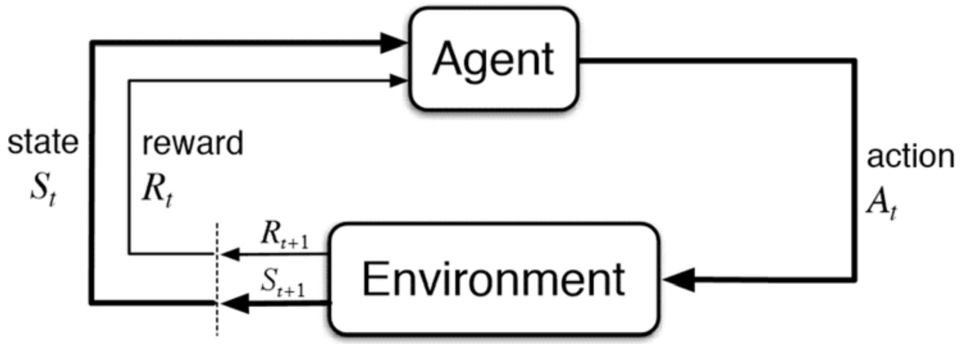


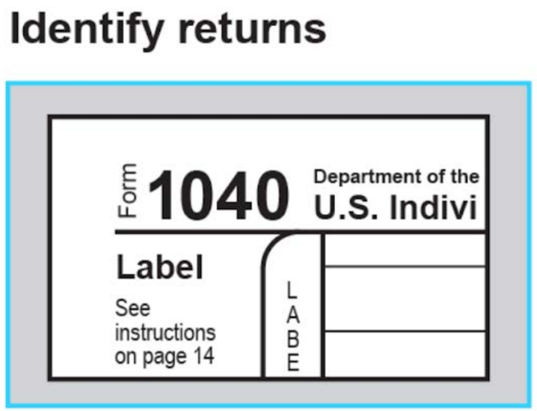
Context	Tax return information (taxpayer, stratum, etc.)
Actions	Selecting returns to audit
Reward	Under-reporting
Learner	Identify policy to maximize cumulative reward

+ Estimate unbiased population statistics (e.g., tax gap, average misreporting)

Sequential Decision-Making In the Real World

Example: IRS



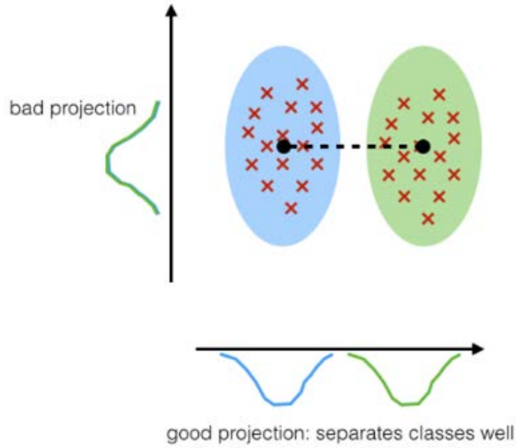


Secondary objective not typical of machine learning literature

+ Estimate **unbiased** population statistics (e.g., tax gap, average misreporting)

Sequential Decision-Making In the Real World

Example: IRS



Risk model

Tempting Solution: Use a regression-based risk-model to do selection and estimation, with no random sampling.

Problems: Sequentially-learned models are known to be biased and there are no theoretically guaranteed ways to *remove* this bias in the low sample regime (yet). (Nie et al., 2018)
Lack of exploration leads to suboptimal feedback loops. (Jiang et al., 2019)

Optimize-and-Estimate Structured Bandits

Machine Learning Literature on Sequential Decision-Making (e.g., bandit algorithms that optimize for reward only)

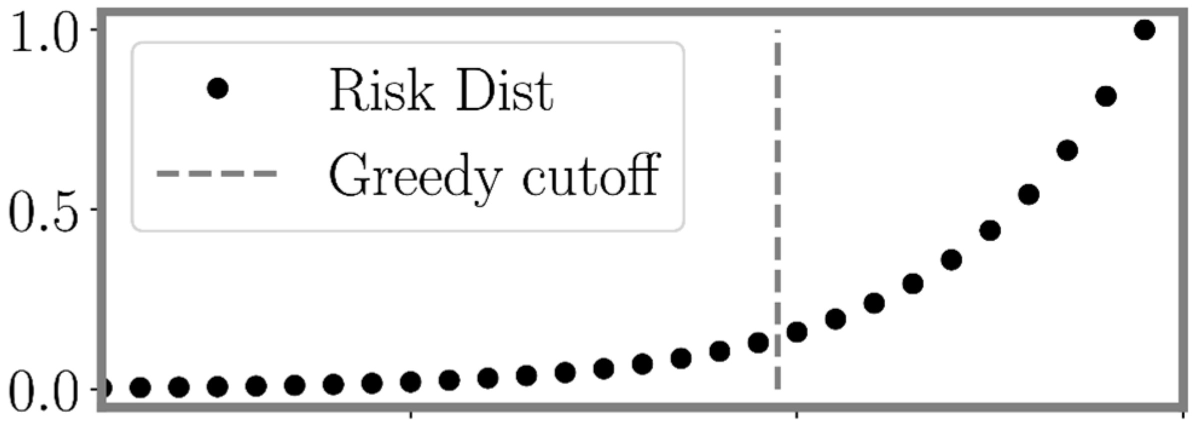
+

Sampling Literature (unbiased estimation of population statistics)

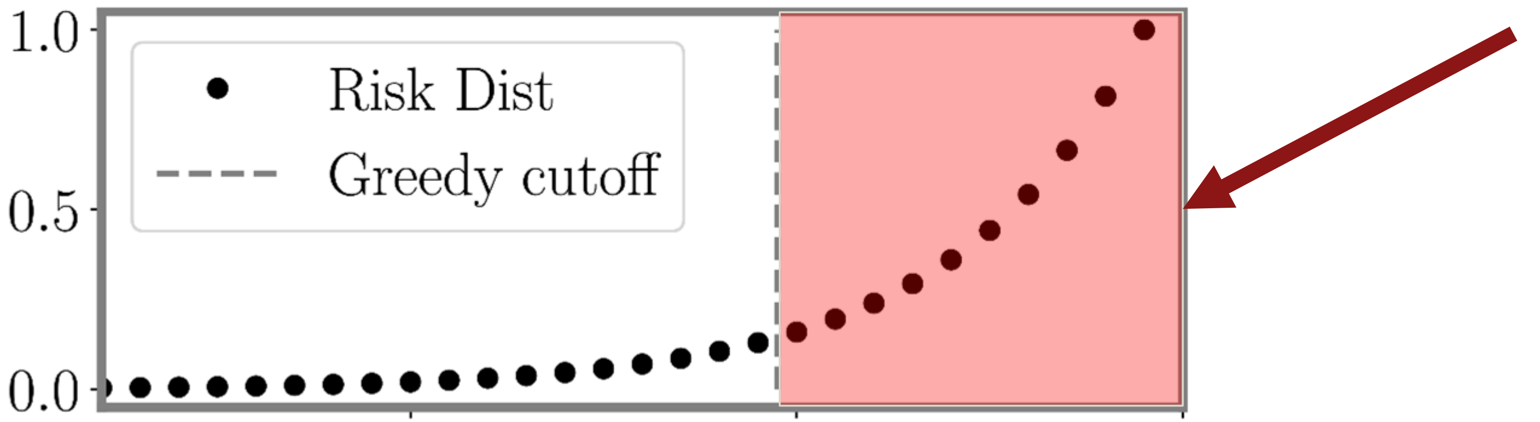
=

Optimize-and-estimate Structured Bandits

Adaptive Bin Sampling



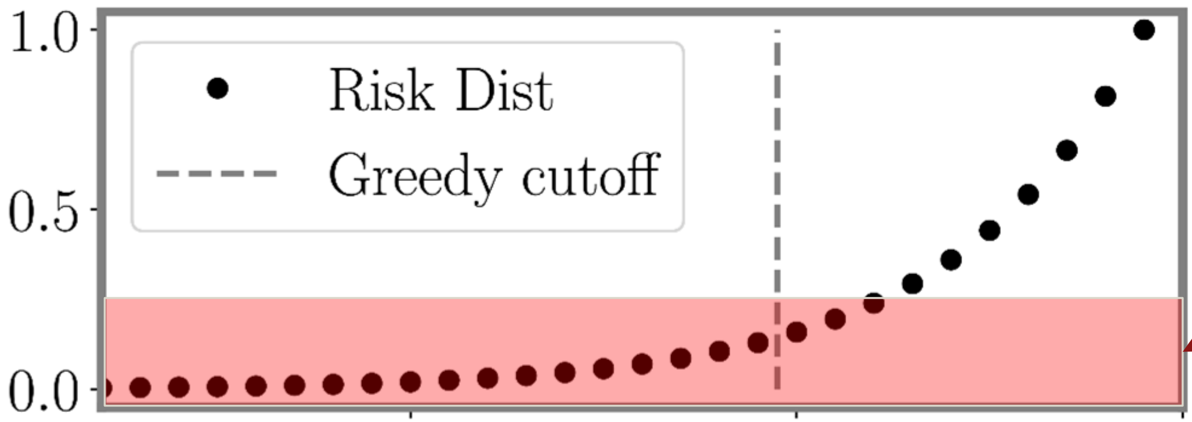
Adaptive Bin Sampling



Greedy selection
(e.g., stylized version of Op audits)

If only use this:
biased model, biased estimate

Adaptive Bin Sampling

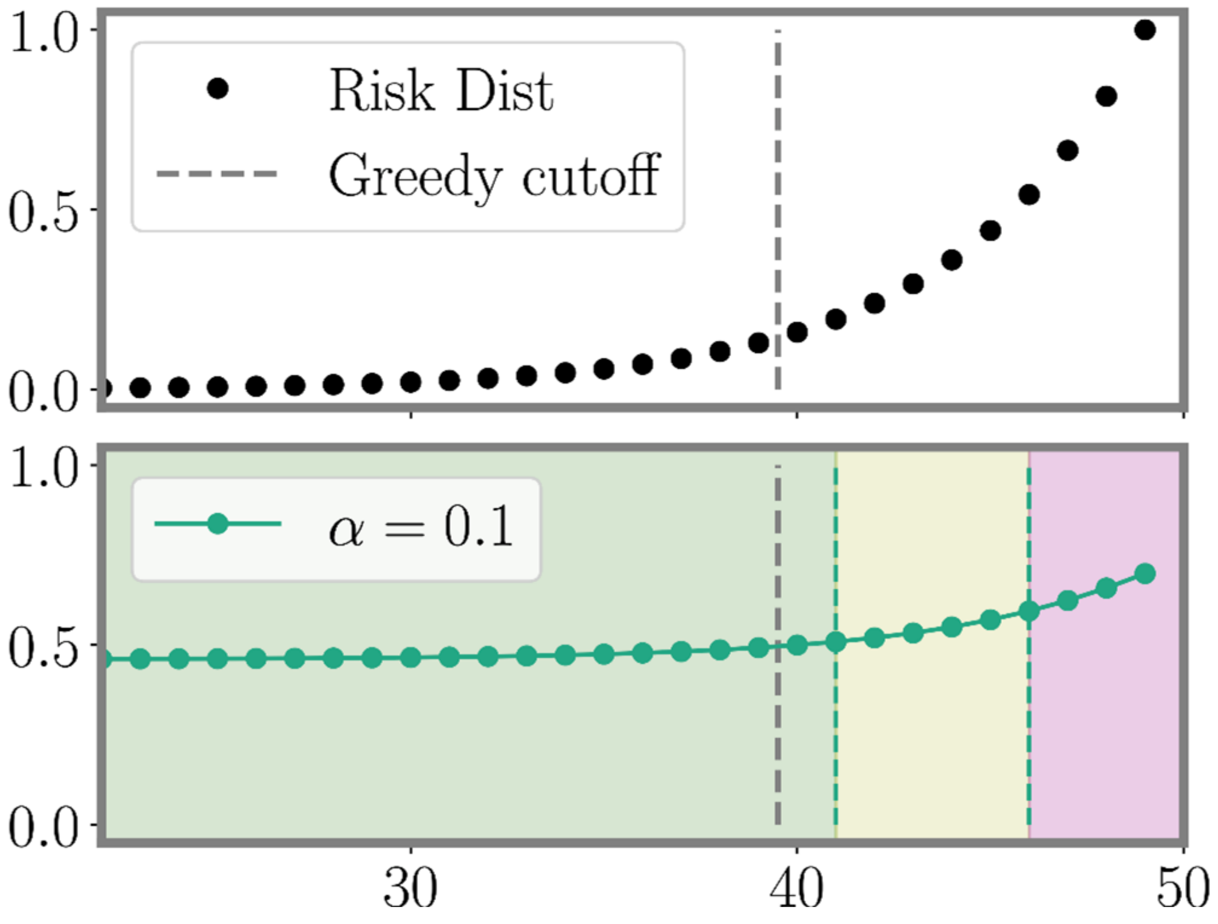


Random selection

If only use this:

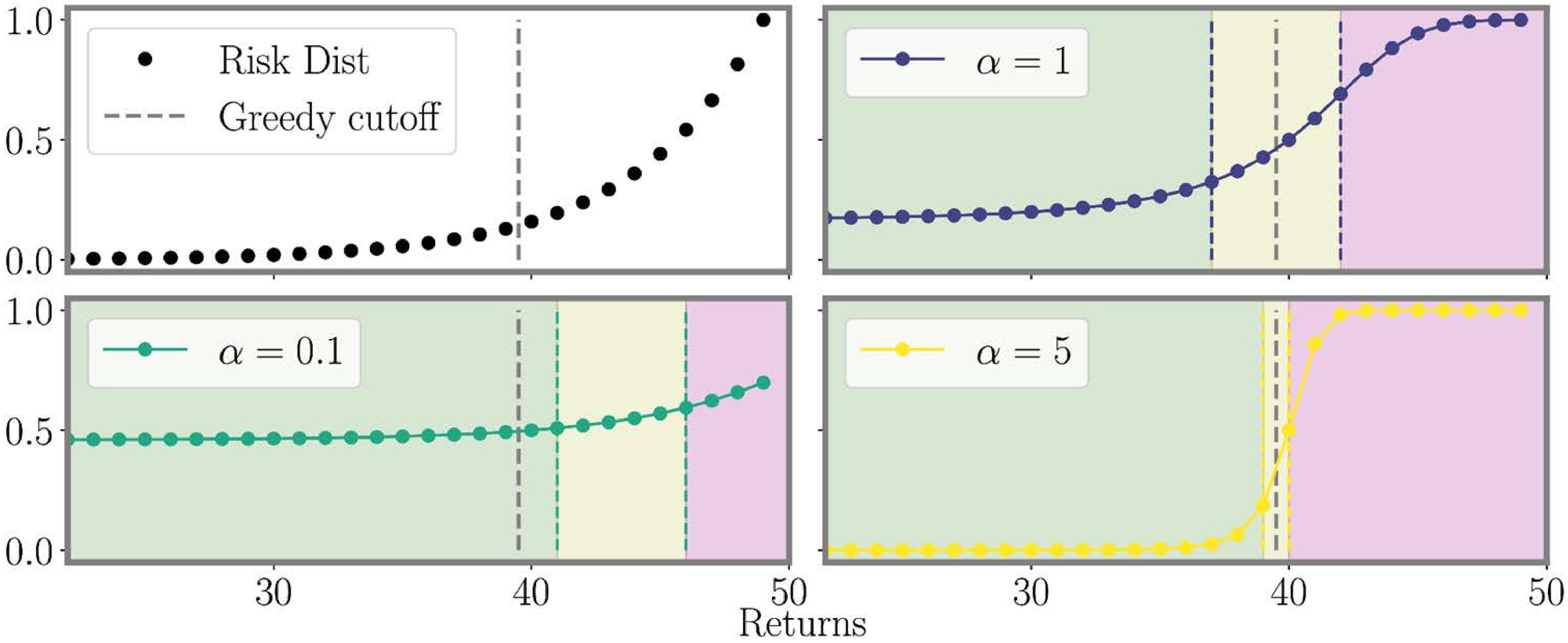
Unbiased estimate,
but sub-optimal and low reward

Adaptive Bin Sampling



Stanford

Adaptive Bin Sampling



Stanford

Adaptive Bin Sampling

Horvitz-Thompson estimator gives **unbiased** estimate.

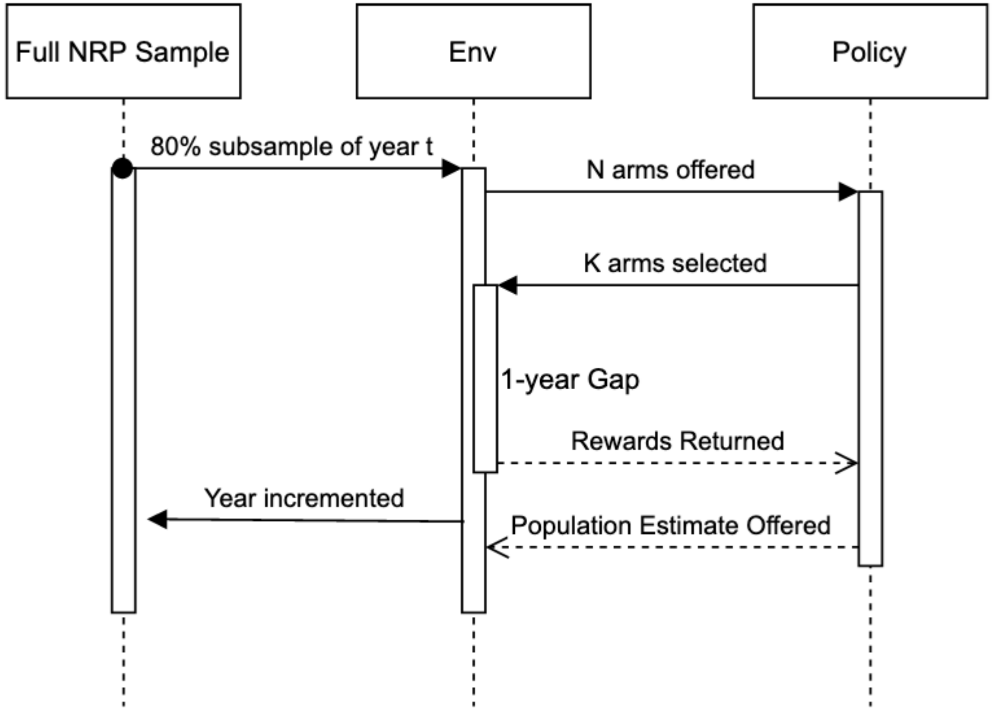
And we have **fine-grained** control over reward-variance trade-off.

$$\hat{\mu}_{HT}(t) = \frac{1}{\sum_a w_a} \sum_{a \in \mathcal{K}} \frac{w_a r_a}{p_a},$$

Experiments

For NRP data years 2006-2014

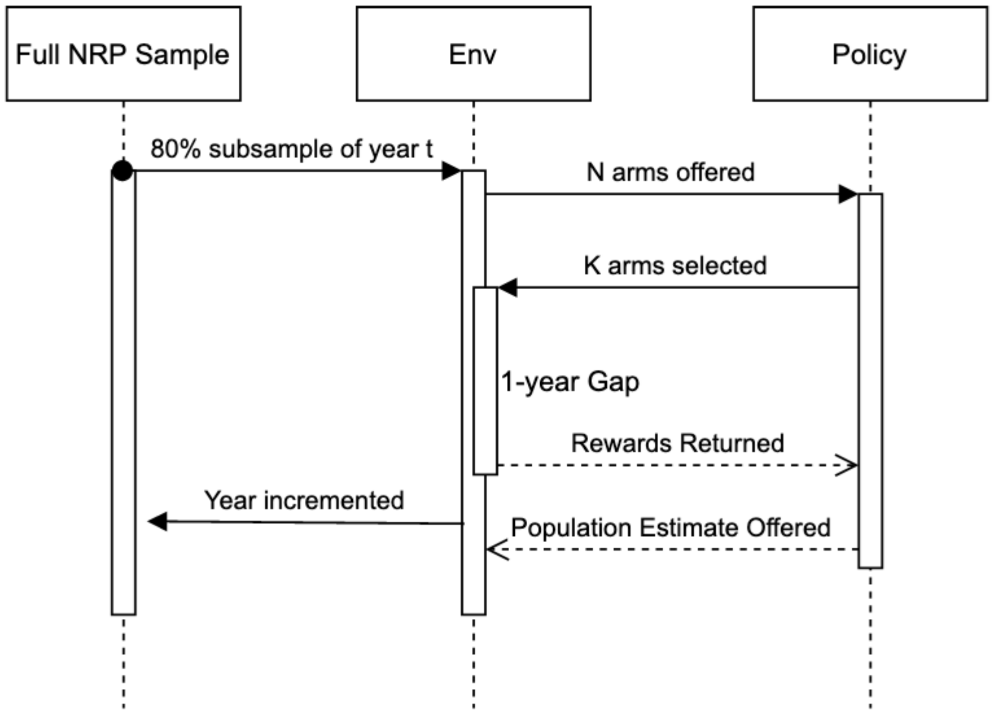
1. Take 80% subsample



Experiments

For NRP data years 2006-2014

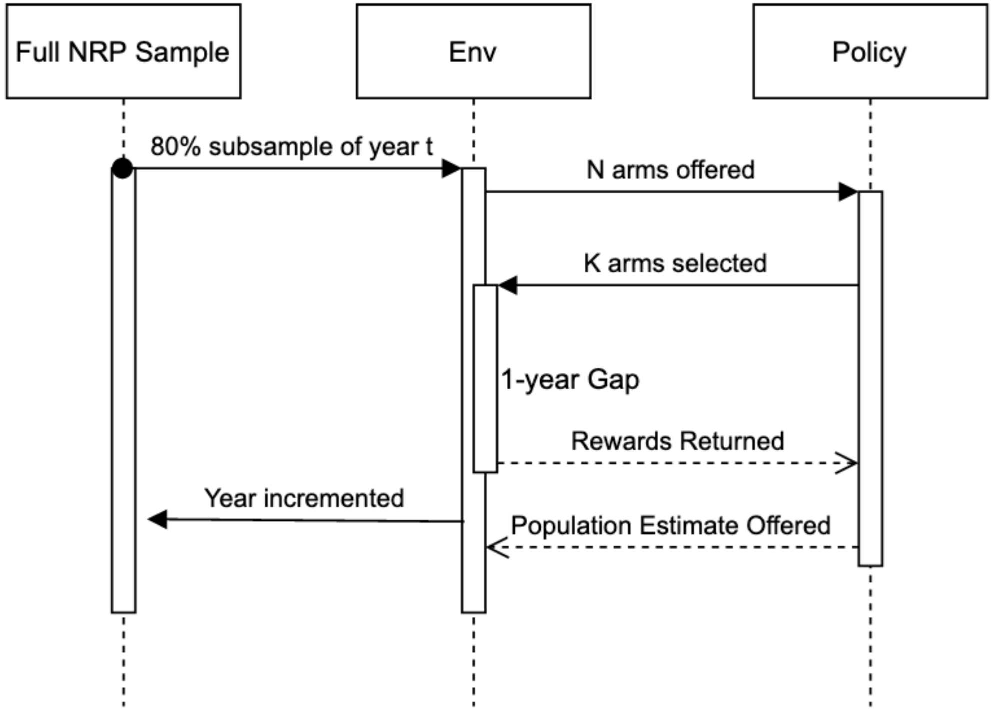
1. Take 80% subsample
2. Give selection policy ~500 covariates from tax return data for each “arm” (tax return) in the sample



Experiments

For NRP data years 2006-2014

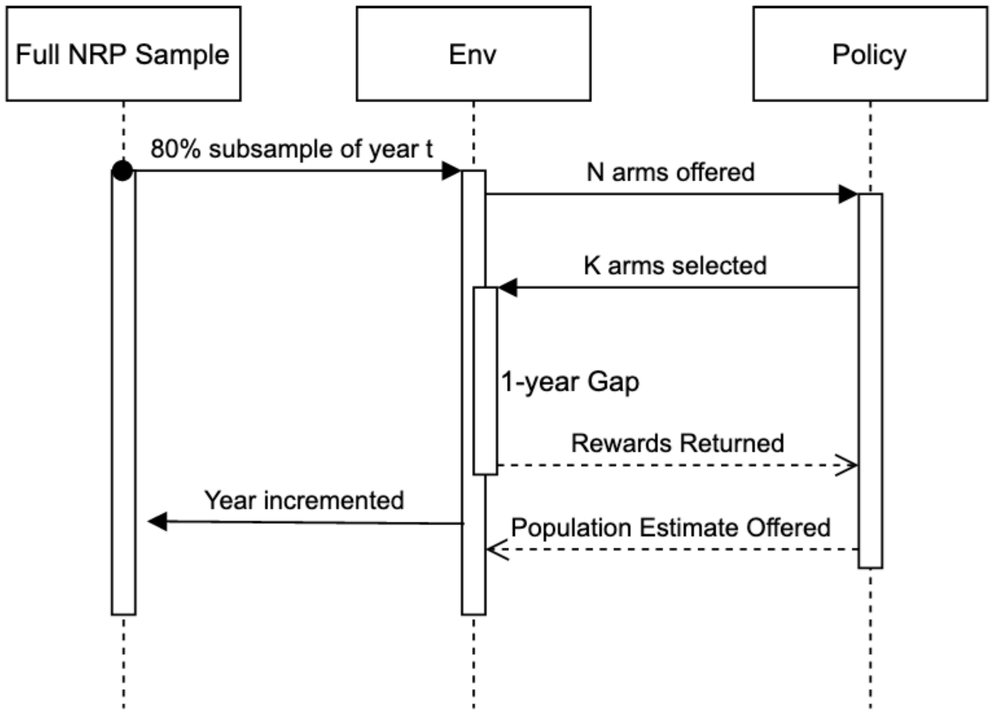
1. Take 80% subsample
2. Give selection policy ~500 covariates from tax return data for each “arm” (tax return) in the sample
3. Selection policy returns arms to audit



Experiments

For NRP data years 2006-2014

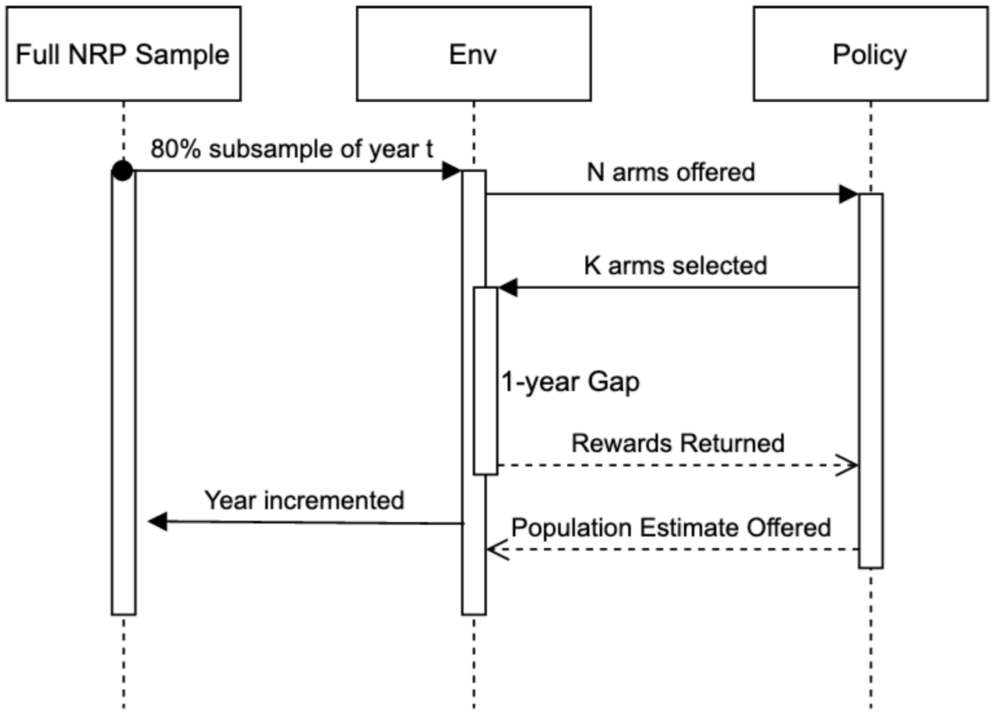
1. Take 80% subsample
2. Give selection policy ~500 covariates from tax return data for each “arm” (tax return) in the sample
3. Selection policy returns arms to audit
4. Simulate a 1 year gap



Experiments

For NRP data years 2006-2014

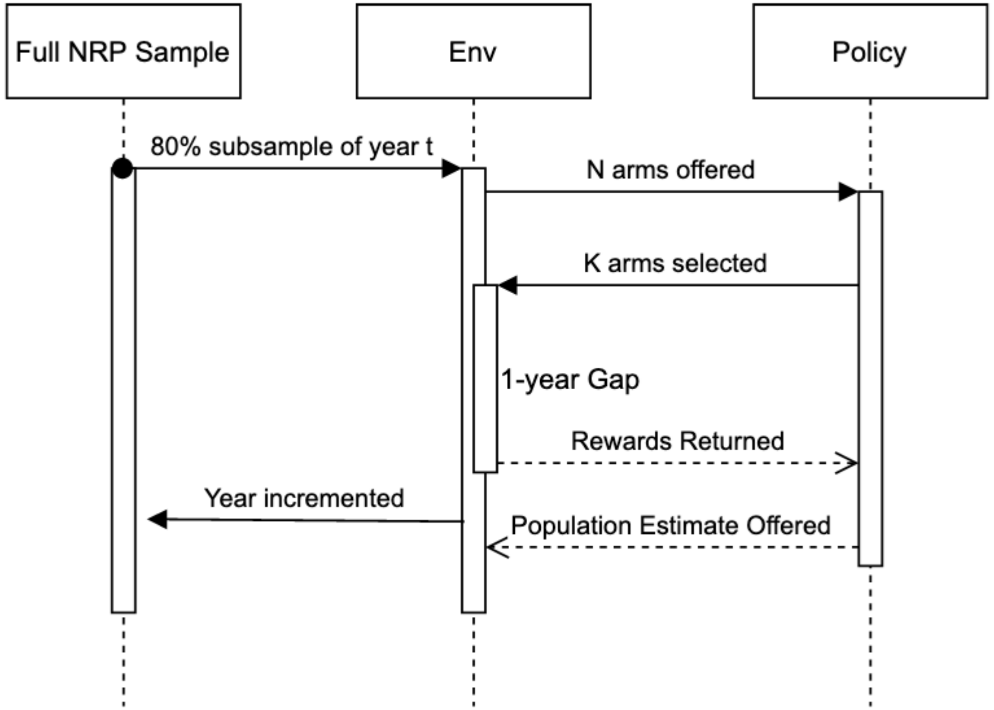
1. Take 80% subsample
2. Give selection policy ~500 covariates from tax return data for each “arm” (tax return) in the sample
3. Selection policy returns arms to audit
4. Simulate a 1 year gap
5. Return the tax adjustment (reward) after that gap



Experiments

For NRP data years 2006-2014

1. Take 80% subsample
2. Give selection policy ~500 covariates from tax return data for each “arm” (tax return) in the sample
3. Selection policy returns arms to audit
4. Simulate a 1 year gap
5. Return the tax adjustment (reward) after that gap
6. Policy makes population estimate



Experiments

Best Reward Settings

	Policy	R	μ_{PE}	σ_{PE}	μ_{NR}
Unbiased Methods	ABS-1	**$41.5M***	**0.4** ✓	31.0	**37.6%**
	ϵ-only	$41.3M*	4.3✓	37.4	38.3%
	ABS-2	$40.5M*	0.6✓	24.5	38.3%
	Random	$12.7M	1.5✓	**14.7**	53.1%

Experiments

10% (ε) random sample,
rest greedy

Best Reward Settings

	Policy	R	μ_{PE}	σ_{PE}	μ_{NR}
Unbiased Methods	ABS-1	**$41.5M***	**0.4** ✓	31.0	**37.6%**
	ϵ-only	$41.3M*	4.3✓	37.4	38.3%
	ABS-2	$40.5M*	0.6✓	24.5	38.3%
	Random	$12.7M	1.5✓	**14.7**	53.1%

Experiments

Best Reward Settings

	Policy	R	μ_{PE}	σ_{PE}	μ_{NR}
Unbiased Methods	ABS-1	**$41.5M***	**0.4** ✓	31.0	**37.6%**
	ϵ-only	$41.3M*	4.3✓	37.4	38.3%
	ABS-2	$40.5M*	0.6✓	24.5	38.3%
	Random	$12.7M	1.5✓	**14.7**	53.1%

Fully random sample every year, rest greedy

Experiments

ABS can yield lower variance, similar reward, lower no-change rate, and retain unbiasedness

Best Reward Settings

	Policy	R	μ_{PE}	σ_{PE}	μ_{NR}
Unbiased Methods	ABS-1	**$41.5M***	**0.4** ✓	31.0	**37.6%**
	ϵ-only	$41.3M*	4.3✓	37.4	38.3%
	ABS-2	$40.5M*	0.6✓	24.5	38.3%
	Random	$12.7M	1.5✓	**14.7**	53.1%

Experiments

Greedy tends to perform well in highly stochastic low-sample regime (which matches our experimental setup). (Bastani et al., 2022 proved this recently.)

Best Reward Settings

	Policy	R	μ_{PE}	σ_{PE}	μ_{NR}
Biased Methods	Greedy	**\$43.6M***	16.4 ✗	8.8	**36.5%**
	UCB-1	\$42.4M*	15.3 ✗	9.4	38.6%
	ϵ-Greedy	\$41.3M*	**6.1** ✗	**7.5**	38.3%
	UCB-2	\$40.7M*	15.6 ✗	10.21	40.7%

Stanford

Experiments

Best Reward Settings

	Policy	R	μ_{PE}	σ_{PE}	μ_{NR}
Biased Methods	Greedy	**$43.6M***	16.4 ✗	8.8	**36.5%**
	UCB-1	$42.4M*	15.3 ✗	9.4	38.6%
	ϵ-Greedy	$41.3M*	**6.1** ✗	**7.5**	38.3%
	UCB-2	$40.7M*	15.6 ✗	10.21	40.7%

Use regression model for both selection and population estimate. Means biased prediction, but slightly more reward and lower variance

Experiments

Best Reward Settings

	Policy	R	μ_{PE}	σ_{PE}	μ_{NR}
Biased Methods	Greedy	**$43.6M***	16.4 ✗	8.8	**36.5%**
	UCB-1	$42.4M*	15.3 ✗	9.4	38.6%
	ϵ-Greedy	$41.3M*	**6.1** ✗	**7.5**	38.3%
	UCB-2	$40.7M*	15.6 ✗	10.21	40.7%

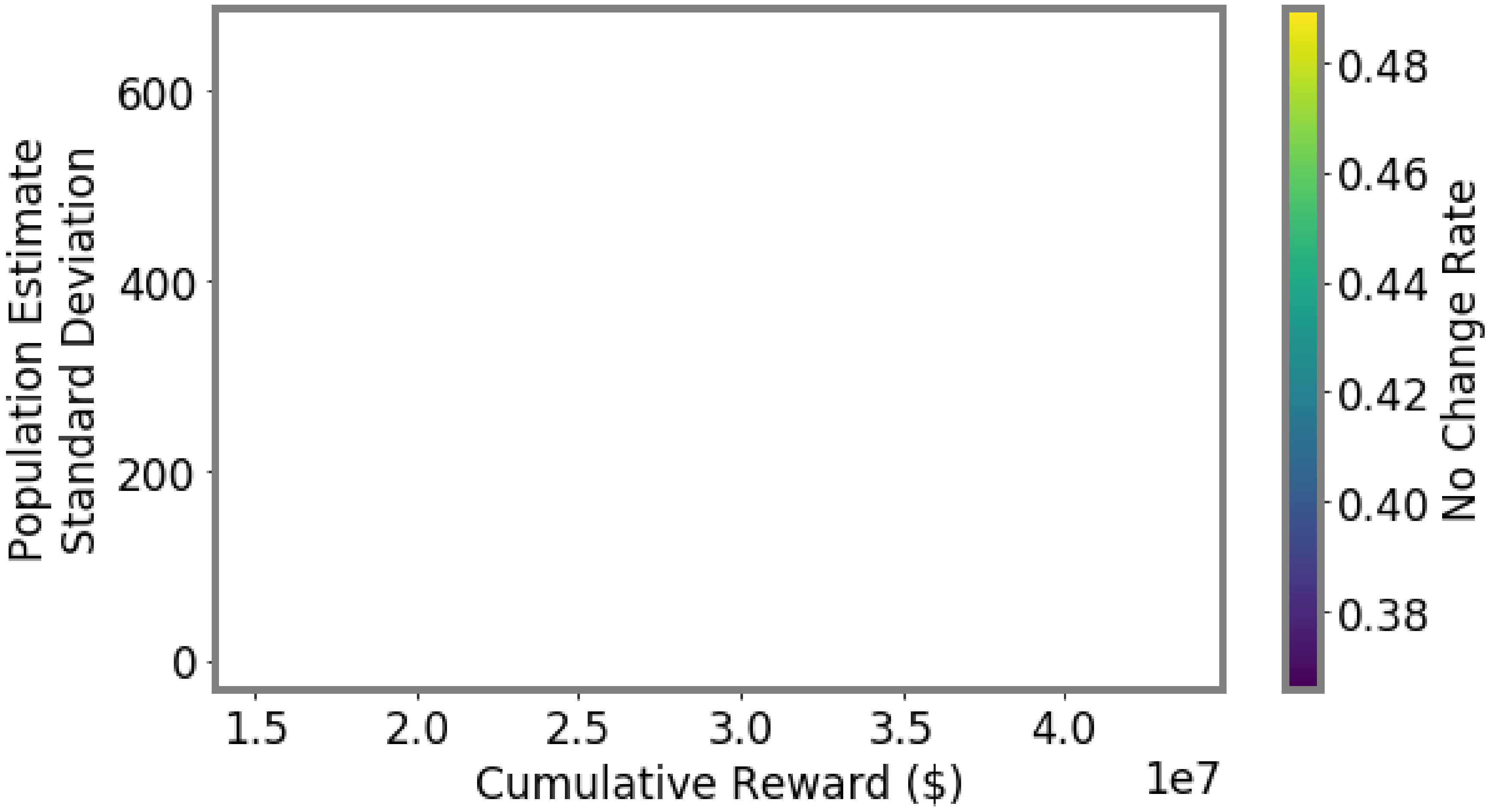
Even some randomness,
reduces bias of model-based estimate,
but not guaranteed.

Experiments

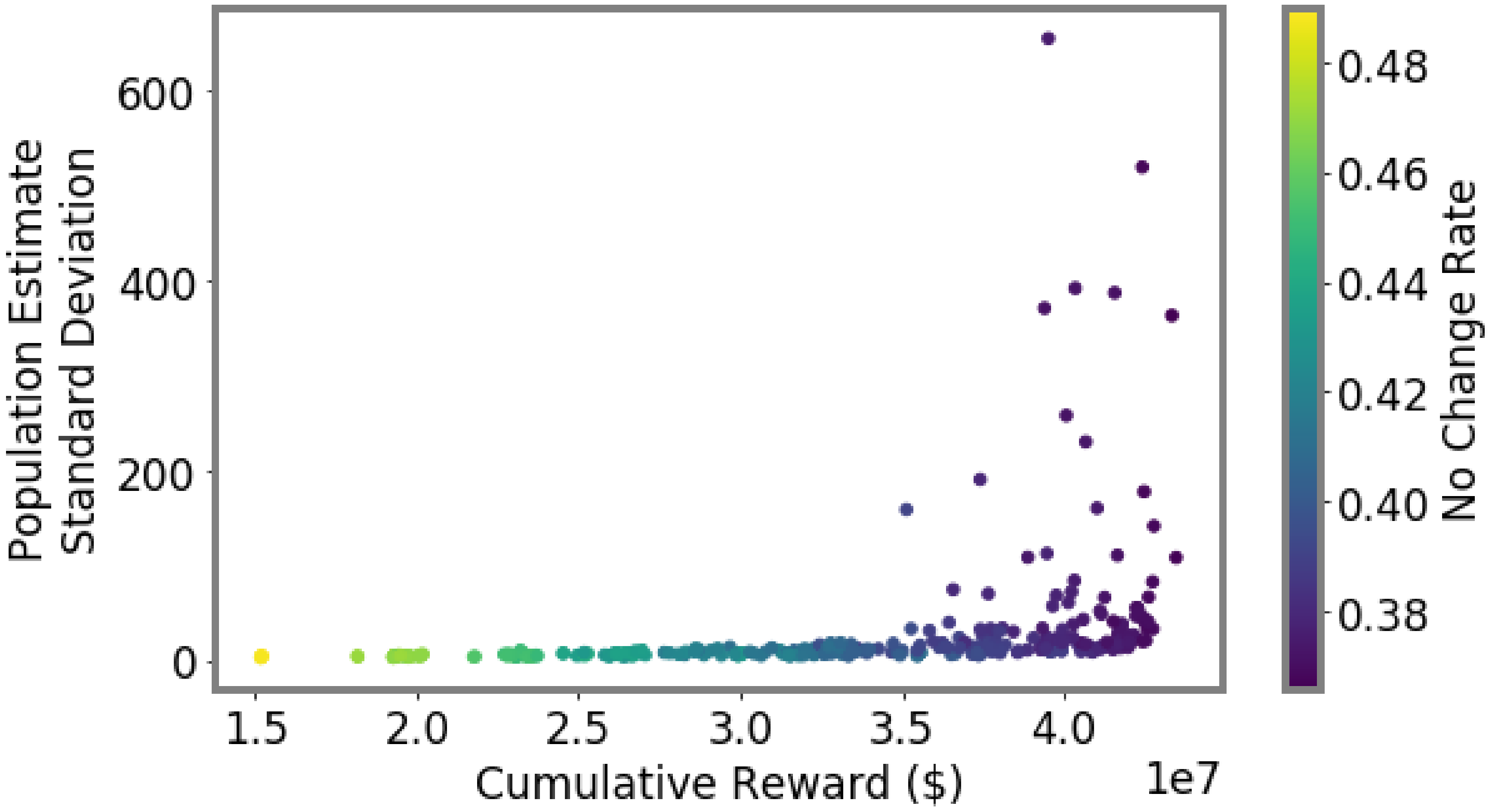
Best Reward Settings

	Policy	R	μ_{PE}	σ_{PE}	μ_{NR}
Unbiased Methods	ABS-1	**$41.5M***	**0.4** ✓	31.0	**37.6%**
	ϵ-only	$41.3M*	4.3✓	37.4	38.3%
	ABS-2	$40.5M*	0.6✓	24.5	38.3%
	Random	$12.7M	1.5✓	**14.7**	53.1%
Biased Methods	Greedy	**$43.6M***	16.4 ✗	8.8	**36.5%**
	UCB-1	$42.4M*	15.3 ✗	9.4	38.6%
	ϵ-Greedy	$41.3M*	**6.1** ✗	**7.5**	38.3%
	UCB-2	$40.7M*	15.6 ✗	10.21	40.7%

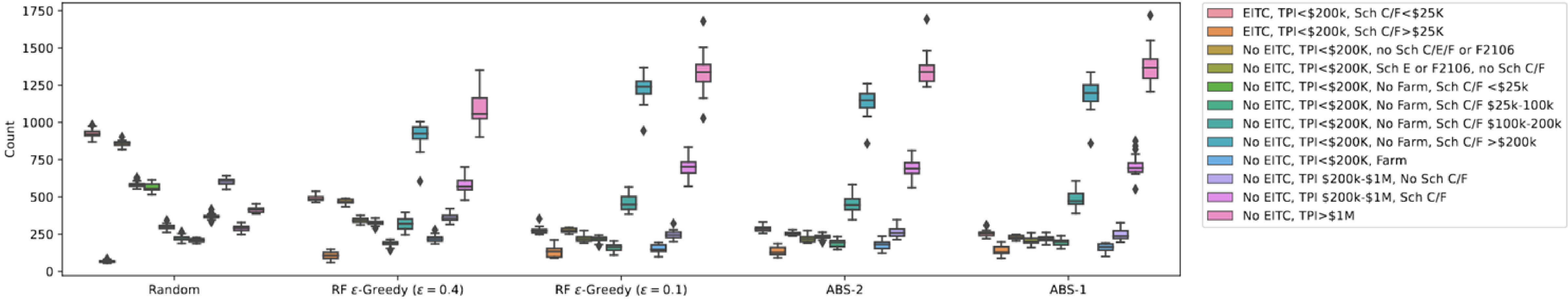
ABS Enables Formal Tradeoff Between Precision and Reward



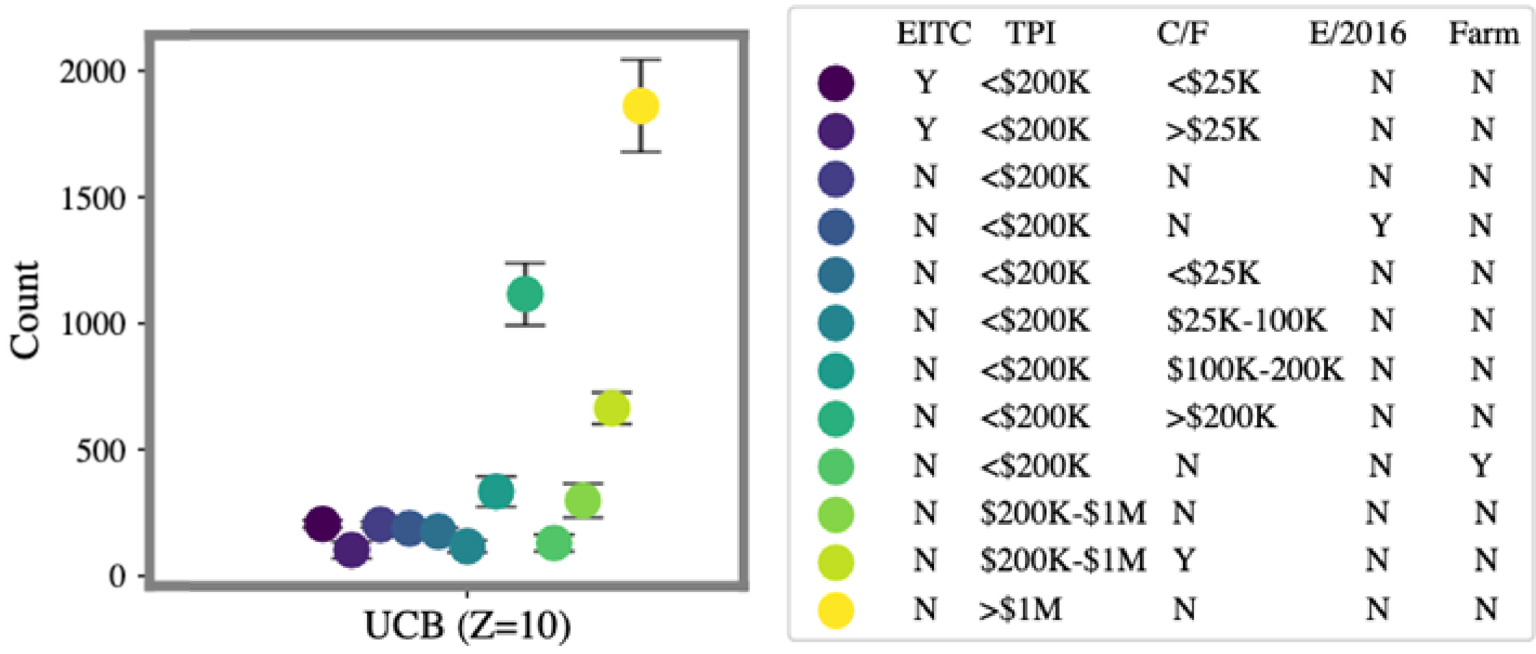
ABS Enables Formal Tradeoff Between Precision and Reward



More optimal methods sample higher incomes



But heteroskedasticity can also drive sampling higher incomes



Takeaways

1. Unbiased estimation of population (e.g., average misreporting) **can still yield returns almost as high as greedy selection, with careful sampling and HT estimation**.
    a. Suggests that a **unified optimize-and-estimate program** could be better and be more efficiently optimized.
2. Model-based population mechanisms are not guaranteed to be unbiased, but bias in practice can be reduced with some randomness.
3. More optimal methods tend to sample higher incomes in our experiments.
4. But heteroskedasticity also drives sampling of higher-incomes in uncertainty-based methods.

June 16, 2022

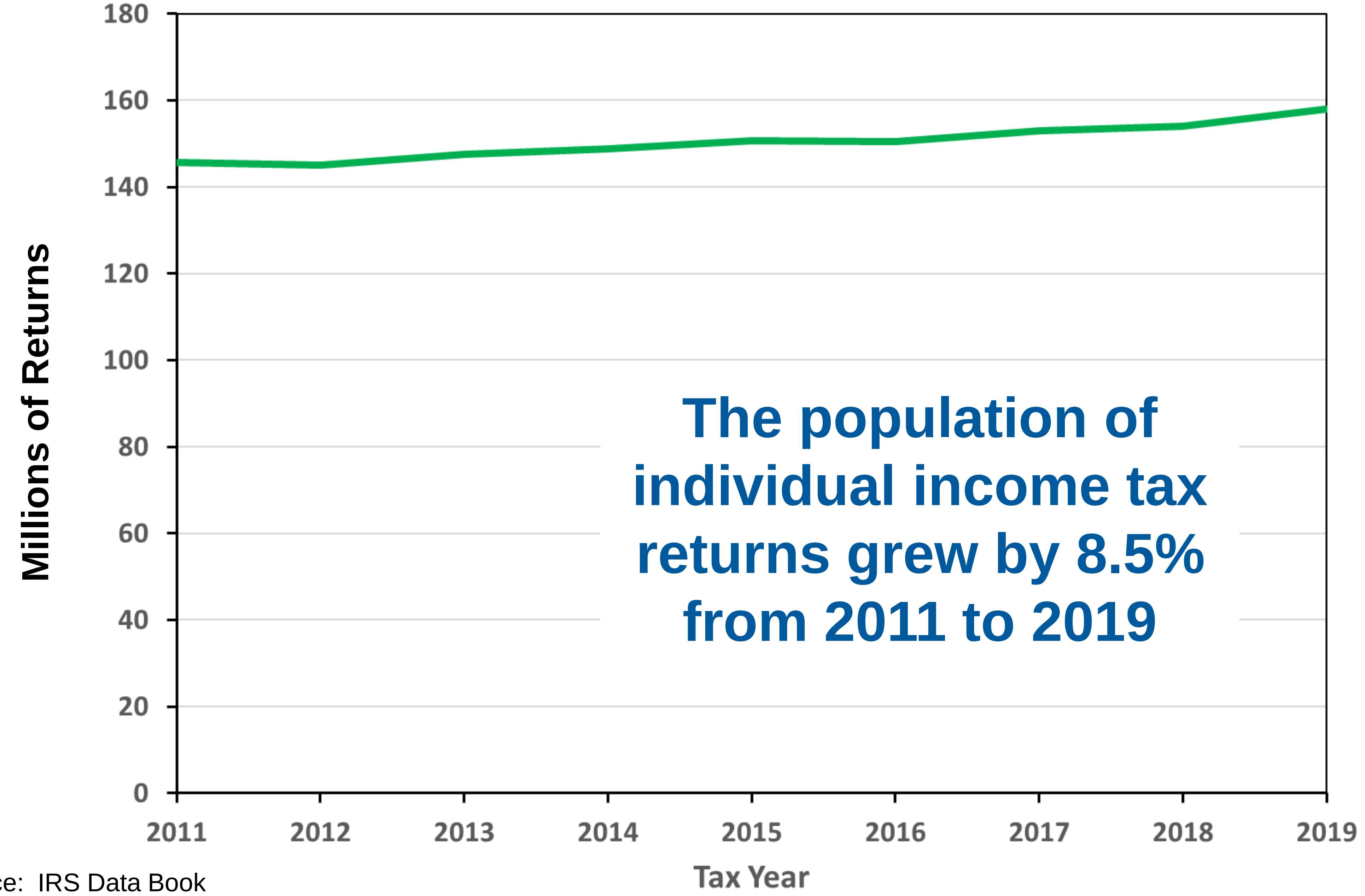
Balancing Audits: Enforcement vs. Measuring Non-Compliance

IRS-TPC Research Conference

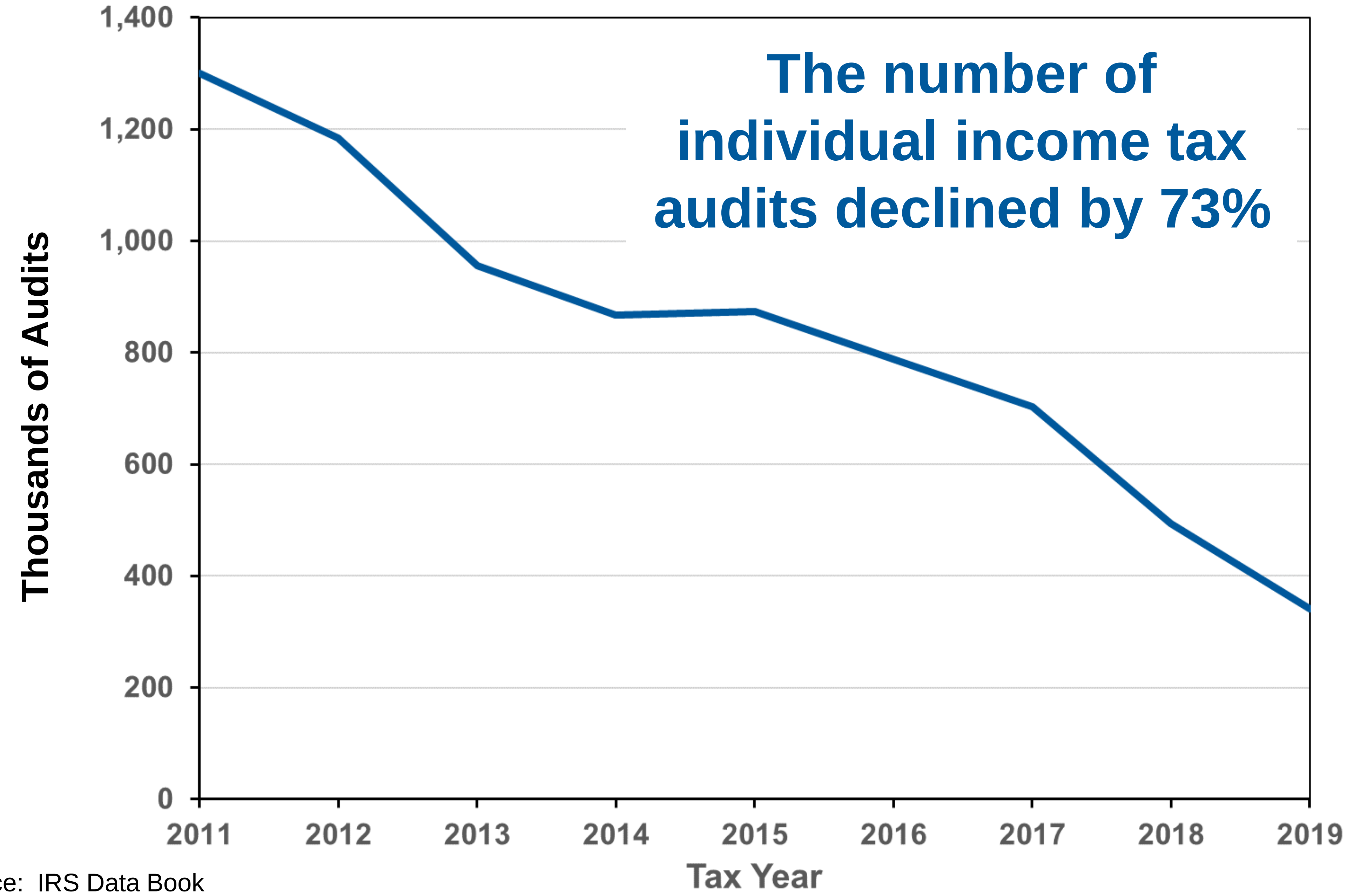
Alan Plumley (IRS, RAAS), Discussant

Caveat: My views and comments do not necessarily reflect the position of the IRS.

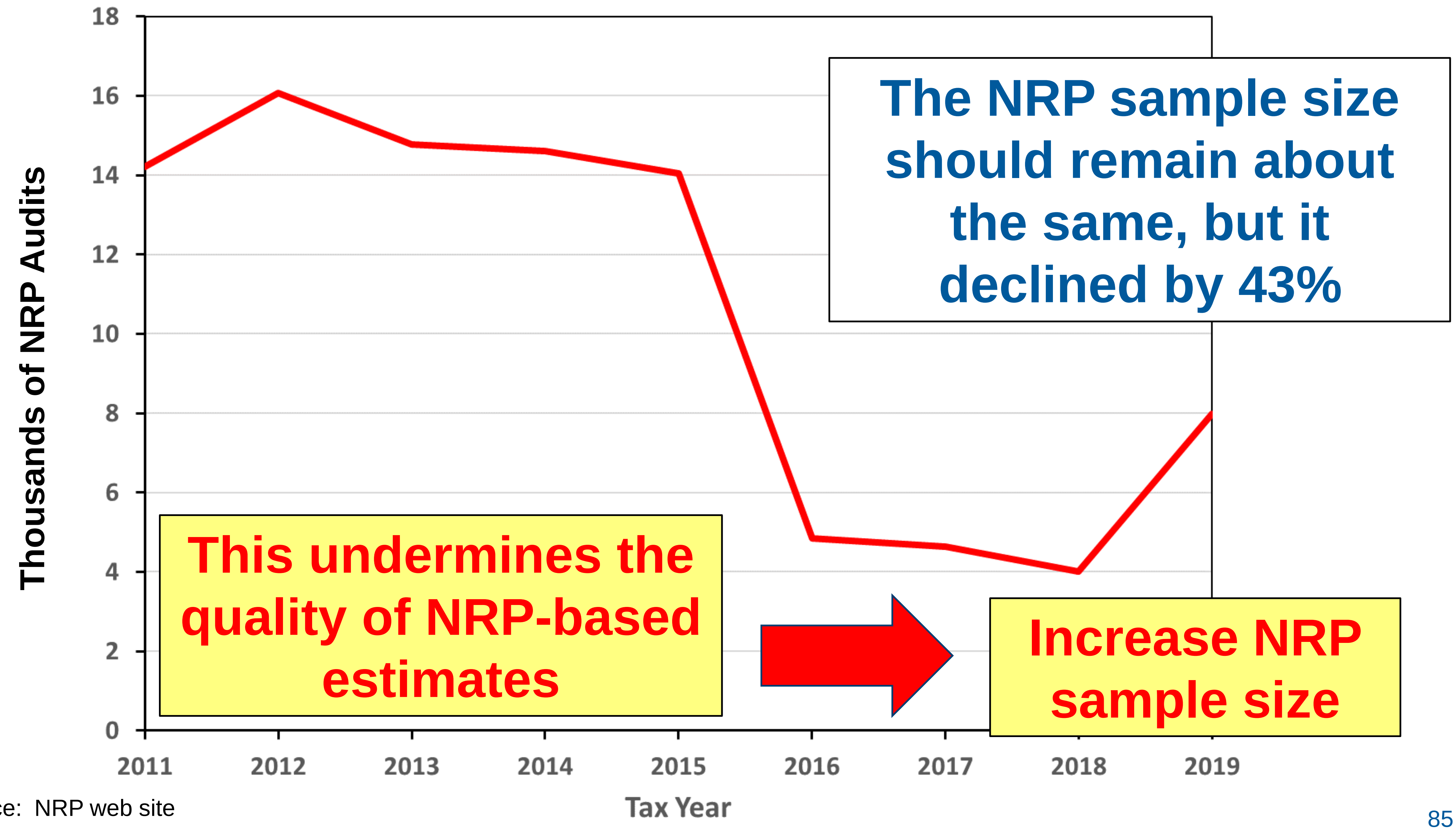
Context



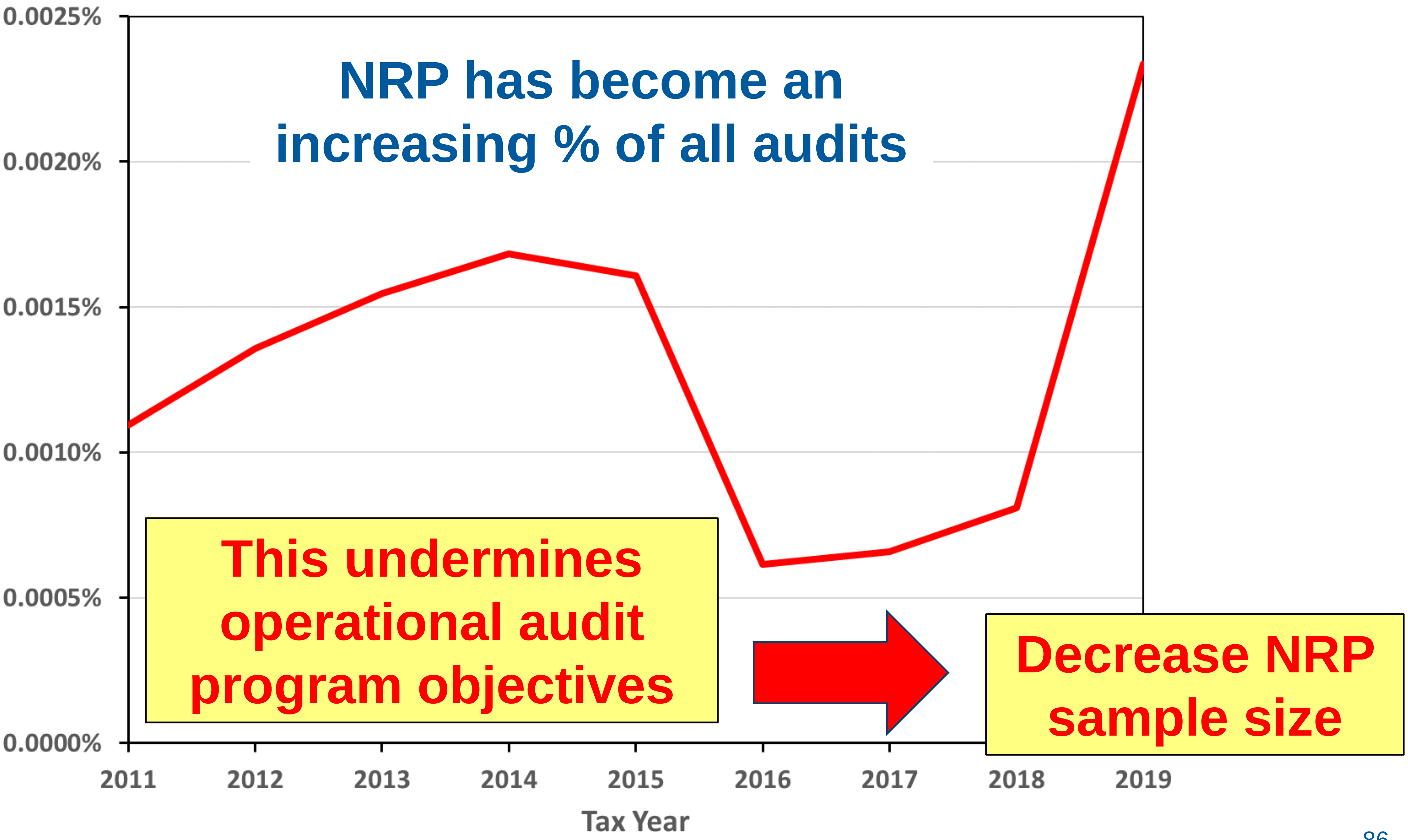
Source: IRS Data Book



Source: IRS Data Book



Source: NRP web site


NRP has become an increasing % of all audits
This undermines operational audit program objectives
Decrease NRP sample size
0.0025%
0.0020%
0.0015%
0.0010%
0.0005%
0.0000%
2011
2012
2013
2014
2015
2016
2017
2018
2019
Tax Year

Can we get the best of both worlds by using operational audits to mitigate the loss of some NRP audits?

Approach
Approach 1: Use OP audits selected for OP reasons

Can we get the best of both worlds by using operational audits to mitigate the loss of some NRP audits?

Approach
Approach 1: Use OP audits selected for OP reasons
Approach 2: Select OP & random audits to meet revenue & measurement objectives simultaneously

Can we get the best of both worlds by using operational audits to mitigate the loss of some NRP audits?

Approach	Papers
Approach 1: Use OP audits selected for OP reasons	Turk, *et al.* Rizzo, *et al.*
Approach 2: Select OP & random audits to meet revenue & measurement objectives simultaneously	Henderson, *et al.*

Can we get the best of both worlds by using operational audits to mitigate the loss of some NRP audits?

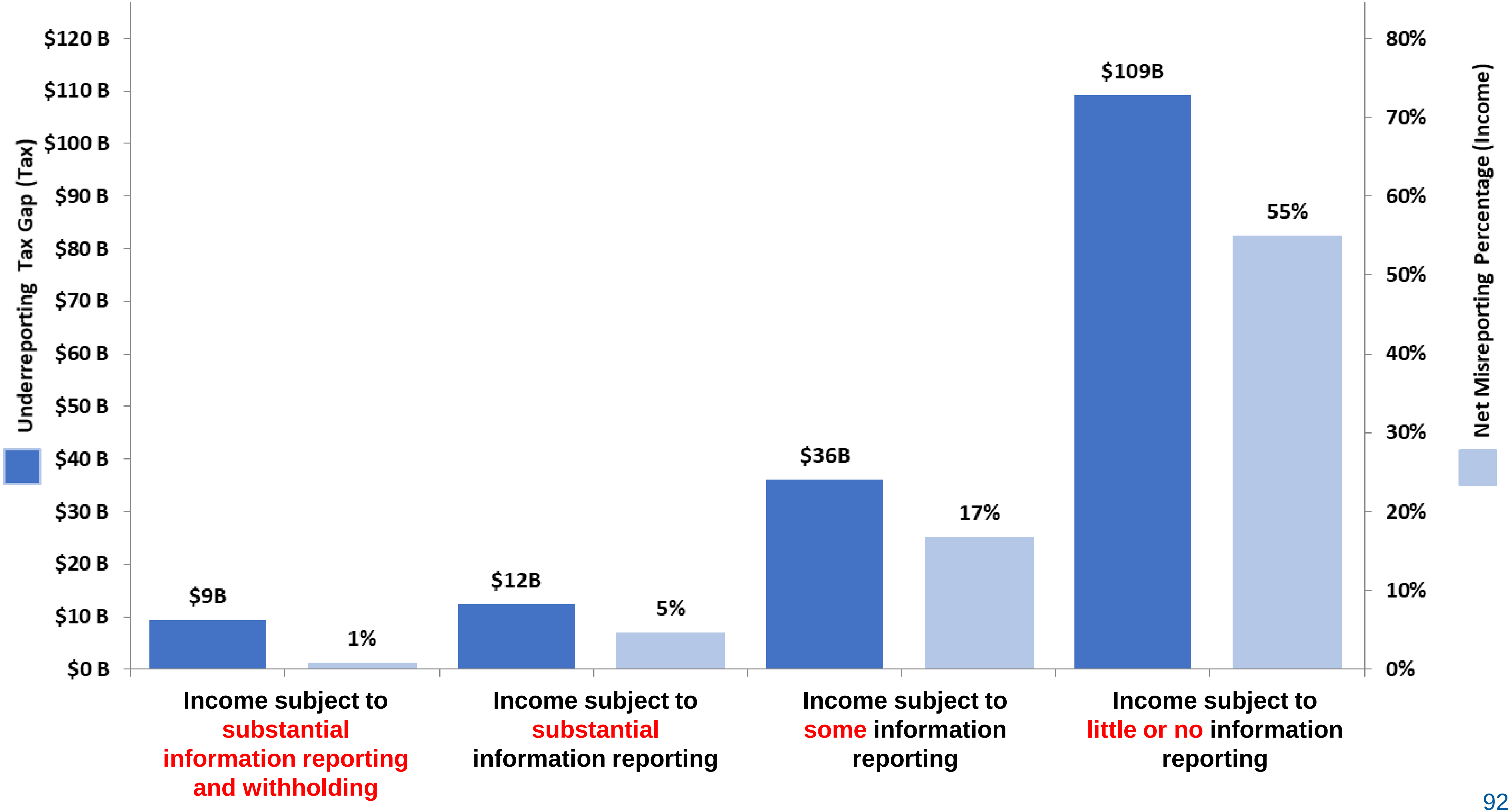
Approach	Papers	Advantage
Approach 1: Use OP audits selected for OP reasons	**Turk, *et al.*** **Rizzo, *et al.***	No change to OP program
Approach 2: Select OP & random audits to meet revenue & measurement objectives simultaneously	**Henderson, *et al.***	Potentially a better balancing of objectives

A Major Concern About All 3 Papers

Presumption: We need unbiased estimates of the gross underreporting tax gap.

Reality: We need unbiased compliance estimates at the line-item level for each year.

Compliance Varies Across Line Items (TYs 2011-13)



A Major Concern About All 3 Papers

Presumption: We need unbiased estimates of the gross underreporting tax gap.

Reality: We need unbiased compliance estimates at the line-item level for each year.

Problem: OP data at the line-item level are very poor. It may be cheaper to increase the number of random audits than to fix operational line-item data.

“Improving Risk Models by Supplementing Random NRP Audits with Non-Random OP Audits Using Statistical Controls for Bias”
Turk, *et al.*

- **Selection Bias Correction:** You account for DIF and DIF^2. You might rather consider using the percentile from the distribution of DIF scores in each Activity Code.
- **Risk Model:** You draw randomly from OP audits and assign a normalized weight of 1 to each. You might rather stratify the OP audit sample using the NRP strata within each Activity Code and normalize the weights the same as for NRP audits.
- **Mix of OP & NRP Audits:**
  - I don’t think there’s a need to predict OP outcomes using NRP or mixed data.
  - Have you tried many independent 50% and 25% samples to construct confidence intervals around your estimates?
  - Is there a way to transform OP data to be fully compatible with NRP data (weights and outcomes) so population compliance stats can be tabulated directly from the mixed data, rather than from a model?
  - Have you tried this for other activity codes?

"Augmenting NRP Tax Change Estimates by Incorporating OP Audit Information: A New RAAS Research Initiative"
Rizzo, *et al.*

- Good that you account for **OP cases that were surveyed**, but imputing tax change only to some of them may be problematic. Given the difference in exam intensity between NRP and OP exams, perhaps positive tax changes need to be adjusted, as well.
- I question the **weights of 1** for the OP audits. They don't all represent the population equally, even after you adjust the NRP weights for them. More needs to be done to control for selection bias.
- Not clear how this approach could be applied **if the NRP sample were much smaller**.
- Using DIF scores to help overcome the selection bias in the OP cases depends on having **a robust DIF model for each stratum**, but that's not guaranteed if the NRP sample continues to shrink over time.

“Integrating Reward Maximization and Population Estimation: Sequential Decision-Making for IRS Audit Selection” Henderson, *et al.*

- **Reward Maximization Objective:** Need to account for cost.
  - Total Reward is subject to a budget constraint and (unlike one-arm bandits) audits vary widely in their cost.
  - Selection should be based on expected **revenue/cost ratio**.
- **No-change rate** is not nearly as bad for taxpayers in random audits as in risk-based audits.
  - Told up front that they were selected at random in order for IRS to maintain its operational selection methods, and
  - Told that their audit could be quite educational for them and even result in a refund.
- Would have liked to have seen a comparison of the metrics between these alternatives and the **status quo** (where the two objectives are separated).
- DIF stands simply for **Discriminant Function.**

Session 2:
Burden vs. Opportunity

The Spiderweb of Partnership Tax Planning

EMILY BLACK
JACOB GOLDIN
RYAN HESS
DANIEL HO
REBECCA LESTER
MANSHEEJ PAUL
ANNETTE PORTZ

Research Question

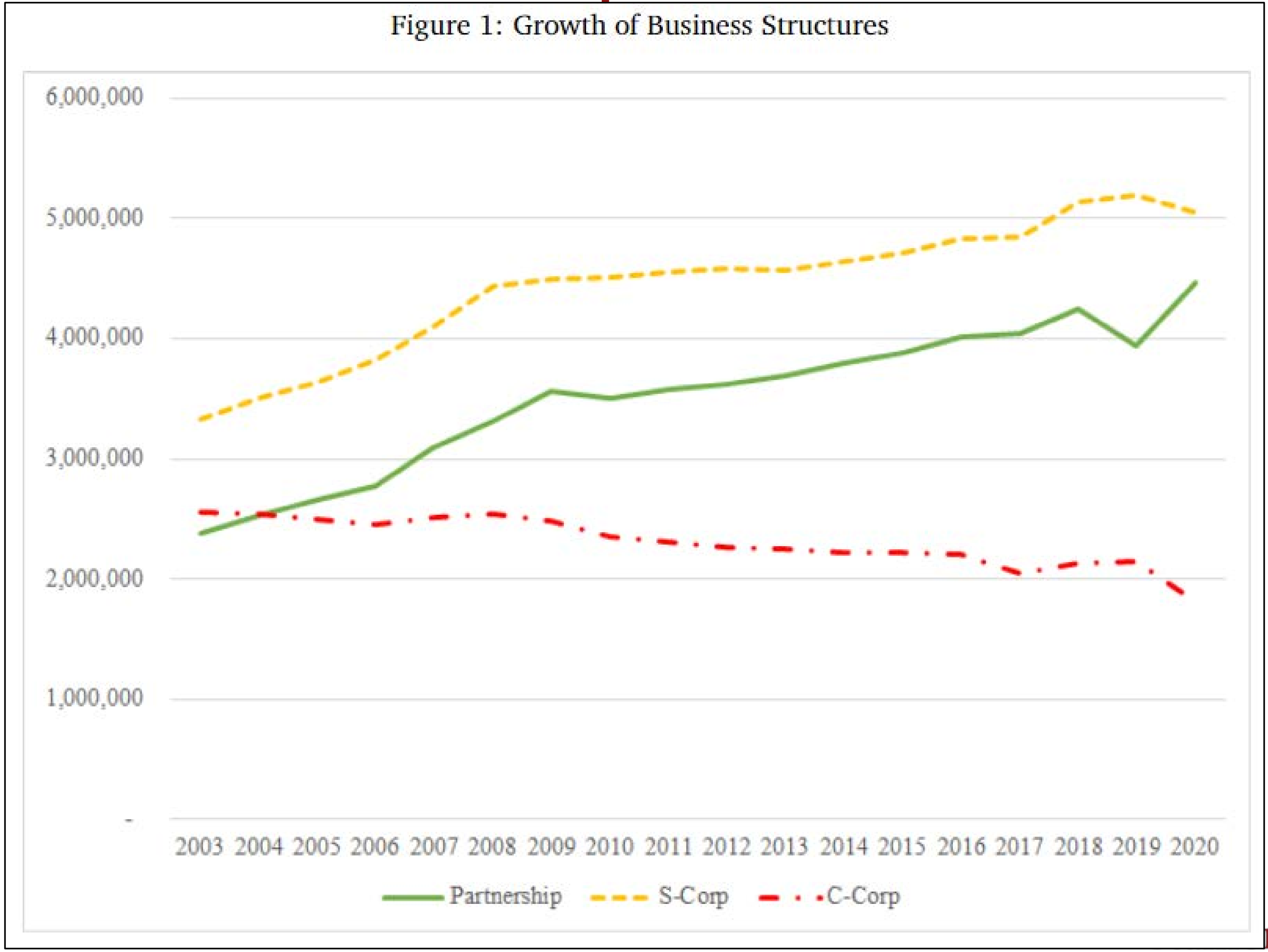
Examine the tax planning choices of pass-through businesses

1. Construct partnership organization structures
   - Examine the size, ownership, and complexity of partnership structures
2. Test partnership features associated with tax planning
   - Utilize audit outcomes to establish tax planning
   - Leverage machine learning models

Motivation

1. Partnership entities are prevalent and economically significant
    - Approximately 4.5 M partnership returns filed in 2020
    - Prior literatures focuses on C corporations
2. Low tax rate on tax on partnership income
    - Estimated tax rates of 15.9 percent (Cooper et al. 2016)
    - Substantial portion of income is unreported

Motivation: Partnership Growth



Motivation: Tax Gap

Total Tax Gap

Total True Tax Liability	Tax Paid Voluntarily & Timely	Gross Tax Gap: Nonfiling	Underreporting	Underpayment	Gross Tax Gap	Enforced & Other Late Payments	Net Tax Gap (Tax Not Collected)
$2,683	$2,242	$39	+ $352	+ $50	= $441	− $60	= $381

Corporations

Corporation Income Tax	Corporation Income Tax	Corporation Income Tax	Corporation Income Tax	Corporation Income Tax	Corporation Income Tax	Corporation Income Tax	Corporation Income Tax
$294	$251	#	+ $37	+ $5	= $42	− $10 (24%)	= $32

Individuals

Individual Income Tax	Individual Income Tax	Individual Income Tax	Individual Income Tax	Individual Income Tax	Individual Income Tax	Individual Income Tax	Individual Income Tax
$1,398	$1,084	$31	+ $245	+ $38	= $314	− $43 (14%)	= $271

Business Income	Non-Business Income	Credits	Income Offsets [1]	Filing Status	Other Taxes [2]	Unallocated Marginal Effects [3]
$110	$57	$42	$20	$5	$1	$10

Literature

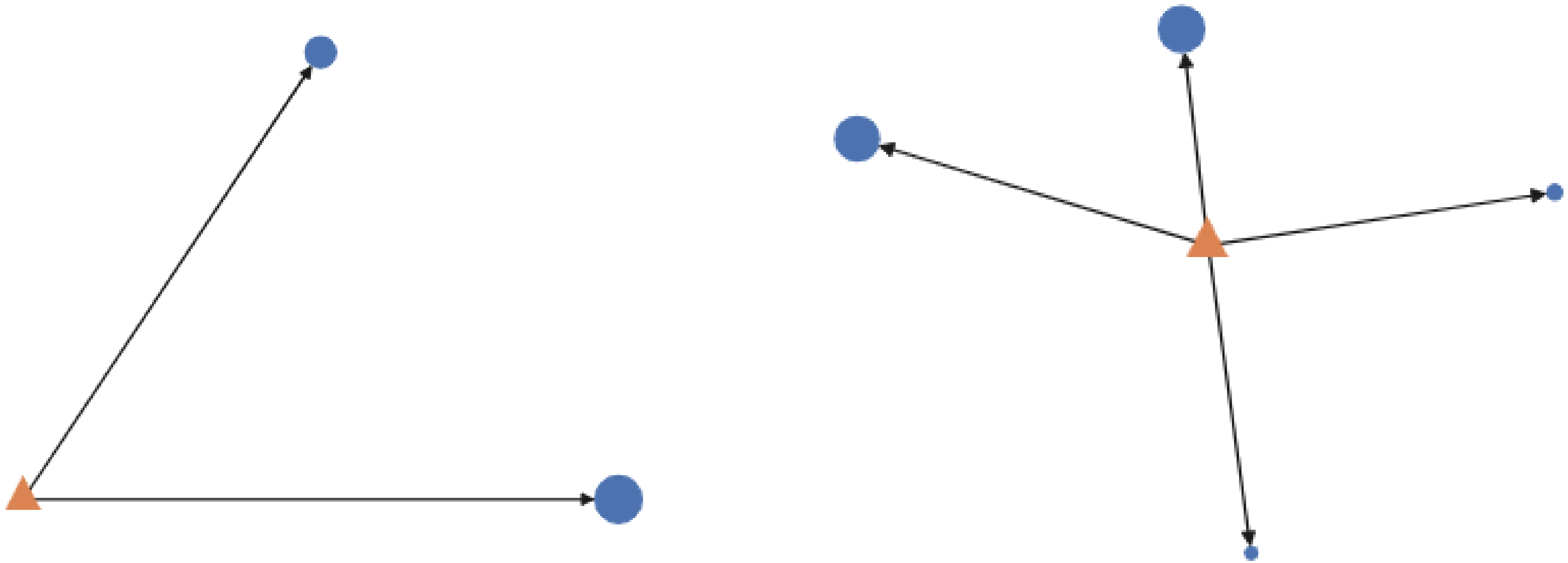
- Partnership structures are opaque, owned by high earners, and have a relatively low federal tax rate (Cooper et al. 2016)
- Significant portion of partnership income flows to a foreign owner (Love 2021)
- Partnership tax evasion is understated (Guyton et al. 2021)

- Corporate structures with embedded partnerships report lower effective tax rates (Agarwal et al. 2021)

Data

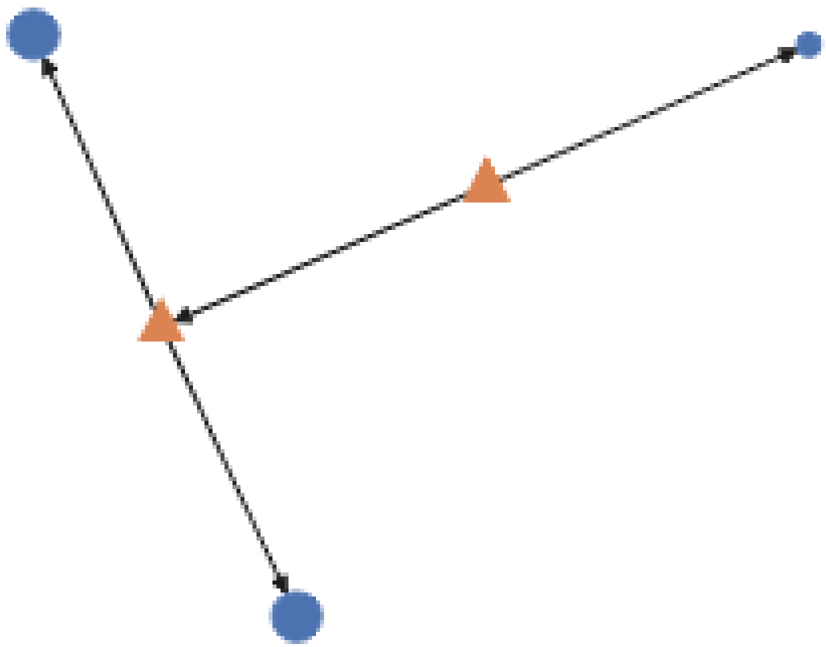
- Identify 3.7 M partnership entities (TY 2014)
  - Represents near universe of partnerships
  - Require non-zero income amounts to create network structures
  - Create network structures and classify organizations:
    - Simple: single partnership wholly owned by individual taxpayers
    - Complex: all other organization structures

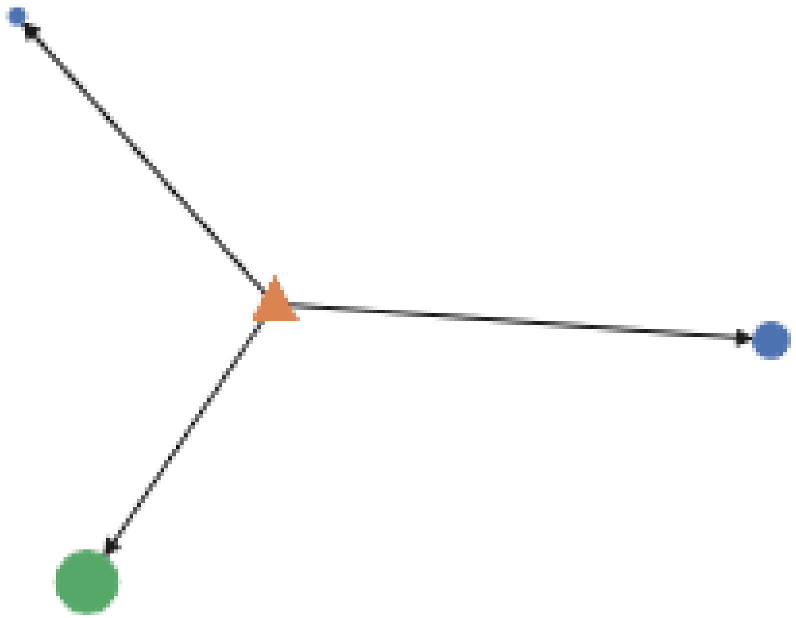
Simple Partnership Organizations



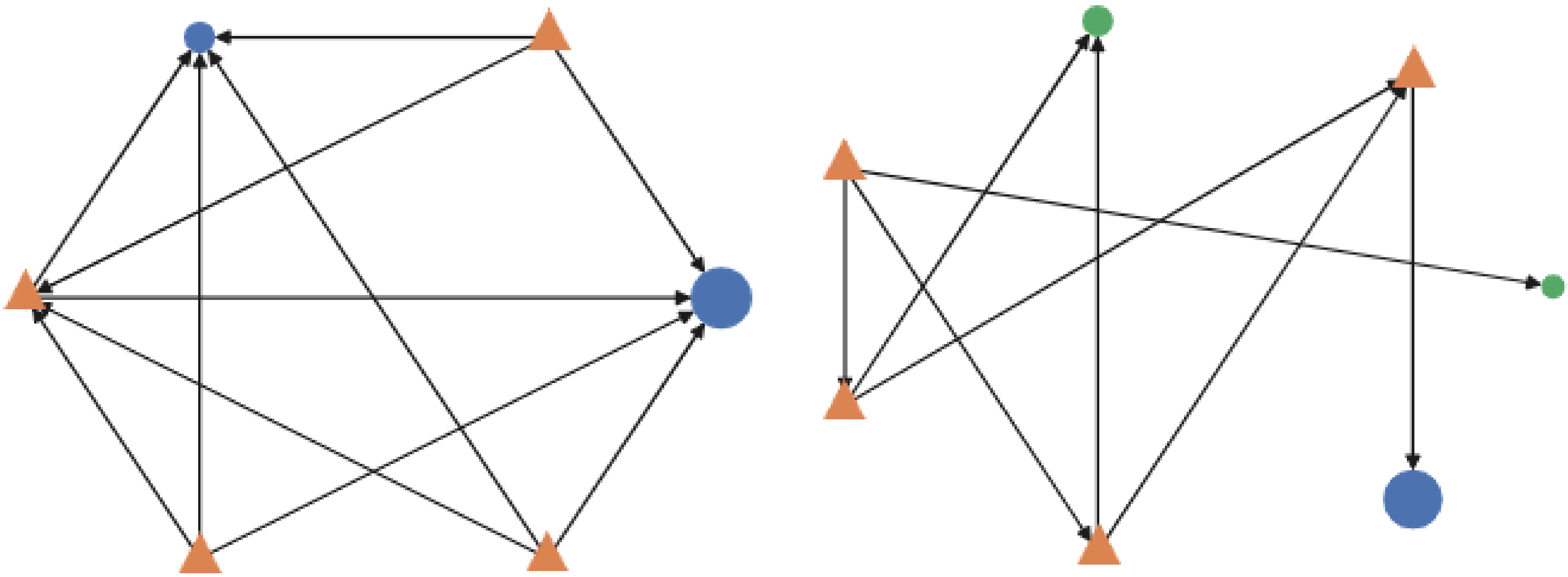
Stanford University

Complex Partnership Organizations

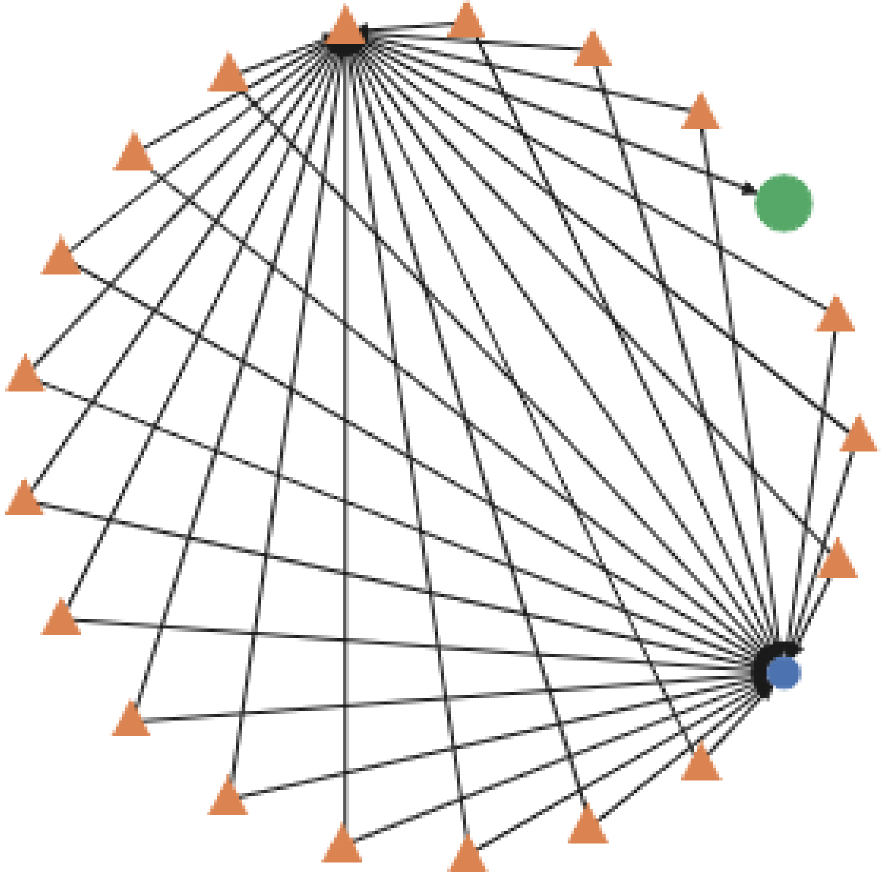


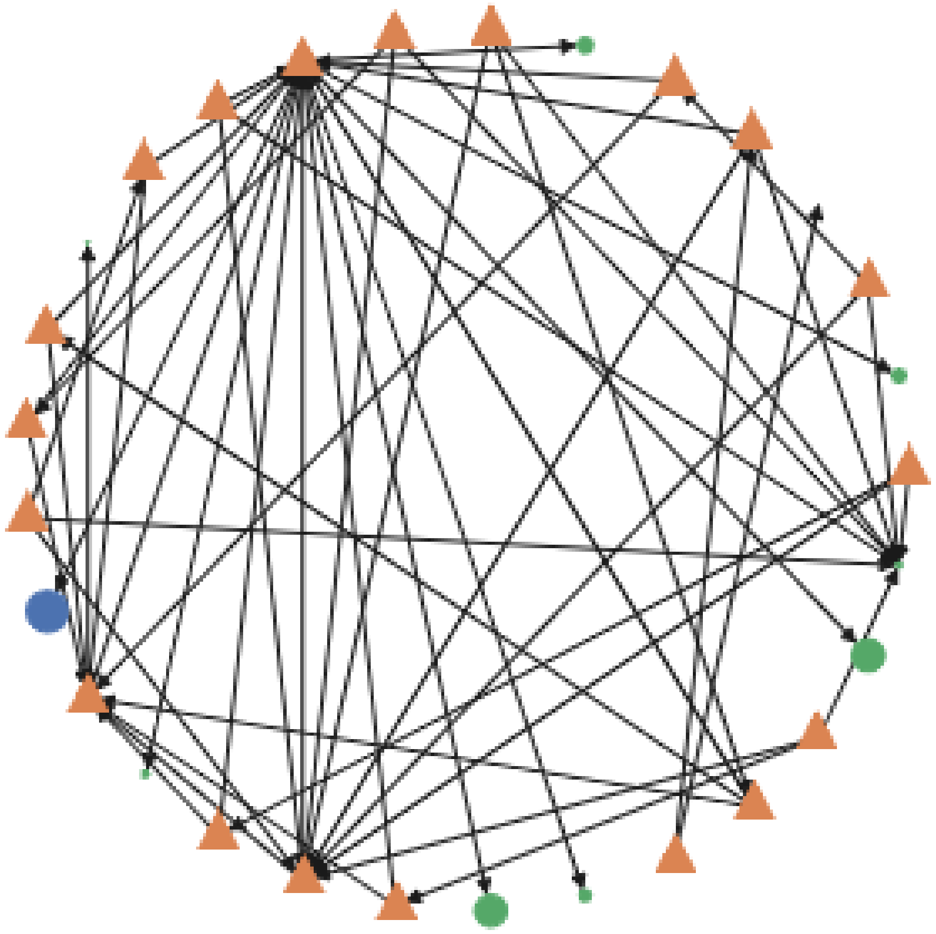


Complex Partnership Organizations



Complex Partnership Organizations





Sample of Partnerships

Simple Partnership Organizations		Complex Partnership Organizations	
Number of Partners	Number of organizations	Number of Partners	Number of organizations
2	1,561,764	2	163,362
3	310,830	3	97,029
4	160,760	4	60,959
5	61,352	5	39,420
6	31,468	6	26,774
7	16,359	7	17,553
8	11,029	8	13,208
9	7,097	9	9,765
10	5,250	10	7,635
11-20	15,755	11-20	31,683
21-30	2,312	21-30	8,223
31-40	603	31-40	3,362
41-50	233	41-50	1,577
51+	419	51+	2,920

Summary Statistics

Number of Partners	Simple Organizations: Total Income – Profitable Partnerships	Simple Organizations: Total Income – Loss Partnerships	Complex Organizations: Total Income – Profitable Partnerships	Complex Organizations: Total Income – Loss Partnerships
All Partnerships	223,690	(44,429)	966,068	(408,104)
2	169,355	(38,168)	1,104,097	(417,711)
3	121,278	(48,968)	556,429	(351,565)
4	892,437	(53,775)	532,041	(409,537)
5	178,771	(67,432)	650,992	(340,395)
6	222,973	(73,265)	665,782	(330,904)
7	222,783	(83,440)	746,573	(415,977)
8	235,887	(83,616)	921,331	(337,868)
9	267,031	(84,277)	675,799	(432,768)
10	315,403	(113,646)	1,170,100	(430,847)
11-20	494,016	(103,438)	1,301,334	(492,880)
21-30	965,977	(122,317)	1,763,281	(656,693)
31-40	2,102,284	(214,254)	2,838,317	(743,120)
41-50	4,315,869	(197,764)	2,734,750	(1,035,758)
51+	16,492,848	(486,452)	10,548,677	(1,258,778)

Empirical Design

- Examine subsample of partnerships that have undergone an audit
- Test accuracy with which we can predict whether there was an audit adjustment imposed

- Utilize ridge regression and random forest models

Model Features

		Audited Partnerships	
		Mean	SD
Partnership	*SALES*	18,107,154.990	449,590,642.500
	FOREIGN TAX	20,027.158	996,730.161
	DEPRECIATION	442,004.207	12,748,818.430
	INTEREST EXP	156,416.542	3,890,482.910
	LOSS	0.506	0.500
Audit	*AGENT RANK*	10.081	3.149
	AUDIT TIME	0.223	3.612
Network	*IN DEG*	0.147	0.778
	OUT DEG	2.625	3.037
	PARTNERS	4.307	30.822
	PARTNERSHIPS	1.858	4.400
	S CORPS	0.149	0.869
	TRUSTS	0.155	1.371
	N	12,627	12,627

Empirical Results

	Ridge ADJUST [0/1]t	Ridge ADJUST [0/1]t	Random Forest ADJUST [0/1]t	Random Forest ADJUST [0/1]t
SALES	(0.057)	(0.061)	✓	✓
FOREIGN TAX	(0.066)	(0.068)	✓	✓
DEPRECIATION	0.041	0.034	✓	✓
INTEREST EXP	(0.016)	(0.012)	✓	✓
LOSS	0.013	0.010	✓	✓
AGENT RANK	1.642	1.651	✓	✓
AUDIT TIME	(2.040)	(2.014)	✓	✓
IN DEG		0.098		✓
OUT DEG		(0.183)		✓
PARTNERS		0.059		✓
PARTNERSHIPS		(0.064)		✓
Training set observations	9,470	9,470	9,470	9,470
Test set observations	3,157	3,157	3,157	3,157
Area under ROC	0.7361	0.7472	0.7388	0.7889
Generalization error	0.0085	0.0061	0.1876	0.1925
Model accuracy	0.7260	0.7092	0.6579	0.7415

Conclusion and Next Steps

- We contribute to our understanding of partnership organization structure and complexity
- Preliminary evidence that network structures are beneficial in effectively identifying partnerships engaging in tax planning
  - Relationship between structures and tax planning exhibits non-linear relationships
- Expand the sample period and scope of features examined
- Deep Neural Networks

Automatic Tax Filing:
Simulating a Pre-Populated Form 1040

Lucas Goodman* Katherine Lim+
Bruce Sacerdote† Andrew Whitten*

* Office of Tax Analysis, U.S. Dept. of Treasury
+ Federal Reserve Bank of Minneapolis
† Dartmouth College

June 2022

Disclaimer: The views expressed in this presentation are those of the authors and do not necessarily reflect the views of the U.S. Department of the Treasury, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

MOTIVATION

- Tax filing is burdensome
  - Compliance costs estimated to be 10% of revenue or $350 billion (Guyton et al. (2008), Fichtner, Gale, and Trinca (2019))
- Filing burden is <u>not</u> proportionate to income
  - Relative time and out of pocket cost is higher for lower AGI taxpayers (Marcuss et al. 2013)
- IRS already knows (some of) our income
  - Information reporting by third parties (e.g. wages, interest/dividends, income from S corps/partnerships, retirement distributions, etc.)

POTENTIAL REMEDY?

IRS fills out your tax return for you

- May reduce filing burden and distribute it more progressively
- Could boost EITC & CTC take-up
  - EITC take-up is 75% (Plueger 2009)
  - 91% of non-take-up is due to non-filing (Plueger 2009)
  - IRS outreach targets filers (Manoli and Turner, 2016)
- Could prompt filling among those owed refunds or delinquent non-filers

This project

- Imagine IRS prepares a tax return using only:
  - Information returns for the year
  - Filing status and dependents from the prior year
- Quantify the share of filing taxpayers for whom this procedure is “successful”
- Provide concrete results on both filers and non-filers for policy maker consideration

Does not:

- Exhaustively discuss or quantify benefits, costs, and behavioral responses of implementing a pre-populated regime

Two approaches to measure success

1. "Upper bound":
   - "Success" are returns **without** a situation where pre-population automatically fails
     - E.g. Schedule C income reported $\neq$ 1099-MISC non-employee compensation
   - Upper bound because we can't enumerate all failure scenarios
1. "Lower bound":
   - Directly estimate tax liability using (in-house) NBER `TAXSIM`.
   - "Success" when calculated liability matches reported tax liability
   - Lower bound because the tax calculator could be richer

Detail

DATA

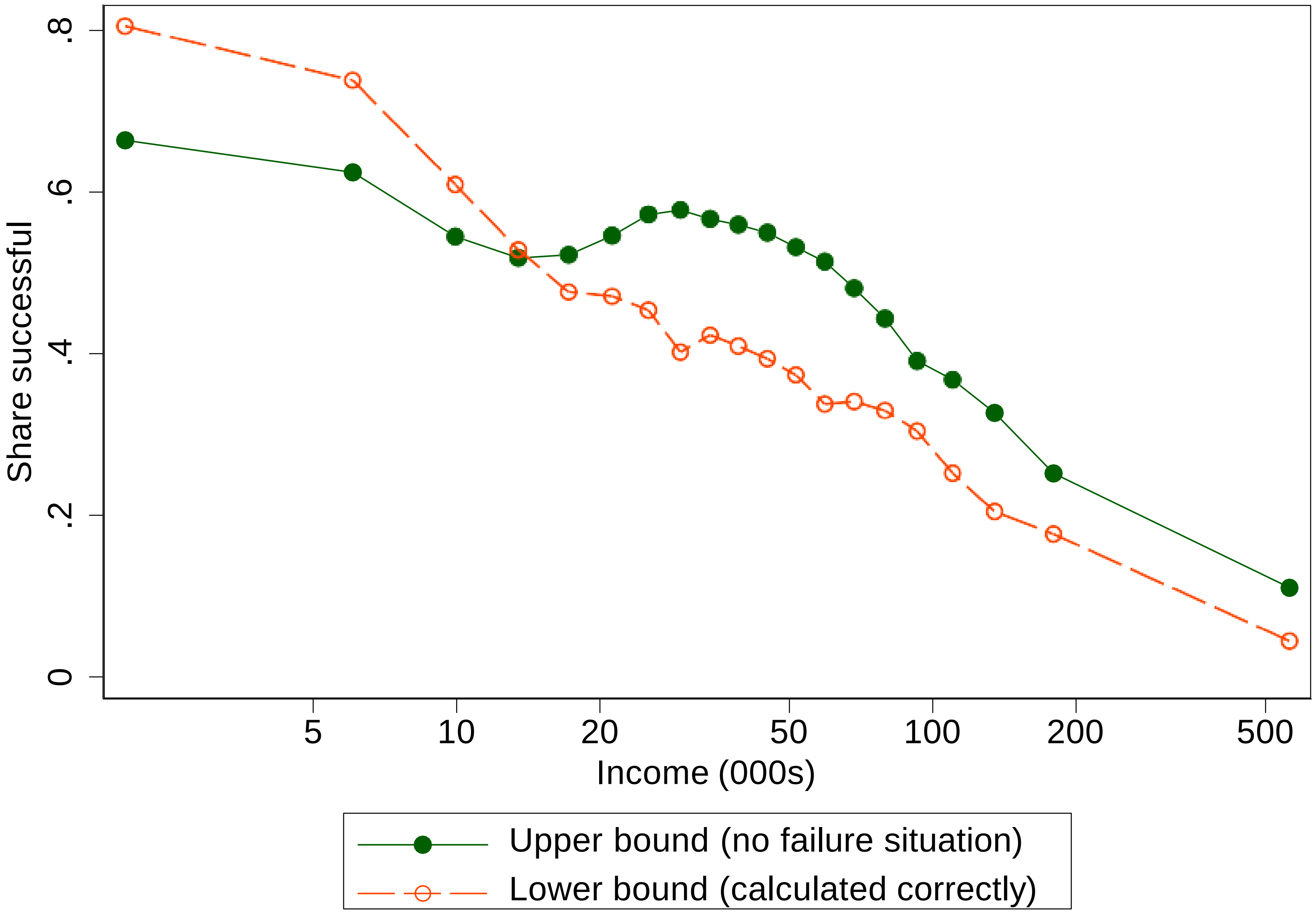
- Procedure is to pre-populate 2019 Form 1040
- **Filer data**: 344,400 tax units in 2019 (using the INSOLE file), matched to 2018 tax returns and 2019 information returns.
- **Non-filer data**: 0.1% sample of 55 million 2019 non-filers

Table: **Success Rates**

Lower bound	
Success rate	**0.406**
Tax difference tolerance	$100
Upper bound	
Success rate	**0.481**
Credit tolerance	$100
Income/deduction tolerance	$500

More

SUCCESS RATES DECLINE WITH INCOME



Upper bound approach failures

- Overall most common failure is Schedule C income; common among all income levels
- Among high income taxpayers itemizing and rents and royalties are important failure situations

Figure Reasons

Characteristics of those successful

Successful pre-populated returns under both methods generally:

- Had lower average and median income and tax liability
- Were less likely to use a paid preparer but 43% of them did use a paid preparer
- Were younger
- Were less likely to have dependents and less likely to be married

Among unsuccessful returns the median number of failure situations was 1

Table

Table: **Success rates for subsets of taxpayers, with subsets defined based on 2019 tax**

	Cumulative share of pop.	Cumulative success rate		Marginal success rate	
		Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)
1. **Narrowest:** Single, no dependents, only wages, no unobs. credits/deductions, income under $100k	0.20	0.78	0.82	0.78	0.82
2. Allow married	0.22	0.76	0.80	0.53	0.54
3. Allow dependents	0.30	0.67	0.74	0.41	0.59
4. Add interest/dividends	0.34	0.67	0.75	0.71	0.82
5. Add Social Security	0.36	0.68	0.75	0.76	0.78

Table: **Success rates for subsets of taxpayers, with subsets defined based on 2019 tax**

	Cumulative share of pop.	Cumulative success rate		Marginal success rate	
		Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)
1. **Narrowest:** Single, no dependents, only wages, no unobs. credits/deductions, income under $100k	0.20	0.78	0.82	0.78	0.82
2. Allow married	0.22	0.76	0.80	0.53	0.54
3. Allow dependents	0.30	0.67	0.74	0.41	0.59
4. Add interest/dividends	0.34	0.67	0.75	0.71	0.82
5. Add Social Security	0.36	0.68	0.75	0.76	0.78

Table: **Success rates for subsets of taxpayers, with subsets defined based on 2019 tax**

	Cumulative share of pop.	Cumulative success rate		Marginal success rate	
		Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)
1. **Narrowest:** Single, no dependents, only wages, no unobs. credits/deductions, income under $100k	0.20	0.78	0.82	0.78	0.82
2. Allow married	0.22	0.76	0.80	0.53	0.54
3. Allow dependents	0.30	0.67	0.74	0.41	0.59
4. Add interest/dividends	0.34	0.67	0.75	0.71	0.82
5. Add Social Security	0.36	0.68	0.75	0.76	0.78

Table: **Success rates for subsets of taxpayers, with subsets defined based on 2019 tax**

	Cumulative share of pop.	Cumulative success rate		Marginal success rate	
		Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)
1. **Narrowest:** Single, no dependents, only wages, no unobs. credits/deductions, income under $100k	0.20	0.78	0.82	0.78	0.82
2. Allow married	0.22	0.76	0.80	0.53	0.54
3. Allow dependents	0.30	0.67	0.74	0.41	0.59
4. Add interest/dividends	0.34	0.67	0.75	0.71	0.82
5. Add Social Security	0.36	0.68	0.75	0.76	0.78

Table: **Success rates for subsets of taxpayers, with subsets defined based on 2019 tax**

	Cumulative share of pop.	Cumulative success rate		Marginal success rate	
		Lower bound	Upper bound	Lower bound	Upper bound
	(1)	(2)	(3)	(4)	(5)
6. Add pension/IRA distributions	0.43	0.67	0.74	0.64	0.70
7. Add gambling, UI, state tax refunds	0.44	0.67	0.74	0.54	0.72
8. Add capital gains	0.47	0.66	0.73	0.59	0.60
9. Add high income	0.52	0.64	0.73	0.48	0.68
10. All income types	0.69	0.54	0.60	0.23	0.19
11. **Broadest**: Eliminate deduction and credit restrictions	1.00	0.41	0.48	0.10	0.22

NON-FILERS

- Draw a 0.1% sample from 54.7m non-filers
- We consider 2019 non-filers for whom there would be a pre-populated return:
  - Has 2019 information return (including 1095 series) with U.S. address
  - Aged 18 to 105 years
  - Drop if part of a 2018 M.F.J. couple where one spouse is a 2019 filer.
- 46.3 million (85%) **appear** not to have a filing obligation
- 8.4 million **appear** to have a filing obligation

Among those **without** filing obligation:

- Median age is 59 years
- 8% have childless EITC >0, 0.7% have child EITC
- 17% have potential refund; median $200

Among those **with** a filing obligation:

- 5% have childless EITC >0, 10% have child EITC
- 43% have potential refund; median $900
- 55% potentially owe; median $1,500

Detail

Additional Issues and Considerations

- Information returns would need to come in earlier
- Shifts compliance/ calculation burden onto the IRS
- Uncertain effects on compliance and revenue
  - Pre-population could serve as a nudge leading to over-compliance and under-compliance relative to current tax liability
  - Could increase salience of tax authority information and change perceived audit probabilities or tax morale

Conclusion and next steps

- Pre-populated returns are an intuitively appealing idea
- It would be successful for 40-50% of the population
- Many implementation obstacles would need to be overcome and we need to be careful about unintended effects

TWO APPROACHES TO MEASURE SUCCESS

- Caveat: these are <u>not</u> true upper and lower bounds
- The lower bound (TAXSIM) success rate can exceed the upper bound (no failure situations)
- Example: upper bound failure because reported wages $\neq$ W-2 wages but \$0 tax liability anyway, so lower bound success

Back

Table: **Success Rates**

	Strictest (1)	(2)	Preferred (3)	(4)	Least strict (5)
Panel A: Lower bound					
Success rate	0.326	0.370	**0.406**	0.458	0.552
Tax difference tolerance	$10	$50	$100	$200	$500
Panel B: Upper bound					
Success rate	0.432	0.448	**0.481**	0.509	0.598
Credit tolerance	$10	$50	$100	$200	$500
Income/deduction tolerance	$100	$200	$500	$1000	$5000

Back

Most common failure situations

Table: **Share with common failure situations**

Situation	Share
Schedule C income does not match 1099-MISC NEC	0.162
Itemized deductions in 2019	0.109
Wages do not match W-2 wages	0.089
Taxable pension/IRA income does not match 1099-R	0.067
Sched E rents/royalties (except from K-1)	0.065
Dependent mismatch from 2018 to 2019	0.061
Child/dependent care credit	0.039
Capital gains income does not match 1099-B	0.036

Table: **Share with common failure situations (cont)**

Situation	Share
S corp/partnership income does not match K-1	0.034
Change in filing status from 2018 to 2019	0.034
1099-R with taxable amount not determined	0.033
EITC dependent mismatch	0.029
Interest does not match 1099-INT	0.028
Certain above-the-line deductions	0.023
Dividends do not match 1099-DIV	0.021
Section 199A deduction complications	0.020
Lifetime learning credit	0.014
Any failure situation	0.519

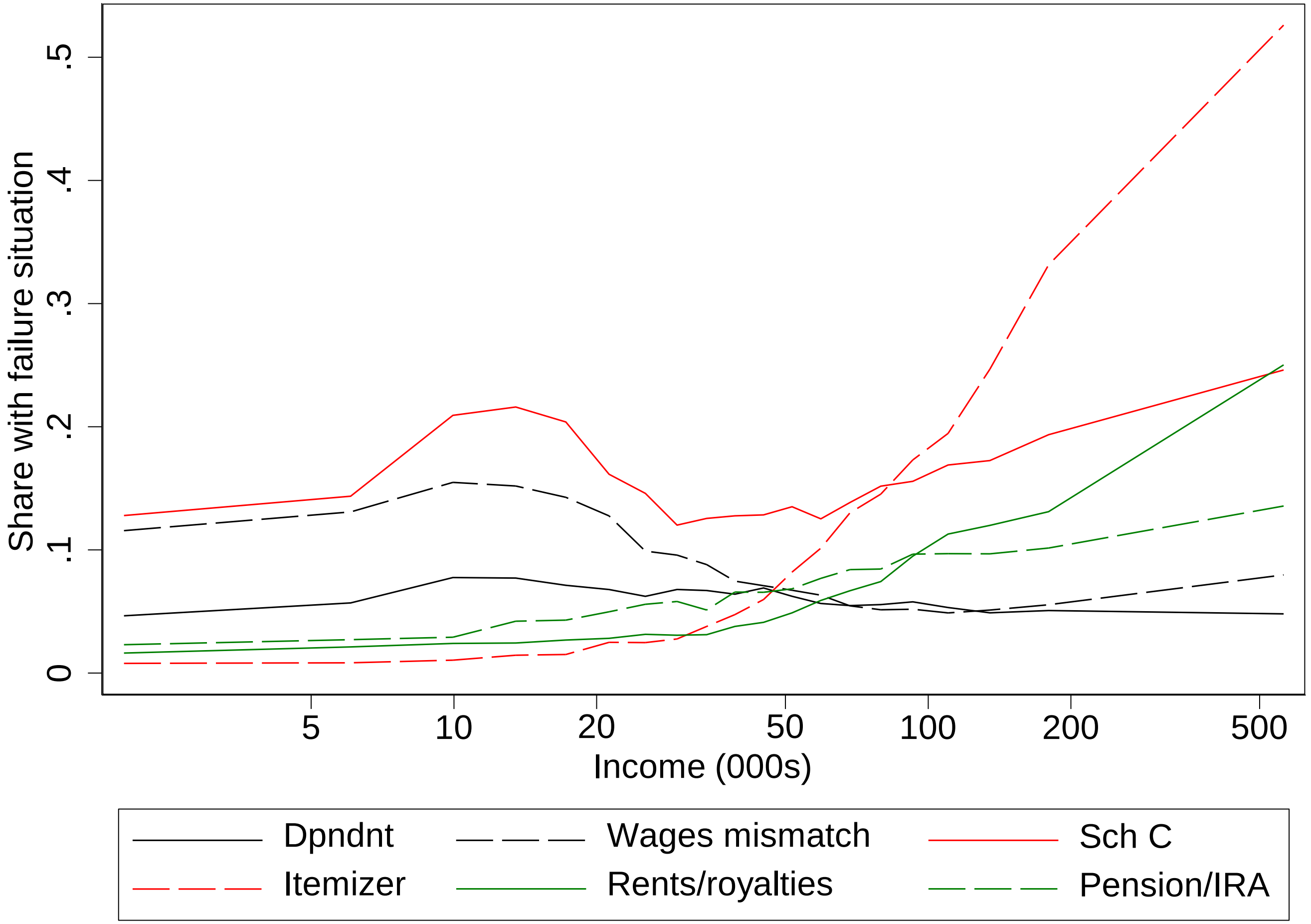
Back

CHARACTERISTICS OF THOSE SUCCESSFUL

	No failure situation? (upper bound)		Correct calculation? (lower bound)		
	✓	X	✓	high	low
	(1)	(2)	(3)	(4)	(5)
Mean					
Married	0.26	0.48	0.24	0.44	0.51
Has dependents	0.24	0.39	0.17	0.46	0.33
Uses paid preparer	0.43	0.58	0.44	0.53	0.63
Primary filer age	42	48	43	45	52
AGI (thousands)	47.2	104.0	38.6	81.1	150.4
Liability: calculated (thousands)	3.8	15.3	2.9	12.8	17.8
Liability: taxpayer-reported (thousands)	3.7	15.1	2.9	8.6	26.6
Liability: calculated less reported (thousands)	0.1	-0.2	0.0	4.2	-8.8
Failure situations	0.0	1.7	0.2	1.3	1.5
Median					
Primary filer age	39	47	38	43	52
AGI (thousands)	33.5	53.0	26.2	49.4	69.2
Liability: calculated (thousands)	1.5	2.8	0.7	3.6	2.5
Liability: taxpayer-reported (thousands)	1.4	2.4	0.7	1.8	5.0
Liability: calculated less reported (thousands)	0.0	0.4	0.0	1.4	-1.3
Failure situations	0	1	0	1	1
Count (millions)	73.1	78.9	61.7	62.2	28.1

Back

Upper bound approach: Failure Rates By Type and AGI



Back

Table: **Non-filers**

	No filing obligation		Has filing obligation	
	Mean	Median	Mean	Median
Age	55	59	47	46
Female	0.50	–	0.33	–
Filed a 2018 return	0.11	–	0.37	–
Pre-populated return with dependents	0.02	–	0.10	–
Calculated childless EITC >0	0.081	–	0.050	–
Calculated child EITC or CTC >0	0.007	–	0.096	–
...with revealed dependents	0.008	–	0.101	–
Calculated AGI	1,172	0	55,371	36,000
Calculated AGI (if potential refund)	3,884	2,900	39,865	28,100
Tax withholding	37	0	4,429	1,700
Tax withholding (if >0)	259	100	5,379	2,300
Has potential refund	0.17	–	0.43	–
Has potential taxes owed	0.00	–	0.55	–
Potential refund (if >0)	384	200	1,483	900
Potential taxes owed (if >0)	–	–	5,015	1,500
Count (millions)	46.3		8.4	

Back

Table: **Summary statistics**

	Mean	Median
Primary filer age	45	43
Married	0.37	–
Has dependents	0.32	–
Dependents	0.57	0
Uses paid preparer	0.51	–
Claims Earned Income Credit	0.17	–
Claims Child Tax Credit	0.30	–
Adjusted gross income	76,650	41,200
Taxable income	59,083	24,100
Tax liability	9,598	1,800
Count (millions)	152.0	

Notes: The table describes our sample of 2019 tax units. Sample weights are used.

- Marcuss et al (NTJ 2013)

Table 3
Individual Compliance Burden ($) by AGI Strata

	Population (Thousands)	Time (Hours)	Average Out Pocket Costs ($)	Average Monetized Burden ($)	Burden/ AGI (%)
Entire Population	142,985	12.54	198	373	6.8
No adjusted gross income	2,577	26.09	243	441	--
1 to 5,000	9,961	7.30	73	127	83.3
5,000 to 10,000	12,278	8.95	97	164	2.2
10,000 to 15,000	12,812	10.34	114	192	1.5
15,000 to 20,000	11,742	11.24	124	210	1.2
20,000 to 25,000	10,173	11.30	128	222	1.0
25,000 to 30,000	8,961	11.46	136	240	0.9
30,000 to 40,000	14,620	11.74	148	268	0.8
40,000 to 50,000	10,991	12.69	164	315	0.7
50,000 to 75,000	18,769	13.44	192	380	0.6
75,000 to 100,000	11,828	14.09	237	480	0.6
100,000 to 200,000	13,945	14.51	328	670	0.5
200,000 and more	4,328	29.79	1,250	2,331	0.5

Table: **Success rate for selected lines on Form 1040**

Line	Unconditional match rate	Conditional match rate
	(1)	(2)
Wages	0.918	0.903
Taxable interest	0.970	0.919
Qualified dividends	0.977	0.905
Income from Sch. 1	0.700	0.241
Taxable IRA and pensions	0.929	0.806
Capital gains	0.893	0.442
AGI	0.507	-
Taxable income	0.521	-
EITC	0.892	0.455
Child tax credit	0.856	0.628

Distribution of the Tax Year 2011-2013 Individual Income Tax and Self-Employment Tax Underreporting Tax Gap

Drew Johns
(IRS, Research, Applied Analytics & Statistics)

The author is an economist in the Knowledge Development & Application Division, Compliance Modeling Lab. The views expressed in this paper do not necessarily represent the views of the Department of the Treasury or the Internal Revenue Service.

IRS-TPC Conference June 16, 2022

Introduction

- To fulfill its mission, the IRS works to promote voluntary compliance:
  - Providing taxpayer services and education
  - Strategically enforcing the law
- IRS employs a balanced compliance enforcement strategy:
  - Enforcement presence across all types of tax returns
  - Returns that are most likely to have significantly underreported taxes are audited at a relatively higher rate
- A variety of compliance enforcement strategies and tools commensurate with the nature of the noncompliance:
  - Math error corrections during return processing
  - Matching return information to third-party information
  - **Examinations**
    - ✓ Campus correspondence examinations
    - ✓ Office examinations
    - ✓ Field examinations

The Tax Gap: Objective

- **The objective of tax gap estimation is to measure taxpayer compliance behavior as it manifests as tax not paid voluntarily and timely**
- **The focus and challenge is to measure actual behavior.**
- **Because the goal is to measure actual behavior, the tax gap concept is inherently retrospective.**
- **Our tax gap estimates reflect tax noncompliance**
  - Tax noncompliance and tax gap estimates reflect both intentional and unintentional errors
  - Do not include tax “avoidance”. Tax avoidance refers to use of legal means to reduce tax liability. In other words, estimates do not reflect a “policy gap”.
  - We do not use the term “evasion”. “Tax evasion” has specific meanings within tax administration reflecting, in general, intentional noncompliance rising to the level of criminality. Some intentional errors might rise to the level of tax evasion, but tax noncompliance/tax gap and tax evasion are not interchangeable terminology.

The Tax Gap: Key Points

The tax gap *is:*

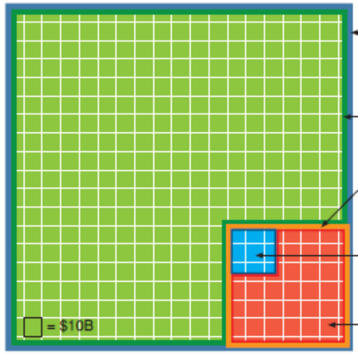
- a tax year (TY) concept, as opposed to a fiscal year concept;
- broadly defined to encompass both tax and refundable and non-refundable tax credits;
- based on all the relevant events that occurred during a tax year and the Internal Revenue Code (IRC) provisions in effect at the time;
- most informative if grounded in data that reflect observed compliance behavior;
- a rough gauge of the extent of overall tax noncompliance;
- a bottom-up approach to compliance measurement;
- a compliance indicator — not an IRS performance measure;
- not necessarily an optimal method for allocating resources

Underreporting Tax Gap

Tax Gap Estimates for Tax Years 2011–2013

(Money amounts are in billions of dollars; estimates are annual average amounts.)

IRS Research, Applied Analytics & Statistics



Estimated Total True Tax Liability*
$2,683B

Tax Paid Voluntarily & Timely
$2,242B 83.6% Voluntary Compliance Rate (VCR)

Gross Tax Gap
$441B

Enforced & Other Late Payments
$60B

Net Tax Gap (Tax Not Collected)
$381B 85.8% Net Compliance Rate (NCR)

Calculating the Net Tax Gap

Nonfiling
Underreporting
\+ Underpayment
Gross Tax Gap
− **Enforced & Other Late Payments**
Net Tax Gap

	Total True Tax Liability	Tax Paid Voluntarily & Timely	Gross Tax Gap: Nonfiling	Gross Tax Gap: Underreporting	Gross Tax Gap: Underpayment	Gross Tax Gap	Enforced & Other Late Payments	Net Tax Gap (Tax Not Collected)
Total	$2,683	$2,242	$39	+ $352	+ $50	= $441	− $60	= $381
By Type of Tax								
Individual Income Tax	$1,398	$1,084	$31	+ $245	+ $38	= $314	− $43 (14%)	= $271
Corporation Income Tax	$294	$251	#	+ $37	+ $5	= $42	− $10 (24%)	= $32
Employment Tax	$920	$839	$6 [4]	+ $69	+ $6	= $81	− $5 (6%)	= $77
Estate Tax	$16	$13	$2	+ $1	+ $<0.5	= $3	− $2 (55%)	= $1

Individual Income Tax underreporting ($245):

Business Income	Non-Business Income	Credits	Income Offsets [1]	Filing Status	Other Taxes [2]	Unallocated Marginal Effects [3]
$110	$57	$42	$20	$5	$1	$10

Corporation Income Tax underreporting ($37):

Large Corporations	Small Corporations
$26	$11

Employment Tax underreporting ($69):

Self-Employment Tax	FICA & Uncollected FICA TAX	Unemployment
$45	$24	$1

NOTES:
* Totals include Excise Tax.
#—No estimate.
Detail may not add to totals due to rounding.
[1] Includes adjustments, deductions, and exemptions.
[2] Includes the Alternative Minimum Tax and taxes reported in the "Other Taxes" section of the Form 1040 except for self-employment tax and unreported social security and Medicare tax (which are included in the employment tax gap estimates).
[3] Is the difference between (1) the estimate of the individual income tax underreporting tax gap where underreported tax is calculated based on all misreporting combined and (2) the estimate of the individual income tax underreporting tax gap based on the sum of the tax gaps associated with each line item where the line item tax gap is calculated based on the misreporting of that item only. There may be differences if the marginal tax rates are different in these two situations.
[4] Self-employment tax only.

Revised 09/2019

Publication 5365 (Rev. 9–2019) Catalog Number 73349K Department of the Treasury **Internal Revenue Service** www.irs.gov

Distribution of Tax Year 2011-2013 Individual Income Tax and Self-Employment Tax Underreporting Tax Gap

Individual Income Tax Underreporting Tax Gap

- Foundation of estimate is National Research Program (NRP) data
  - NRP: selects stratified random sample of filed tax returns for examination; audited under consistent set of procedures.
  - NRP is a rich data source so have more detail than for other estimates.
- Methodology accounts for income undetected in examinations
  - Econometric technique called Detection Controlled Estimation (DCE) estimated at the line-item level for income items
  - Microsimulation with tax calculator to reflect realistic distribution of undetected income

Underreporting Tax Gap Measures (aggregate level)

- Net misreported amount (NMA): the net dollar amount misreported on a return or schedule line item in the favor of taxpayers
- Net misreporting percentage (NMP): The NMA divided by the sum of the absolute values of the amounts that should have been reported

Activity Codes

- IRS stratifies tax returns based on tax return characteristics into mutually exclusive groups for examination planning and tracking.

Activity Code	Definition
270	EITC > 0; TGR <$25,000
271	EITC > 0; TGR >=$25,000
272	TPI<$200,000 and No Schedule C, E, F, or 2106
273	TPI<$200,000; No Schedule C/F; E or 2106 present
274	Schedule C; TGR < $25,000; TPI < $200,000
275	Schedule C; $25,000<=TGR<$100,000; TPI<$200,000
276	Schedule C; $100,000<=TGR< $200,000; TPI<$200,000
277	Schedule C; $200,000<=TGR; TPI< $200,000
278	Schedule F Not Classified Elsewhere; TPI<$200,000
279	No Schedule C/F; $200,000=< TPI < $1,000,000
280	Schedule C/F; $200,000<= TPI < $1,000,000
281*	TPI >= $1,000,000

TPI = Total Positive Income

TGR = Total Gross Receipts

EITC = Earned Income Tax Credit

*In Filing Year 2022, the IRS is in the process of changing Form 1040 Activity Codes from 12 to 14 Activity Codes. This will be done by splitting Activity Code 281 into three new Activity Codes:

282: TPI >= $1,000,000 and < $ 5,00,000

283: TPI >= $ 5,000,000 and < $10,000,000

284: TPI >= $10,000,000

Tax Gap Distribution by Activity Code

Distribution of TY 2011 – 2013
Individual Income Tax and SE Tax Underreporting Tax Gap
Compared to FY 2019 Examinations Closed, by Activity Code

		TY 2011 - 2013 Tax Gap		FY 2019 Exams Closed[1]		
Activity Code	**Calendar Year 2018 Returns**	**Examiner Determined**	**DCE Adjusted**	**Exams Closed**	**Exam Time**	**Exam Cost**
270	16.2%	28.5%	20.3%	42.1%	12.8%	10.0%
271	1.3%	7.3%	8.0%	2.4%	5.0%	5.0%
272	54.9%	15.5%	12.5%	14.6%	8.2%	7.6%
273	10.3%	12.5%	13.3%	13.3%	11.9%	11.4%
274	7.6%	8.8%	12.0%	10.5%	8.7%	8.0%
275	2.3%	5.9%	8.0%	4.4%	5.4%	5.5%
276	0.6%	3.0%	3.6%	2.2%	4.9%	5.1%
277	0.5%	4.5%	5.0%	1.5%	6.3%	6.7%
278	0.8%	1.9%	2.9%	0.4%	1.1%	1.2%
279	3.7%	4.3%	5.3%	3.0%	7.5%	8.1%
280	1.5%	6.0%	6.5%	3.5%	14.3%	15.8%
281	0.4%	1.7%	2.6%	2.1%	13.9%	15.7%
Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

[1] – Includes exams of returns that were not filed.

Distribution of TY 2011 – 2013 Individual Income Tax and SE Tax Underreporting Tax Gap Compared to FY 2020 Examinations Closed, by Reported TPI

		TY 2011 - 2013 Tax Gap		FY 2020 Exams Closed[1]		
Reported TPI Level[2]	**TY 2018 Returns Filed**	**Examiner Determined**	**DCE Adjusted**	**Exams Closed**	**Exam Time**	**Exam Cost**
$0 or less[4]	0.4%	0.5%	0.6%	12.5%	9.7%	10.3%
$1 under $25,000	32.1%	32.3%	28.3%	36.7%	13.8%	11.7%
$25,000 under $50,000	23.8%	21.7%	19.9%	13.3%	9.5%	8.6%
$50,000 under $75,000	14.1%	12.1%	12.7%	8.4%	7.1%	6.7%
$75,000 under $100,000	9.1%	7.9%	9.0%	6.8%	6.7%	6.4%
$100,000 under $200,000	14.4%	15.0%	16.6%	13.6%	18.5%	18.5%
$200,000 under $500,000	4.8%	6.9%	8.3%	4.6%	13.5%	14.4%
$500,000 under $1,000,000	0.8%	2.0%	2.2%	1.7%	7.0%	7.6%
$1,000,000 under $5,000,000	0.4%	1.3%	2.0%	1.8%	9.8%	10.9%
$5,000,000 under $10,000,000	[3]	0.2%	0.3%	0.3%	1.6%	1.8%
$10,000,000 or more	[3]	[3]	0.1%	0.3%	2.7%	3.0%
Total	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

[1] – Includes nonfiler examinations.
[2] – TPI in this table for TY 2018 returns filed and FY 2020 Exams Closed reflects the IRS calculation of reported TPI calculated during return processing. This may differ slightly from the tax gap tax calculator calculation of reported TPI_{TG} used for the TY 2011 – 2013 Tax Gap estimates.
[3] – Less than 0.05 percent or $0.05 billion
[4] – About 95% of examinations closed in the $0 or less TPI level are nonfiler examinations. These exams technically should be associated with the $37 billion TY 2011 – 2013 individual income tax plus SE tax nonfiler tax gap, not the underreporting tax gap.

Tax Gap Distribution by DCE Adjusted TPI

Distribution of *DCE Adjusted* TY 2011-2013 Individual Income Tax and SE Tax Underreporting Tax Gap: Arrayed by Level of Reported TPI_{TG} and DCE Adjusted TPI_{TG}

	Arrayed by Reported TPI_{TG}			Arrayed by DCE Adjusted TPI_{TG}		
TPI_{TG} Level	**$ Billions**	**Share**	**NMP**	**$ Billions**	**Share**	**NMP**
$0 or less	$1.8	0.6%	105.2%	$0.5	0.2%	[1]
$1 under $25,000	$82.1	28.3%	83.6%	$26.6	9.2%	55.2%
$25,000 under $50,000	$57.7	19.9%	46.4%	$41.3	14.2%	41.8%
$50,000 under $75,000	$36.9	12.7%	29.0%	$35.7	12.3%	30.2%
$75,000 under $100,000	$26.3	9.0%	21.1%	$27.2	9.3%	22.4%
$100,000 under $200,000	$48.3	16.6%	15.2%	$66.5	22.9%	19.9%
$200,000 under $500,000	$24.2	8.3%	9.6%	$51.9	17.9%	17.8%
$500,000 under $1,000,000	$6.3	2.2%	5.4%	$18.7	6.4%	14.2%
$1,000,000 under $5,000,000	$5.8	2.0%	3.8%	$17.0	5.8%	10.3%
$5,000,000 under $10,000,000	$0.8	0.3%	2.3%	$3.9	1.3%	9.9%
$10,000,000 or more	$0.2	0.1%	0.3%	$1.1	0.4%	1.7%
Total	**$290.4**	**100.0%**	**20.5%**	**$290.4**	**100.0%**	**20.5%**

[1] – More than 1,000%
DCE – Detection Controlled Estimation
NMP – Net Misreporting Percentage is the ratio of the aggregate net misreported amount over the sum of the absolute values of the amounts that should have been reported.
TPI_{TG} – Total Positive Income (Tax Gap) Model Calculation

Tax Gap Distribution by Examiner Determined TPI

Distribution of *Examiner Determined* TY 2011-2013 Individual Income Tax and SE Tax Underreporting Tax Gap: Arrayed by Level of Reported TPI_{TG} and Examiner Determined TPI_{TG}

	Arrayed by Reported TPI_{TG}			Arrayed by Examiner Determined TPI_{TG}		
TPI_{TG} Level	**$ Billions**	**Share**	**NMP**	**$ Billions**	**Share**	**NMP**
$0 or less	$0.7	0.5%	116.1%	$0.6	0.5%	[1]
$1 under $25,000	$45.0	32.3%	65.4%	$27.3	19.6%	51.3%
$25,000 under $50,000	$30.2	21.7%	30.1%	$24.9	17.9%	27.0%
$50,000 under $75,000	$16.9	12.1%	15.7%	$17.2	12.4%	16.3%
$75,000 under $100,000	$11.0	7.9%	10.0%	$12.7	9.2%	11.6%
$100,000 under $200,000	$20.9	15.0%	7.2%	$27.4	19.7%	9.2%
$200,000 under $500,000	$9.6	6.9%	4.1%	$17.0	12.2%	6.9%
$500,000 under $1,000,000	$2.8	2.0%	2.4%	$4.9	3.5%	4.2%
$1,000,000 under $5,000,000	$1.8	1.3%	1.2%	$5.7	4.1%	3.8%
$5,000,000 under $10,000,000	$0.3	0.2%	0.9%	$0.9	0.6%	2.5%
$10,000,000 or more	[2]	[2]	[2]	$0.4	0.3%	0.5%
Total	**$139.1**	**100.0%**	**10.9%**	**$139.1**	**100.0%**	**10.9%**

[1] – More than 1,000%
[2] – Less than 0.05 percent or $0.05 billion
DCE – Detection Controlled Estimation
NMP – Net Misreporting Percentage is the ratio of the aggregate net misreported amount over the sum of the absolute values of the amounts that should have been reported.
TPI_{TG} – Total Positive Income (Tax Gap) Model Calculation

Key Points

- The tax gap associated with taxpayers who claimed EITC is significantly higher than the tax gap associated with the EITC line item.
- Examination measures based on the number of examinations closed are a poor proxy for the allocation of examination resources.
- Allocation of examination resources and distributional analysis should differentiate between the reported income of taxpayers and the true (unobserved) income of taxpayers.
    - Examination strategies that focus exclusively on examining taxpayers based on their self-reported high income are likely to miss some of the most egregious noncompliance.
- More than half of the estimated individual income tax and SE tax underreporting tax gap is unlikely to be detected during an examination.
    - The tax gap estimates do not necessarily reflect the amount of revenue that could be obtained through expanding examination coverage.

Data Limitations and Concerns

- Subject to sampling error, **measurement error** and modeling error.
- **Measurement error** refers to the correctness and completeness of the examiner's determination of what should have been reported.
    - Tax gap estimation assumes that the recommended adjustments made by the examiners are correct and appropriate.
    - Also assumes that there may be income that examiners did not detect that impacts the completeness of the examiner's determination (DCE).
- A tax gap estimate of a specific issue may not be possible even though that issue is accounted for in the tax gap estimates.
    - Data may not be collected at that level of detail.
    - The issue may be too rare to provide an estimate with sufficient precision.
- Emerging and complex issues
    - **Substitution effects**, is there growth in noncompliance or a change in the type?
    - **Nonadditive**, to what extent do issues overlap (digital assets vs. international activities vs. illegal activities)?

BREAK
The program will resume momentarily

Session 3:
Improving Audit Outcomes: Thinking Inside the Box

Graph-based machine learning for case selection and population segmentation

IRS/TPC Joint Research Conference

June 16, 2022

IRS

The goals are familiar – even if the methods are not

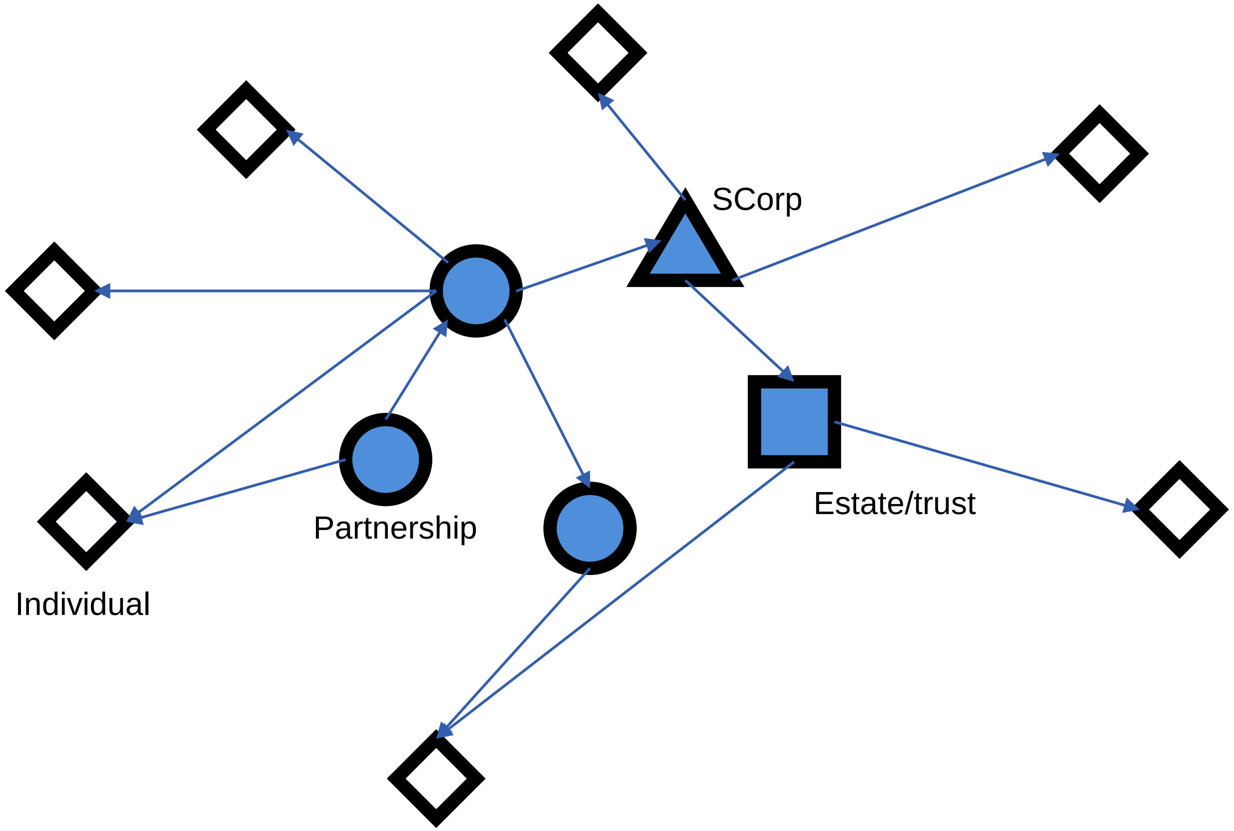
Stratify to improve statistical certainty and guide sampling

Identify emergent non-compliance

Select and construct case work

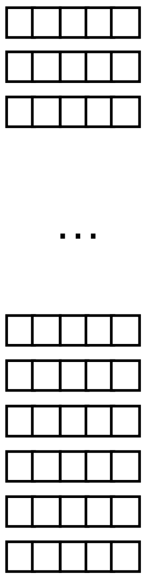
Discover links between entities not otherwise connected in IRS data resources

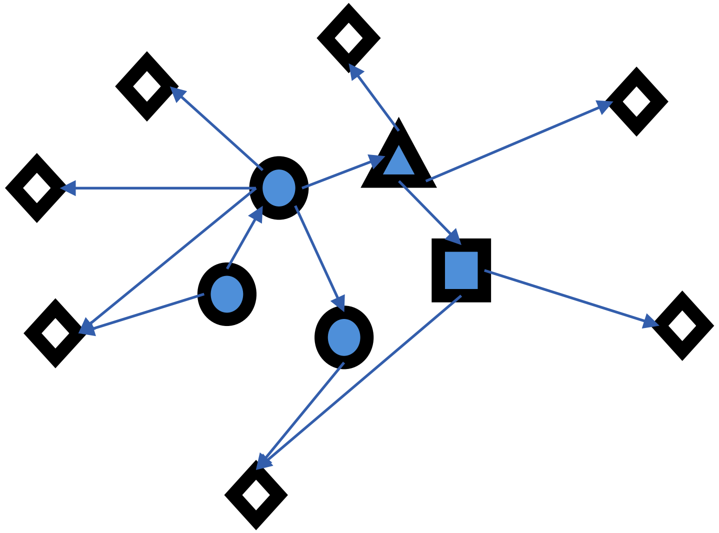
Pass-through networks contain signal of tax avoidance, uncertainty, and noncompliance



Agarwal, A., Chen, S., & Mills, L. (2020). Entity Structure and Taxes: An Analysis of Embedded Pass-Through Entities. *The Accounting Review*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3743370

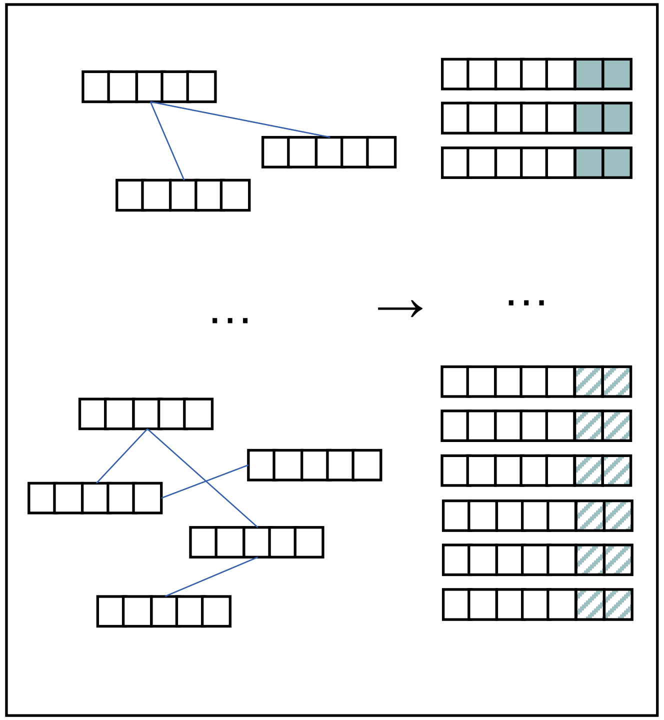
“Traditional” entity vectors don’t collect the story



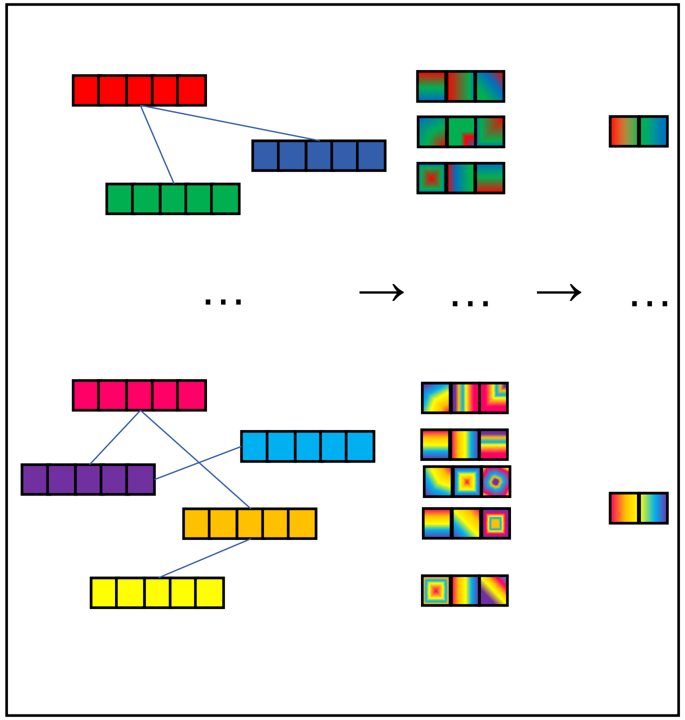


… but variable network structures impair construction of a general fixed-length vector

So we either build network features or learn them



Manual engineering is necessary – and will be necessary – for the foreseeable future. See the fraud literature.



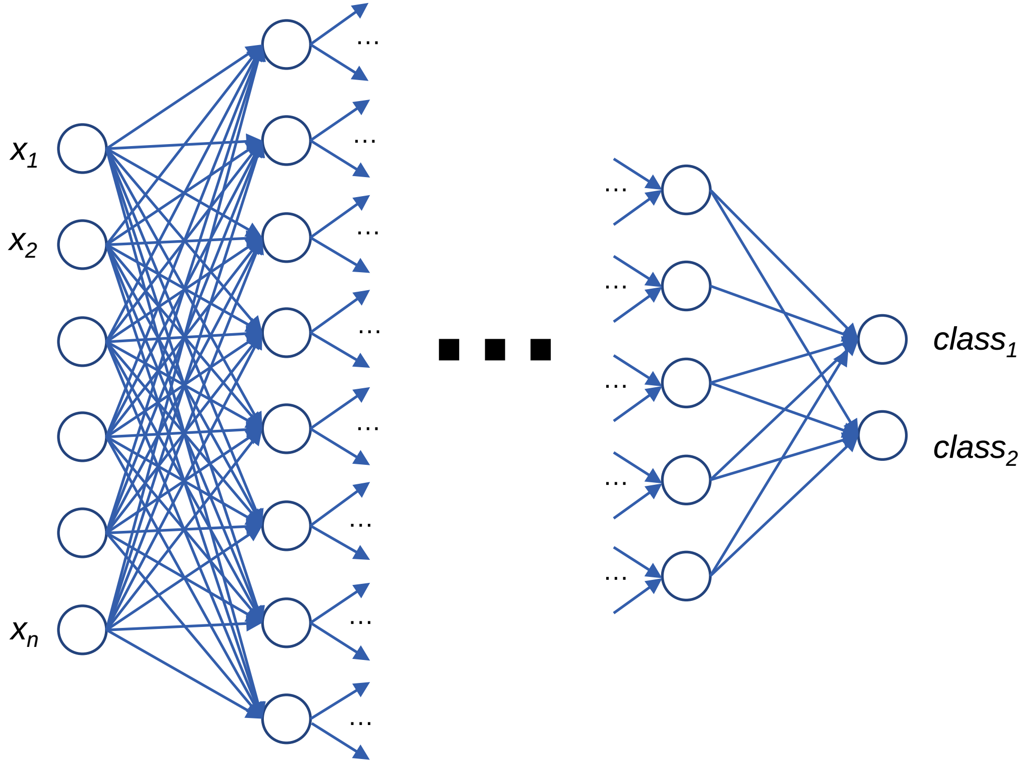
Our focus is here to stretch the technology inside IRS.

Neural nets are the base technology

$$f(x;\beta) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

becomes

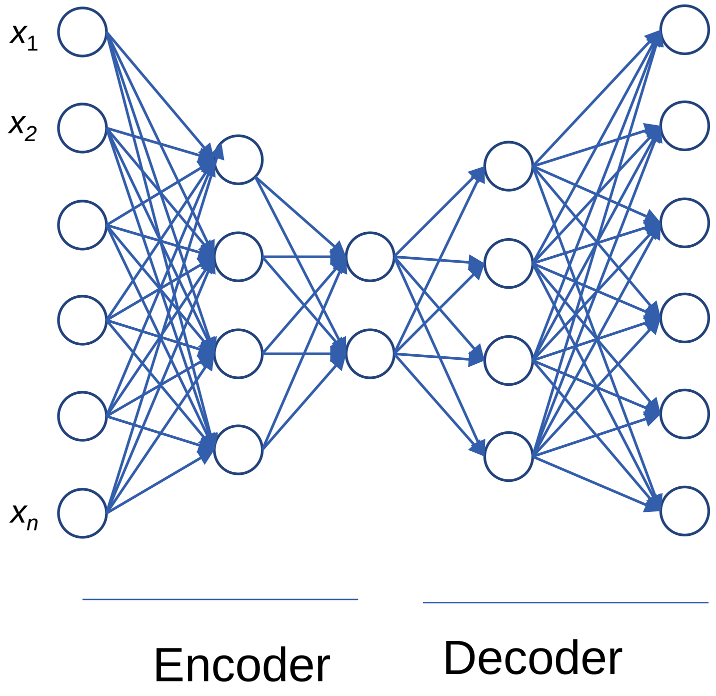
$$f(x; b, w, h(.))=\ldots$$



Graph neural networks extend this to learn fixed-length vector representations from fixed node features, fixed edge features, and G(V,E)

$$f(x, G(V,E); b, w, h(.), g(.))=\ldots$$

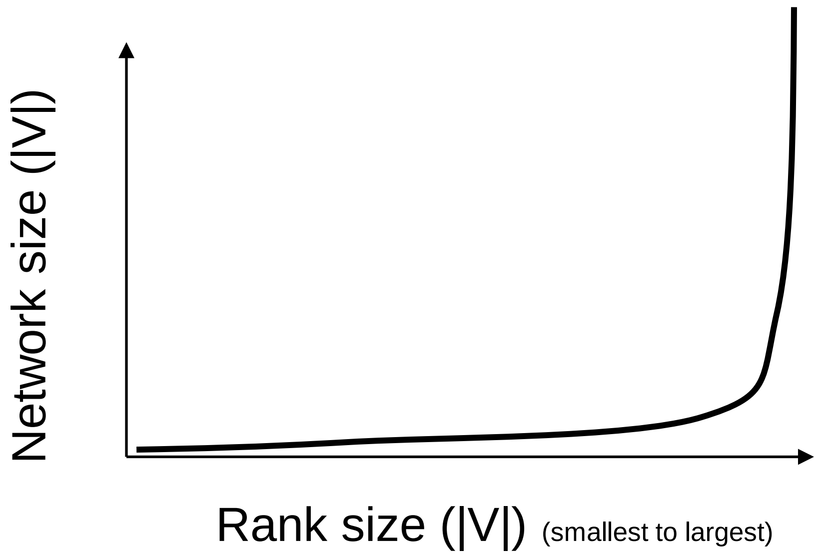
An autoencoder reduces the full vector (and graph) into lower dimensional representations

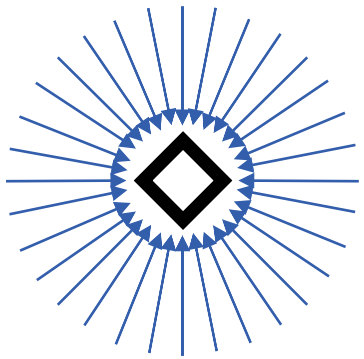


- Fit with the output layer of the decoder
- Use full encoder/decoder pair for anomaly detection (open question from early hints: are anomalous networks audited less frequently?)
- Use just the encoder to build vectors for other ML methods (though regularization is different for the two use-cases)

So we ~~either~~ build network features ~~or~~ learn them

BOTH **AND**





Many entities receive

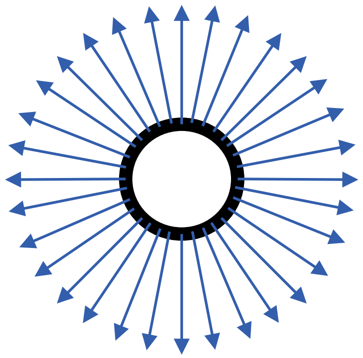
>10k

K1s.

Many PTE pay more than

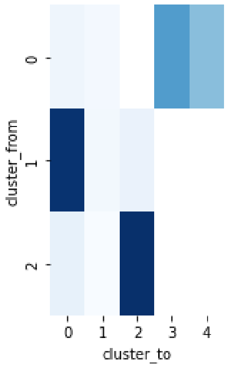
100k

entities.

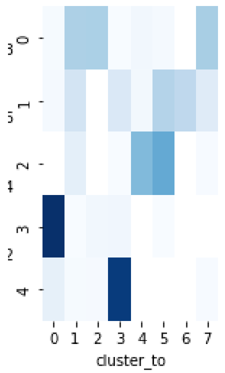


2020 (so far)

Practical implementation drives the investigation as much as the model development



vs



Examiners are <u>busy</u>.

- As Research, we're aiming for a light footprint until value is clear.
- NN are problematic for interpretability.

The question space is <u>huge</u>.

- Emphasizes re-useable, shareable code – packages and services.
- Adherence to open-source state-of-the-practice eases transition and recruiting
- Emphasizes KM and dissemination.
- Emphasizes expertise within IRS.

Automated Discovery of Tax Schemes using Genetic Algorithms

*Scheme is defined as a sequence of transactions between entities and/or actions by entities

Siggy Scott[1], Camrynn Fausey[1], Karen Jones[1], Geoff Warner[1], and Hahnemann Ortiz[2]

June 18, 2022

[1]The MITRE Corporation
[2]Internal Revenue Service

Internal Revenue Service

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD®

Seeking New Tax Gap Solutions

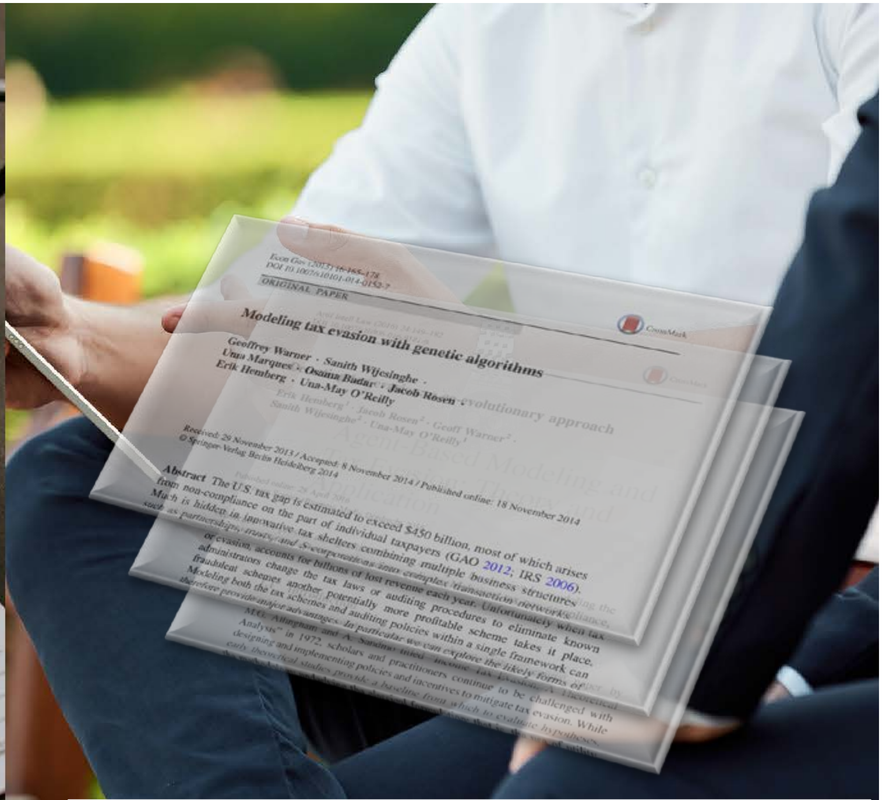
The 2021 Gap is ~$600B[1]

- Perhaps **$60B** from corporate taxes[2]
- And another **$50B** from pass-through income[3]
- May reach $1T when accounting for crypto, etc.[4]

Challenges

- The landscape is continuously changing
- Data collection is slow

1. Natasha Sarin, "The Case for a Robust Attack on the Tax Gap," U.S. Department of the Treasury, 2021.
2. Extrapolated from IRS-1415, "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011–2013," 2019.
3. Extrapolated from GAO-14-453, "Partnerships and S Corporations: IRS Needs to Improve Information to Address Tax Noncompliance," 2014.
4. "IRS chief says $1 trillion in taxes goes uncollected every year," *Reuters*, 2021.



We are building on research that points to as a way of **proactively anticipating** changes in taxpayer behavior

Problem: Lag in Reaction

Traditional Approach

- Tax-paying entities **start planning** as soon as a tax change is announced
- It takes at least **2 years** for the IRS to accumulate **sufficient data** to analyze new behaviors…
- … and to respond with an effective **filter** to direct (increasingly) scarce audit resources.

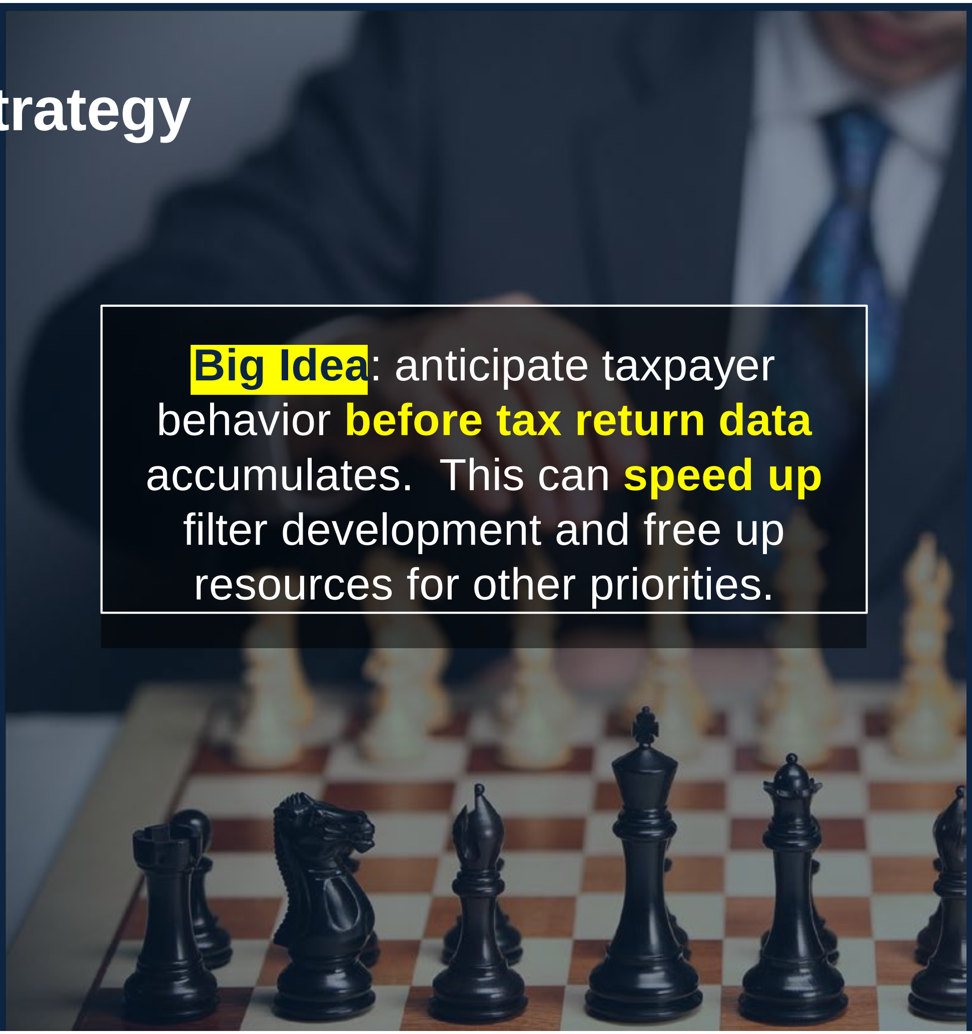
Example

- **2017**: TCJA introduces **Base-Erotion Anti-Abuse Tax** (BEAT) to combat off-shore profit-shifting.
- **2017–:** taxpayers develop strategies to minimize liability.
- **2022**: **no filter yet exists** to address BEAT compliance.

Building a Proactive Audit Strategy

Automated Discovery of Tax Schemes*

- **Model taxpayer behavior** in simulation
- Analyze the **space of decisions** that are available to taxpayers and policy makers…
- … and look for effective responses that **preempt new strategies** <u>before</u> historical data accumulates.



Big Idea: anticipate taxpayer behavior **before tax return data** accumulates. This can **speed up** filter development and free up resources for other priorities.

Background and Motivation

- Simulation offers a means to study policy and explore causal assumptions in advance of accumulated data.
- However, simulation models are often niche and domain-specific, requiring significant resources and time to modify for a new domain or study focus.
- Our research describes a modeling process and an evolutionary behavior discovery software framework (EBD) in the tax policy domain that is generalizable enough to lighten the effort needed for application to a diversity of complex tax policy scenarios of interest.

Evolutionary Behavior Discovery (EBD) Research Objectives

Research objective

1. Demonstrate extensibility of our EBD by modeling a variety of taxpayer scenarios.
2. Validate our EBD approach by showing it finds advantageous behaviors across taxpayer scenarios.

How objectives were achieved

We modeled multiple IRC changes introduced under the *Tax Cuts and Jobs Act* (TCJA), ***illustrating framework flexibility*** across diverse and complex scenarios.

We provided insight into how a taxpayer could minimize their tax liability and avoid detection, which ***enables LB&I to explore potential filters.***

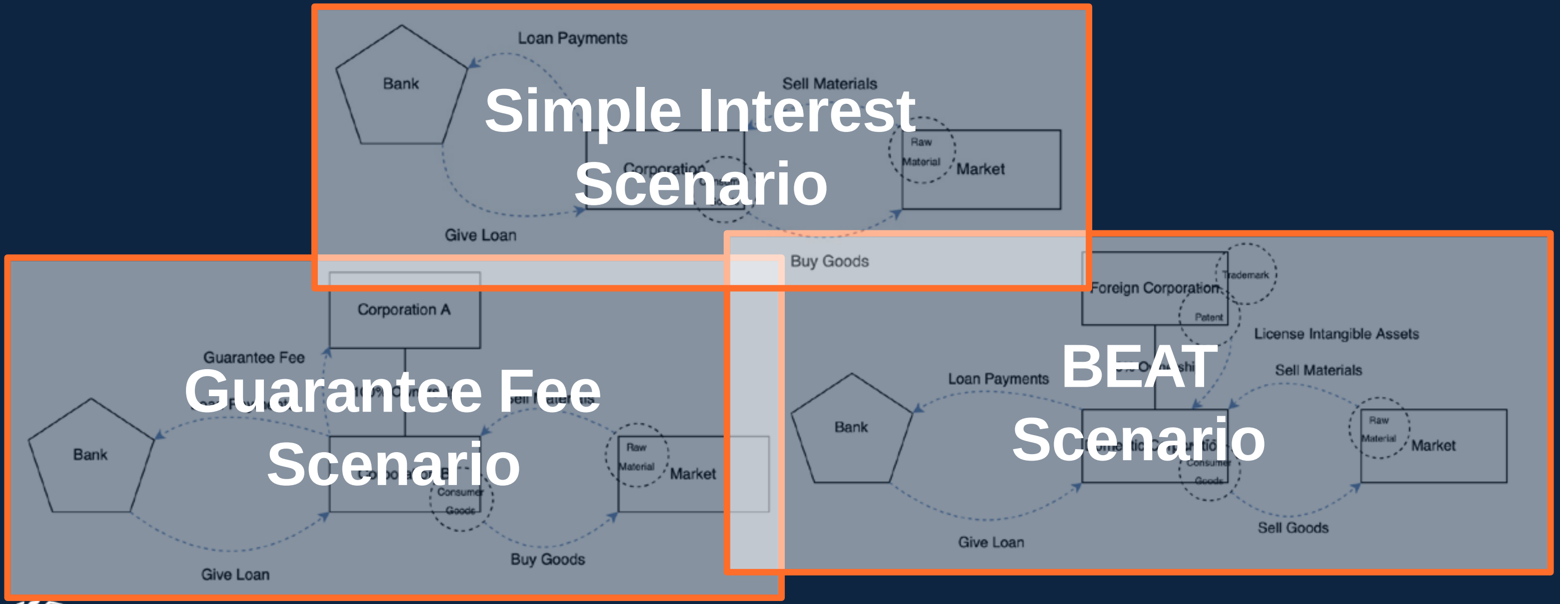
We will be showcasing the EBD's 1) **modular extensibility** and 2) **ability to discover tax planning behavior** through our ***BEAT simulation***.

MITRE

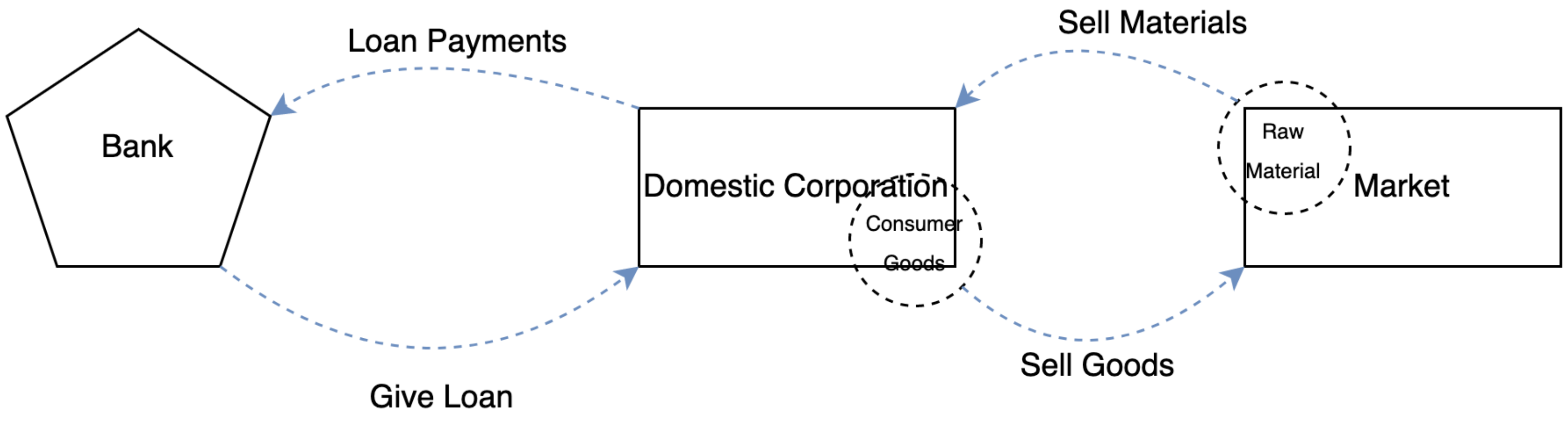
Modeling Framework

Scenarios

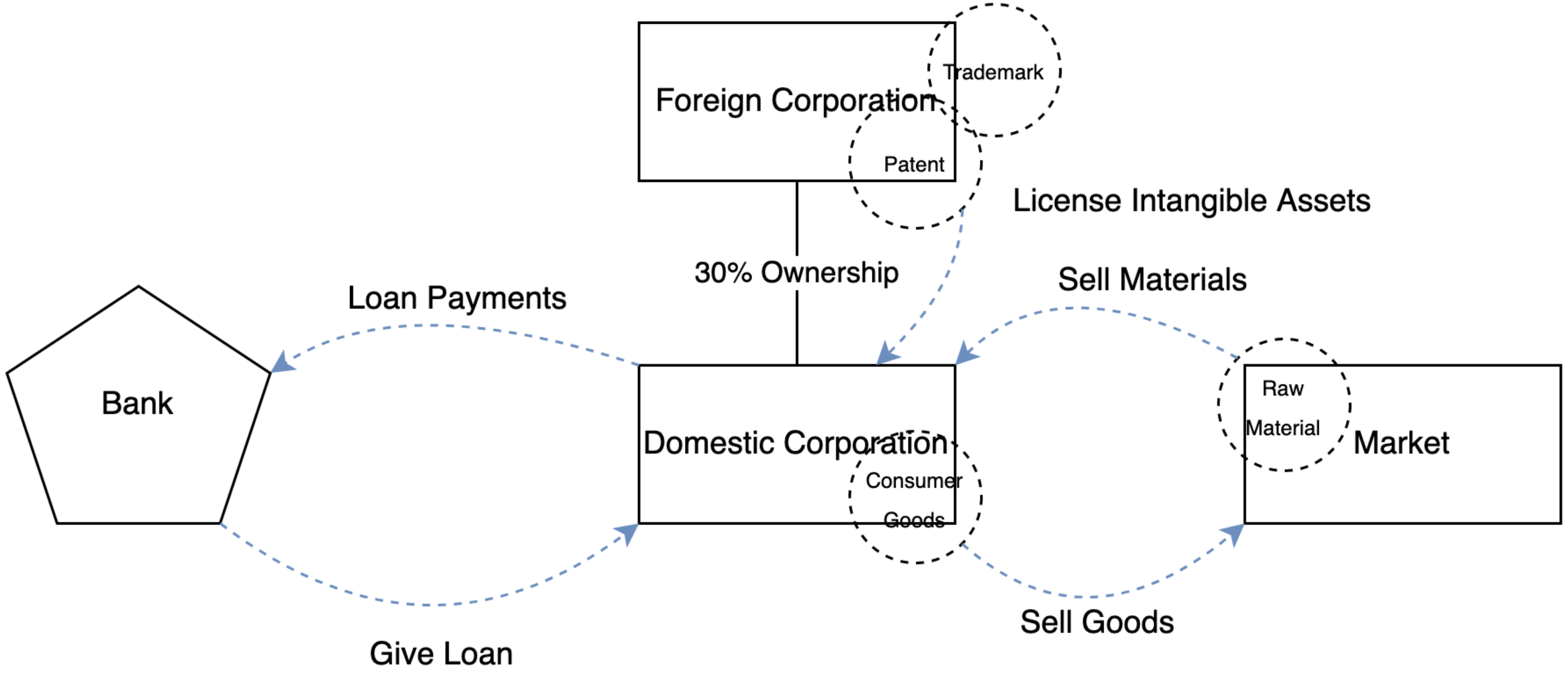
- We have used a suite of **several tax system scenarios**
  - A way to validate and ***build-out the extensibility*** of our modeling framework



Building Out a BEAT Scenario

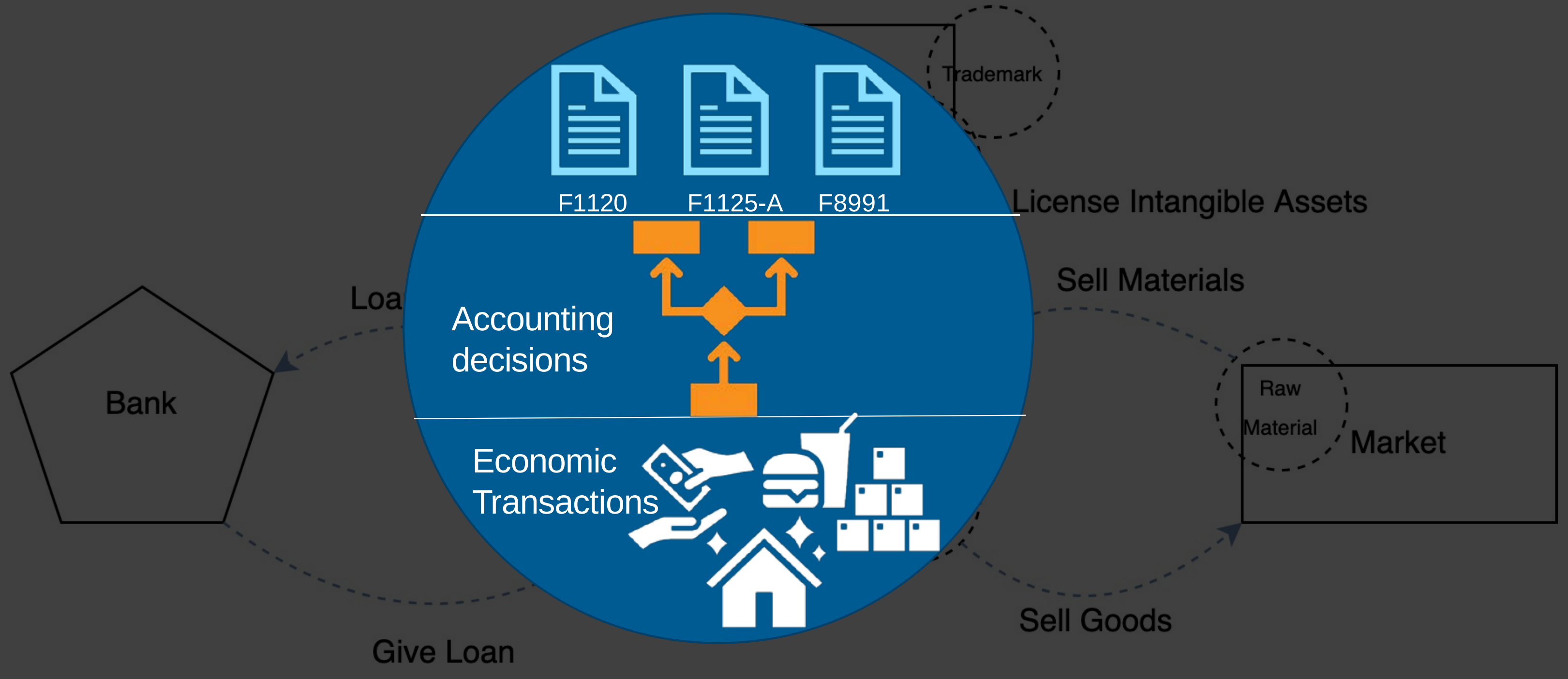


Building Out a BEAT Scenario



MITRE

Building Out a BEAT Scenario



Simulation Preparation

- We modularized and generalized the atomic components across scenarios of interest.
- Discrete components enables a developer to follow a relatively straightforward process to extend or augment the simulation.
- With a more generalized structure, new simulation feature development can scale efficiently.

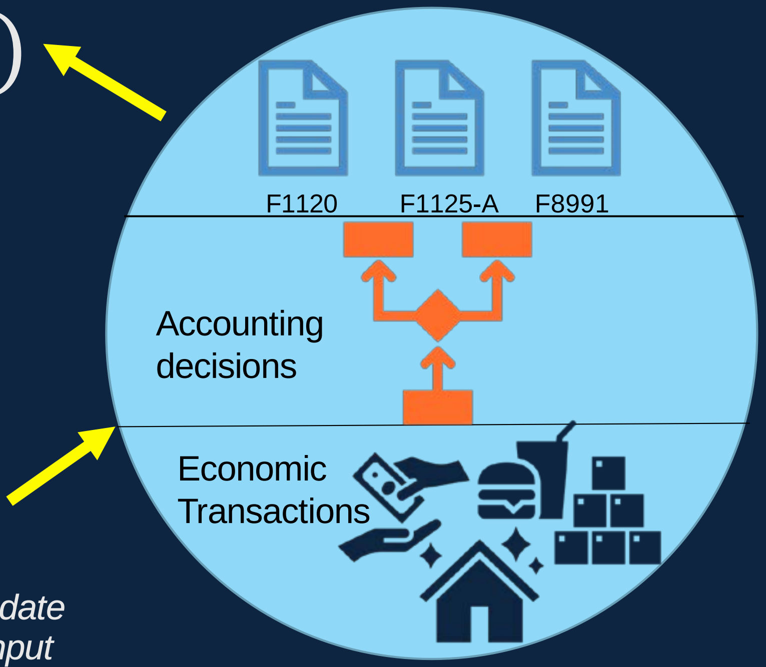
User Stories/ Hypotheses → Defined Scenario → Required Features → Atomic Components → Development Process

Evolutionary Behavior Discovery

Behavior Discovery Simulation

Simulation returns fitness scores

Simulation

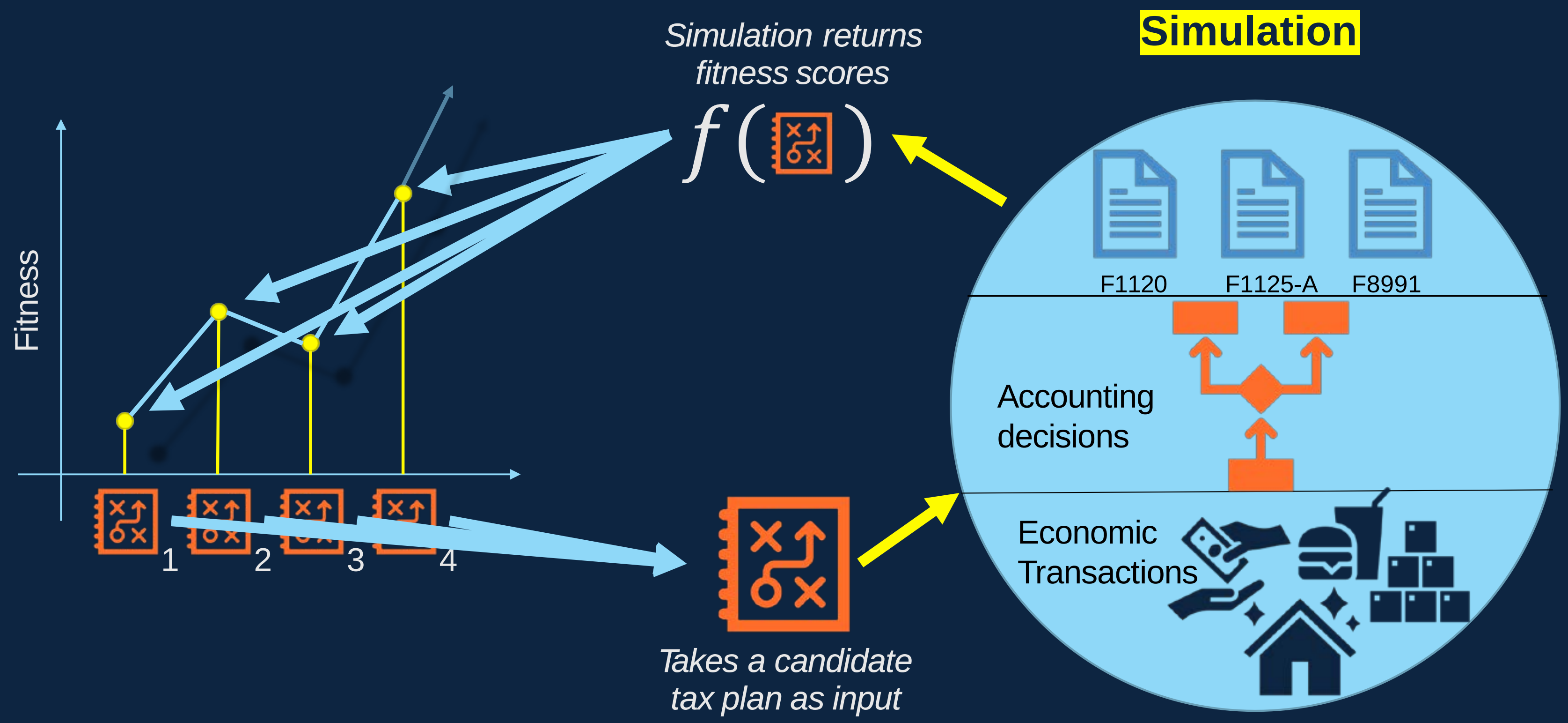


If we can **compute a score** over a space of inputs…

then we can use a **search algorithm** to maximize that score

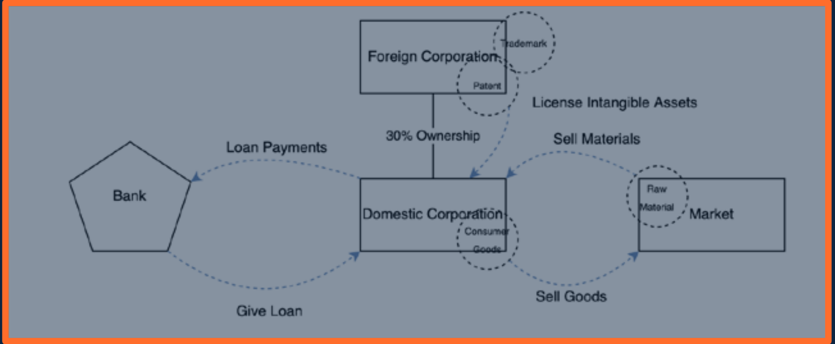
Takes a candidate tax plan as input

Behavior Discovery Simulation



BEAT Scenario – Preliminary Results

- Algorithm discovered a way for the entity to **avoid**:
  - Additional tax liability due to BEAT and
  - Detection by integrating a risk function (audit filter)

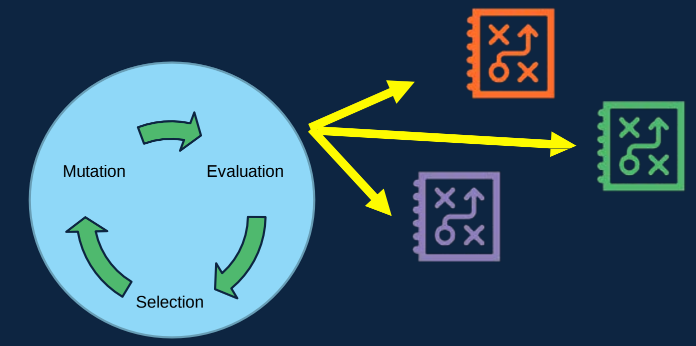


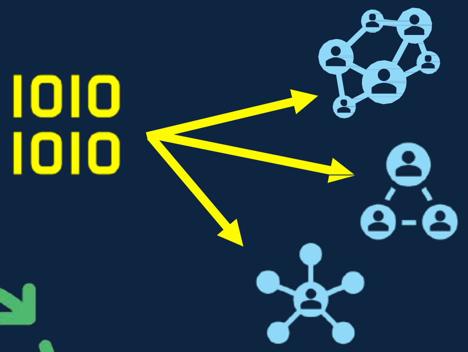
Action	Target Entity	Parameters
Trade	Foreign Corporation	Royalty Contract 1
Payment	Foreign Corporation	$1,050
Trade	Foreign Corporation	Royalty Contract 2
Payment	Foreign Corporation	$280
Trade	Market	Buy RawMaterial for $5,000
Produce	—	Produce a ConsumerGood
Trade	Market	Sell ConsumerGood for $7,000
Trade	Market	Buy RawMaterial for $5,000
Trade	Market	Buy RawMaterial for $5,000
Produce	—	Produce a ConsumerGood
Trade	Market	Sell ConsumerGood for $7,000
Produce	—	Produce a ConsumerGood
Trade	Market	Sell ConsumerGood for $7,000
Payment	Foreign Corporation	$2,100 on Royalty 1
Payment	Foreign Corporation	$560 on Royalty 2

MITRE

Discussion

1. Validation of the evolutionary discovery approach as a method for predicting taxpayer behavior
2. Demonstration of model extensibility across tax policy scenarios using modular components
   - Insights from each scenario drove down the time required for subsequent scenario development
3. Creation of reusable common resources for future needs (e.g., output results report templates, experimental notebooks, and user story templates)





MITRE

Limitations

- Computational inefficiency
- Large levels of tax domain knowledge required
- Some of current mechanics require maturation

Future Research

- Model verification and validation
- Extension of framework to other tax policy scenarios
- Documentation for future users of the system: the end user, an intermediate user, and a fundamental developer.

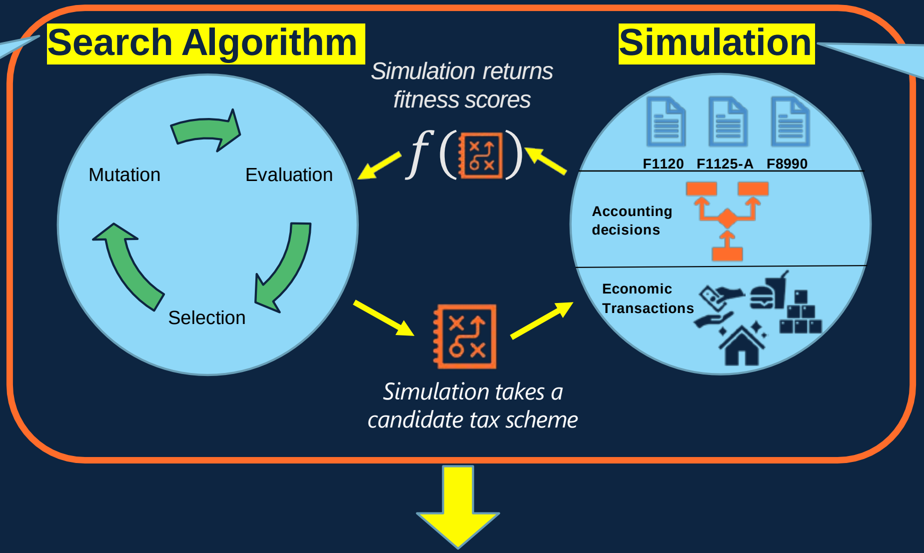
Contact us:
Siggy Scott - escott@mitre.org
Camrynn Fausey, Ph.D. - cfausey@mitre.org
Karen Jones - kajones@mitre.org
Geoff Warner, Ph.D. - gwarner@mitre.org
Hahnemann Ortiz - hahnemann.ortiz@irs.gov

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™

Appendix

Evolutionary Behavior Discovery

Proposes sequences of actions



Evaluates consequences of actions

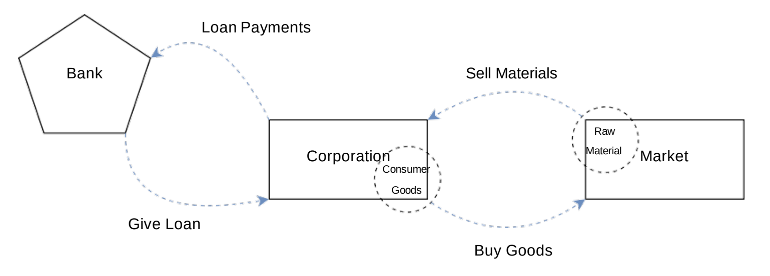
Our simulation framework captures **essential elements** of the tax process:

- Business entities
- Trades, goods, inventory
- Banks & loans
- Key values reported on tax forms

A **discovery method** for understanding how taxpayer behavior might **evolve** to lower tax liability while avoiding detection in response to new or proposed laws.

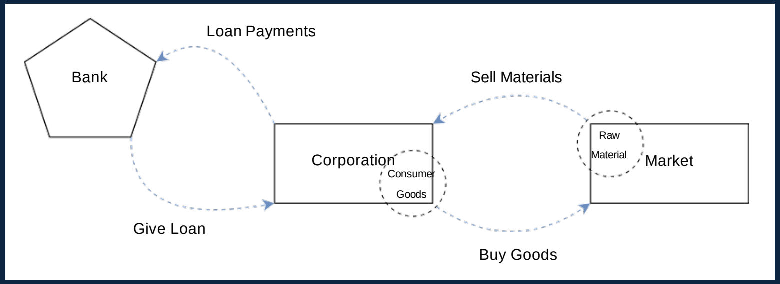
Simple-Interest Scenario – Set up

- **Fitness function** maximizes entities **cash-on-hand** and minimizing its **tax liability** and outstanding **debt**
- We include a **penalty term** that discourages the evolutionary algorithm from creating action sequences with a large number of "**dead actions**." Quantitatively, we express these incentives as a linear combination



Simple-Interest Scenario

- Can we ***model interactions*** between a corporation and a market and between a corporation and a bank?
- Evolutionary ***algorithm discovered*** that a corporation needs to take out a loan to buy raw material and sell processed goods to earn profit in order to pay down their loan.



Now that we have a base scenario resulting in expected behaviors, we can start adding other modular components for additional complexity.

Simple-Interest Scenario – Preliminary Results

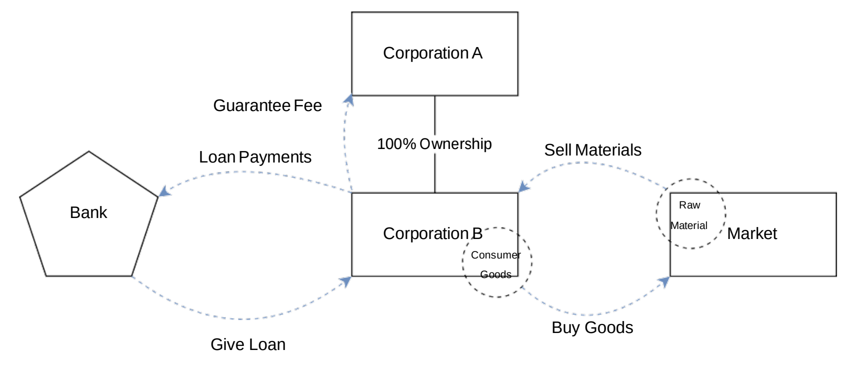
- Evolutionary algorithm discovers that it needs to take out a loan to buy raw material and sell processed goods to earn profit in order to pay down loan

Action	Target Entity	Parameters
TakeOutLoan	Bank	$100,000, interest rate 0.05
Payment	Bank	$10,000
Payment	Bank	$10,000
Trade	Market	Buy RawMaterial for $8,000
Payment	Bank	$10,000
Payment	Bank	$10,000
Payment	Bank	$10,000
Trade	Market	Buy RawMaterial for $8,000
Payment	Bank	$10,000
Payment	Bank	$10,000
Payment	Bank	$10,000

MITRE

Guarantee-Fee Scenario

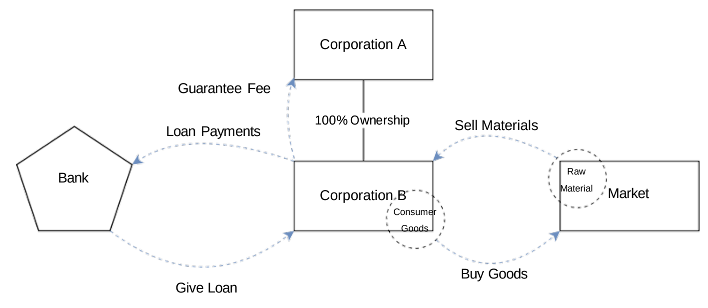
- Can we ***expand*** the simple interest scenario by introducing an alternative decision for a corporation to secure a loan, adding a ***noncompliant decision space***?
- Evolutionary ***algorithm discovered*** that a corporation needs to take a loan by paying a related entity a guarantee fee to buy raw material in order to ***increase its tax deduction*** compared to if it took a loan directly from a bank.



Now that we have an expanded base scenario resulting in expected noncompliant behaviors, we can add 1) new components or 2) nuances to current components, resulting in complex dynamics and an expanded noncompliant decision space.

Guarantee-Fee Scenario – Set up

- **Fitness function** maximizes entities **cash-on-hand** and minimizing its **tax liability** and outstanding **debt**
- We incentivize the algorithm to focus less on paying off debt than we did in the previous scenario, and more on minimizing tax liability



MITRE

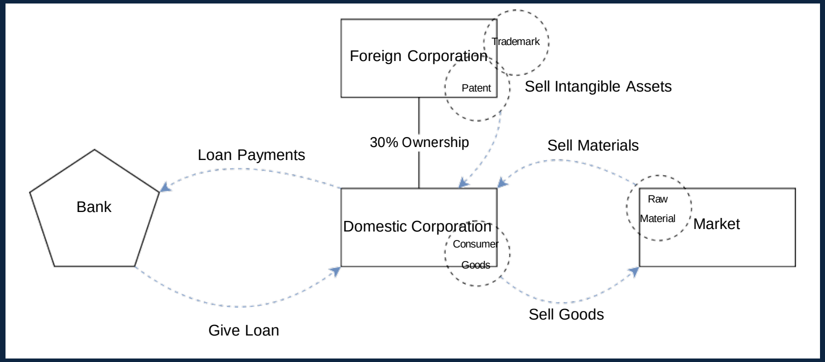
Guarantee-Fee Scenario – Preliminary Results

- Evolutionary algorithm discovers that it needs to take a loan by paying a related entity a guarantee fee to buy raw material, resulting in an increased deduction that could be filed on the tax return compared to a loan directly from a bank.

Action	Target Entity	Parameters
TakeOutLoan	Bank	$100,000 guaranteed loan at 3% interest
Payment	Corporation A	Pay a $2,000 guarantee fee
Trade	Market	Buy RawMaterial for $5,000
Produce	—	Produce a ConsumerGood
Trade	Market	Sell ConsumerGood for $15,000
Trade	Market	Buy RawMaterial for $5,000
Produce	—	Produce a ConsumerGood
Trade	Market	Sell ConsumerGood for $15,000
Payment	Bank	$10,000
Trade	Market	Buy RawMaterial for $5,000
Payment	Bank	$10,000

Base Erosion Anti-Abuse (BEAT) Scenario

- Can we ***add dynamics and entity decisions*** by ***expanding*** the simple interest and guarantee fee scenarios by introducing complex transactions that could increase the ***noncompliant decision space***?
- Evolutionary ***algorithm discovered*** a way for the entity to avoid**:**
  - Additional tax liability due to BEAT and
  - Detection by integrating a risk function (audit filter)



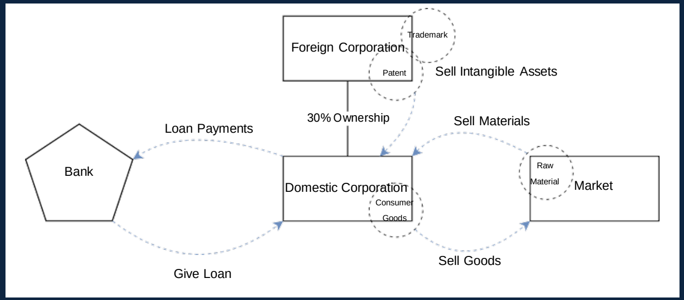
Example Audit Filter Criteria:

- Entities where a foreign corporation owns 30% of a domestic corporation
- The domestic corporation exhibits > X % difference in line items of interest (e.g., cost of goods sold) when comparing to prior tax returns

MITRE

Base Erosion Anti-Abuse (BEAT) Scenario – Set up

- **Fitness function** maximizes entities **cash-on-hand** and minimizing its **tax liability**
- Simulation is run for 3 simulated years (2016-2018)



June 16, 2022

Operationalizing the Indirect Effect of Audits

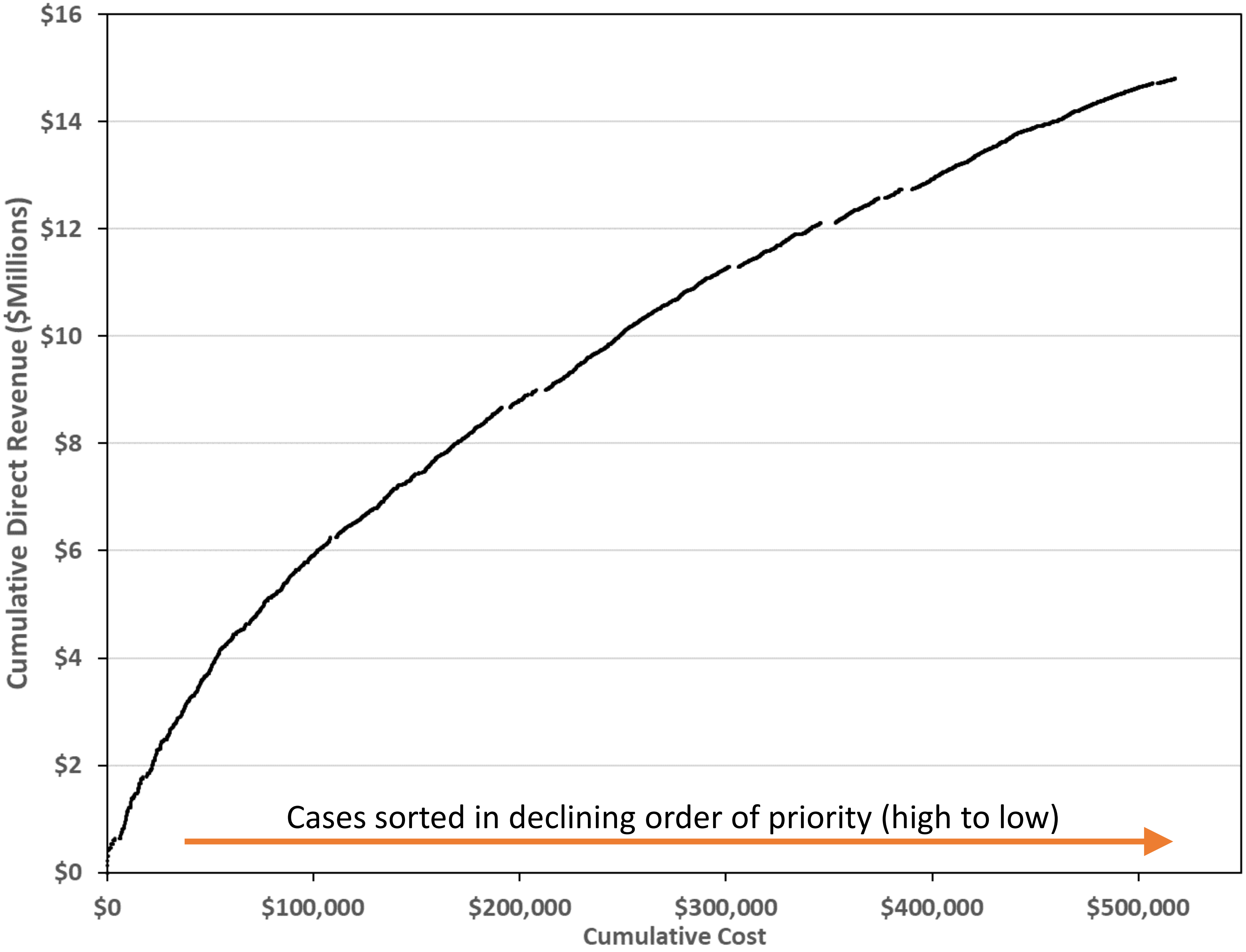
IRS-TPC Research Conference on Tax Administration

Alan Plumley *and Daniel Rodriguez (IRS), and Leigh Nicholl (The MITRE Corporation)*

Research, Applied Analytics & Statistics

KNOWLEDGE DEVELOPMENT & APPLICATION

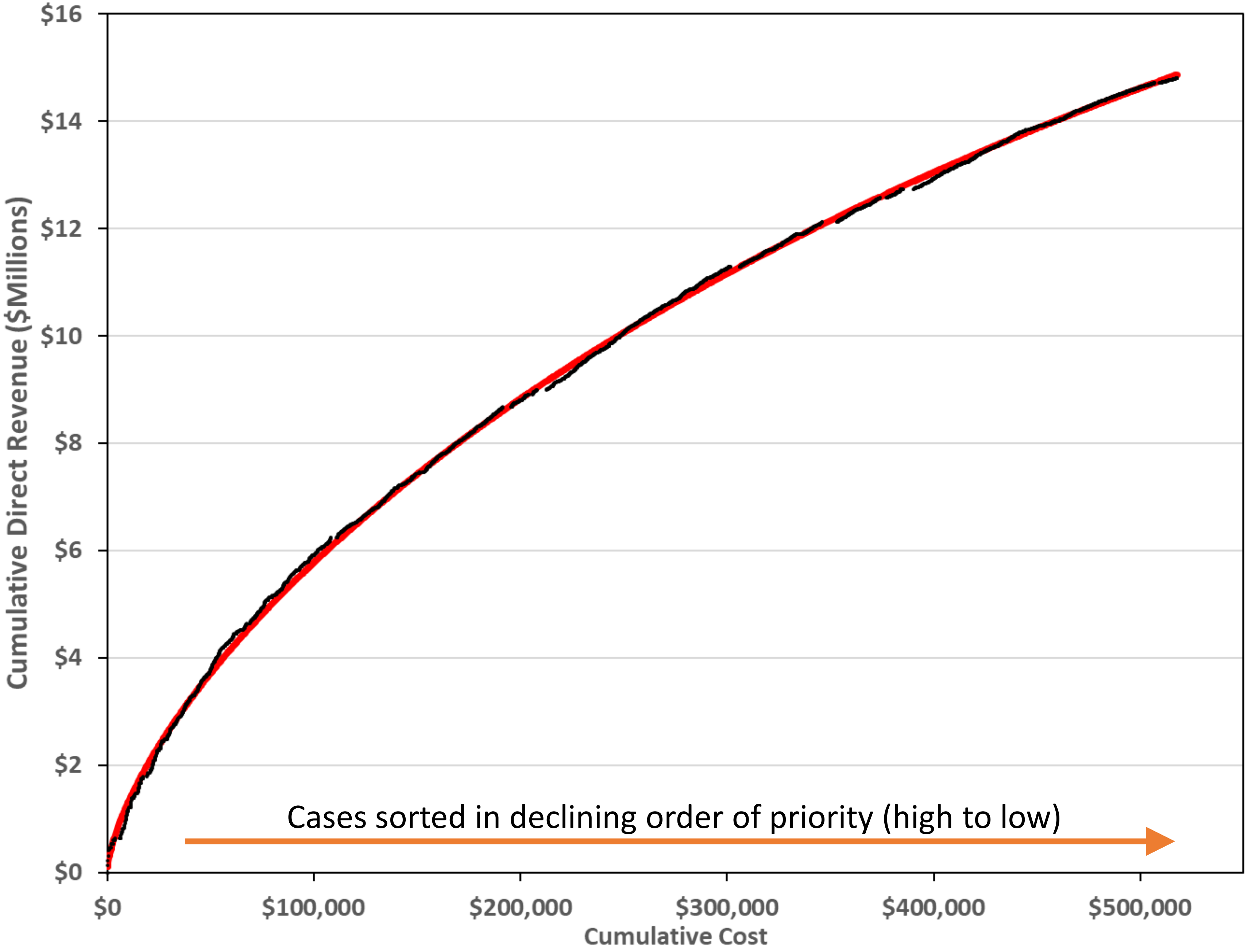
MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™



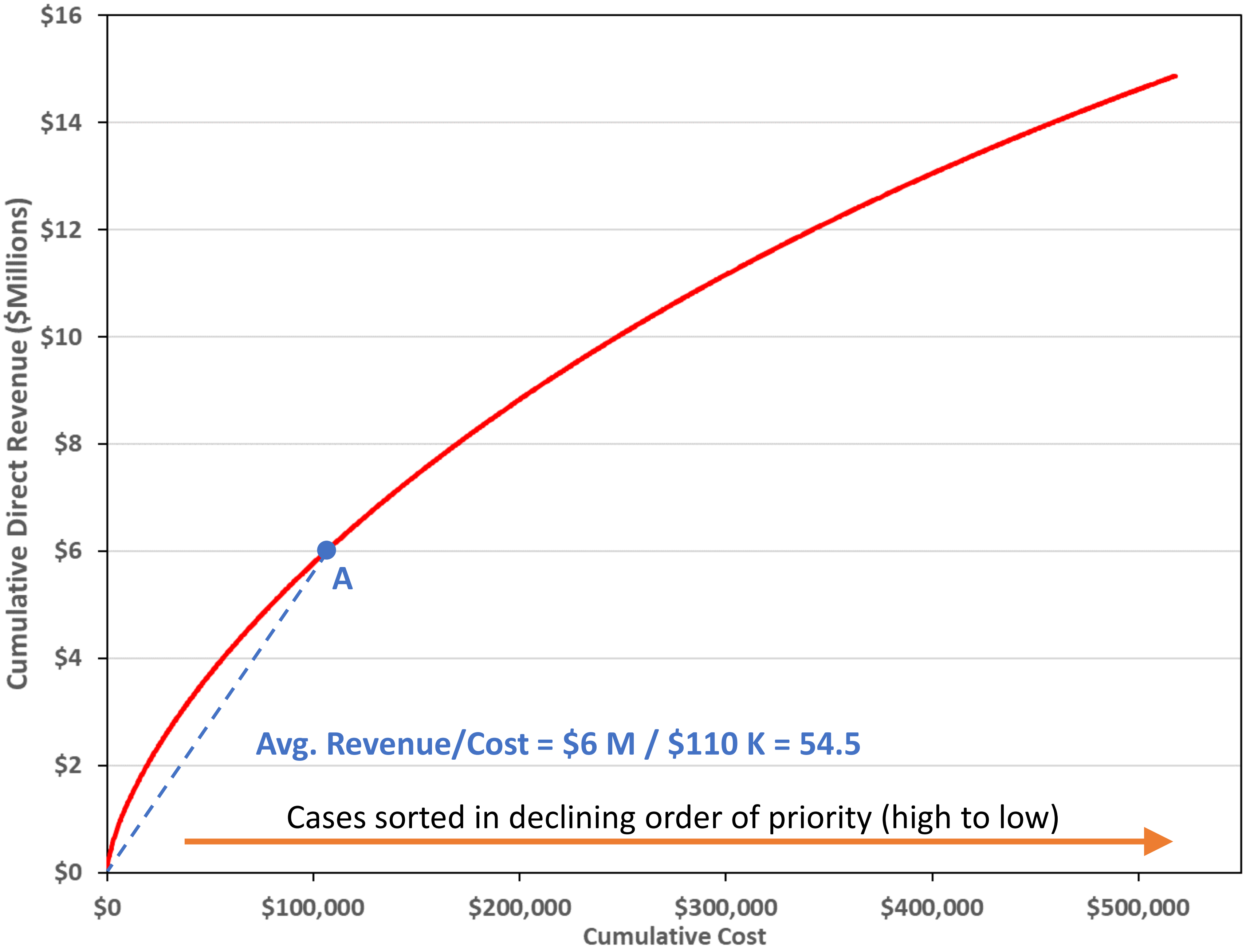
Typical plot of Correspondence Audit outcome (**direct tax revenue**) vs. the **cost** to achieve that outcome for one audit category

Direct tax revenue: The additional tax paid by the audited taxpayer for the year under audit

Cost: The full life-cycle cost of Examination, Appeals, Counsel, and Collection activity related to the audit



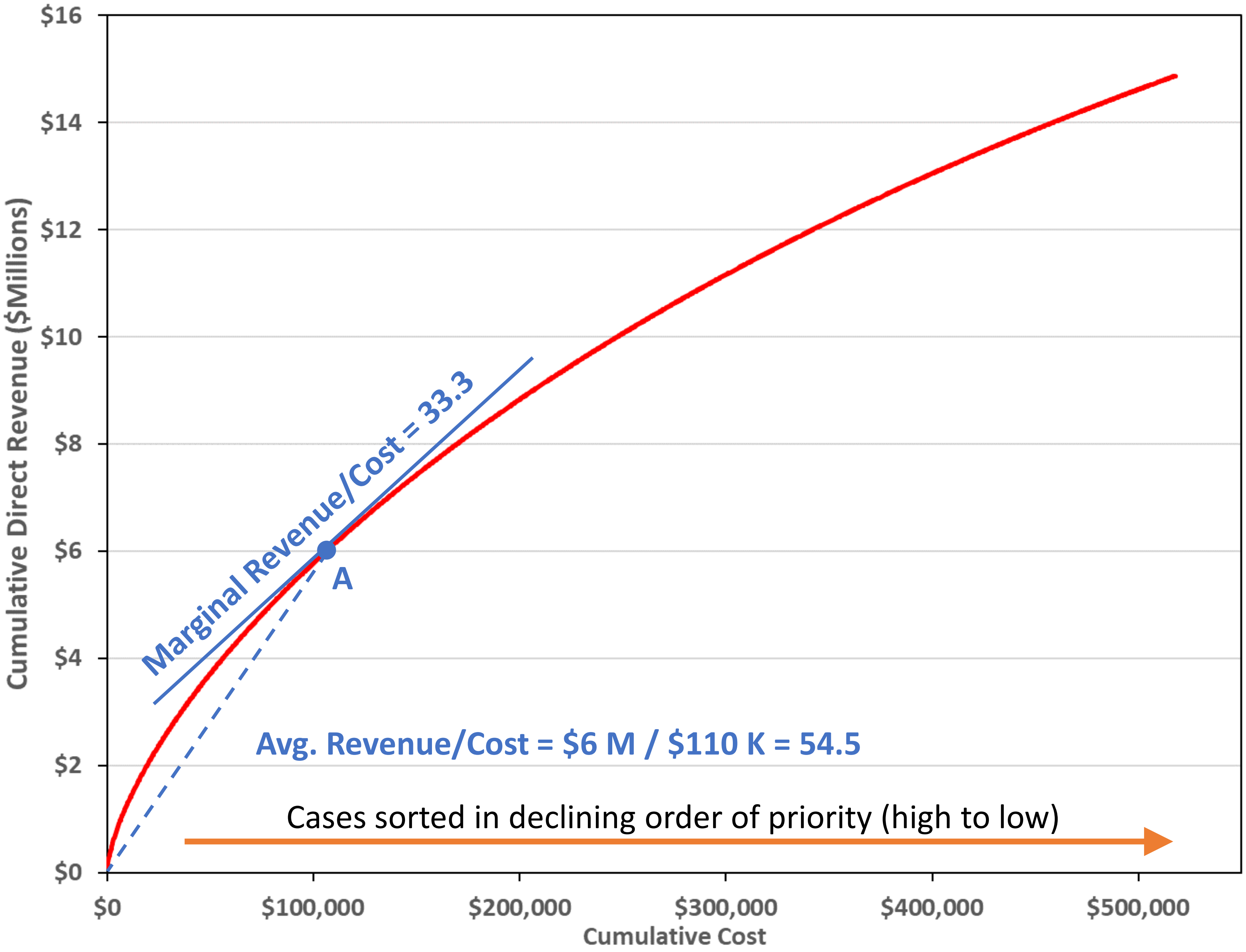
We fit a curve through the revenue vs. cost data



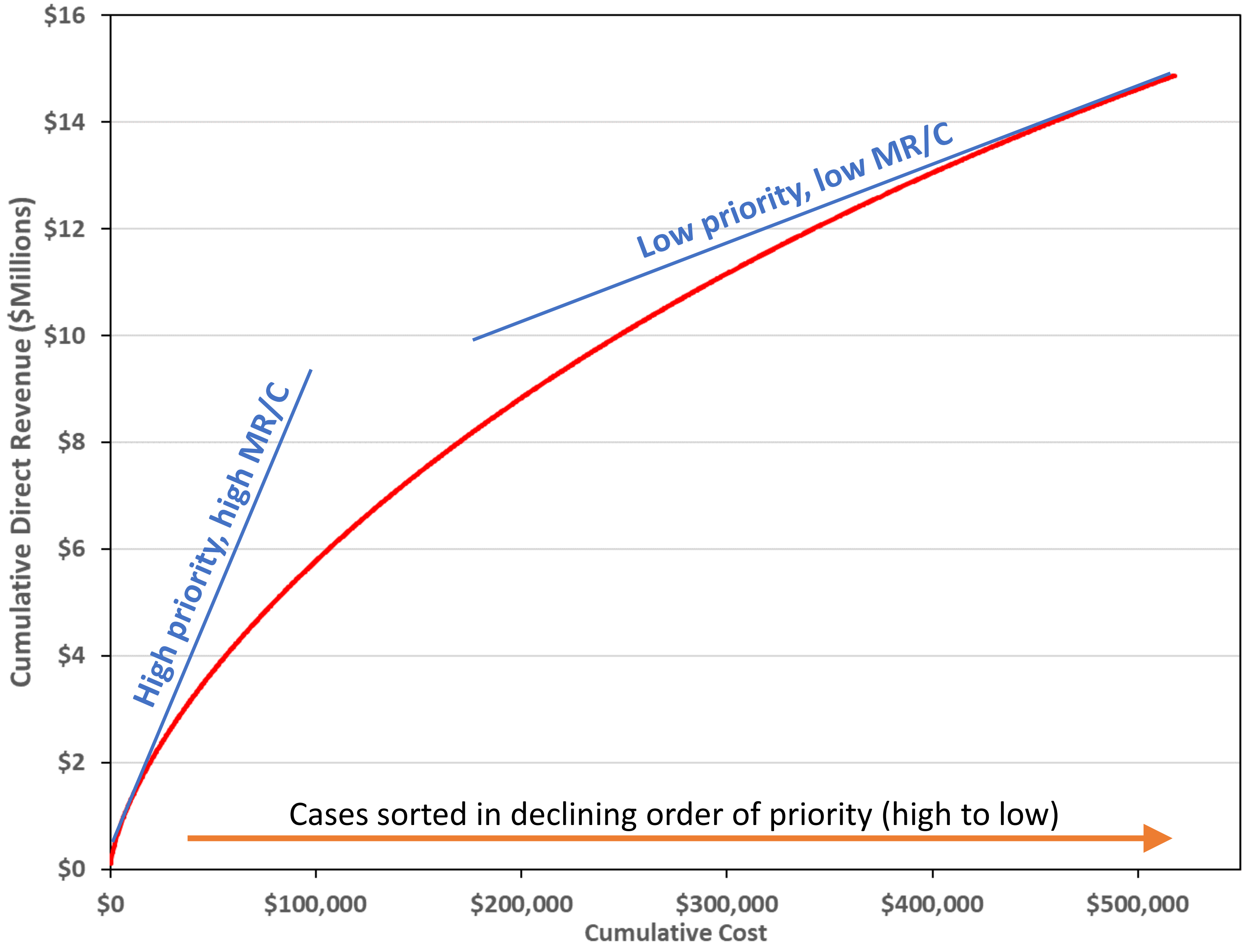
The fitted curve alone

At any point A on the curve, **average revenue/cost** is the slope of the line from the origin to point A

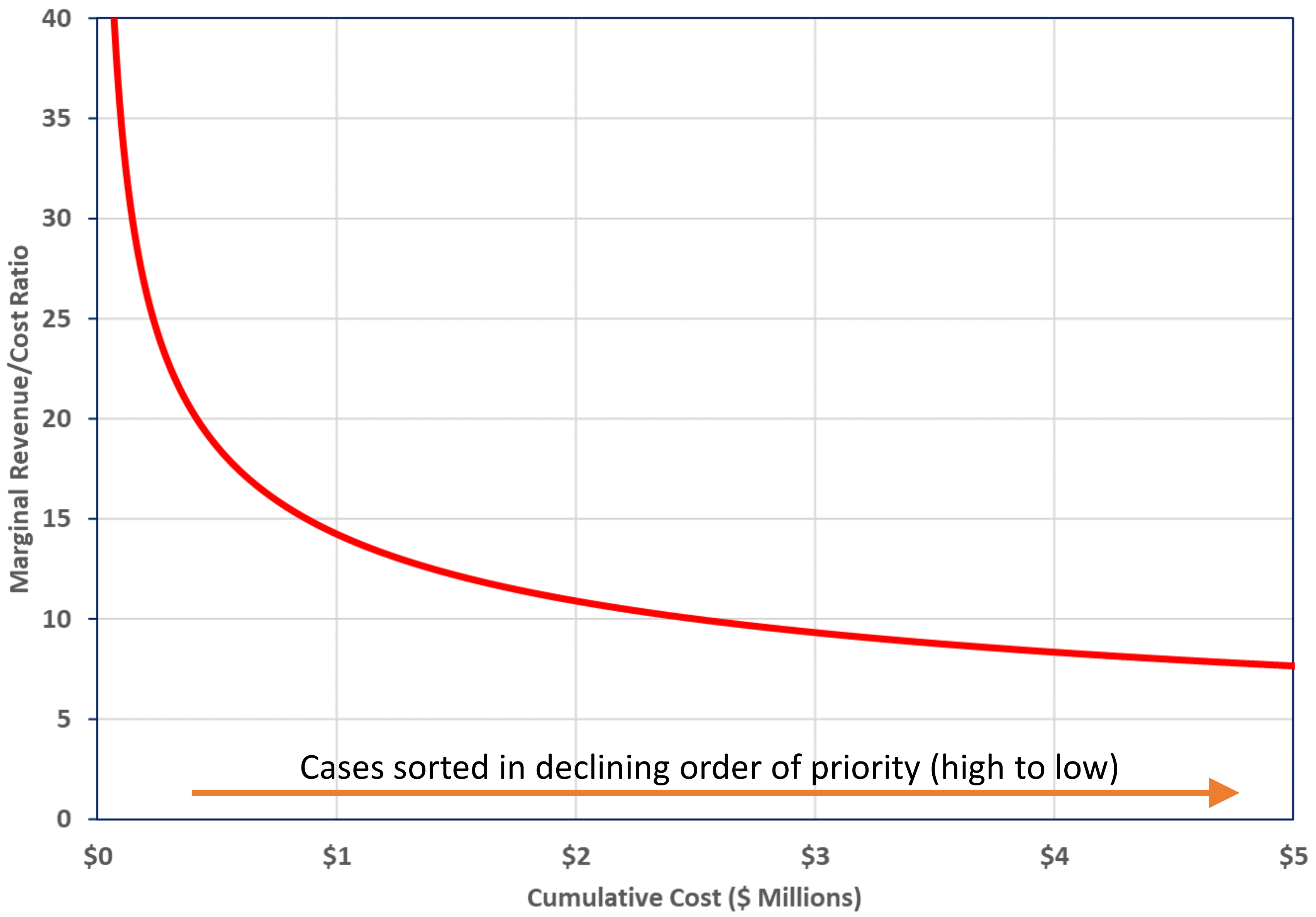
Average revenue/cost typically declines with increasing cumulative cost



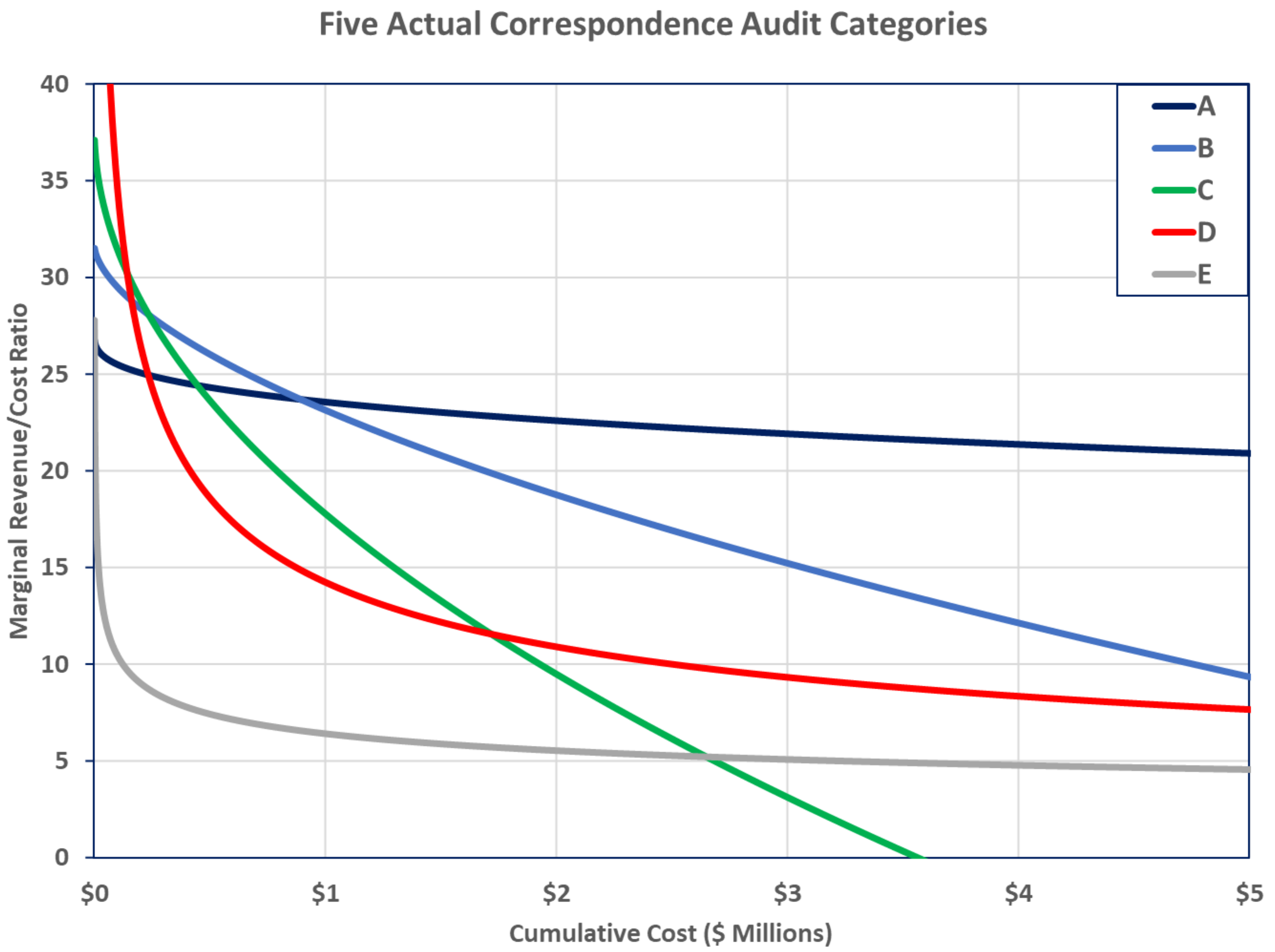
When deciding how to allocate the budget, what's important is how the ***marginal*** revenue/cost ratio (the slope of the curve) varies with cumulative cost. It tells us at any given point on the curve **how much additional (or less) revenue to expect if one more (or less) dollar of budget were allocated** to this category.



If case selection is better than random, then Marginal Revenue/Cost declines as the budget allocation increases.



This is for the same audit category; the only difference is that the Y-axis is now MR/C rather than Cumulative Revenue

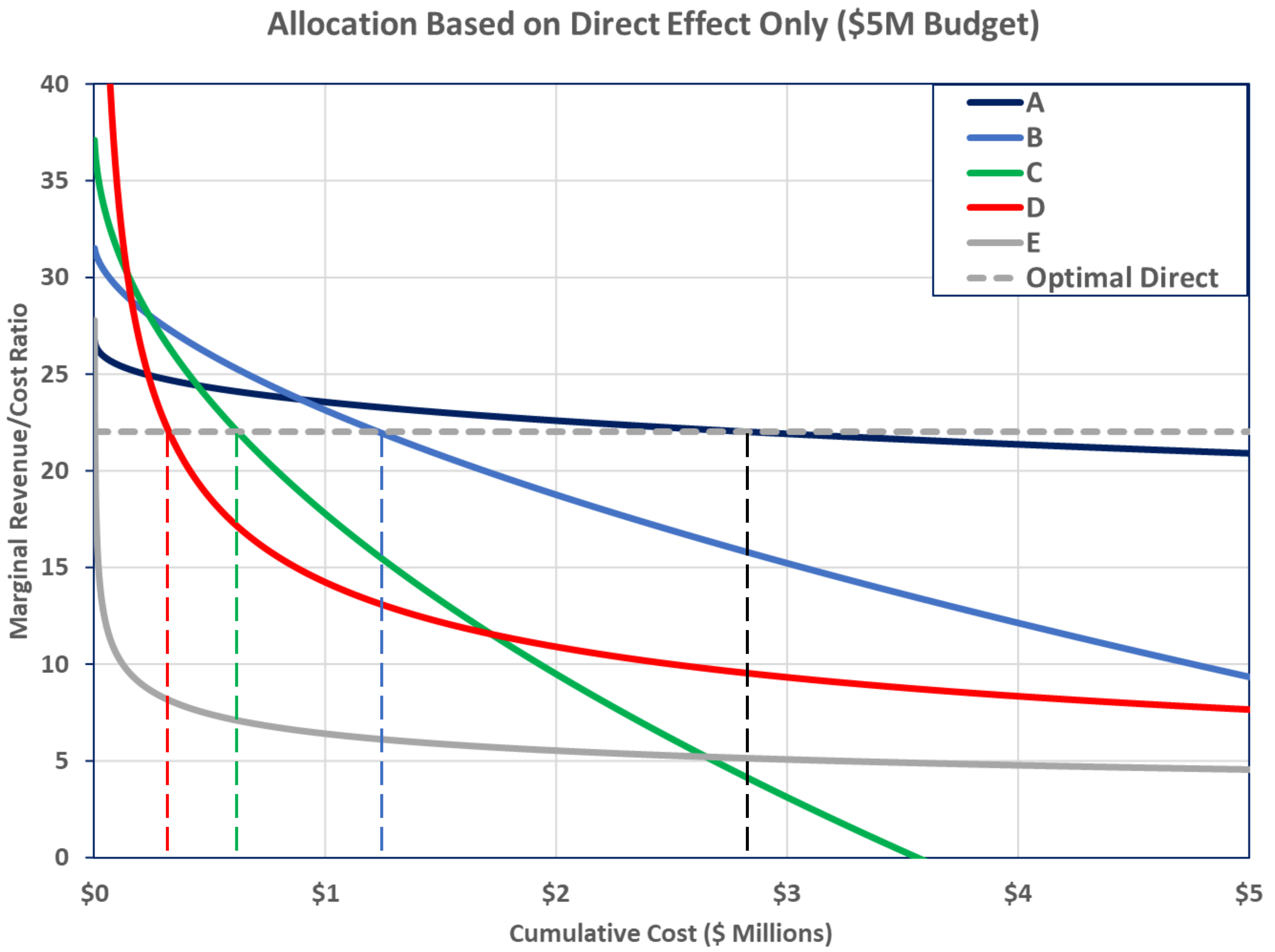


How should a $5 M budget be allocated to these 5 categories to maximize revenue (assuming that budget is the only constraint)?

Revenue is maximized when:

1. **All of the budget** is allocated; and
2. The **marginal revenue/cost ratio is equalized** across all unconstrained categories.

If the MR/C ratio is not equalized, then additional revenue could be achieved by shifting resources from categories with lower MR/C to the category with the highest MR/C.



If all categories share the same MR/C, then the budget for each category is where the dashed line intersects that category's curve.

Category	Optimal Allocation of $5 M	Direct Revenue ($M)
A	$2.82	$65.9
B	$1.23	$31.5
C	$0.62	$16.9
D	$0.32	$11.6
E	$0.003	$0.1
Sum	**$5.00 M**	**$126.0**

Now, how do we incorporate indirect effects?

In a prior study,* we estimated the average **specific indirect effect** for multiple correspondence audit categories. The specific indirect effect is the subsequent change in reporting behavior of taxpayers who have experienced an enforcement activity.

We have updated and added to those. Here are the results for our 5 categories.

Category	Avg. Indirect Revenue
A	$4,479.57
B	$4,282.58
C	$1,465.31
D	$118.50
E	$501.66

* *Nicholl, Lykke, McGill, and Plumley (2020 IRS-TPC Annual Research Conference)*

Now, how do we incorporate indirect effects?

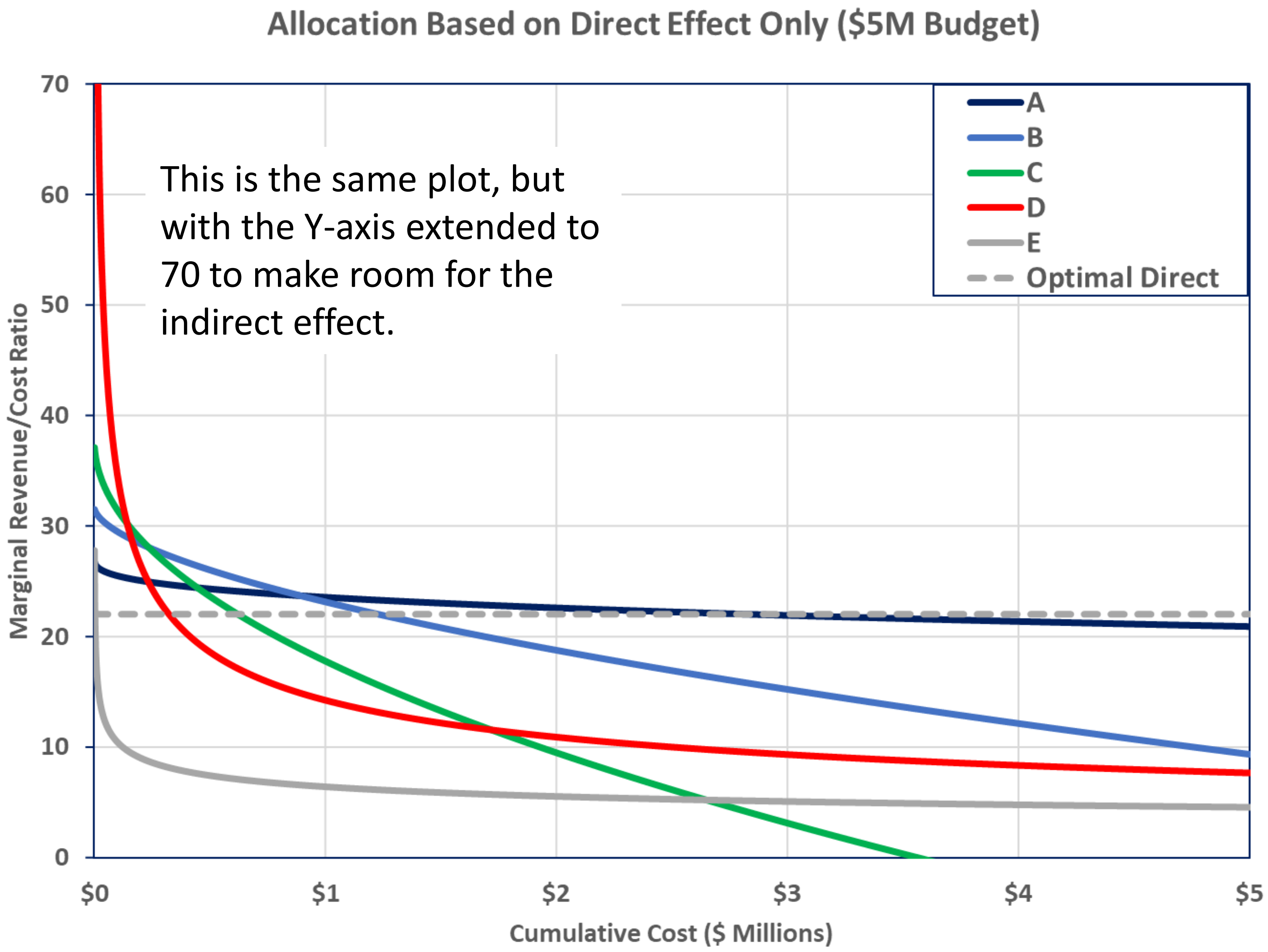
In a prior study,* we estimated the average **specific indirect effect** for multiple correspondence audit categories. The specific indirect effect is the subsequent change in reporting behavior of taxpayers who have experienced an enforcement activity.

We have updated and added to those. Here are the results for our 5 categories.

Category	Avg. Indirect Revenue	Avg. Cost	Avg. Indirect Revenue/Cost
A	$4,479.57	$158.14	28.3
B	$4,282.58	$161.02	26.6
C	$1,465.31	$156.86	9.3
D	$118.50	$158.63	0.7
E	$501.66	$109.47	4.6

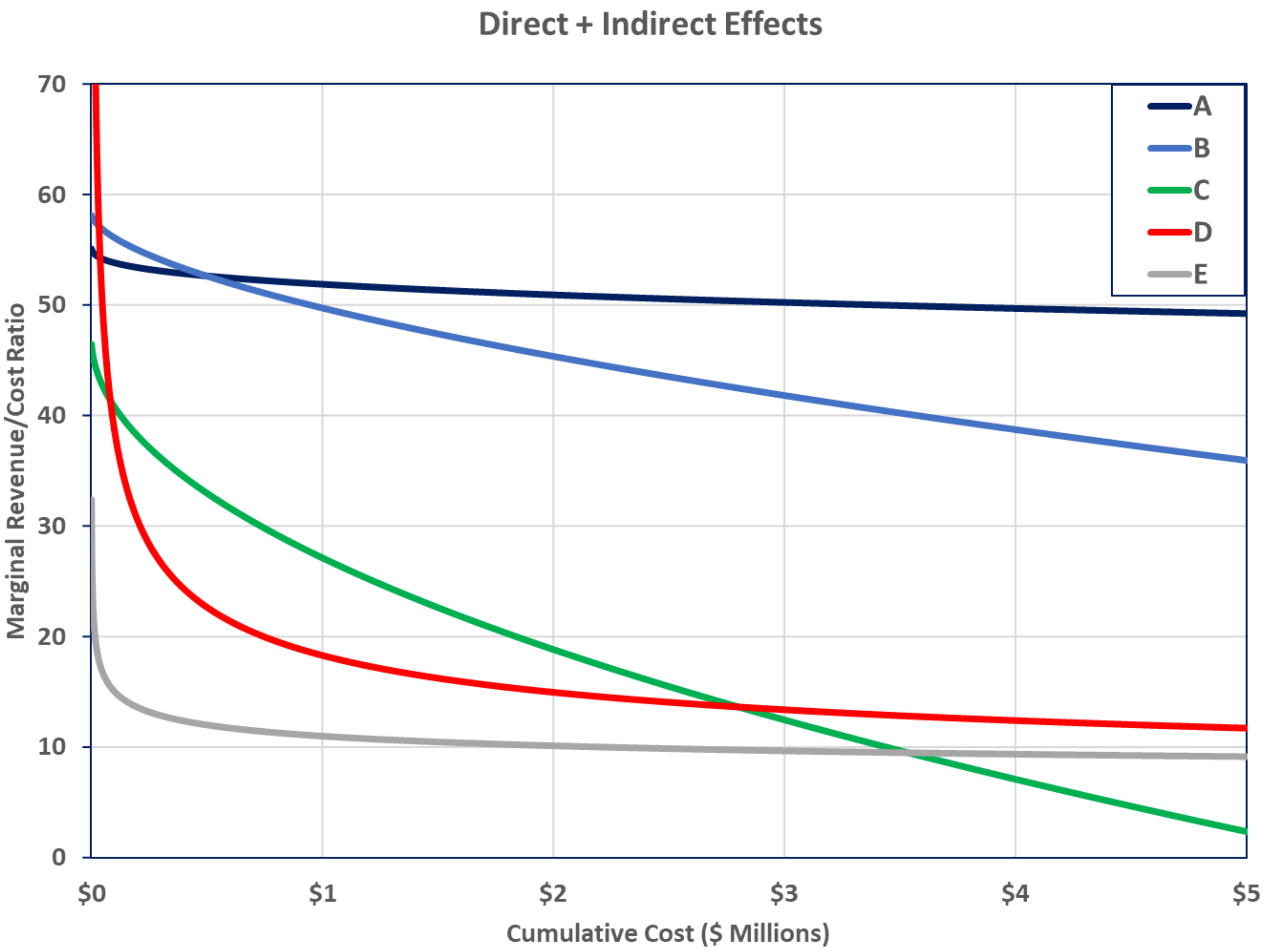
We know the average cost in each category, so we can compute the average indirect revenue/cost ratio for each.

* *Nicholl, Lykke, McGill, and Plumley (2020 IRS-TPC Annual Research Conference)*

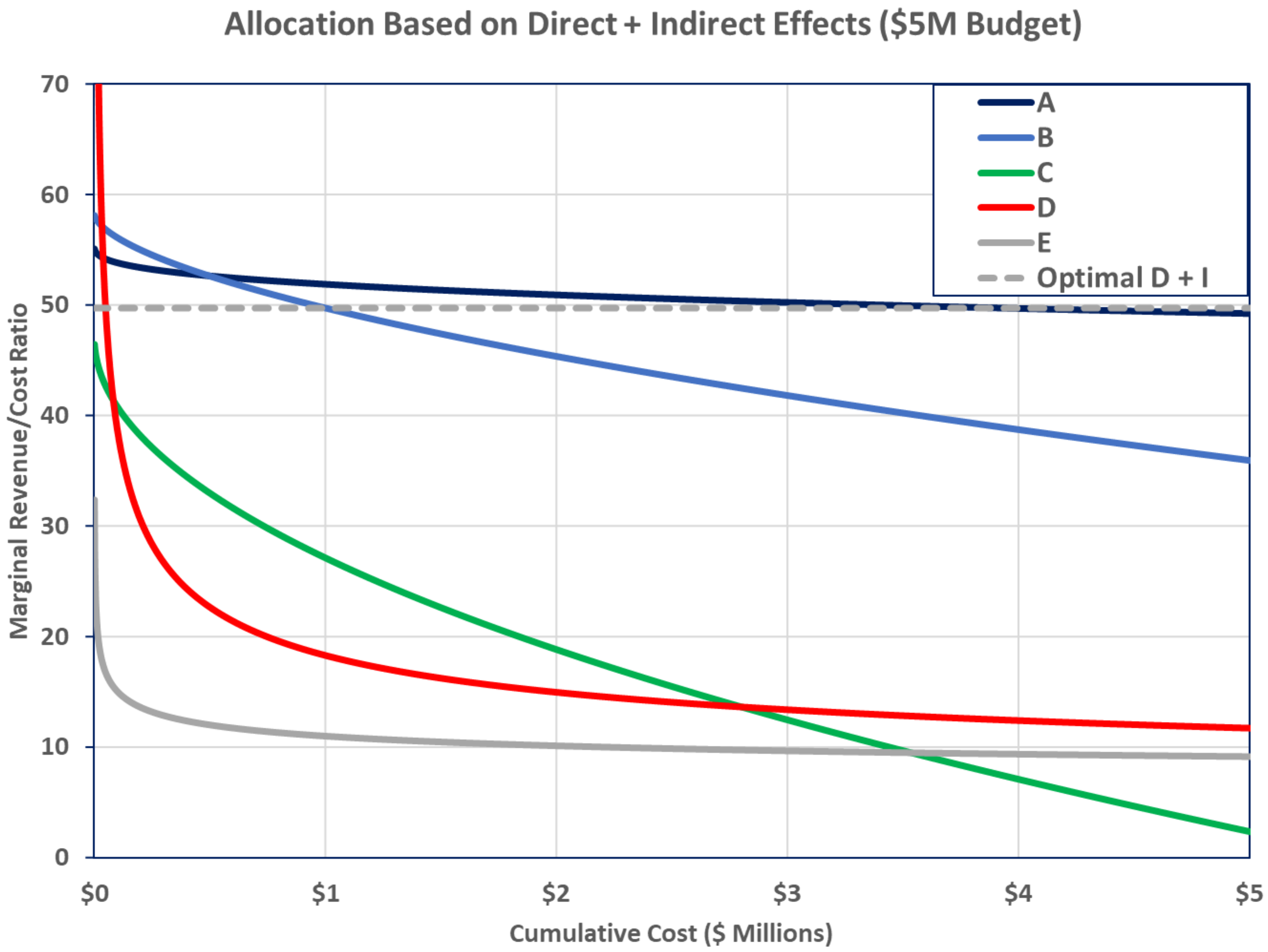


How can we take the indirect effects into account?

Because the indirect effects we've estimated are the *average* effect for all cases within a category, adding the indirect effect merely shifts the direct MR/C curves upwards…



Because the estimated indirect effect varies across the 5 categories, each curve is shifted upward by a different amount.



If all categories share the same MR/C, then the budget for each category is where the dotted line intersects that category's curve.

Category	Optimal Allocation of $5 M	D + I Revenue ($ M)
A	$3.96	$202.1
B	$1.00	$53.1
C	$0.00	$0.0
D	$0.04	$3.8
E	$0.00	$0.0
Sum	**$5.00 M**	**$258.9**

Impact of Accounting for Indirect Effects, $5 M Budget

$ in Millions	Optimal for Direct Effect Only		Optimal for Direct + Indirect Effect		Difference in Revenue
	Optimal Budget Allocation	Revenue	Optimal Budget Allocation	Revenue	
A	$2.82	$65.87	$3.96	$90.42	$24.55
B	$1.23	$31.50	$1.00	$26.42	-$5.08
C	$0.62	$16.94	$0	$0	-$16.94
D	$0.32	$11.58	$0.04	$3.24	-$8.34
E	$.003	$0.084	$0.0001	$0.006	-$0.077
Total Direct	$5.00	$125.97	$5.00	$120.08	-$5.89

Impact of Accounting for Indirect Effects, $5 M Budget

$ in Millions	Optimal for Direct Effect Only		Optimal for Direct + Indirect Effect		Difference in Revenue
	Optimal Budget Allocation	Revenue	Optimal Budget Allocation	Revenue	
A	$2.82	$65.87	$3.96	$90.42	$24.55
B	$1.23	$31.50	$1.00	$26.42	-$5.08
C	$0.62	$16.94	$0	$0	-$16.94
D	$0.32	$11.58	$0.04	$3.24	-$8.34
E	$.003	$0.084	$0.0001	$0.006	-$0.077
Total Direct	$5.00	$125.97	$5.00	$120.08	-$5.89
A	$2.82	79.98	$3.96	$112.04	$32.05
B	$1.23	32.69	$1.00	$26.70	-$5.99
C	$0.62	5.80	$0	$0	-$5.80
D	$0.32	0.24	$0.04	$0.03	-$0.21
E	$.003	0.014	$0.0001	$0.005	-$0.009
Total Indirect	$5.00	$118.73	$5.00	$138.77	$20.04
Total D+I	$5.00	$244.70	$5.00	$258.85	$14.15

Impact of Accounting for Indirect Effects, $5 M Budget

$ in Millions	Optimal for Direct Effect Only		Optimal for Direct + Indirect Effect		Difference in Revenue
	Optimal Budget Allocation	Revenue	Optimal Budget Allocation	Revenue	
A	$2.82	[illegible]	[illegible]	[illegible]	$24.55
B	$1.23	[illegible]	[illegible]	[illegible]	-$5.08
C	$0.62	[illegible]	[illegible]	[illegible]	-$16.94
D	$0.32	[illegible]	[illegible]	[illegible]	-$8.34
E	$.003	$0.084	$0.0001	[illegible]	-$0.077
Total Direct	$5.00	$125.97	$5.00	$120.08	-$5.89
A	$2.82	79.98	$3.96	$112.04	$32.05
B	$1.23	32.69	$1.00	$26.70	-$5.99
C	$0.62	5.80	$0	$0	-$5.80
D	$0.32	0.24	$0.04	$0.03	-$0.21
E	$.003	0.014	$0.0001	$0.005	-$0.009
Total Indirect	$5.00	$118.73	$5.00	$138.77	$20.04
Total D+I	$5.00	$244.70	$5.00	$258.85	$14.15

Accounting for indirect effects will always reduce the total direct revenue achievable compared with if we had allocated only on the basis of the direct effect (in this example, a drop of $5.89M)

Impact of Accounting for Indirect Effects, $5 M Budget

$ in Millions	Optimal for Direct Effect Only		Optimal for Direct + Indirect Effect		Difference in Revenue
	Optimal Budget Allocation	Revenue	Optimal Budget Allocation	Revenue	
A	$2.82	[illegible]	[illegible]	[illegible]	$24.55
B	$1.23	[illegible]	[illegible]	[illegible]	-$5.08
C	$0.62	[illegible]	[illegible]	[illegible]	-$16.94
D	$0.32	[illegible]	[illegible]	[illegible]	-$8.34
E	$.003	$0.084	$0.0001	[illegible]	-$0.077
Total Direct	$5.00	$125.97	$5.00	$120.08	-$5.89
A	$2.82	79.98	$3.96	$112.04	$32.05
B	$1.23	[illegible]	[illegible]	[illegible]	-$5.99
C	$0.62	[illegible]	[illegible]	[illegible]	-$5.80
D	$0.32	[illegible]	[illegible]	[illegible]	-$0.21
E	$.003	0.014	$0.0001	[illegible]	-$0.009
Total Indirect	$5.00	$118.73	$5.00	$138.77	$20.04
Total D+I	$5.00	$244.70	$5.00	$258.85	$14.15

Accounting for indirect effects will always reduce the total direct revenue achievable compared with if we had allocated only on the basis of the direct effect (in this example, a drop of $5.89M)

But this is accompanied by a larger increase in the indirect revenue (in this example, an increase of $20.04M for a net gain of $14.15M)

Impact of Accounting for Indirect Effects, $5 M Budget

$ in Millions	Optimal for Direct Effect Only		Optimal for Direct + Indirect Effect		Difference in Revenue
	Optimal Budget Allocation	Revenue	Optimal Budget Allocation	Revenue	
					$24.55
					-$5.08
					-$16.94
					-$8.34
		$0.084	$0.0001		-$0.077
		$125.97	$5.00	$120.08	-$5.89
		79.98	$3.96	$112.04	$32.05
					-$5.99
					-$5.80
					-$0.21
		0.014	$0.0001		-$0.009
Total Indirect	$5.00	$118.73	$5.00	$138.77	$20.04
Total D+I	$5.00	$244.70	$5.00	$258.85	$14.15

This is completely analogous to the current practice of imposing minimum coverage constraints—sacrificing some direct revenue in order to promote even greater indirect revenue (voluntary compliance)—but here we are doing it with empirical estimates of indirect effects rather than subjective constraints.

Accounting for indirect effects will always reduce the total direct revenue achievable compared with if we had allocated only on the basis of the direct effect (in this example, a drop of $5.89M)

But this is accompanied by a larger increase in the indirect revenue (in this example, an increase of $20.04M for a net gain of $14.15M)

How to Handle Special Cases

1. If we estimate a negative indirect effect

- Negative suggests that audited taxpayers become *more noncompliant* following the audit.
- May be true, but test the analysis carefully
    - o Check trends in both the test (audited) and control (not audited) groups:
        - ✓ Reporting of items adjusted in the audit
        - ✓ Total income and total tax
    - o Are the control and test groups not similar?
    - o Does the analysis control for necessary factors?

How to Handle Special Cases

2. If we estimate a **huge** indirect effect

- This would shift most of the budget to this category.
- May be true but test the analysis carefully.
- If true, it may be reasonable to impose a *maximum* budget allocation constraint to this category.
- Re-estimate with more years of data.

What About the General Indirect Effect?

Specific effect is the change in behavior of **an audited taxpayer** following the audit.

- Doesn't depend on the number of audits conducted.

General effect is the change in behavior of taxpayers who are **not audited.**

- It *would* likely depend on the number of audits conducted, so it wouldn't simply move the MR/C curve upward by the same amount everywhere.
- This could be larger than the specific effect.
- May affect both resource allocation *and* case selection.
- Until we have solid empirical estimates, continue using minimum coverage constraints to roughly account for the general effect.

Session 3: Improving Audit Outcomes: Thinking Inside the Box

Discussion of Papers and Future Directions

June 16, 2022

Common Themes Across the Three Papers

- **Graph-Based Machine Learning Methods for Case Selection and Population Segmentation**
  - Matt Olson, Ben Howard, Devika Mahoney-Nair (MITRE); Annette Portz (IRS, RAAS)
- **Automated Discovery of Tax Schemes Using Genetic Algorithms**
  - Karen Jones, Camrynn Fausey, Eric O. Scott, Geoff Warner, Sanith Wijesinghe (MITRE); Hahnemann Ortiz (IRS, LB&I)
- **Incorporating the Specific Indirect Effect of Correspondence Audits Into IRS Resource Allocation Decisions**
  - Alan Plumley, Daniel Rodriguez (IRS, RAAS); Leigh Nichol (MITRE)

Common Themes

- Continued advances in data availability and computing power create new opportunities for the IRS to use analytics to guide enforcement action.
- The value of data-driven insights is enhanced through "human in the loop" processes, which accelerate machine learning through the addition of subject matter expertise.
- New data and analytic methods are enhancing the IRS's ability to detect noncompliance that was previously unobservable (or unknown).

Indirect Effect of Correspondence Audits:
Summary and Key Insights

- IRS audits generate additional tax revenue from audit taxpayers (*direct effect*), but may also influence the future compliance of audit taxpayers (*specific indirect effect*) or unaudited taxpayers (*general indirect effect*)

- The resource allocation that maximizes total revenue (direct + indirect) differs from the allocation that maximizes direct revenue. This paper examines how the IRS could use estimates of specific indirect effects to better allocate audit resources.

- The ultimate objective of the IRS is to maximize total tax revenue collected, subject to a budget constraint. While the IRS has other objectives (e.g., fairness, taxpayer experience), these should be viewed as constraints on revenue maximization.

- Accounting for specific indirect effects could increase audit revenue significantly. In a simple example (5 correspondence audit types, $5M budget), the authors estimate a 5.8% increase in total audit revenue (from $244.7M to $258.9M).

Indirect Effect of Correspondence Audits:
Topics for Consideration and Discussion

- It’s unfortunate that we need to think of the “budget constraint” as fixed. While an improved allocation method increases enforcement revenue by 5.8%, an additional $1 (in the authors’ example) would generate nearly $50!

- Exam planning at the IRS is subject to many constraints (e.g., resource availability across geographic regions and skill categories, minimum required coverage). The methods outlined in the paper will help to quantify the impact of these constraints, but changing the constraints may prove challenging.

- General indirect effects are difficult to quantify, but what does the evidence tell us regarding the relative size of general vs. specific indirect effects, and what are the resulting implications for the allocation of enforcement resources?

- How might taxpayers react to changes in the allocation of enforcement resources? How would it impact the shape of various Marginal Revenue/Cost curves?

Indirect Effect of Correspondence Audits:
Future Areas of Investigation

- Use behavioral research methods to amplify the value created through improved allocation of enforcement resources:
  - Conduct laboratory research to more fully understand the indirect effect of audits
  - Design enforcement treatments to increase their specific (or general) indirect effect
- Expand optimization approach to include other types of enforcement activity, while recognizing dependencies across functions (e.g., Exam, Collection, Appeal)
- Consider synergies with prior research that examines indirect effects across networks of taxpayers (e.g., taxpayers using the same preparer, employers in the same zip code).

Graph-ML for Case Selection and Segmentation: Summary and Key Insights

- The authors examine the use of graph-based machine learning (specifically, graph neural network autoencoders) to support two near-term goals:
  - Guide the selection of Global High Wealth (GHW) audit cases, particularly in situations where the history of audits is too thin to support development of supervised models
  - Improve IRS estimates of the tax gap associated with noncompliant GHW taxpayers
- This work is spurred by the idea that signals of noncompliance are embedded in the networks surrounding GHW taxpayers (e.g., related pass-through entities).
- Using autoencoders is a good solution to not having enough labeled data. It enables more direct comparison of approaches (i.e., graph NN vs non-graph NN) by avoiding the extra complication of how accurate training labels are.

Graph-ML for Case Selection and Segmentation:
Topics for Consideration and Discussion

- The experimental approach of comparing a “vanilla” autoencoder and a GNN-augmented auto encoder is a very clear and straightforward way to demonstrate the added benefit of graph structure.
- The authors express concerns about the autoencoder loss function they are going to use, and that it is “likely unsuitable for the full scope of development over time”.
  - Are there other loss functions that might be more suitable? For example, would it be worth investigating loss functions that specifically try to determine how much of the graph structure can be re-created from the embeddings?

Graph-ML for Case Selection and Segmentation: Future Areas of Investigation

- Will future work transition towards supervised methods (as you accumulate labeled training data from annotators)? Or will you explore more advanced graph autoencoders, like variational graph autoencoders or other new models?

- There could be valuable opportunities for collaboration with supervised GNN work RAAS is doing in the area of Identity Theft (IDT) detection.

- Incorporating human-in-the-loop sampling strategies (such as diversity sampling and uncertainty sampling) to pick samples for human review, in addition to looking at the disagreement between their two models, might amplify the impact of human annotated training samples.

Discovery of Tax Schemes Using Genetic Algorithms: Summary and Key Insights

- Detecting new and novel tax schemes is inherently challenging, as the IRS has no past examples to guide the development of detection strategies. These schemes may contribute substantially to the tax gap until they are eventually discovered.

- The authors seek to automate the discovery of tax schemes using an Exploratory Behavior Discovery (EBD) software framework, which involves two iterative steps:
  1. Use evolutionary algorithms to model changes in the structure and behavior of tax entities—and transactions among these entities—within a defined set of rules.
  2. Simulate the outcomes that result from this evolution, evaluate the resulting tax benefits, and use the results to guide further evolution.

- The authors focus on schemes surrounding Base Erosion and Anti-Abuse Tax (BEAT) and limitations on business interest expense, but the approach could be applied to any area to support the development of new audit filters or detection methods.

Discovery of Tax Schemes Using Genetic Algorithms:
Topics for Consideration and Discussion

- Is it feasible to create a simulation framework that is robust enough to discover new types of noncompliance? What level of fidelity is needed in the characteristics and behavior of the entities, and the rules of their environment?
- To what extent can the simulated noncompliance be observed in the data available to the IRS? What other data might enhance this approach?
- What actions can the IRS take using insights generated through simulation?
  - Accelerate the design and implementation of new examination campaigns
  - Provide insight to auditors working related compliance issues
  - Recommend areas of focus for studies such as the National Research Program
  - Accelerate the issuance of Rulings, Procedures, Internal Revenue Bulletins
  - Anticipate the consequences of new tax policy (perhaps before it's implemented)

Discovery of Tax Schemes Using Genetic Algorithms: Future Areas of Investigation

- Explore various ways of structuring “human in the loop” processes to accelerate model development and refinement, and the use of model findings
  - Use SME input to define entities, transactions, and other simulation elements
  - Use known cases of fraud as “starting points”
  - Provide compliance experts with “case profiles” for exploratory research or audits
- Consider how to balance the fidelity and complexity of the simulation, in order to generate findings that are impactful, actionable, and timely
- Continue to promote the development of a reusable and extensible code base
  - Documentation for various types of users
  - Patterns that can be used to simulate new compliance domains
  - Ability to extend existing patterns and/or integrate across patterns

Session 4:
Why Do Taxpayers Comply?

June 16, 2022

To File or Not to File? What Matters Most?

IRS-TPC Research Conference

Mark Payne (IRS)

With Brian Erard (B. Erard & Associates) and
Tom Hertz, Pat Langetieg, and Alan Plumley (IRS)

What is Nonfiling?

- Failing to file a **required** tax return **on time**
  - **“Required”** if:
    - Gross income > $ threshold set by law for your filing status or
    - Net income from self-employment > $433
  - **“On time”** if filed by the filing deadline (as extended)
- Accounts for **around 9%** of the individual income tax gap
- Total nonfiling = late filing + not filing at all

Why Study the Drivers of Nonfiling?

- May lead to better interventions to prevent nonfiling
- May lead to useful legislative changes

Why Is It So Hard To Study?

- Without a tax return, IRS often lacks the information necessary to establish whether a return was required (e.g., filing status, tax unit composition, self-employment income)
- Every source of data (e.g., IRS or Census) is deficient in some important ways.

Our Prior Approach

Combine two distinct samples of taxpayers having a legal filing obligation:

- A tax administrative sample of approximately 76,000 randomly selected federal income tax **filers** for Tax Year 2010; and
- A sample from the CPS-ASEC of approximately 113,000 individuals representative of **the entire population** (both filers and nonfilers).

Estimate a qualitative response model using the “calibrated probit” methodology

- Relies on differences between characteristics of filers and the characteristics of the overall population of filers and nonfilers in two data sets instead of differences between filers and nonfilers in a single data set (standard probit)

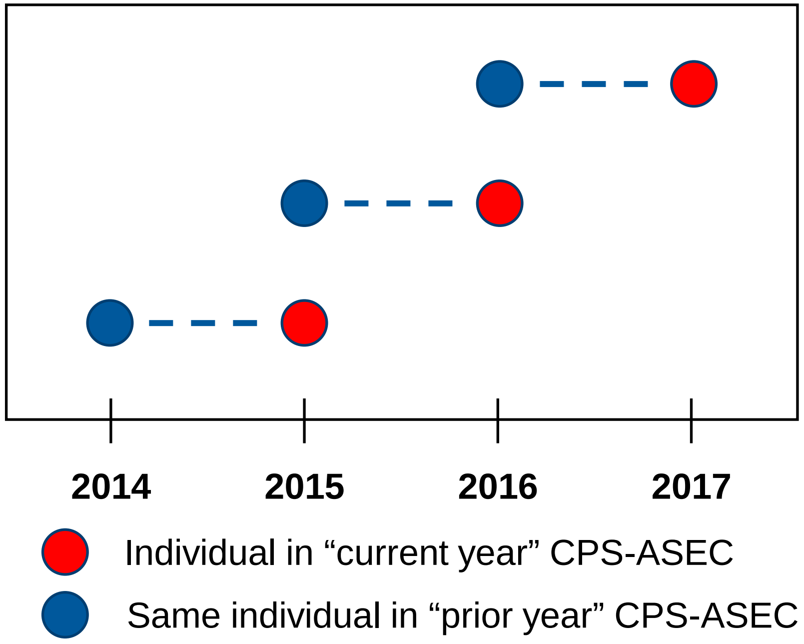
Innovative Data for Current Study

Best of both worlds:

- Detailed demographic data from **Census samples** (to estimate filing threshold, assemble tax units) linked to…
- Detailed micro information on income and fact of filing from **tax administrative data** (to identify requirement to file)

Limited pooled panel structure:

- Most individuals are in CPS-ASEC two consecutive years (half in first year, other half in their second year)
- Better est. of filing req. for prior year



Data Methods

Weights: inverse probability weights for person-level records derived from probit model predicting probability of being in the matched panel from full 16 and over population found in tax administrative data

Probit model for weighting:

- decile position of the amount of each income line item
- geographic region
- age group
- current and prior year filing (timely, late, nonfiler)

Combine person-level records into tax units and assign filing status and children using CPS-ASEC information.

Create mock tax returns for current year and prior year using income reported on 3rd-party information documents

Sample Counts

Total Sample

Years	Timely CY	Not Timely CY	Total
2014-2015	18,010	1,831	19,840
2015-2016	18,200	2,053	20,250
2016-2017	17,520	1,621	19,140
Total	**53,500**	**5,500**	**59,000**

Timely Prior Year

Years	Timely CY	Not Timely CY	Total
2014-2015	17,540	597	18,140
2015-2016	17,790	741	18,530
2016-2017	17,080	569	17,650
Total	**52,500**	**2,000**	**54,500**

High Income

Years	Timely CY	Not Timely CY	Total
2014-2015	4,487	239	4,726
2015-2016	4,582	263	4,844
2016-2017	4,607	262	4,869
Total	**13,500**	**1,000**	**14,500**

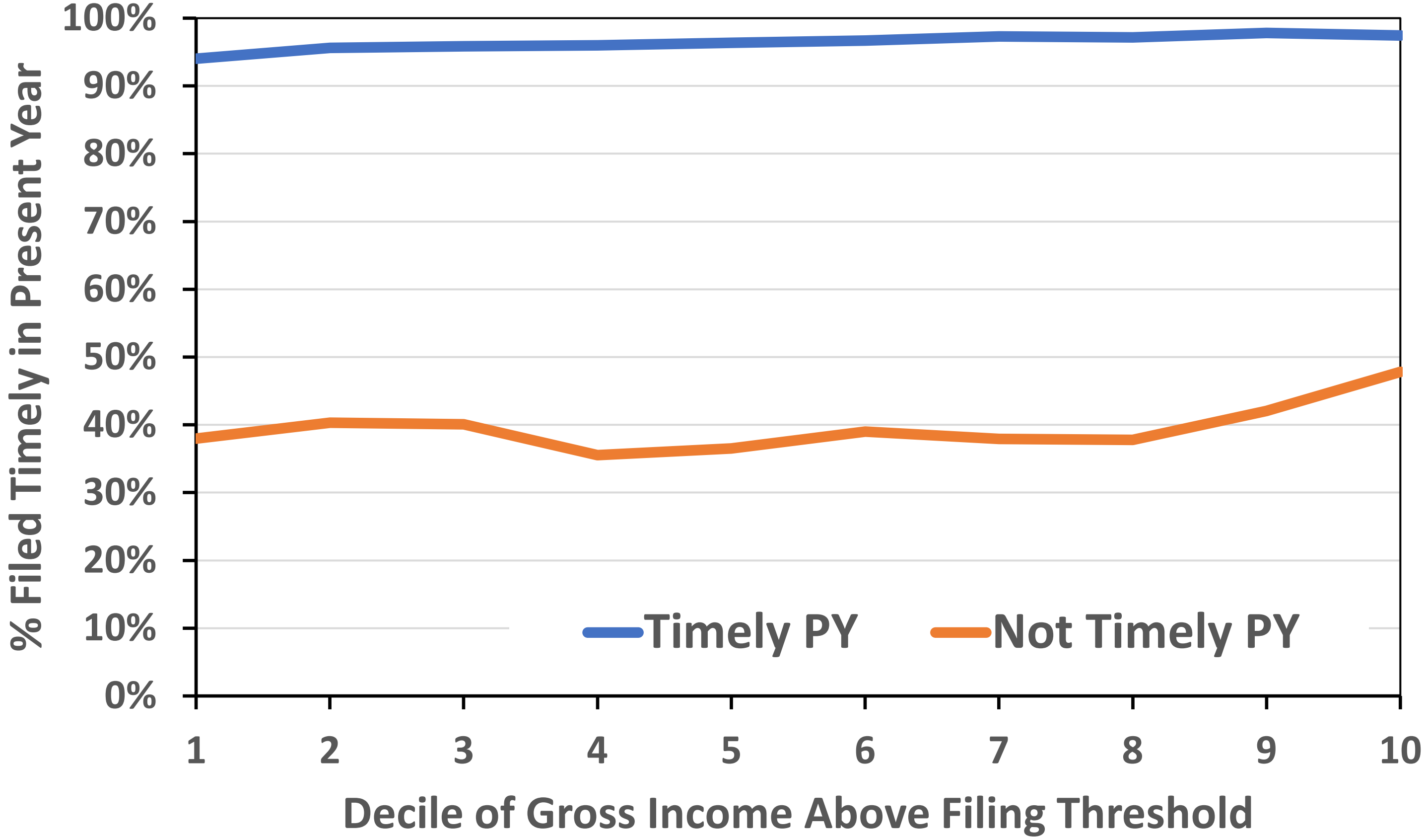
Not Timely Prior Year

Years	Timely CY	Not Timely CY	Total
2014-2015	667	1,027	1,694
2015-2016	618	1,094	1,711
2016-2017	636	876	1,512
Total	**1,900**	**3,000**	**4,900**

Timely Filing and Nonfiling Tend to Persist from One Year to the Next

Filing Behavior		Current Year			
		Filed Timely	Filed Late	Did Not File	Total
Prior Year	Filed Timely	**96.5%**	2.0%	1.5%	100.0%
	Filed Late	**57.1%**	33.2%	9.7%	100.0%
	Did Not File	28.7%	3.3%	**67.9%**	100.0%
	Total	90.7%	3.2%	6.1%	100.0%

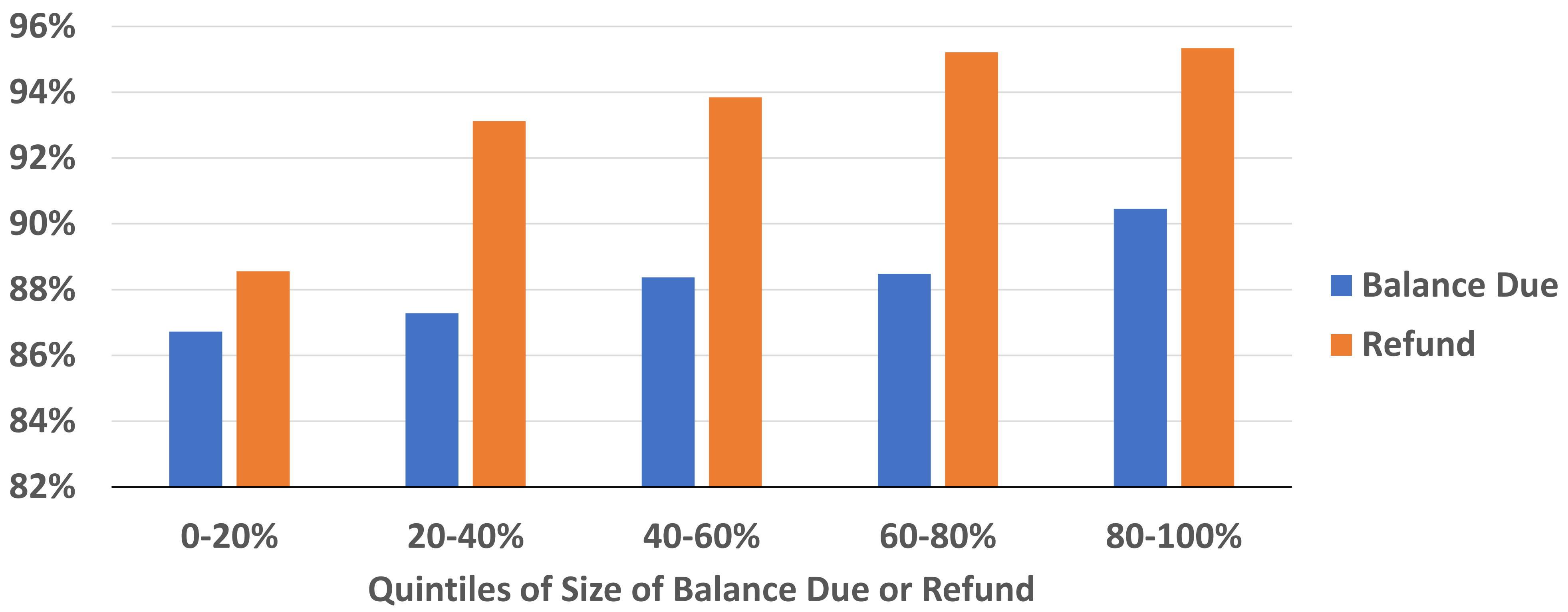
Timely Filing Rate by Decile of Gross Income Above Filing Threshold



Decile	Timely PY	Not Timely PY
1	94.0%	37.9%
2	95.6%	40.3%
3	95.9%	40.1%
4	96.0%	35.5%
5	96.4%	36.5%
6	96.7%	39.0%
7	97.3%	37.9%
8	97.1%	37.8%
9	97.8%	42.0%
10	97.4%	47.8%

- Not filing more common in lower income deciles over the filing threshold
- Timely filing much lower for those who did not file in the prior year even at the upper deciles

Timely Filing by Balance Due/Refund Quintiles



- Taxpayers with a balance due are on average less likely to file than those due a refund.
- Likelihood of filing significantly higher as the balance due or refund amount increases. This is likely because balance due/refund amounts are correlated with income.
- Late filing does not seem to be significantly correlated with the balance due or refund amount.

Logistic Models Predicting Timely Filing in the Current Year

$$\Pr(TF = 1) = \frac{e^{\delta' x}}{1+e^{\delta' x}}, \delta' x$$

Given large impact of prior-year filing behavior, we estimate separate logit specifications for tax units that did and did not file timely in the prior year.

Also, particular interest in understanding factors affecting nonfiling among “high income” (>$100,000) taxpayers

Three subsamples:

1. Timely Filed Prior Year (TPY) (“Stop-Filer” analysis)
2. Not Timely Filed Prior Year (NTPY) (“Start-Filer” analysis)
3. High Income (HI)

Results: Timely Filed in the Prior Year

Dep. Var. = Timely Filed in Current Year

	Coefficient	Marginal Effect
year2016	-0.190	-0.006
year2017	0.042	0.001
gi_req_py	0.078	0.003
singlefemale	0.416	0.014
married	0.630	0.021
age	-0.053	-0.002
agesq	0.001	0.000
midatlantic	0.538	0.018
logrefundcred	0.063	0.002
githresh_dec1	-1.096	-0.036
githresh_dec2	-0.658	-0.022
githresh_dec34	-0.570	-0.019
githresh_dec56	-0.353	-0.012
githresh_dec78	-0.139	-0.005
plur_retire_cy	-0.482	-0.016
plur_seinc_cy	-0.965	-0.032
plur_other_cy	-0.971	-0.032
plur_investnc_cy	-0.114	-0.004
addkids	-0.215	-0.007
dropkids	-0.383	-0.013
burden_tpi	-0.647	-0.021
unemployed_cy	-0.340	-0.011
Intercept	4.430	

N = 54,500;
Pseudo R-Square = 0.031

Significant at the 1% level

Significant at the 5% level

Timely filing less likely for:

- Tax Year 2016
- Income near the filing threshold
- Largest amount of income from sources other than wages
- Fewer children eligible for EITC than previous year
- Unemployed part of current year

Timely filing more likely for:

- Married and single females
- Residents of Mid Atlantic
- Larger eligible amount of refundable credits

Results: Not Timely in the Prior Year

Dep. Var. = Timely Filed in Current Year

	Coefficient	Marginal Effect
late_py	**1.709**	**0.300**
gi_req_py	**-1.565**	**-0.295**
singlefemale	0.163	0.028
married	**0.282**	**0.050**
logrefundcred	**0.059**	**0.010**
oneeicchild	**-0.439**	**-0.075**
twoeicchild	**-0.315**	**-0.055**
threeeicchild	-0.323	-0.056
gtthreeeicchild	**-0.810**	**-0.135**
age	**-0.098**	**-0.005**
agesq	**0.001**	
githresh_dec1	**-0.658**	**-0.114**
githresh_dec2	-0.298	-0.053
githresh_dec34	-0.224	-0.040
githresh_dec56	-0.188	-0.034
githresh_dec78	-0.246	-0.044
plur_retire_cy	**-0.296**	**-0.054**
plur_seinc_cy	**-0.973**	**-0.165**
plur_other_cy	**-1.226**	**-0.202**
plur_investnc_cy	-0.328	-0.059
Intercept	**3.198**	

N = 4,900;
Pseudo R-Square = 0.213

Significant at the 1% level

Significant at the 5% level

Timely Filing more likely for:

- Late filers in prior year as opposed to nonfilers
- Married
- Larger amounts of refundable credits

Timely Filing less likely for:

- Required in the prior year
- Having qualifying children for EITC
- Gross income within the first decile above the filing threshold
- Plurality of income from source other than wages

Results: “High Income”

(Gross Income>$100,000)

Dep. Var. = Timely Filed in Current Year

	Coefficient	Marginal effect
late_py	**1.685**	**0.053**
timely_py	**5.183**	**0.162**
singlefemale	0.115	0.004
married	0.260	0.008
oneeicchild	-0.024	-0.001
twoeicchild	**0.297**	**0.009**
threeeicchild	-0.154	-0.005
gtthreeeicchild	-0.066	-0.002
age	**-0.121**	**-0.004**
agesq	**0.001**	**0.000**
notaxstate	-0.286	-0.009
logrefundcred	0.077	0.002
plur_retire_cy	**-0.538**	**-0.017**
plur_seinc_cy	**-1.055**	**-0.033**
plur_other_cy	-0.652	-0.020
plur_investnc_cy	**-0.469**	**-0.015**
gotmarried	0.193	0.006
gotdivorced	**-1.063**	**-0.033**
addkids	-0.214	-0.007
dropkids	-0.203	-0.006
burden_tpi	-0.479	-0.015
unemployed_cy	**-0.602**	**-0.019**
Intercept	1.460	

N = 14,500;
Pseudo R-Square = 0.3495

Significant at the 1% level
Significant at the 5% level

Timely filing more likely for:

- Prior year timely and late filers
- Two children

Timely filing less likely for:

- Largest share of income is from retirement, self-employment, and investment income
- Those who got divorced or lost a spouse
- Unemployed part of the current year

Future Work

- **Include more years** to analyze effects of policy changes (e.g., economic stimulus credits, expansion of EITC) and longer-term prior filing behavior
- With more data, **analyze drivers of late filing** or “filed within calendar year” rather than timely filing
- Investigate additional drivers: e.g., **changes in income amounts and sources**
- Check model estimates focused on **somewhat different populations** (e.g., required by gross income or SE income test (SE from Form 1099Misc)
- Analyze drivers of filing behavior **using IRS administrative data** and study effects of IRS **enforcement** actions

Economic Influencers of Enforcement Revenue and Operational Implications

Jess Grana[1], Lucia Lykke[1], Sam Schmitz[1], and Ron Hodge[2]

June 16, 2022

[1]The MITRE Corporation

[2]Internal Revenue Service

Can we incorporate *external conditions* into IRS resource planning?

Traditional approach

Historical revenue values

Expected value of revenue

Enforcement Process Model

Allocate FTEs to parts of the enforcement process

May not apply to the future
→ **Suboptimal allocation**

Approach explored in this study

External predictors

- Changing economy
- Changing taxpayer population
- Changing IRS resources

Forecasted value of revenue

Enforcement Process Model

Allocate FTEs to parts of the enforcement process

MITRE

Research Questions

1. What are the high-level drivers of TERC?

- We forecast components of TERC based on:
  - Economic indicators
  - IRS resource levels
  - Taxpayer characteristics
- *“Macro” model to understand aggregate fluctuations*

2. How can we allocate resources based on forecasted tax revenue?

- We forecast revenue per case at the compliance step level
- We update resource allocation based on revenue forecasts
- *“Micro” model at compliance step level*

Macro Model Summary

Predict TERC based on economy, taxpayer characteristics, and IRS resources

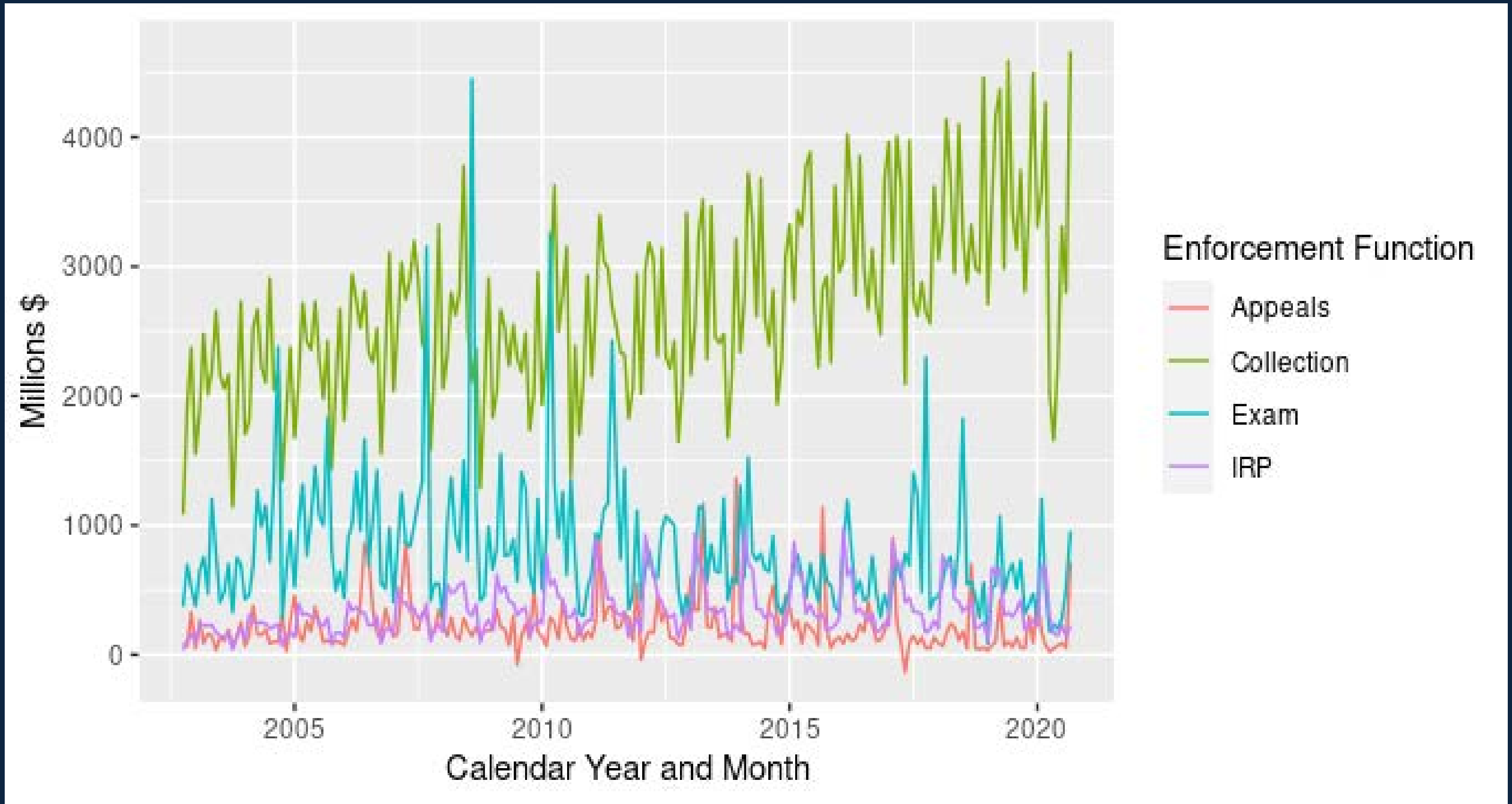
What drives Total Enforcement Revenue Collected (TERC)?

- ***Economic drivers*** affecting total tax?
- ***IRS resources*** affecting enforcement capacity?
- ***Taxpayer characteristics*** affecting compliance?

TERC falls into 4 categories:

- Exam
- Appeals
- Collection
- Information Returns Processing (IRP)

TERC by Enforcement Function, 2000 – 2021.



Source: MITRE

TERC “Influencers”

Concept:

These can affect total tax, voluntary compliance rate, enforced/late payments

Economic Indicators	Consumer price index, seasonally adjusted (1982-1984=100)
	NASDAQ Composite index close level (Thou)
	Personal bankruptcies (Thou)
	Business bankruptcies (Thou)
	Unemployment rate (%)
	New single-family houses sold (Thou units)
	Consumer confidence index (1966Q1=100)
	Gross domestic product (Trill $)
	New housing authorized by building permits (Mill units)
	Rental expenditures (Bill $)
IRS Resources	FTE, business systems modernization (Thou)
	FTE, exams and collections (Thou)
	FTE, filing and account services (Thou)
	FTE, investigations (Thou)
	FTE, prefiling taxpayer assistance & education (Thou)
	FTE, regulatory (Thou)
	FTE, shared services and support (Thou)
Taxpayer Attributes	Percent of US population age 65 and above (%)
	Weekly median earnings ($)
	Share of net worth held by top 1% (%)

MITRE

Macro Model: Key Takeaways

- Forecast TERC components using:
  - Time lags of influencer variables
  - Month and year trends
- Different revenue functions require different models
  - Appeals, Collection, Exam and IRP are different
- Causal stories are difficult to tell
  - IRS resources play a part, but not all that matters
  - CPI, bankruptcies, and other influencers matter
- Selection of time lags matter

MITRE

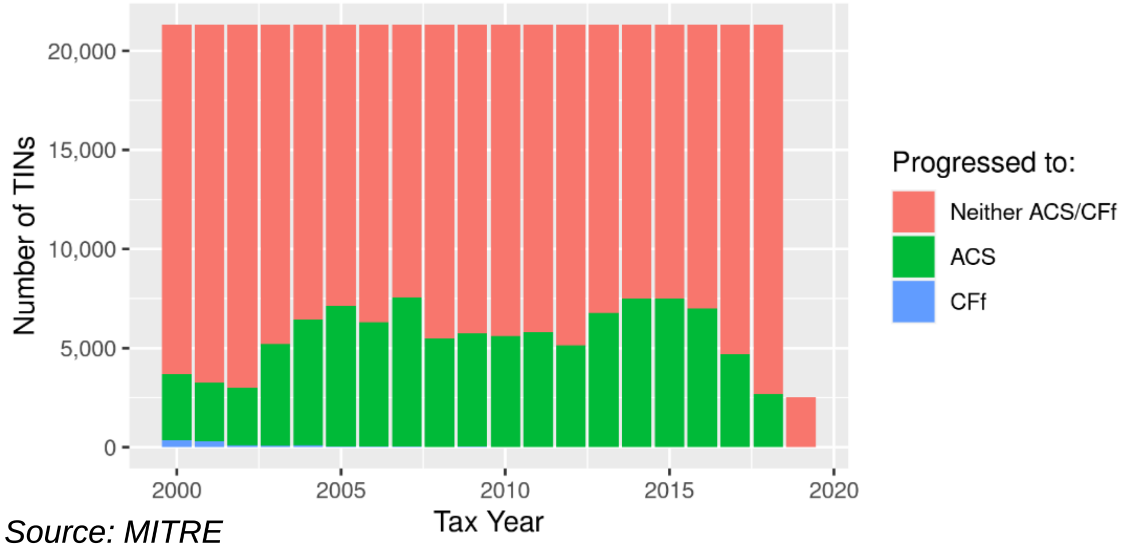
Micro Model Summary

Model the probability of individual taxpayers moving from one compliance step to the next and their predicted revenue at each step

Predict Taxpayer Progression and Revenue

1) Which taxpayers advance through each compliance step?

Multinomial logit regression of compliance step outcomes

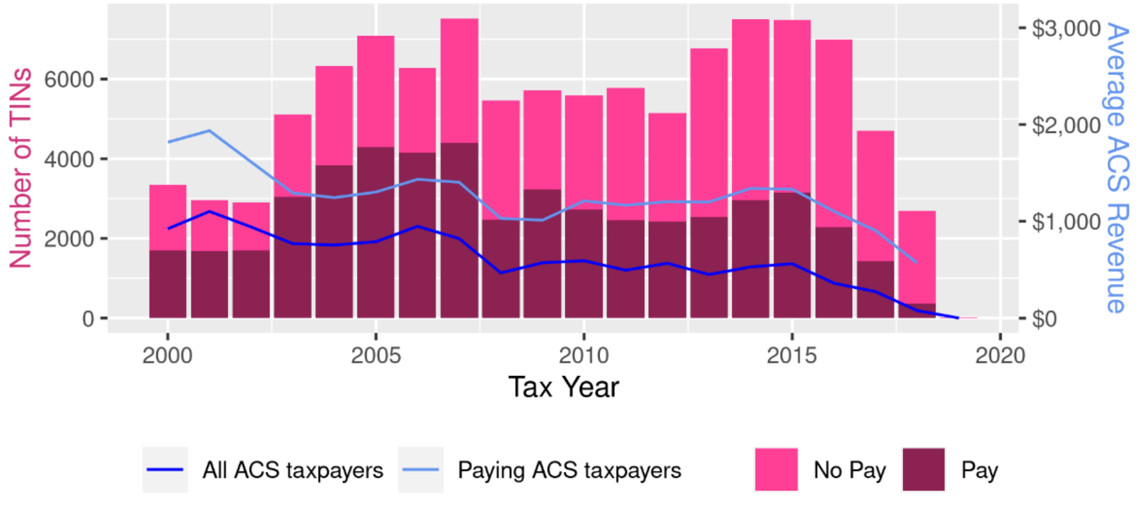


Source: MITRE

10% Random Sample of Correspondence Audits

2) How much revenue can we expect once they advance to that compliance step?

Zero-inflated model of amount paid



Source: MITRE

Note: Average revenue among TINs who progressed to ACS Step

Amount Paid per ACS Case

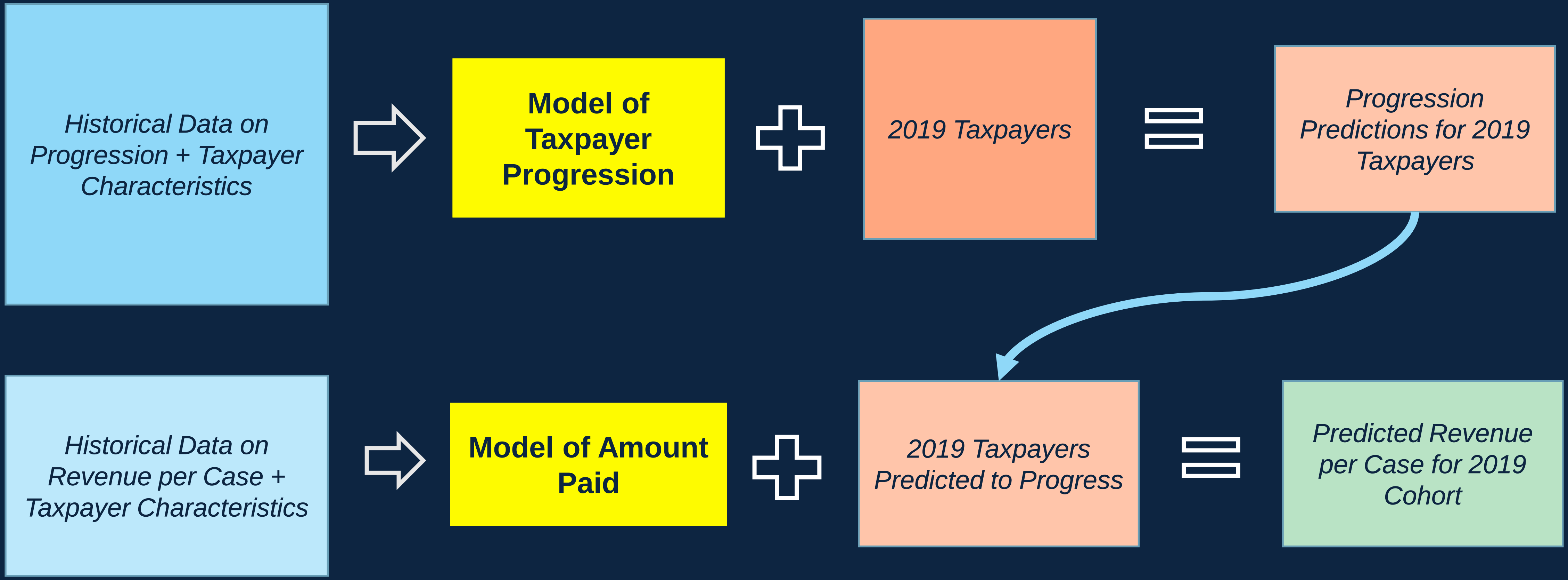
MITRE

Takeaways

- What increases probability of progression to step
  - ACS: Preparer assistance, child tax credit amount, rents and royalties loss
  - Field Coll: Preparer assistance, child tax credit amount, casualty and theft loss
- What increases probability of payment in step
  - ACS: Child tax credit amount, preparer assistance, total tax, Schedule C net profit or loss
  - Field Coll : Preparer assistance
- What increases amount paid in step
  - ACS: Child tax credit amount , rents and royalties loss, preparer assistance, total tax
  - Field Coll : AGI, preparer assistance, rents and royalties loss, total tax

MITRE

Forecasting Revenue per Case



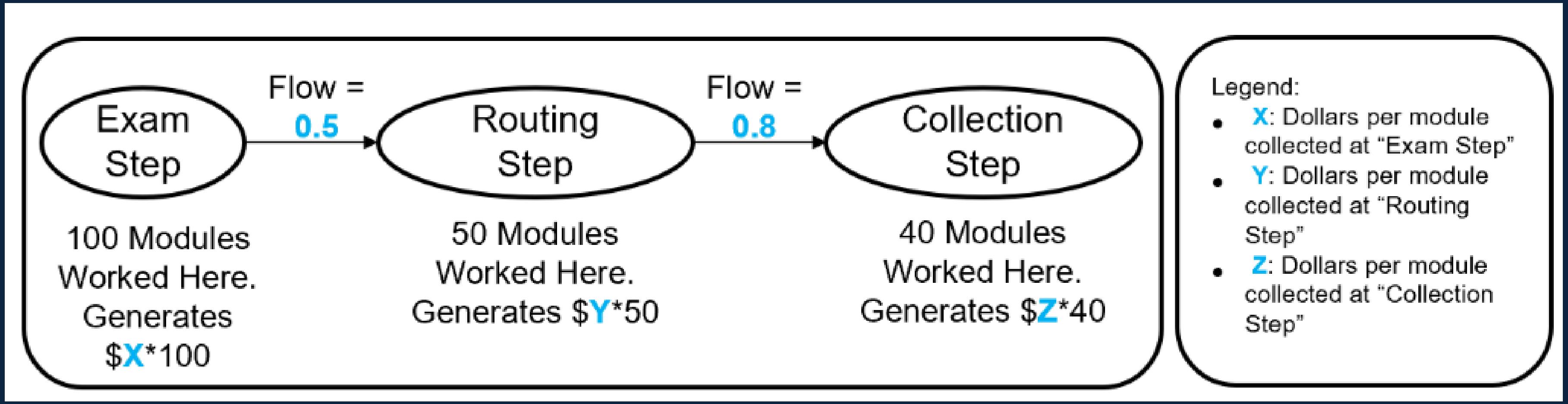
MITRE

Resource Allocation

Apply forecasted revenue values to an optimization model, resulting in new FTE allocations

Applying to Resource Allocation

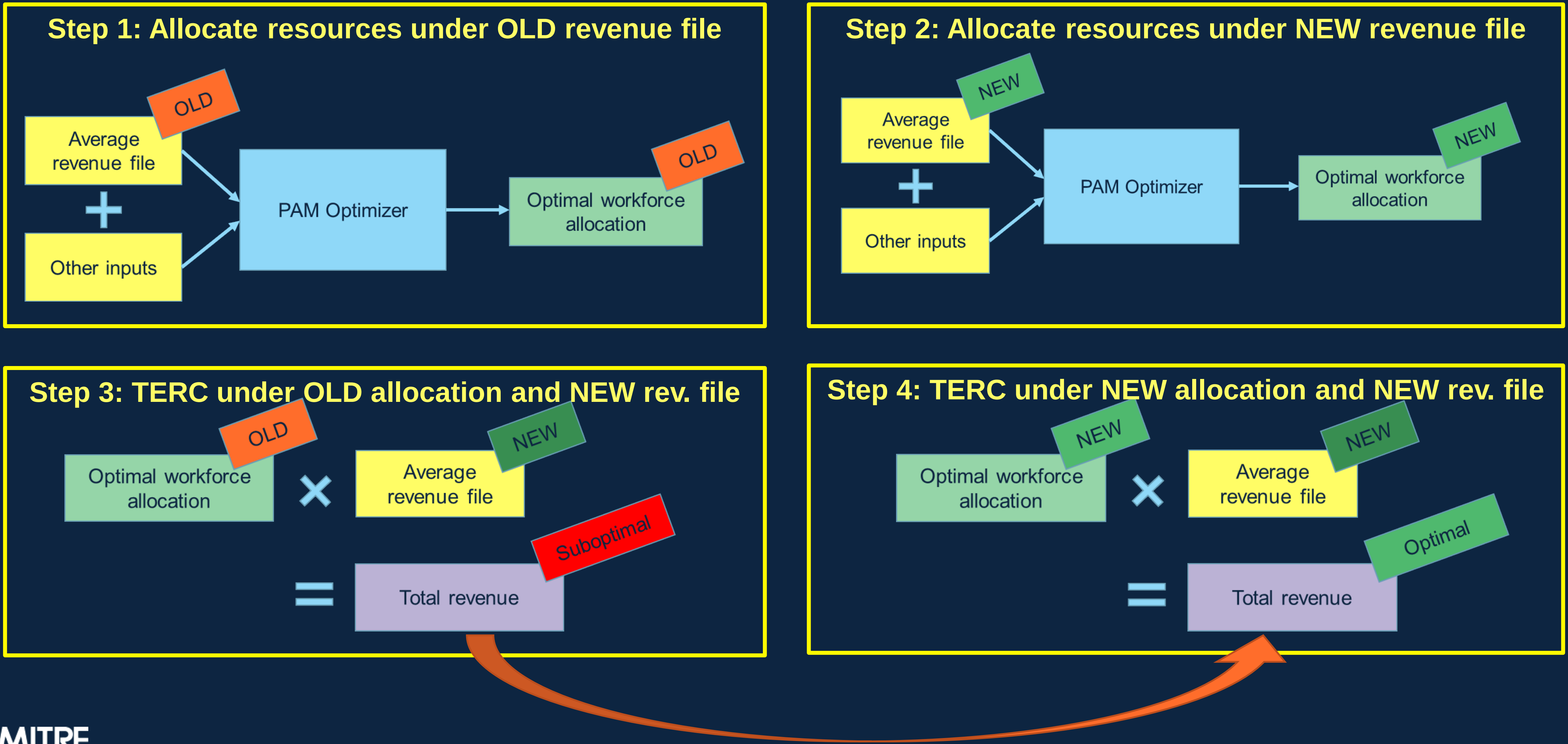
- Program Assessment Model (PAM) allocates FTE across the SB/SE compliance process
  - Maximizes revenue subject to operational constraints like level of service
  - Accounts for interdependencies in compliance process
- PAM takes revenue per case as inputs



Source: MITRE

- We update XYZ inputs

Calculating the Effect on TERC



What did we find?

Commodity and Step	Old Revenue Parameter	New Revenue Parameter	FTE Moved from Step	FTE Moved Across System	Diff in Revenue
corr_exam: ACS	$603	$570	-4	4	$188,391
corr_exam: CFf	$1,394	$1,265	0	0	$0

Little or no change in TERC (param change very small)…

MITRE

Sensitivity Analysis

Conduct sensitivity analysis (+/-50% change in revenues)…

Commodity	Step	Old Revenue Parameter	New Revenue Parameter	FTE Moved from Step	FTE Moved Across System	Diff in Revenue
exam_employment	SpecEx_Employment	$7,740	$3,870	-163	163	$49,102,248
aurMed	AUR	$400	$200	-53	53	$18,600,495
aurMed	Collection_Bal_Due	$625	$313	0	44	$14,989,252
exam_gift	SpecEx_Gift	$44,500	$22,250	-32	32	$11,670,594
baldue_med	ACS	$1,086	$543	-215	215	$4,367,040
corr_exam	Queue	$151	$76	0	8	$2,250,374
imf_NF_med	Queue	$231	$116	0	11	$1,998,898

Discussion

- Main contribution
  - Equip resource allocation to *anticipate changing external conditions* that affect revenue
- Proof-of-concept models
  - Macro model explaining/predicting high-level shifts in TERC
  - Micro model predicting revenue at compliance step
- Many potential extensions
  - Adding more influencer variables
  - Replicating to other compliance steps
  - Using different lags in macro model or different model specification

Contact Us

Jess Grana cheny@mitre.org
Lucia Lykke llykke@mitre.org
Sam Schmitz sschmitz@mitre.org
Ron Hodge Ronald.H.HodgeII@irs.gov

Non-Monetary Sanctions as Tax Enforcement Tools:

Evaluating California's Top 500 Program

Chad Angaretis
CA FTB (Retired)

Brian Galle
Georgetown Law
& SEC

Paul R. Organ
Michigan

Allen Prohofsky
CA FTB

March 2022

CELS

Disclaimer

Any views expressed in this presentation are those of the authors and not official positions of the California Franchise Tax Board or the SEC.

Tax debt is a big issue

- The tax gap is large
  - Federal U.S. tax gap estimated at $600B+
  - State income tax gap in CA for 2018 estimated at $20-25B
- Part of that gap is liability assessed but unpaid
  - About 25% of U.S. federal tax gap (remainder is underreporting)
- Various tools are used to encourage resolution of unpaid liability
  - Primarily financial: liens, levies, interest and penalties
- One lesser-used and lesser-studied tool: **collateral sanctions**
  - “Name and shame” lists
  - Licenses suspensions

We study California's Top 500 program

STATE OF CALIFORNIA
Franchise Tax Board

File | Pay | Refund | Forms

home / about ftb / newsroom / top 500 past due balances

Top 500 past due balances

About the delinquent taxpayer list

< Newsroom

- **Top 500 past due balances**
- Personal income tax list
- Corporate income tax list
- If you're on the list

Related Content

- Respond to a letter

One of our main responsibilities is to collect state income tax and corporate franchise tax. Sometimes, people don't pay their taxes. Those who don't pay their state income taxes contribute to California's tax gap — the difference between taxes owed and taxes paid.

For 2018, the estimated annual tax gap for California is $20 billion to $25 billion.

The law

FTB is required by law to publish a list of the 500 largest tax delinquencies in excess of $100,000 twice a year and update the list when names are removed. The list is replaced when updated. Updates are ongoing, with major updates occurring twice a year.

The intent of this list is to encourage those who are on the list (or may be placed on the list) to pay their taxes.

Recent highest published CA balances

This is a list of the 500 largest tax delinquencies over $100,000. By law, we must publish this list at least twice a year.

Last updated: 10/06/2020

Name	Address	Subtotal	Total	Lien	License	Status	Number
Moreland, Peggy J & Terry L	Bakersfield, CA 93306	$5,306,836.86	$5,306,836.86	01/24/2008	Contractor's State License Board	Expired	856954
					Contractor's State License Board	Expired	362166
Cooksey, Jimmy D	Bowling Green, KY 42104	$2,403,194.62	$2,403,194.62	04/25/2008			
Amin, Joseph & Sharona	Beverly Hills, CA 90210	$1,730,698.65	$1,730,698.65	04/14/2014	Board of Pharmacy	Active	0034252
Patrick, William L & Susan K	Cody, WY 82414	$1,648,546.31	$1,648,546.31	05/31/2019			

https://www.ftb.ca.gov/about-ftb/newsroom/top-500-past-due-balances/personal-income-tax-list.html

Recent lowest published CA balances

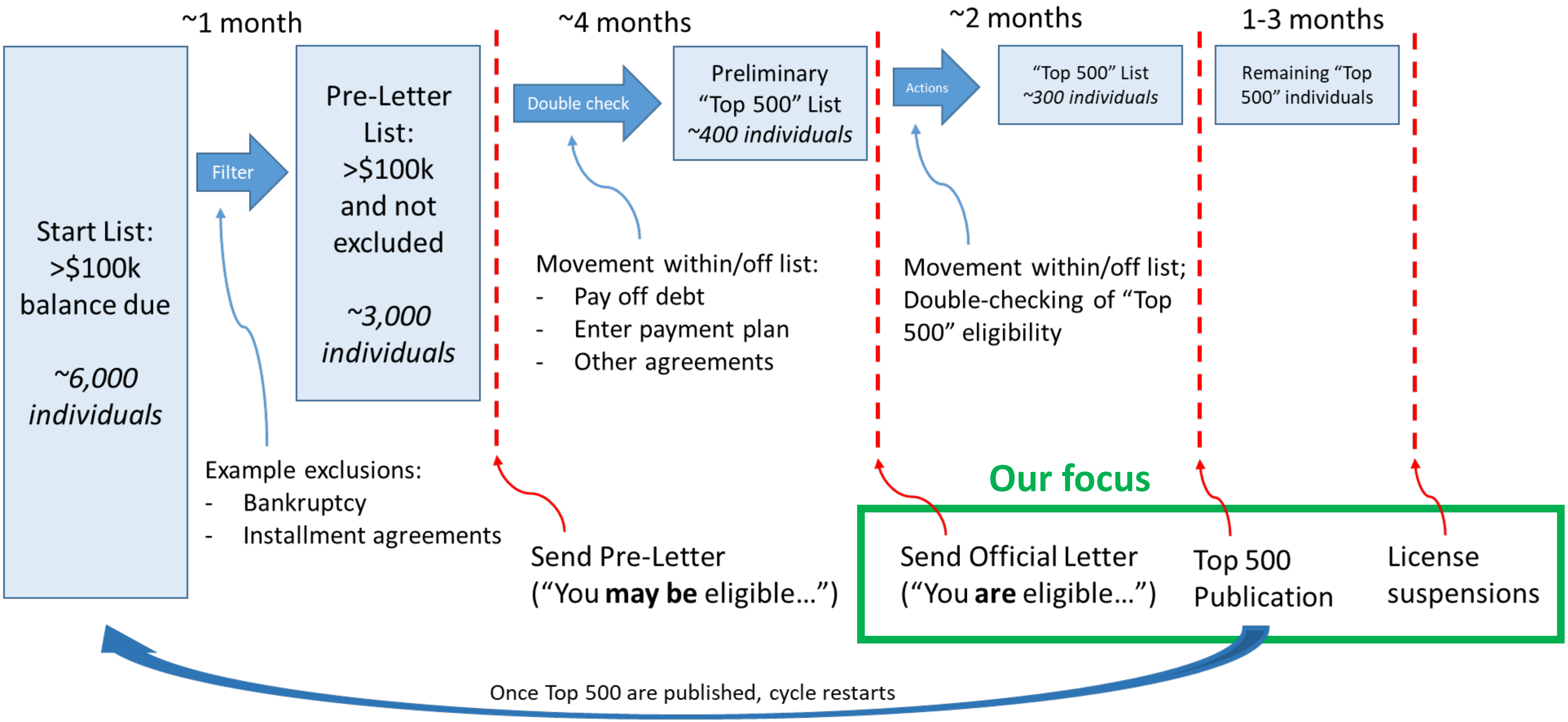
This is a list of the 500 largest tax delinquencies over $100,000. By law, we must publish this list at least twice a year.

Last updated: 10/06/2020

Name	Address	Subtotal	Total	Lien	License	Status	Number
Nguyen, Peter	Santa Ana, CA 92704	$157,856.33	$157,856.33	06/18/2019			
Peppard, Jeff D	Santa Barbara, CA 93109	$167,405.18	$167,405.18	03/17/2016	Board of Dental Examiners	Active	0037003
Gilmore, Joel R	Semmes, AL 36575	$193,166.57	$193,166.57	12/10/2013			
Lee, Jung S	Valencia, CA 91381	$193,971.49	$193,971.49	09/13/2006			
Schmid, Earl	Williamsburg, OH 45176	$194,384.70	$194,384.70	12/16/2014	Contractor's State License Board	Active	853323

https://www.ftb.ca.gov/about-ftb/newsroom/top-500-past-due-balances/personal-income-tax-list.html

Note: our analysis focuses only on **individual taxpayers**, not businesses



Official letter

Arguably, the response to the official letter is a response to the threat of a *bundle* of two sanctions:

1. Certain publication
2. Potential license suspension

Notice of Public Disclosure of Tax Delinquency

Notice Date:
Taxable Years:

Account Number:
Balance Due:
Pay By:

Revenue and Taxation Code (R&TC) Section 19195 directs the Franchise Tax Board (FTB) to publicly disclose a list of the 500 largest state income tax delinquencies. These delinquencies must total in excess of $100,000 and be subject to a recorded notice of state tax lien. We intend to post this list on our website at **ftb.ca.gov**.

Your account qualifies for this disclosure and Internet posting. If you do not pay your tax liability or take other action described below, we may add to a list posted on our website:

- **Your name and address.**
- Your occupational or professional licenses with type, status, and license numbers.
- The liened amount owed and the earliest date a notice of state tax lien was recorded.

Your inclusion on the list may lead to the denial or suspension of your licenses, including driver's licenses, under Business and Professions Code Section 494.5, and will preclude you from entering into contracts for the acquisition of goods or services with California state agencies under Contract Code Section 10295.4.

To avoid public disclosure of tax delinquency, you must do one of the following within 30 days of the notice date:

- **Pay your balance due. You may be required to make payments electronically.** Go to **ftb.ca.gov** and search for **mandatory epay**. If your estimated tax or extension payment exceeds $20,000 or your tax liability exceeds $80,000 for any taxable year beginning on or after January 1, 2009, **you must make all future payments electronically**, regardless of the taxable year. Payments made by other means will result in a penalty of 1 percent of the amount paid, unless your failure to pay was for reasonable cause and not willful neglect (R&TC Section 19011.5). If you are not required to pay electronically, enclose the above part of this notice and mail it with a check or money order for the total amount due payable to the Franchise Tax Board. Write your full name and account number on your payment. **Use the enclosed return envelope and mail to:** FRANCHISE TAX BOARD, PO BOX 3065, RANCHO CORDOVA, CA 95741-3065. No additional penalties or interest accrue on the existing liability if we receive full payment within 15 days of the notice date.
- **Arrange to pay your balance due**. To determine if you qualify for installment payments, call us at 888.426.8555.

Partial payment (even a reduction of the balance due below $100,000) will not preclude you from being on the list. If your name appears on the list, FTB will continue to pursue collection actions. Call 888.426.8555 if you believe you should not be on the list, have questions, paid the balance due, made payment arrangements, otherwise resolved the balance due, think you do not owe the balance due, or filed bankruptcy.

Get FTB 1131, *Franchise Tax Board Privacy Notice*, at **ftb.ca.gov** and search for **privacy notice**.

Get FTB 1140, *Personal Income Tax Collections Information*, at **ftb.ca.gov** and search for **1140**.

Internet and Telephone Assistance
Website: **ftb.ca.gov**
Telephone: 888.426.8555 from within the United States
916.845.7874 from outside the United States
TTY/TDD: 800.822.6268 for persons with hearing or speech impairments

FTB 4192 PIT PC (REV 10-2012)

Pre-letter notifies of program existence; makes clear that taxpayer will be notified if ultimately in Top 500

Data

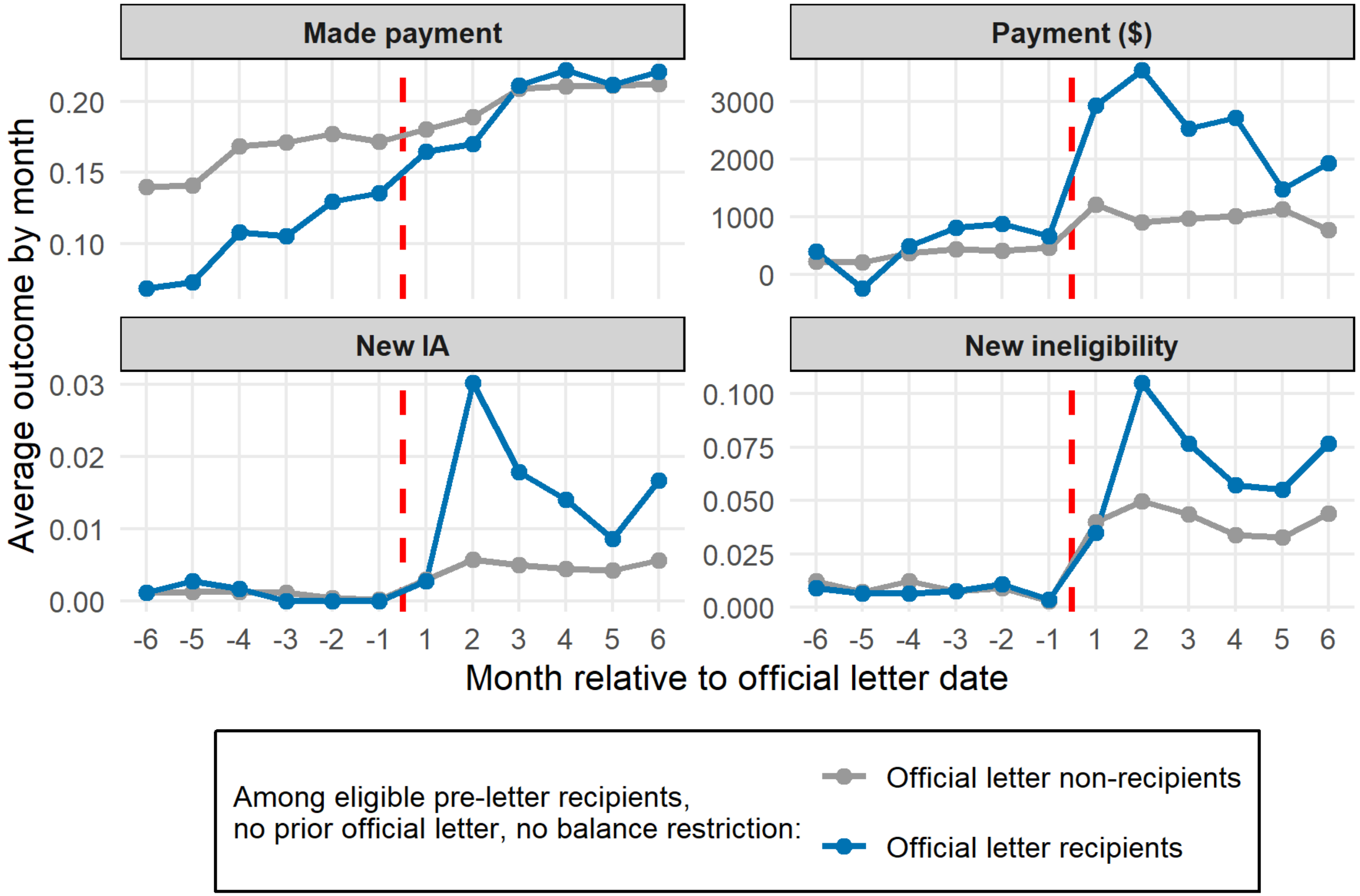
- Study includes publication cycles during 2013-2017
  - 10 Top 500 cycles (April and October publication)
- Admin data on debt/balance due over time
- Outcomes:
  - Payments made, and when
  - "Status codes" which allow identification of new Installment Agreements
- Linked to CA state income tax returns
  - Indication of on-time or late filing, or no filing
  - Among filers, CA AGI, Wages, presence of business income, and residency

Summary stats

Empirical strategies

- Identification
  - Ideally, use exact spreadsheet CART used, ranked by amount: RDD at #500
  - Or exploit quasi-random variation in this cutoff value
  - Problem: tiny % of outstanding debt is in this range (mean debt: ~$859k)
- Effect of warning letter (official letter)
  - Graphical: Time series of treated vs. untreated
  - Regression: Assume non-confoundedness by other events affecting (eligible) letter recipients at date of letter; omit repeat list appearances (robust to inclusion)
  - Heterogeneity by filing behavior, income amount and type
- Effect of publication
  - Graphical: Time series of treated vs. untreated (selection!)
  - Heterogeneity by professional license holding

Warning letter: Graphical approach (full range)



Compares among those:
- eligible for official letter (based on most recent status change);
- who received pre-letter that cycle;
- who have not received a prior official letter;
- with balance >$100K.

Back

Warning letter: Diff-in-Diff results (full range)

Main spec includes:

- Letter date +/- 3 months
- eligible for official letter (using most recent status)
- received pre-letter that cycle
- no prior official letter
- balance > $100K

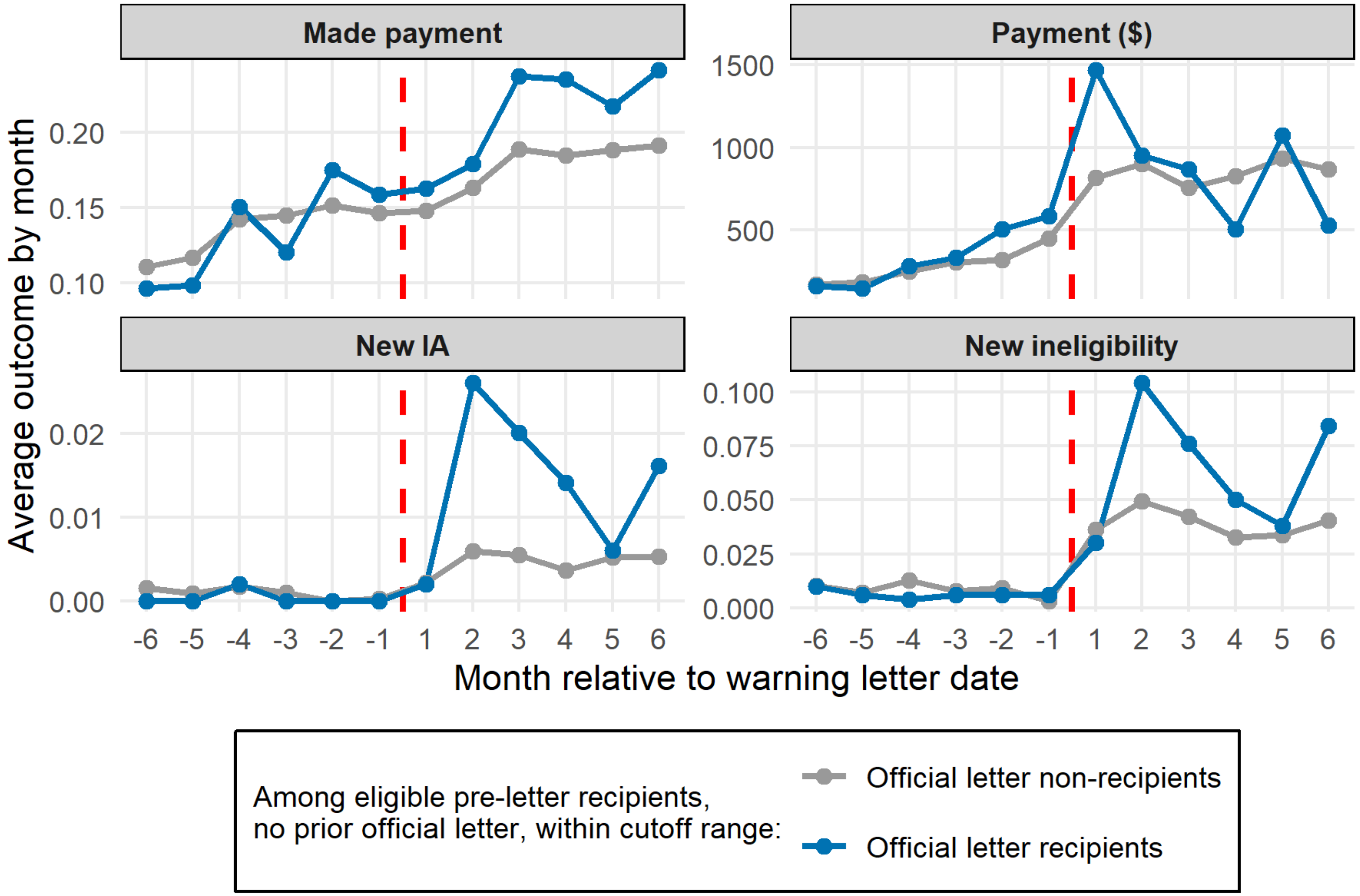
We still find significant positive effects for making payment, starting IA, and becoming ineligible.

We also now find significant positive payment amount effects.

	Dependent variables:			
	Made payment	Payment amount ($)	New IA	New ineligibility
Official letter * Post	**0.0392*** (0.0063)**	**1621.5*** (471.06)**	**0.013*** (0.0018)**	**0.0271*** (0.0035)**
Official letter	-0.05*** (0.0073)	(220.47) (216.96)	-0.0005*** (0.0001)	0.0007 (0.0012)
Post	0.0193*** (0.0016)	590.31*** (82.80)	0.004*** (0.0003)	0.0379*** (0.0009)
Balance ($ thousands)	0*** 0.0000	0.2771 0.1811	0*** 0.0000	0.0000 0.0000
April publication	-0.0414*** (0.0027)	-497.45*** (99.07)	-0.0007** (0.0003)	-0.0001 (0.0009)
Intercept	0.196*** (0.0047)	629.87*** (69.95)	0.001*** (0.0002)	0.0066*** (0.0006)
Observations	126,444	126,444	126,444	126,444
R2	0.0048	0.0018	0.0042	0.0167
Mean dep var.	0.1803	842.2733	0.0031	0.0268

Back

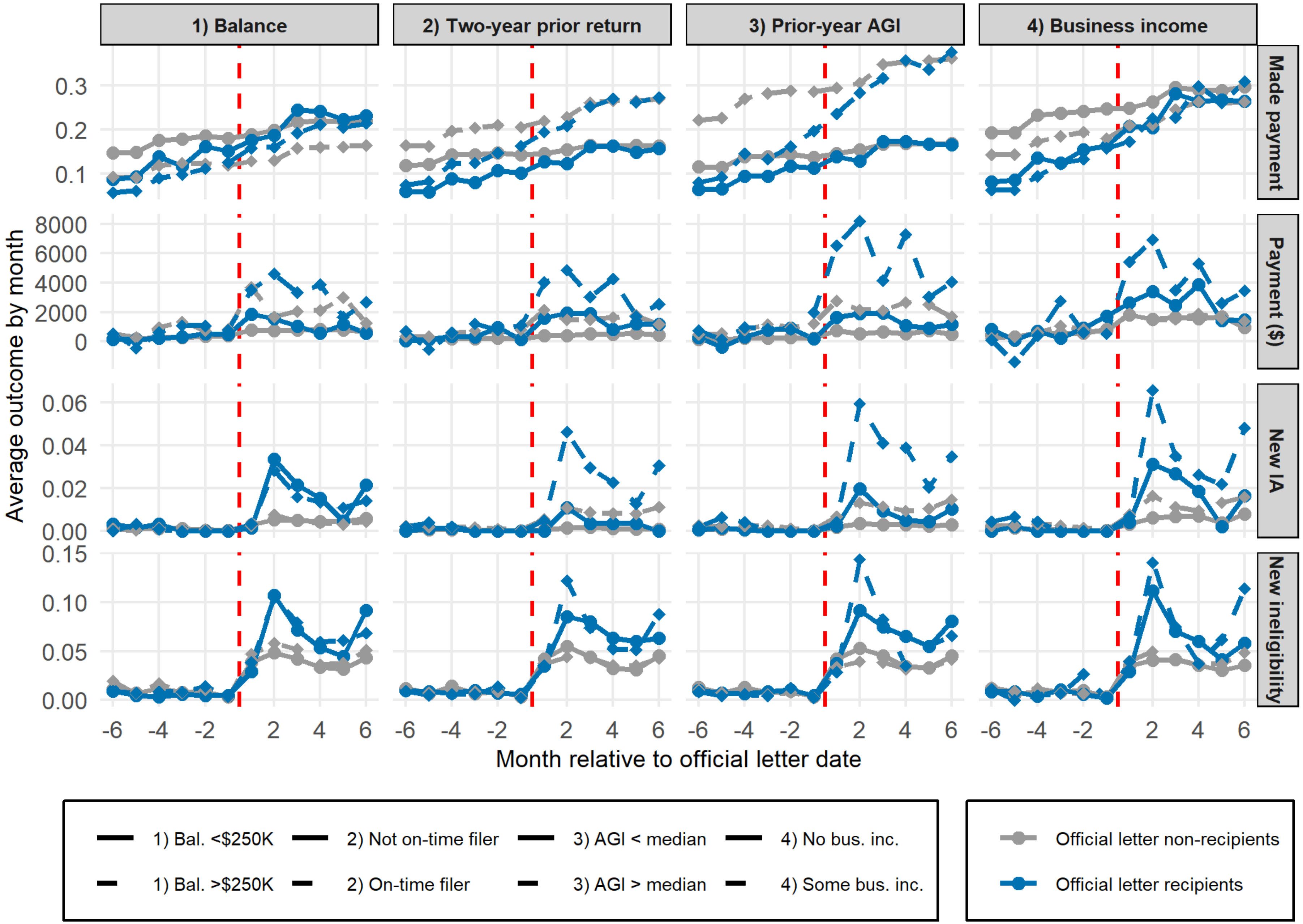
Warning letter: Graphical approach (near-cutoff only)



Compares among those:
- eligible for official letter (based on most recent status change);
- who received pre-letter that cycle;
- who have not received a prior official letter;
- **with balance roughly $150-225K (within range of cutoffs)**

Warning letter: More in paper

- Robustness checks — Data filters - Graphs | Data filters - Regs | Time windows
- Expand beyond cutoff range
  - Quasi-random argument is strongest within the cutoff range — Graphs | Regs
  - We argue that even above the range, letter receipt is unexpected
- Heterogeneity
  - Use data from CA income tax returns — Data | Graph - IA
  - On-time filers, higher AGI, business income have larger response — Regs | Graph - All
- Long-term payment effects
  - Cumulative payment effect, for use in Cost-Benefit Analysis — Graphs | Regs
- Future income effects
  - Positive point estimates, very imprecisely estimated — Results



1) Balance
2) Two-year prior return
3) Prior-year AGI
4) Business income
Made payment
Payment ($)
New IA
New ineligibility
Average outcome by month
Month relative to official letter date
1) Bal. <$250K
1) Bal. >$250K
2) Not on-time filer
2) On-time filer
3) AGI < median
3) AGI > median
4) No bus. inc.
4) Some bus. inc.
Official letter non-recipients
Official letter recipients

Back

Warning letter: Cumulative payment effect

This specification includes:

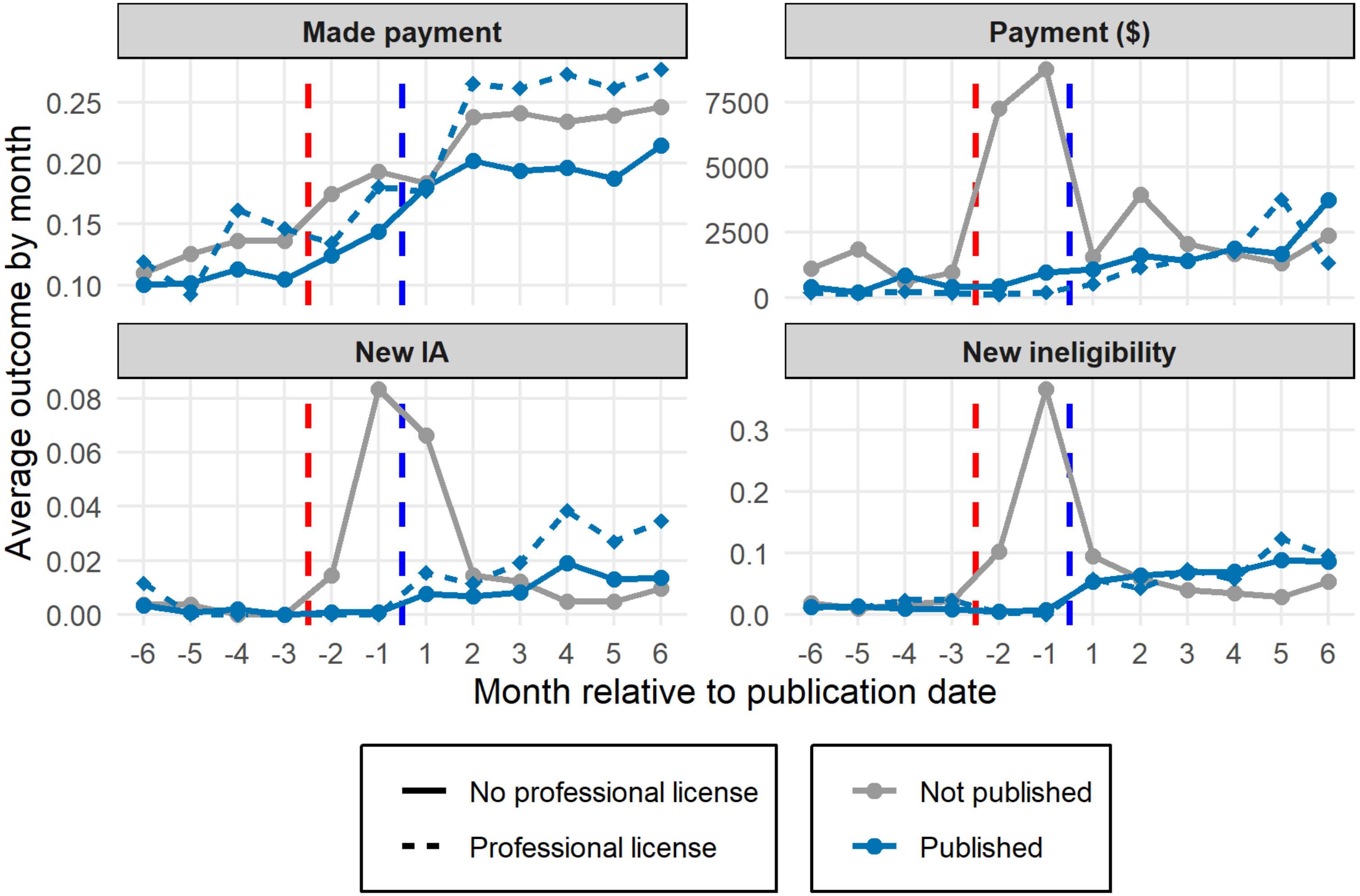
- Letter date +/- 3 months
- eligible for official letter (using most recent status)
- received pre-letter that cycle
- no prior official letter
- balance > $100K

Using full range of observations, not just cutoff range, to more fully capture the Top 500 effect overall

	Dependent variables: Cumulative payment amount post-official-letter ($)			
	3 months	6 months	12 months	24 months
Official letter	**5388.73*****	**8332.44*****	**13608.8*****	**18283.3*****
	(1,318.31)	**(1,746.36)**	**(2,234.60)**	**(2,628.68)**
Balance	1.13	1.85	2.61*	2.63*
($ thousands)	(0.74)	(1.15)	(1.50)	(1.59)
April publication	-2048***	-2487.13***	-2406.15***	-2034.9**
	(510.48)	(684.55)	(740.21)	(831.59)
Intercept	3869.5***	6838.05***	10802.19***	15299***
	(405.42)	(562.87)	(703.61)	(835.56)
Observations	21,074	21,074	21,074	21,074
R^2	0.0039	0.0049	0.0070	0.0075
Mean dep var.	3,626	6,835	11,519	16,620

Back

Publication: Some evidence for license effect



Among letter recipients, we compare published vs. non-published

Selection is very important to consider here!

Estimation of private costs

- Subjective costs of disclosure for those who are published
  - For those with ability-to-pay, arguably this cost is zero
    - Could have avoided by starting an IA, chose not to
  - For those without ability-to-pay, cost could be non-zero
    - Upper bound would be the cost to compliers of avoiding publication (those who take action to avoid publication arguably have the highest subjective costs of publication)
    - Among compliers, average payments over two-years post-letter are $46,000
  - Key question – what share have ability to pay?
    - Those in hardship are supposed to be excluded, but screening may be imperfect, or it may be costly to demonstrate hardship
    - Conservative estimate: share of taxpayers with AGI < median ($40K) = 50%
    - Thus have 161 initial non-compliers per year who we estimate to have $46,000 of subjective costs = **$7.4M upper bound**
- Gains from foregone avoidance
  - At margin, these = compliance costs for compliers ($7.2M in payments = $7.2M in foregone avoidance costs)

Policy implications

- Estimate of additional revenue:
  ~400 new taxpayers receiving warning letter each year
  X $18K in additional payments after two years
  $7.2M in additional payments induced by letter each year
  Note that this is a lower bound (IAs are longer-term; ignores publication effect)
  Also include pre-letter effects (small positive effects on large group) of **$3.2M**
- Estimate of administrative cost: **$1.6M**
- Costs = $10.4M in payments (debate on whether to include); $7.4M in publication disutility, -$7.2M in foregone avoidance costs
- Keen and Slemrod (2017) welfare test:

$$\boldsymbol{\phi(Revenue - Admin\ Cost) - Private\ Cost > 0}$$

  - Assume ϕ (marginal value of public spending) is greater than one (1.5)
  - $\phi(\$10.4m - \$1.6m) - \$17.8m + \$7.2m = (\phi)\$8.8m - \$10.6m$ = $\$2.6m$

Back

Conclusions

- Warning letter induces positive compliance responses
  - Time series graphs show effects; Regressions give quantitative estimates
  - Notable increase in probability of starting an Installment Agreement
  - Meaningful increase in long-run cumulative payments
    - Additional $7K/person paid over two years; $18K when including full range of balances
- Publication may also have small effect
  - Tentative evidence of small positive effect of professional license suspension
- Cost-benefit analysis yields favorable assessment of program overall
  - Consideration of private costs is especially important for collateral sanctions
  - Appendix includes detailed notes
- Comments welcome! Brian.galle@Georgetown.edu

Appendix

Contribution

Welfare analysis

Contribution to literature

- “Name and shame” papers
  - Perez-Truglia and Troiano (2018), “Shaming tax delinquents”
    - Letter RCT to those on lists in three states (KS, KY, WI)
    - Sent on UM letterhead, independent from tax agencies
    - Findings: letters increased “compliance” (odds of subsequent list) for those with smaller debts (<$2500); financial reminders increased payment at all debt levels
  - Dwenger and Treber (2019), effects of first year of business tax delinquents program in Slovenia
- License suspension: Kenchington and White (2021)
- Also relevant: broader literature on disclosure
  - Hasegawa et al. (2013); Bo, Slemrod, and Thoresen (2015); Hoopes, Robinson and Slemrod (2018); Slemrod, Ur Rehman, and Waseem (2019)
- Our paper
  - Evidence on sustainable effects (admin data, over many years)
  - (Lack of) income effects
  - Effects on those with large balance
  - License suspensions

Back

Pre-letter

Information Regarding Public Disclosure of Tax Delinquency

Notice Date:
Taxable Years:

Account Number:
Balance Due:
Pay By:

Revenue and Taxation Code (R&TC) Section 19195 directs the Franchise Tax Board (FTB) to publicly disclose a list of the 500 largest state income tax delinquencies. These delinquencies must total in excess of $100,000 and be subject to a recorded notice of state tax lien. We intend to post this list on our website at **ftb.ca.gov**.

Your account may qualify for this disclosure and Internet posting. If we determine that you are among the 500 largest tax delinquencies, we will send you a notice by certified mail advising you of the inclusion on this list of your name, address, and any occupational or professional licenses with status. At the time of publication, **your occupational, professional, and driver licenses issued by a California agency will be submitted for suspension**. You will be prohibited from contracting with any California state agency for the acquisition of goods or services. **Final determination of the top 500 names eligible for publication is pending confirmation of any resolutions or other qualifying circumstances for exclusion, including payment or arrangement for payment of tax liabilities.**

Pay your debt in full. You may be required to make payments electronically. Go to **ftb.ca.gov** and search for **mandatory epay**. If your estimated tax or extension payment exceeds $20,000 or your tax liability exceeds $80,000 for any taxable year beginning on or after January 1, 2009, you must make all future payments electronically, regardless of the taxable year. Payments made by other means will result in a penalty of 1 percent of the amount paid, unless your failure to pay was for reasonable cause and not willful neglect (R&TC Section 19011.5). If you are not required to pay electronically, enclose the above part of this notice and mail it with a check or money order for the total amount due payable to the Franchise Tax Board. Write your full name and account number on your payment. **Use the enclosed return envelope and mail to:** FRANCHISE TAX BOARD, PO BOX 3065, RANCHO CORDOVA, CA 95741-3065. No additional penalties or interest accrue on the existing liability if we receive full payment within 15 days of the notice date.

Call 888.426.8555 if you have questions, need assistance, think you do not owe this amount, paid the balance due, or filed bankruptcy.

Get FTB 1131, *Franchise Tax Board Privacy Notice*, at **ftb.ca.gov** and search for **privacy notice**.

Get FTB 1140, *Personal Income Tax Collections Information*, at **ftb.ca.gov** and search for **1140**.

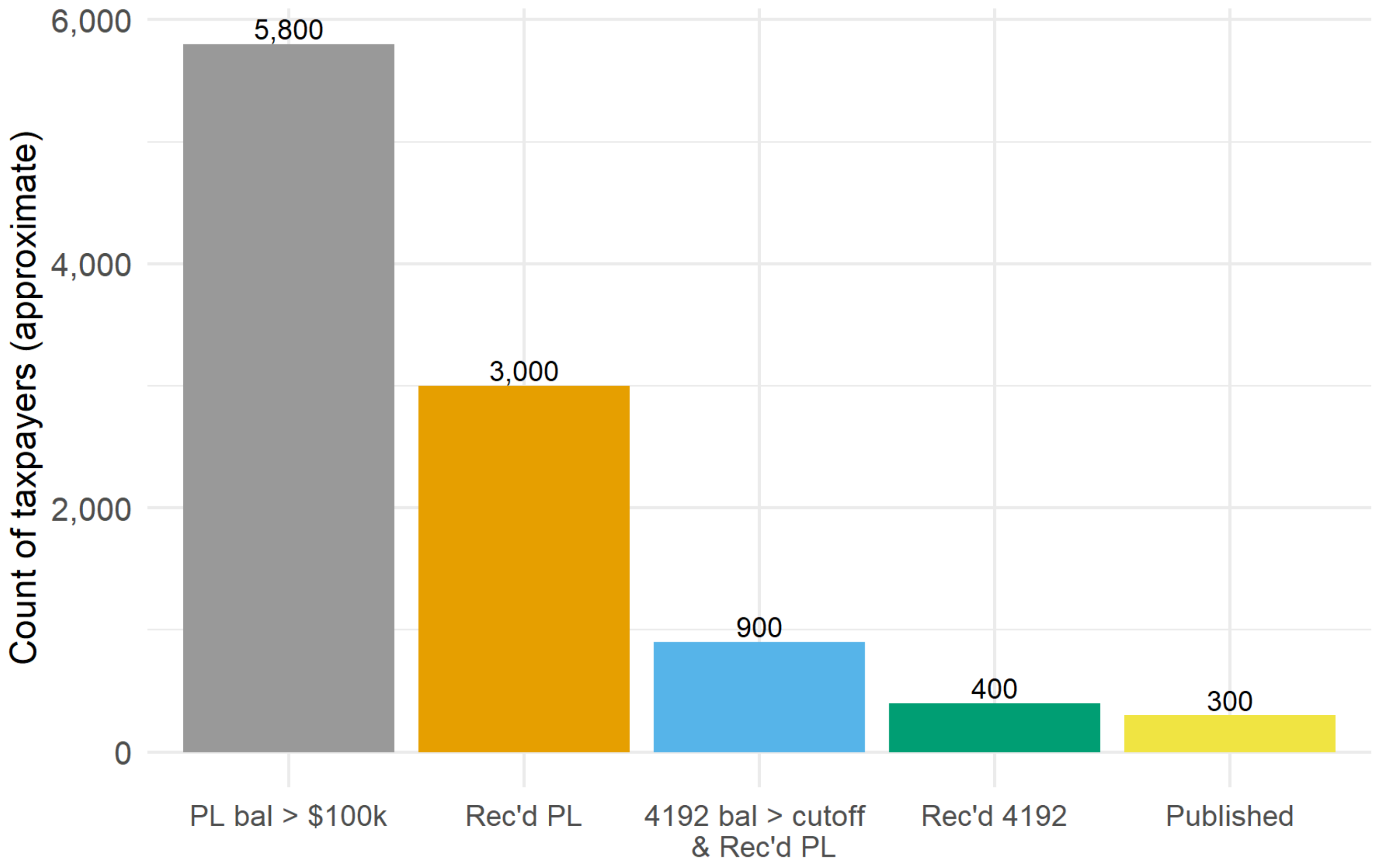
Internet and Telephone Assistance

Website:	**ftb.ca.gov**
Telephone:	888.426.8555 from within the United States
	916.845.7874 from outside the United States
TTY/TDD:	800.822.6268 for persons with hearing or speech impairments

FTB 3703 PIT PC (REV 03-2012)

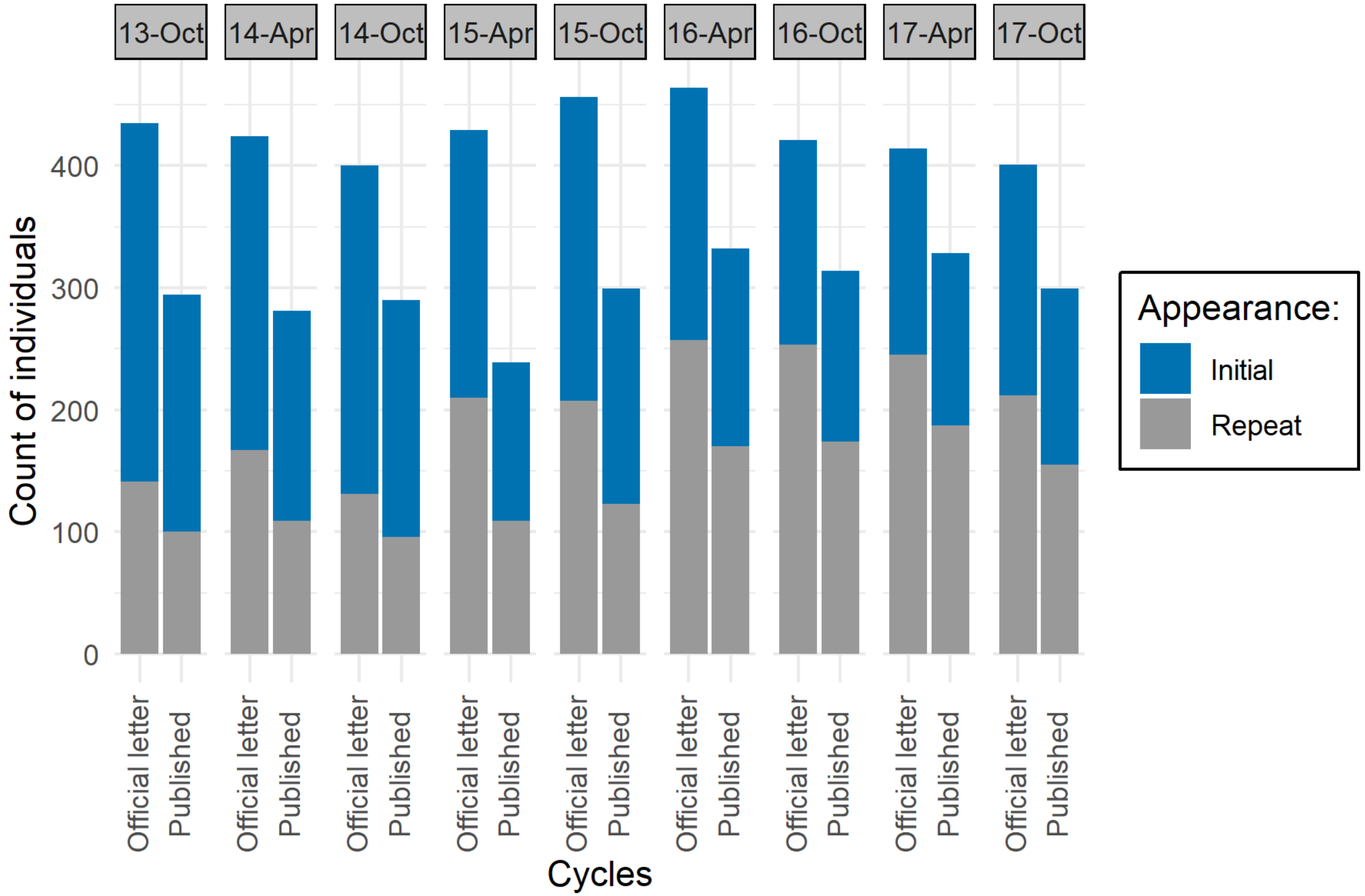
Back

From pre-letter to publication: typical cycle



Note: our analysis focuses only on **individual taxpayers**, not businesses

Lists are somewhat "sticky"



Summary stats – official letter recipients

	Mean	Std. Dev.	P5	P25	Median	P75	P95
Panel A: Among all official letter recipients							
Balance due as of official letter ($ thousands)	859	7,069	181	245	324	539	1,796
Filed on-time return for two-years prior tax year (1/0)	0.57	0.50	0.00	0.00	1.00	1.00	1.00
Filed return for prior tax year (1/0)	0.55	0.50	0.00	0.00	1.00	1.00	1.00
Among those with filed returns for prior tax year:							
AGI ($ thousands)	-250	3,913	-1,100	2	40	152	884
Wages ($ thousands)	71	699	0	0	0	36	213
Has business income (1/0)	0.49	0.50	0.00	0.00	0.00	1.00	1.00
CA resident (1/0)	0.96	0.19	1.00	1.00	1.00	1.00	1.00
Panel B: Among first-time official letter recipients							
Balance due as of official letter ($ thousands)	606	2,012	172	229	300	493	1,512
Filed on-time return for two-years prior tax year (1/0)	0.56	0.50	0.00	0.00	1.00	1.00	1.00
Filed return for prior tax year (1/0)	0.53	0.50	0.00	0.00	1.00	1.00	1.00
Among those with filed returns for prior tax year:							
AGI ($ thousands)	-111	3,509	-825	2	41	165	983
Wages ($ thousands)	98	943	0	0	0	43	272
Has business income (1/0)	0.49	0.50	0.00	0.00	0.00	1.00	1.00
CA resident (1/0)	0.97	0.17	1.00	1.00	1.00	1.00	1.00

Skewed balance distribution

About 50% filing

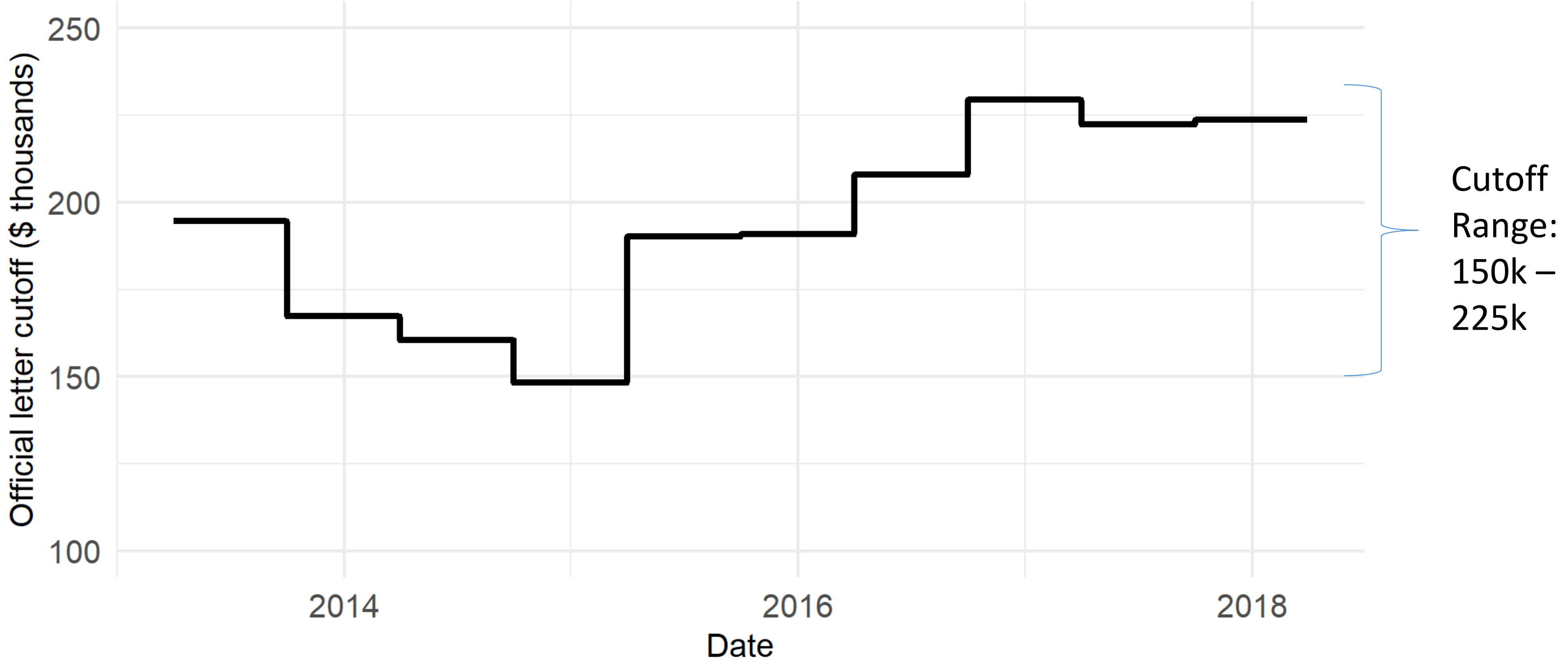
Large variation in AGI

~50% have bus. inc.

Almost all CA residents

Back

Varying cutoff values for warning letter



This motivates our main approach: **within the range of cutoff values, warning letter receipt is quasi-random**

Warning letter: Diff-in-Diff results

Main spec includes:

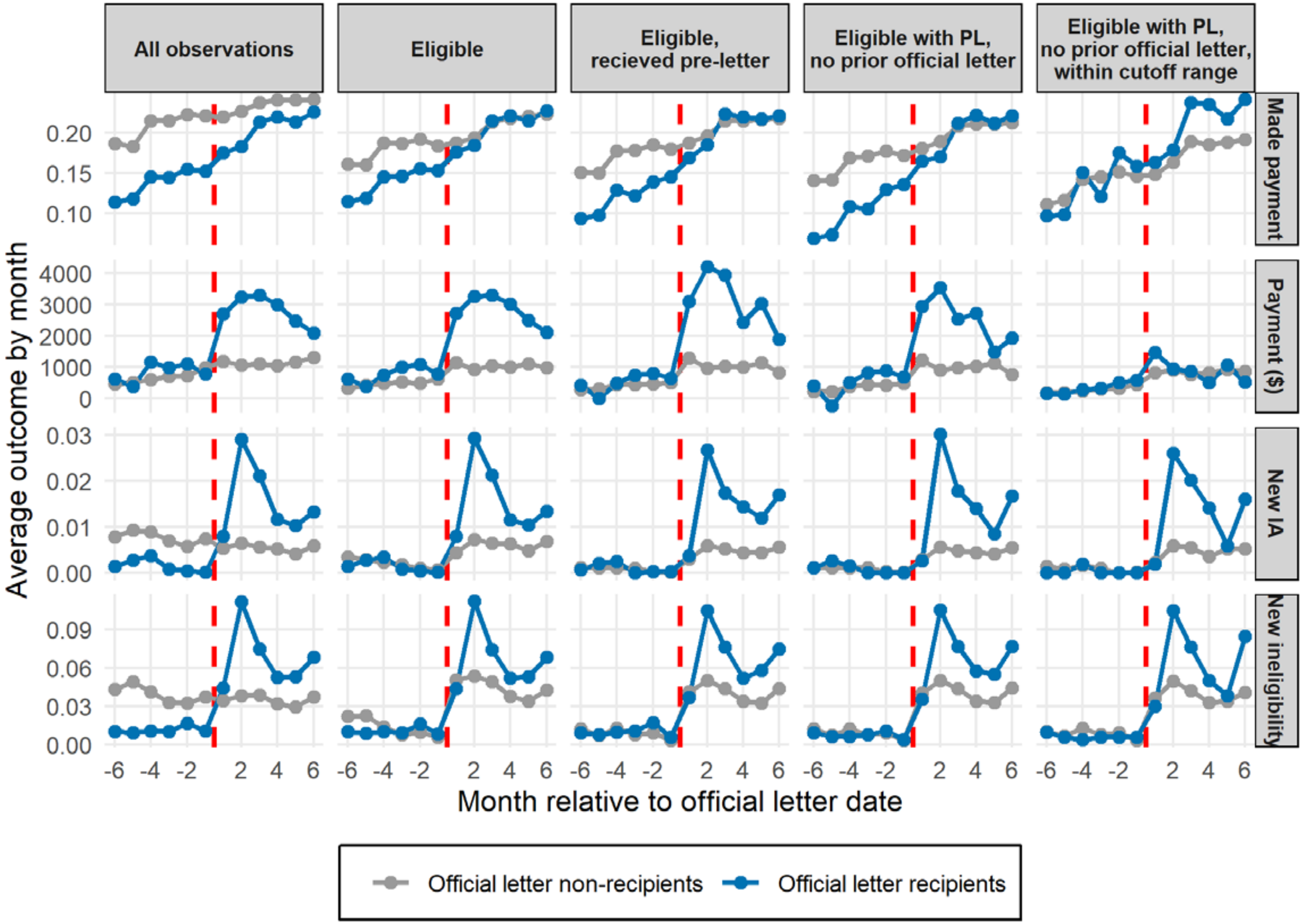
- Letter date +/- 3 months
- eligible for official letter (using most recent status)
- received pre-letter that cycle
- no prior official letter
- **balance roughly $150-225K (within range of cutoffs)**

Results are robust to:

- Data filters (pre-letter, eligibility, prior letter)
- Balance filters (narrower cutoff range, excluding balance control)
- Lengthening time window

	Dependent variables:			
	Made payment	Payment amount ($)	New IA	New ineligibility
Official letter * Post	**0.0223*** **(0.0118)**	**155.77** **305.11**	**0.0119***** **(0.0032)**	**0.0283***** **(0.0066)**
Official letter	0.0010 (0.0146)	73.73 173.93	-0.0004 (0.0003)	-0.0004 (0.0022)
Post	0.0192*** (0.0028)	467.612*** 77.52	0.0041*** (0.0005)	0.036*** (0.0016)
Balance ($ thousands)	-0.0002 (0.0002)	0.93 2.03	0.0000 (0.0000)	0.0000 (0.0000)
April publication	-0.0441*** (0.0049)	-253.07*** 83.20	-0.0014*** (0.0005)	-0.0001 (0.0015)
Intercept	0.1974*** (0.0440)	319.08 356.70	0.0030 (0.0022)	0.0099 (0.0066)
Observations	37,848	37,848	37,848	37,848
R2	0.0047	0.0014	0.0041	0.0157
Mean dep var.	0.1583	606.2506	0.0029	0.0258

Warning letter: Data filters



Controlling for eligibility is important; Other filters marginally affect patterns; Restricting to cutoff range affects payment outcomes, not IA or ineligibility

Back

Warning letter: Robustness

	Dependent variables:			
	Made payment	Payment amount ($)	New IA	New ineligibility
Main specification	**0.0223***	**155.77**	**0.0119*****	**0.0283*****
	(0.0118)	305.11	(0.0032)	(0.0066)
Robustness to data filters:				
Pre-letter: Don't require same-cycle pre-letter	0.0262**	105.67	0.0111***	0.03***
	(0.0105)	251.18	(0.0031)	(0.0062)
Eligibility: Exclude those with any past-month ineligible status change	0.0251	-98.75	0.0079**	0.0159**
	(0.0153)	316.79	(0.0036)	(0.0075)
Eligibility: Don't apply any eligible status filter	0.0161	221.16	0.0181***	0.0611***
	(0.0117)	303.78	(0.0033)	(0.0067)
Eligibility: Drop any non-recipients above that-cycle cutoff value	0.0216*	75.15	0.0119***	0.0279***
	(0.0119)	311.62	(0.0033)	(0.0066)
Prior Top500 experience: Don't restrict to first-time recipients	0.0165	319.36	0.0119***	0.0273***
	(0.0116)	315.10	(0.0030)	(0.0063)
Balance range: Use narrower range of included balances	0.0193	21.75	0.0138***	0.0307***
	(0.0135)	321.73	(0.0038)	(0.0075)
Balance control: Don't control for balance due	0.0223*	155.77	0.0119***	0.0283***
	(0.0118)	305.11	(0.0032)	(0.0066)
Time trend control: Include relative month covariate	-0.0479*	585.08	-0.0037	-0.0164
	(0.0271)	1010.72	(0.0057)	(0.0158)

Back

Warning letter: Robustness

	Dependent variables:			
	Made payment	Payment amount ($)	New IA	New ineligibility
Main specification	**0.0223***	**155.77**	**0.0119*****	**0.0283*****
	(0.0118)	305.11	(0.0032)	(0.0066)
Robustness to observation window:				
Restrict to one month pre/post	0.0021	516.19	-0.0001	-0.0090
	(0.0170)	742.71	(0.0021)	(0.0088)
Restrict to two months pre/post	-0.0027	190.21	0.0101***	0.0246***
	(0.0135)	442.42	(0.0038)	(0.0084)
Expand to four months pre/post	0.0273**	28.56	0.0115***	0.0277***
	(0.0116)	233.87	(0.0028)	(0.0055)
Expand to five months pre/post	0.0312***	58.24	0.0095***	0.0232***
	(0.0112)	211.83	(0.0023)	(0.0046)
Expand to six months pre/post	0.0367***	-6.27	0.01***	0.0267***
	(0.0111)	181.18	(0.0021)	(0.0043)

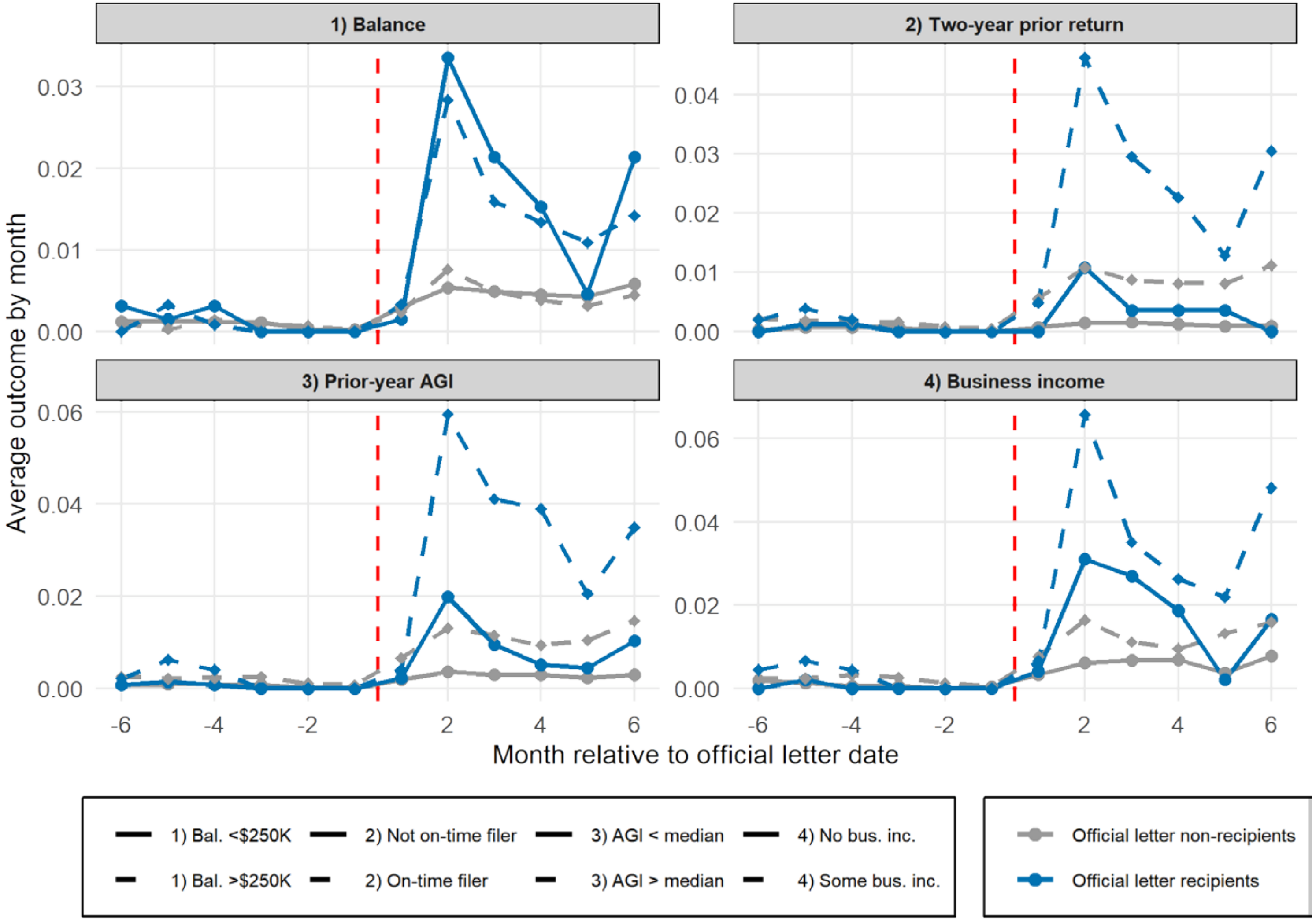
Back

Warning letter: Heterogeneity

- Is treatment effect larger for some groups of taxpayers?
  - Low vs. high balance:
    - High balance may respond more because they will be higher on the list
  - On-time filed two-year prior tax return:
    - On-time filers may respond more because they are more compliant to begin with
  - Low vs. high income:
    - High income may respond more because they are more able
  - Business vs. no business income:
    - Business income may respond more because they can more easily generate new income
- Data from CA income tax returns
  - OTF using two-year prior tax year (2011 tax year for 2013 T500 cycles)
  - Others using one-year prior tax year (2012 tax year for 2013 T500 cycles)

Back

Warning letter: Heterogeneity (IA outcome)



Stronger response for on-time filers, higher-income, and those with business income

Back

Warning letter: Heterogeneity

Run model of the form:

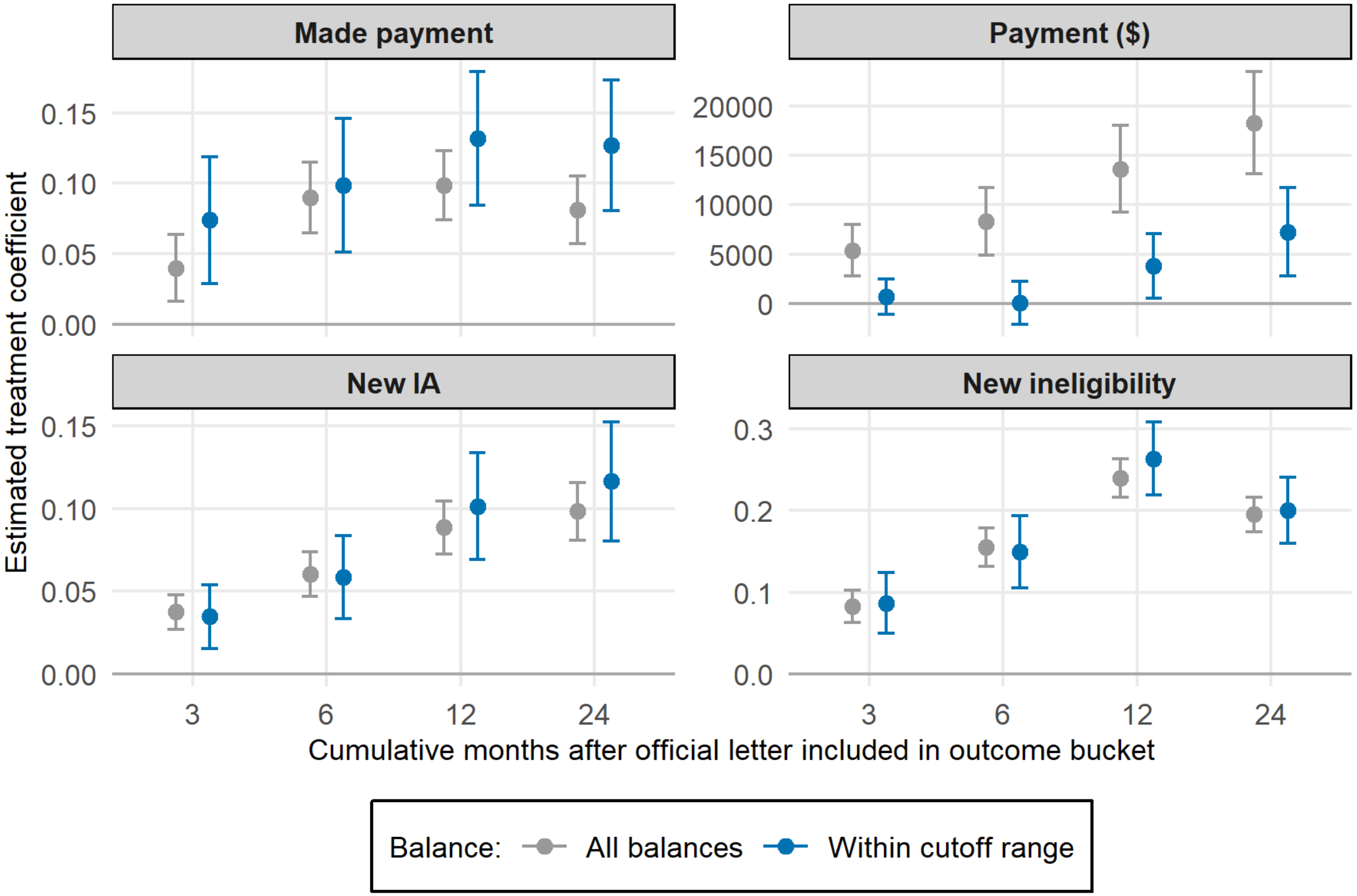
$$Outcome = Treat + Post + Characteristic + TreatXPost + TreatXCharacterstic + PostXCharacteristic + \boldsymbol{TreatXPostXCharacteristic}$$

Table reports the estimates on the coefficient for the final, triple-interaction term

	Dependent variables:			
	Made payment	Payment amount ($)	New IA	New ineligibility
On-time filer	**0.0074**	**-125.51**	**0.017*** **	**0.0353*** **
(1/0)	(0.0230)	600.10	(0.0061)	(0.0130)
Among those with filed prior-year returns:				
Has business income	**0.0134**	**1864.61** **	**0.0187**	**0.0262**
(1/0)	(0.0382)	914.62	(0.0122)	(0.0202)
AGI	**0.0151**	**834.08**	**0.0066** **	**0.0109**
($ millions)	(0.0217)	1260.68	(0.0030)	(0.0154)
Above median AGI	**0.0693** **	**-759.18**	**0.0224** **	**0.0281***
(1/0)	(0.0304)	875.86	(0.0098)	(0.0150)
Non-negative AGI	**-0.0016**	**-725.57**	**0.0018**	**-0.0079**
(1/0)	(0.0380)	1100.72	(0.0151)	(0.0270)

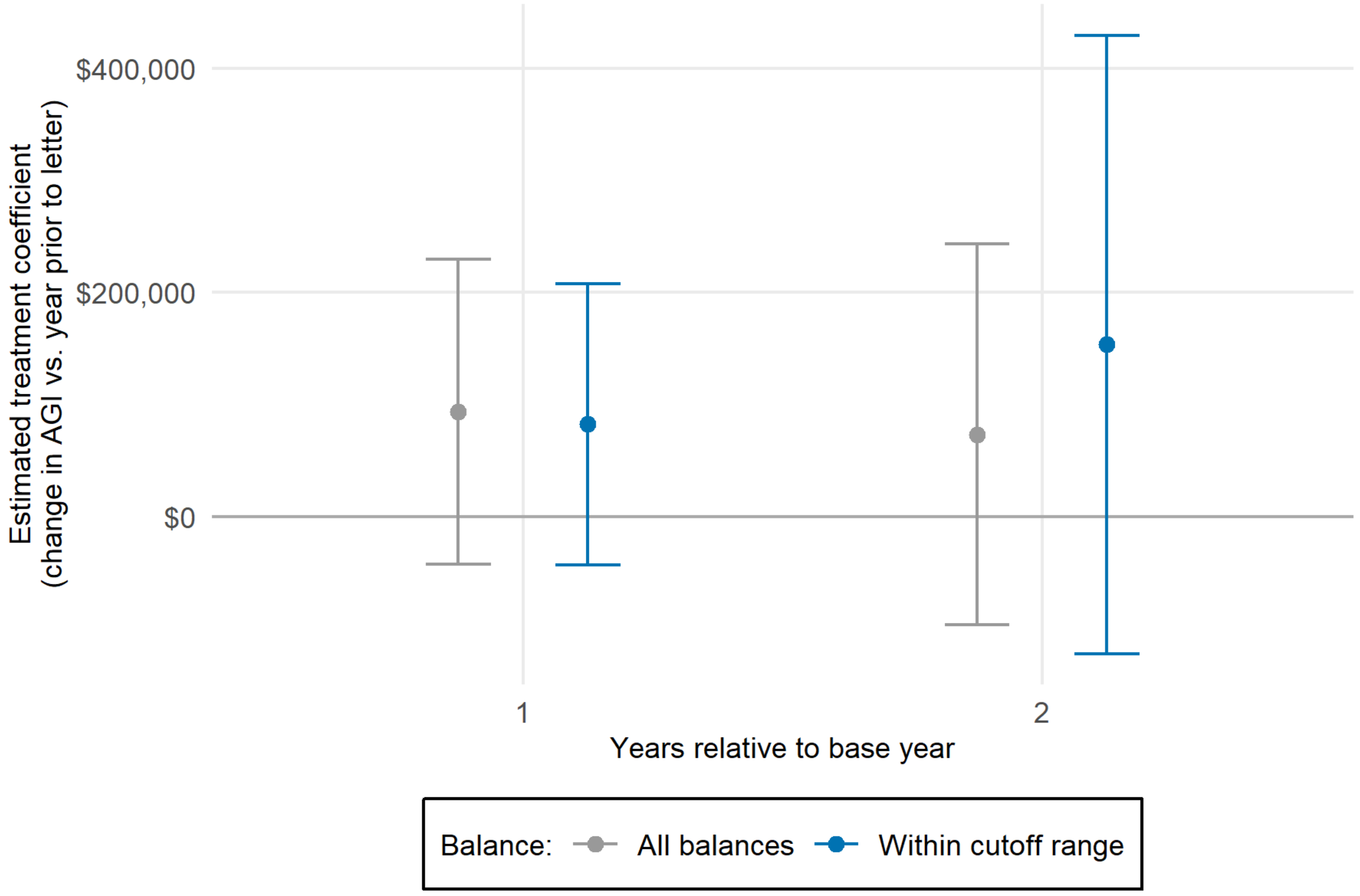
Back

Warning letter: Cumulative outcomes



Back

Warning letter: Future income effects



Point estimates are large and positive, but very imprecise: cannot rule out zero or negative effects

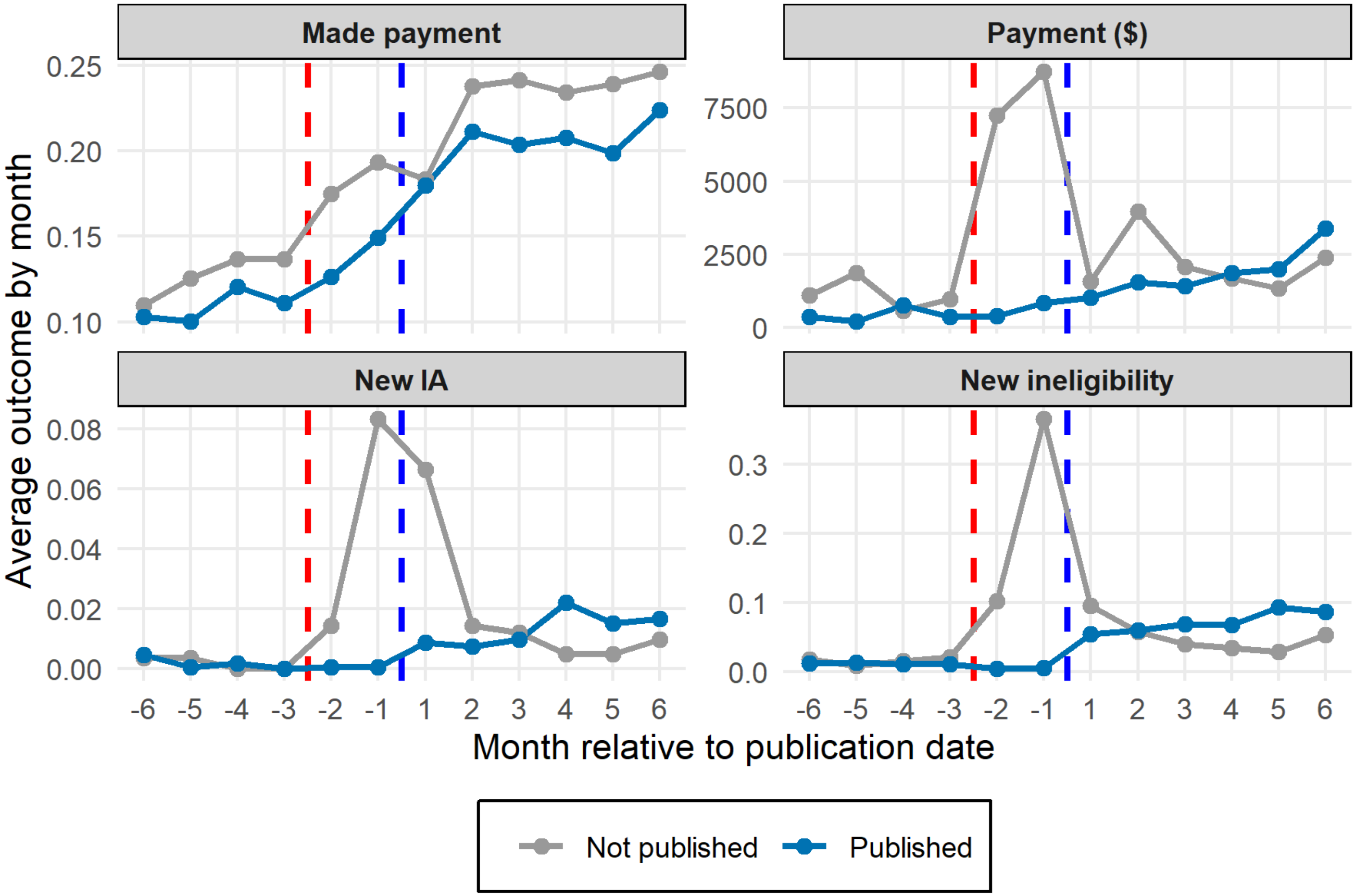
Back

Summary stats – published taxpayers

	Mean	Std. Dev.	P5	P25	Median	P75	P95
Panel A: Among all published taxpayers							
Balance due as of official letter ($ thousands)	915	8,373	180	244	320	540	1,662
Filed on-time return for two-years prior tax year (1/0)	0.53	0.50	0.00	0.00	1.00	1.00	1.00
Filed return for prior tax year (1/0)	0.52	0.50	0.00	0.00	1.00	1.00	1.00
Among those with filed returns for prior tax year:							
AGI ($ thousands)	-320	3,736	-1,055	1	37	131	840
Wages ($ thousands)	42	171	0	0	0	29	186
Has business income (1/0)	0.49	0.50	0.00	0.00	0.00	1.00	1.00
CA resident (1/0)	0.96	0.20	1.00	1.00	1.00	1.00	1.00
Panel B: Among first-time published taxpayers							
Balance due as of official letter ($ thousands)	606	2,170	171	230	301	496	1,448
Filed on-time return for two-years prior tax year (1/0)	0.53	0.50	0.00	0.00	1.00	1.00	1.00
Filed return for prior tax year (1/0)	0.51	0.50	0.00	0.00	1.00	1.00	1.00
Among those with filed returns for prior tax year:							
AGI ($ thousands)	-239	3,934	-817	2	37	135	825
Wages ($ thousands)	46	167	0	0	0	36	212
Has business income (1/0)	0.49	0.50	0.00	0.00	0.00	1.00	1.00
CA resident (1/0)	0.97	0.18	1.00	1.00	1.00	1.00	1.00

Similar takeaways to official letter recipients

Effect of publication



Policy implications

- Cost-benefit analysis comparing revenue raised to admin. cost
  - Welfare analysis should also consider private cost (Keen and Slemrod (2017))
- We estimate longer-term cumulative outcomes:
  $$Cumulative\ Outcome_i \sim \beta_0 + \beta_1 Letter_i + \beta_2 Balance_i + \beta_3 April\ Cycle_i + \varepsilon_i$$
- Not diff-in-diff as earlier, but for long-term outcomes this should not matter:
  - Eligibility for letter receipt is conditional on no IA and no ineligibility prior to receipt
  - Payment amounts are small prior to warning letter

Research, Applied Analytics, and Statistics

TAX POLICY CENTER
URBAN INSTITUTE & BROOKINGS INSTITUTION

12th Annual IRS/TPC Joint Research Conference on Tax Administration

June 16, 2022