A Report on "A Welfare Analysis of Tax Audits Across the Income Distribution" by Boning et al. (2025)

Reviewer 2

February 11, 2026

v1

isitcredible.com

Disclaimer

This report was generated by large language models, overseen by a human editor. It represents the honest opinion of The Catalogue of Errors Ltd, but its accuracy should be verified by a qualified expert. Comments can be made here. Any errors in the report will be corrected in future revisions.

> I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.
>
> Plato, *The Apology of Socrates*, 21d

> To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.
>
> Karl Popper, 'Knowledge and the Shaping of Reality'

Overview

Citation: Boning, W. C., Hendren, N., Sprung-Keyser, B., and Stuart, E. (2025). A Welfare Analysis of Tax Audits Across the Income Distribution. *Quarterly Journal of Economics*. Vol. 140, No. 1, pp. 63–112.

URL: https://doi.org/10.1093/qje/qjae037

Abstract Summary: This paper estimates the returns to IRS audits across the income distribution, finding that audits of high-income taxpayers yield significantly more revenue per dollar spent than audits of low-income taxpayers, especially when accounting for individual deterrence effects. The analysis uses a Marginal Value of Public Funds framework to assess the welfare consequences of expanded audits relative to other revenue-raising policies.

Key Methodology: Empirical analysis using comprehensive internal IRS accounting information and audit-level enforcement logs, exploiting a decline in audit rates to estimate marginal returns, and using random audits from the National Research Program (NRP) to estimate individual deterrence effects, all within a Marginal Value of Public Funds (MVPF) framework.

Research Question: Do audits of high-income taxpayers generate more revenue per dollar spent on tax enforcement, and what are the welfare consequences of tax audits across the income distribution compared to alternative revenue-raising policies?

Summary

Is It Credible?

This article presents a comprehensive analysis of the returns to IRS audits, utilizing internal administrative data to argue that expanding enforcement on high-income taxpayers yields substantial revenue and welfare gains. The authors construct a detailed accounting of audit costs—incorporating overhead, management, and fringe benefits often omitted in prior literature—and compare these against revenues derived from both the initial audit and subsequent changes in taxpayer behavior. The central claim is striking: audits of taxpayers in the 90th–99th income percentiles generate a return of more than 12:1, driven largely by a "deterrence multiplier" where the net present value of future tax payments exceeds the initial audit revenue by a factor of 3.2 (p. 63). The authors further argue that because the Marginal Value of Public Funds (MVPF) for these audits (estimated at 1.15) is lower than that of tax rate increases on high earners (typically estimated around 1.5 to 2.0), enforcement is a more efficient redistributive tool than raising statutory rates (p. 106).

The credibility of these headline figures depends heavily on the robustness of the deterrence estimates. The 3.2x multiplier is derived from random audits conducted via the National Research Program (NRP). While the authors validate this multiplier against a sample of operational audits, finding a broadly consistent factor of 2.5 (p. 93), the application of a single average multiplier across the income distribution introduces uncertainty. Specifically, the analysis lacks the statistical power to precisely estimate deterrence effects for the top 1% of earners, and the estimates for the top 0.1% are based on extrapolation rather than direct observation (p. 98). The unwinsorized data for the top 0.1% actually show a massive negative point estimate for deterrence (-117.68), driven by outliers, though the winsorized estimate is positive (4.11) (p. 95). Consequently, while the conclusion that audits are highly prof-

itable seems robust for the affluent (90th–99th percentiles), the precise returns for the ultra-wealthy (top 0.1%) rely on the assumption that their behavioral response mirrors that of the broader population.

The article's assessment of costs and marginal returns is rigorous but rests on specific accounting and structural assumptions. The authors estimate that marginal costs are significantly lower than average costs due to economies of scale, positing that central overhead is largely fixed (specifically, 27% fixed versus 73% variable) (p. 87). If this accounting convention understates the variable costs of a major enforcement expansion—such as the training and retention of specialized auditors required for complex high-income returns—the calculated return on investment would decrease. Furthermore, the finding that marginal audits yield returns similar to average audits relies on the observation that revenue per audit remained stable while audit rates fell by 40% between 2010 and 2014 (p. 83). While the authors interpret this as evidence that the IRS was not prioritizing audits based on revenue maximization (and thus returns did not diminish as rates fell), it is also possible that the stability reflects a changing compliance environment or other unobserved factors that offset the selection effect.

Finally, the welfare analysis is notably conservative in one respect but contingent on normative judgments in another. The calculation omits general deterrence—the spillover effect of audits on the non-audited population—which the authors acknowledge would likely lower the MVPF further, strengthening their case for audits (p. 109). However, the MVPF framework treats the payment of evaded taxes and the time burden of compliant taxpayers as social costs. The conclusion that audits are welfare-enhancing holds even under conservative assumptions about these burdens, but it is sensitive to how a social planner values the "no change" audits that constitute a significant portion of enforcement activity at the top (p. 102). Overall, while the precise magnitude of the returns is subject to estimation uncertainty regarding the very wealthy, the qualitative evidence that high-income audits are a highly efficient

revenue source appears credible.

The Bottom Line

Boning et al. provide compelling evidence that IRS audits of high-income taxpayers generate revenue far in excess of their costs, primarily through long-term improvements in compliance. While the specific total return on investment for the top 0.1% relies on extrapolation for the deterrence component, and the cost estimates depend on specific accounting assumptions, the core finding—that enforcement is a more efficient revenue-raising tool than tax hikes for the affluent—appears robust. The analysis likely understates the total benefits by omitting general deterrence, suggesting the case for expanded audits is even stronger than presented.

Potential Issues

Omission of general deterrence effects renders the welfare analysis incomplete: The article's welfare analysis is limited to the effects of audits on the audited individuals themselves (individual deterrence) and does not measure general deterrence—the effect of an enforcement regime on the compliance of the non-audited population. This is a significant omission, as general deterrence is a primary channel through which tax enforcement is expected to generate revenue and influence behavior. The authors acknowledge this limitation, suggesting that including spillover effects on others would likely "substantially increase the returns to audits" (p. 109). In the Online Appendix, they elaborate that under their baseline theoretical assumptions, general deterrence has no first-order effect on government revenue, but that relaxing these assumptions would create a positive fiscal effect that would lower the Marginal Value of Public Funds (MVPF), making audits appear even more efficient (Online Appendix, pp. 92–93). While the authors argue that this omission makes their estimates conservative, it remains a central limitation; the article provides a partial equilibrium analysis of a general equilibrium phenomenon, and its welfare conclusions are therefore incomplete.

The external validity of deterrence effects is debatable: The article's central estimate of long-run behavioral change—a multiplier finding that the net present value of future revenue is 3.2 times the initial audit revenue—is derived from the IRS's National Research Program (NRP), which involves random, intensive audits for research purposes (p. 90). This multiplier is then applied to the article's main sample of standard, operational audits, which are targeted based on suspected noncompliance. The behavioral response to being audited for research versus being audited for suspected wrongdoing may differ. The authors acknowledge this potential mismatch and defend their approach by arguing it is likely "a slightly conservative assumption" because individuals targeted for cause may have larger behavioral re-

sponses (p. 96, fn 40). More importantly, they provide empirical support for this choice with a robustness check on a sample of operational audits, which finds a deterrence multiplier of "approximately 2.5, consistent with the results for NRP audits" (p. 93; Online Appendix Figure A.XIV). While this check provides reassurance, the fact remains that the article's primary behavioral parameter is estimated from a treatment and population that do not perfectly match the policy context it aims to inform.

The use of a single, average deterrence multiplier in the main analysis masks significant heterogeneity and uncertainty: The article's headline results on the total returns to marginal audits (e.g., a 12.5:1 return for the 90th–99th percentiles) are derived by applying a single, average individual deterrence multiplier of 3.2 across all income groups (p. 96). This choice obscures the fact that the article's own estimates of this multiplier, presented in Table I, are highly variable and uncertain across income groups (pp. 94–95). For instance, the point estimate for the 90th–99th percentile is 6.29, while for the 80th–90th percentile it is 2.04. The confidence intervals are extremely wide, particularly for high-income groups; for the 99th–99.9th percentile, the 95% confidence interval is [-2.57, 11.78]. The authors acknowledge this uncertainty, stating they "cannot reject the hypothesis that audits generate 5× additional revenue from deterrence, nor can we reject that they have large negative effects" for the top 1% (p. 96, fn 39). They justify using the average multiplier based on its "greater precision" (p. 98). However, the authors also present an alternative analysis using income-specific multipliers, which shows that using the average multiplier is a conservative choice for high-income groups, as their specific point estimates are higher (p. 98, Figure VII, Panel B). This modeling decision in the main analysis creates a misleading impression of uniformity in a key behavioral parameter where the data suggest there is none.

Deterrence effects for the top 0.1% of taxpayers are based on extrapolation, not direct evidence: The article lacks the statistical power to precisely estimate the individual deterrence effect for the highest-income taxpayers, a point the authors explicitly

concede: "We omit the top 1% from our primary results because we do not have the necessary power to precisely estimate the deterrence multiplier" (p. 98, fn 42). Despite this, the article presents illustrative calculations of total returns for these groups (e.g., 36.0:1 for the top 0.1%) by applying the average 3.2x multiplier derived from the rest of the population (p. 98, fn 41). This is a strong and untested assumption, as the nature of noncompliance and the behavioral response to audits may be different at the highest income levels. The article's unwinsorized data for the top 0.1% show a massive negative point estimate for deterrence (-117.68), highlighting the extreme variance and the fragility of any conclusion for this group without winsorization (p. 95, Table I). The article is transparent about this limitation, but readers should be aware that the findings for the very highest earners are based on extrapolation rather than direct estimation.

Cost estimates rely on specific accounting assumptions: A core contribution of the article is its comprehensive measurement of audit costs, which relies on internal IRS accounting data. The method involves allocating non-direct costs (e.g., management, IT, overhead) to individual audits "in proportion to the audit's direct wage costs" (p. 74). Similarly, the distinction between marginal and average costs rests on an assumption that "central overhead costs are fixed" while other costs are variable (p. 87). These are accounting conventions, not empirically demonstrated relationships. The authors acknowledge these are assumptions and argue their choices are conservative, as alternative methods would tend to lower the costs for high-income audits and thus increase their estimated return on investment (p. 74). They also validate their marginal cost assumption against external data, noting it is "broadly consistent with, but slightly more conservative than, estimates from existing IRS budget requests" (p. 87). While the authors' handling of this issue is transparent and reasonable, the fact remains that the article's quantitative cost figures, and all subsequent return-on-investment and welfare calculations, are contingent on these specific modeling choices.

The analysis is static and does not model dynamic policy effects: The article's analysis is based on a "steady state" that existed for returns filed between 2010 and 2014 (p. 109). The marginal cost estimates explicitly omit the transitional costs of a rapid expansion of the audit force, such as hiring and training, which would likely lower returns in the short run. The authors acknowledge this, noting that their estimates are for a "steady state" and that training costs may be higher in the short run for a large expansion, referencing CBO methodology that incorporates rising auditor productivity over time (p. 87, fn 29). Furthermore, the analysis does not model the potential for a strategic "arms race," where a sustained increase in enforcement could induce taxpayers and their advisors to invest in more sophisticated and harder-to-detect evasion techniques. While the article finds stable returns during a period of audit reduction, it is uncertain whether these returns would persist under a large and sustained audit expansion.

The welfare analysis rests on specific normative assumptions: The article's welfare conclusions are derived using the Marginal Value of Public Funds (MVPF) framework. This framework has embedded normative assumptions. For example, it calculates the private welfare cost of an audit by summing the taxpayer's compliance burden and the total tax revenue collected ($R + B$), effectively treating the payment of legally owed taxes as a private cost from the audited individual's perspective (p. 100). The framework also monetizes the burden on compliant taxpayers who are audited but found to owe no additional tax; the article finds that the rate of such "no change" audits reaches "upward of 60% in the top 1%" (p. 102). The authors use this framework transparently and explicitly discuss how different "welfare weights" can be applied to compliant versus noncompliant taxpayers to incorporate concerns for fairness (p. 99). However, the article's headline welfare comparisons are contingent on the specific, and debatable, value judgments inherent in this standard public finance methodology.

The interpretation of stable audit returns during budget cuts is debatable: The

article argues that because average revenue per audit remained stable between 2010 and 2014 while audit rates fell 40%, the marginal audits that were cut must have had similar returns to the average (p. 83). This is interpreted as evidence that the IRS was not prioritizing audits based on expected revenue. An alternative explanation is that two opposing forces cancelled each other out: the IRS did cut its least promising audits (which would raise the average return of the remainder), but this was offset by a changing economic environment or improved tax evasion technology that made non-compliance harder to detect. The authors attempt to rule out this alternative by showing that returns to random (NRP) audits did not decline over the same period, which is evidence against a worsening compliance environment (p. 86). While this strengthens their interpretation, the conclusion is not immune to this alternative explanation.

Several methodological and data choices limit transparency and comparability: The article employs several methodological choices that, while justified, affect the transparency and comparability of the results. First, there is a temporal mismatch between the primary audit data (2010–2014 returns) and the taxpayer burden data (from a 2023 survey on 2019–2021 returns) (pp. 75, 101). The authors note that average burden has increased over time, suggesting that applying the more recent, higher burden figures to older revenue data likely biases the MVPF upward, making their findings conservative (p. 102, fn 47). Second, the article adopts a non-standard definition of an "audit," bundling multi-year examinations into a single event, which complicates direct comparison with official statistics (p. 71, fn 6). Third, the analysis is restricted to audits from the IRS's Small Business/Self-Employment (SB/SE) division, which excludes audits from other key divisions like Large Business and International (LB&I) (p. 70, fn 5). While the SB/SE division does handle high-income individuals, this choice excludes certain types of audits of complex business entities. The authors are transparent about these choices, but they are important context for interpreting the findings.

Minor presentation and clerical issues exist: The article contains several minor clerical issues. The time window for the long-run deterrence analysis is described as "14 years of postaudit revenue" in the text (p. 91) but as "up to 13 years postaudit" in the notes to Table I (p. 95), a likely difference in whether the audit year itself is included in the count. The choice of a 3% discount rate is justified by the timing of costs and revenues but is not subjected to a sensitivity analysis (p. 78). Finally, there is a trivial rounding difference of 0.01 in the calculation of the average overhead cost multiplier presented in Appendix Table A.IV (Online Appendix, p. 80). These minor issues do not affect the substantive conclusions of the article.

Future Research

Measurement of general deterrence: Future work should attempt to quantify the general deterrence effects of tax enforcement, which this study omits. Research could exploit randomized controlled trials or natural experiments where audit probabilities are visibly increased for specific geographic regions or professional sectors to measure the compliance response of non-audited taxpayers. Incorporating these spillovers would provide a more complete picture of the Marginal Value of Public Funds for enforcement activities.

Behavioral responses of the ultra-wealthy: Future research should focus specifically on identifying the deterrence effects for the top 0.1% of the income distribution, where this study lacked statistical power. Given the complexity of high-net-worth returns, researchers could utilize administrative data to examine how specific enforcement interventions—such as audits of pass-through entities or offshore accounts—affect the subsequent tax planning and reporting strategies of the ultra-wealthy, rather than relying on extrapolations from the broader population.

Dynamic analysis of enforcement costs: Future work should model the dynamic costs and returns of audit expansions, moving beyond the steady-state analysis presented here. This could involve estimating the "learning curve" costs of hiring and training new auditors, as well as modeling the potential "arms race" between tax authorities and evasion technologies. Such an analysis would better inform the optimal pace and scale of funding increases for tax administration.

isitcredible.com