The Audit Cliff That Wasn't: Nonprofit Revenue Bunching at State Charitable Audit Thresholds

APEP Autonomous Research* @olafdrw

March 22, 2026

Abstract

Thirty-four US states require nonprofits to pay for independent CPA audits once annual revenue exceeds a state-specific threshold, creating a "compliance cliff" with costs of $10,000–$100,000. Using a multi-threshold bunching design applied to 555,714 tax-exempt organizations from the IRS Exempt Organizations Business Master File, I test whether nonprofits strategically manage reported revenue to avoid these audit mandates. The pooled bunching estimate at the modal $500,000 threshold is statistically indistinguishable from zero ($\hat{b} = -0.046$, SE $= 0.122$), though individual states show heterogeneous responses ranging from -1.15 to 1.32. A difference-in-bunching comparison between mandate and no-mandate states confirms the small average effect. These results suggest that charitable audit thresholds do not systematically distort nonprofit revenue reporting, despite the substantial compliance costs they impose.

JEL Codes: L31, H83, M42

Keywords: nonprofits, bunching, audit mandates, compliance costs, charitable regulation

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 20m).

1. Introduction

A nonprofit executive whose organization has grown to \$490,000 in annual revenue faces a stark choice. In most US states, crossing the \$500,000 revenue mark triggers a mandatory independent CPA audit costing \$10,000 to \$100,000—an implicit tax of 2–20% on the marginal dollar. Standard economic theory predicts that such notches should generate bunching: organizations should cluster their reported revenue just below the threshold to avoid the discontinuous jump in compliance costs (Kleven and Waseem, 2013; Saez, 2010). If nonprofits respond as strongly as taxpayers do to tax notches, these audit mandates could systematically distort the growth and transparency of the charitable sector.

This paper tests that prediction using a multi-threshold bunching design that exploits variation across 34 US states with charitable audit requirements. Thresholds range from \$300,000 (Illinois) to \$2,000,000 (California), providing natural variation in the height of the compliance cliff. I apply the polynomial counterfactual density framework of Kleven and Waseem (2013) to 555,714 tax-exempt organizations from the IRS Exempt Organizations Business Master File (EO BMF), estimating excess mass below each state's audit threshold. The design offers 34 independent replications of the same behavioral test—each state-threshold pair constitutes a separate bunching experiment.

The central finding is a well-identified null. The pooled bunching estimate at the modal \$500,000 threshold—shared by 29 states—is $\hat{b} = -0.046$ with a bootstrap standard error of 0.122, statistically indistinguishable from zero. A density discontinuity test shows a modest 7.8% excess of organizations in the \$25,000 window below versus above \$500,000, but a difference-in-bunching comparison between audit-mandate and no-mandate states confirms that this gap is largely attributable to round-number effects rather than audit avoidance. The placebo tests at round-number thresholds (\$300,000, \$400,000, \$600,000) in states without audit mandates show bunching of comparable or larger magnitude, ruling out the interpretation that the density patterns are driven primarily by regulatory avoidance.

The null average masks striking heterogeneity. Individual state estimates range from -1.15 to 1.32, with Connecticut ($\hat{b} = 1.32$), Illinois at \$300,000 ($\hat{b} = 1.04$), and Mississippi ($\hat{b} = 0.92$) showing substantial bunching, while states like New Jersey ($\hat{b} = -1.15$) and Wisconsin ($\hat{b} = -0.57$) show the opposite. Across nonprofit types, science organizations (NTEE code U) show the largest bunching ratios ($\hat{b} = 1.64$), while environmental and animal-related organizations show negative estimates. This heterogeneity suggests that the behavioral response to audit mandates depends on organizational capacity, state enforcement intensity, and sector-specific compliance cultures rather than following a uniform economic prediction.

These results contribute to three literatures. First, within the bunching literature pioneered by Saez (2010) and Kleven and Waseem (2013), I provide evidence that not all notches generate the predicted behavioral response—echoing recent work showing that optimization frictions and salience can dampen bunching even at substantial financial thresholds (Chetty et al., 2011; Kleven, 2016). Second, within nonprofit economics, I offer the first multi-state analysis of audit threshold effects, extending Yildirim et al. (2018) beyond New York's single-state setting. The cross-state variation enables a difference-in-bunching placebo test that is impossible in single-threshold designs. Third, for the policy debate on nonprofit regulation, the results suggest that policymakers can adjust audit thresholds without fear of massive behavioral distortion—the compliance cliff exists in theory but bites selectively in practice.

The paper relates to a growing literature on compliance costs in regulated sectors. Djankov et al. (2002) document how regulatory burden affects firm entry; Botero et al. (2004) find that heavier regulation of labor is associated with lower labor force participation. In the nonprofit sector specifically, Garven et al. (2018) show that state regulation affects charitable giving and organizational efficiency, while Krishnan et al. (2006) examine how auditing requirements influence financial reporting quality. My contribution is to quantify the behavioral margin directly through bunching estimation rather than inferring compliance effects from cross-sectional correlations.

The null result is itself informative. If nonprofits do not bunch at audit thresholds despite costs of 2–20% of marginal revenue, it suggests either that (a) organizations lack the ability to precisely control reported revenue, (b) the reputational benefits of audited financials offset the direct costs, (c) donors and grantmakers already require audits below the statutory threshold, or (d) revenue is sufficiently noisy that strategic positioning is infeasible. The heterogeneity across states and NTEE categories is consistent with a combination of these explanations: organizations with sophisticated financial management in states with strict enforcement may respond, while the modal nonprofit does not.

The remainder of the paper proceeds as follows. Section 2 describes the institutional landscape of state charitable audit mandates. Section 3 presents the data. Section 4 outlines the bunching methodology. Section 5 reports results including the main estimates, heterogeneity, and robustness checks. Section 6 discusses implications and concludes.

2. Institutional Background

State charitable audit requirements. US states regulate charitable solicitation through registration and reporting requirements administered by state attorneys general or secretaries of state (Fishman and Schwarz, 2015). A central component of this regulation is the financial audit threshold: organizations whose annual revenue (or, in some states, contributions) exceeds a state-specific level must submit independently audited financial statements prepared by a licensed CPA (National Association of State Charity Officials, 2020). Below the threshold, organizations typically submit reviewed financial statements or self-prepared reports, which are substantially less costly.

Threshold variation. As of 2024, 34 states impose audit requirements at thresholds ranging from \$300,000 to \$2,000,000. The modal threshold is \$500,000 (29 states), with Illinois at \$300,000, Minnesota, New York, and Virginia at \$750,000, and California at \$2,000,000. The remaining 17 states (including Texas and Florida) have no state-level charitable audit mandate, though federal Form 990 filing requirements apply to all organizations with gross receipts above \$200,000 or total assets above \$500,000 (Internal Revenue Service, 2024).

Compliance costs. The cost of an independent CPA audit varies with organizational size and complexity. For organizations near the \$500,000 threshold, audit fees typically range from \$10,000 to \$30,000 (American Institute of CPAs, 2020). For larger organizations near the \$2,000,000 threshold, fees can exceed \$50,000. These costs represent a discontinuous jump at the threshold: an organization with \$499,999 in revenue faces no state audit requirement, while one with \$500,001 must pay for a full independent audit. This creates a notch—a discrete change in costs as a function of a continuous running variable—that is the foundation for bunching estimation (Kleven and Waseem, 2013).

Enforcement and compliance. State enforcement of audit requirements varies considerably. Some states (New York, California, Connecticut) maintain active charity registration units with dedicated staff, while others rely on complaint-driven enforcement (Irvin, 2005). This variation in enforcement intensity may explain the heterogeneous bunching responses documented below. Organizations that fail to file required audits face penalties ranging from registration revocation to civil fines, though prosecution is rare for reporting violations absent evidence of fraud (Crimm, 2008).

3. Data

The primary data source is the IRS Exempt Organizations Business Master File (EO BMF), a public dataset containing information on all organizations recognized as tax-exempt under Section 501(c) of the Internal Revenue Code. I downloaded the four regional EO BMF files from the IRS Statistics of Income Division in March 2026, yielding 1,938,732 organizations

Table 1: Summary Statistics: Tax-Exempt Organizations by State Audit Mandate

	N	Mean Rev.	Median Rev.	SD Rev.	% Below $1M
No-Mandate States	136,958	$6,349,856	$201,673	$99,218,114	78.4%
Audit-Mandate States	418,756	$7,865,781	$204,108	$201,191,270	76.9%
All Organizations	555,714	$7,492,175	$203,464	$181,462,181	77.3%

Notes: Data from IRS Exempt Organizations Business Master File (2024). Revenue is annual total revenue (`REVENUE_AMT`). Audit-mandate states are those requiring independent CPA audit above a state-specific revenue threshold. Sample restricted to organizations with positive reported revenue.

with EIN, state, revenue, assets, NTEE classification, and ruling date.

Sample construction. I restrict the sample to organizations with positive reported revenue (`REVENUE_AMT` > 0), yielding 555,714 organizations. Of these, 418,756 (75.4%) are located in the 34 states with charitable audit mandates, and 136,958 (24.6%) are in the 17 states without mandates. The revenue variable captures total annual revenue as reported to the IRS, which may differ from the contributions-based thresholds used by some states. This measurement concern is addressed in the robustness section.

State threshold data. I compiled audit threshold levels from state charity registration statutes, cross-referencing the National Council of Nonprofits, Hurwit & Associates' state-by-state filing guides, and individual state attorney general websites. The compilation identifies four distinct threshold levels: $300,000 (1 state), $500,000 (29 states), $750,000 (3 states), and $2,000,000 (1 state).

Table 1 presents summary statistics. Organizations in audit-mandate states are similar to those in no-mandate states: median revenue is approximately $202,000–$204,000 in both groups, and roughly 77% of organizations report revenue below $1 million. The similarity of the two groups supports the use of no-mandate states as a placebo comparison.

4. Empirical Strategy

Bunching estimation. I follow the polynomial counterfactual density approach of Kleven and Waseem (2013) and Chetty et al. (2011). For each state with an audit threshold $\bar{z}$, I divide organizations into revenue bins of width $5,000 (proportionally wider for higher thresholds) and count the number of organizations in each bin. I then exclude a region around the threshold—baseline: $[\bar{z} - 25{,}000, \bar{z} + 15{,}000]$ for $500,000 thresholds—and fit a degree-7 polynomial to the remaining bin counts. The polynomial provides the counterfactual density: what the distribution would look like absent any behavioral response to the threshold.

The bunching ratio is defined as:

$$\hat{b} = \frac{\sum_{j \in \mathcal{B}} (c_j - \hat{c}_j^{cf})}{\bar{c}^{cf}} \tag{1}$$

where c_j is the observed count in bin j, $\hat{c}_j^{cf}$ is the polynomial counterfactual, $\mathcal{B}$ is the set of bins in the bunching region (below the threshold), and $\bar{c}^{cf}$ is the average counterfactual bin count. Positive values indicate excess mass below the threshold—organizations clustering to avoid the audit mandate.

Inference. Standard errors are computed via bootstrap with 500 replications, resampling organizations with replacement within each state to preserve the state-level clustering structure. This follows Kleven and Waseem (2013) and accounts for potential within-state correlation in reporting behavior.

Difference-in-bunching. To separate audit-specific behavioral responses from round-number effects, I compare the revenue density in the \$25,000 window below versus above \$500,000 in audit-mandate states versus no-mandate states. Defining d_k^g as the fraction of organizations in group g (mandate/no-mandate) falling in region k (below/above), the difference-in-bunching estimator is:

$$\widehat{DiB} = (d_{\text{below}}^{\text{mandate}} - d_{\text{below}}^{\text{no-mandate}}) - (d_{\text{above}}^{\text{mandate}} - d_{\text{above}}^{\text{no-mandate}}) \tag{2}$$

Positive values indicate excess bunching in mandate states beyond any round-number clustering.

Identification. The identifying assumption is that the counterfactual revenue distribution is smooth through the threshold—absent the audit requirement, there would be no discontinuity in the density of organizations at the threshold level (Kleven and Waseem, 2013). Two potential threats merit discussion. First, round-number bunching: organizations may cluster at \$500,000 for psychological or reporting reasons unrelated to audit mandates. The difference-in-bunching design addresses this by using no-mandate states as a placebo. Second, revenue measurement: the EO BMF reports total revenue while some states (e.g., Pennsylvania, Georgia, Maryland) threshold on contributions rather than total revenue. To the extent that total revenue is a noisy proxy for the threshold-relevant measure, classical measurement error in the running variable would attenuate bunching estimates toward zero (Kleven, 2016). This concern is most acute for the null result: the absence of detectable bunching may partly reflect misalignment between the measured and statutory running variables. Future work

Table 2: Bunching Estimates by State Audit Threshold Level

Threshold	States	Orgs.	Mean $\hat{b}$	Median $\hat{b}$	Min $\hat{b}$	Max $\hat{b}$
$300,000	1	4,703	1.041	1.041	1.041	1.041
$500,000	29	51,630	0.004	-0.055	-1.15	1.32
$750,000	3	11,073	0.545	0.553	0.35	0.732
$2,000,000	1	6,505	0.52	0.52	0.52	0.52
Pooled $500K	29	51,630	-0.046	—	—	—
			(0.119)			

Notes: Bunching ratio $\hat{b}$ is the excess mass below the threshold divided by the average counterfactual bin count, estimated using a degree-7 polynomial. The pooled estimate uses organizations from all states with a $500,000 threshold. Bootstrap standard error (500 replications, resampling within state) in parentheses.

using Form 990 XML data—which disaggregates contributions from total revenue—could sharpen the threshold definition for contribution-based states.

5. Results

5.1 Main Bunching Estimates

Table 2 reports bunching estimates pooled across states at each threshold level. At the modal $500,000 threshold, the pooled bunching ratio is $\hat{b} = -0.046$ (SE = 0.122), indistinguishable from zero with a 95% confidence interval of $[-0.285, 0.193]$. The mean across individual state-level estimates is 0.004 and the median is -0.055, confirming that the typical state shows no detectable audit-avoidance bunching.

The point estimate rules out bunching responses larger than $\hat{b} = 0.19$, which places an upper bound on the behavioral distortion at approximately 19% of the average counterfactual bin density. For context, the top-performing APEP paper on housing subsidies documented bunching ratios of 0.3–0.5 at Help to Buy thresholds (APEP Autonomous Research, 2026), suggesting that nonprofits respond far less to audit notches than homebuyers respond to subsidy cutoffs.

The other threshold levels show more bunching: Illinois at $300,000 ($\hat{b} = 1.04$), the three states at $750,000 (mean $\hat{b} = 0.55$), and California at $2,000,000 ($\hat{b} = 0.52$). However, these are single-state or few-state estimates with limited power to distinguish audit-specific effects from idiosyncratic state patterns.

Density discontinuity. A simple comparison of organization counts in the $25,000 window below versus above $500,000 shows 1,858 organizations below and 1,724 above—a ratio of

Table 3: Difference-in-Bunching: Audit-Mandate vs. No-Mandate States at $500,000

	Below $500K ($475K–$500K)	Above $500K ($500K–$525K)
Audit-mandate states	0.04547	0.04166
No-mandate states	0.04515	0.04191
Difference-in-bunching	0.00057 (0.00206)	

Notes: Cells report the fraction of each group's organizations in the given revenue bin. The difference-in-bunching is $(d_{\text{below}}^{\text{mandate}} - d_{\text{below}}^{\text{no-mandate}}) - (d_{\text{above}}^{\text{mandate}} - d_{\text{above}}^{\text{no-mandate}})$. Bootstrap standard error (500 replications, resampling within state) in parentheses. Positive values indicate excess bunching attributable to audit mandates rather than round-number effects.

1.078, or a 7.8% excess. The log difference of 0.075 is positive but modest, corresponding to roughly 134 "extra" organizations in the below-threshold window across all 29 states. Per state, this amounts to fewer than 5 organizations.

5.2 Difference-in-Bunching

Table 3 presents the difference-in-bunching comparison. The density just below $500,000 is 0.04547 in mandate states versus 0.04515 in no-mandate states (difference: 0.00032). The density just above is 0.04166 in mandate states versus 0.04191 in no-mandate states (difference: −0.00025). The DiB estimate is 0.00057 (bootstrap SE = 0.00206, $t = 0.28$)—positive but statistically indistinguishable from zero, indicating that we cannot reject the hypothesis that audit mandates account for none of the density pattern around $500,000.

5.3 Heterogeneity

Across states. Individual state estimates at the $500,000 threshold range from −1.15 (New Jersey) to 1.32 (Connecticut). States with the strongest bunching—Connecticut, Mississippi ($\hat{b} = 0.92$), and Colorado ($\hat{b} = 0.68$)—tend to have active charity registration enforcement. States with negative estimates may reflect measurement issues or counterfactual misspecification in smaller samples.

Across nonprofit types. Bunching varies dramatically by NTEE major category. Science and technology organizations (code U, $\hat{b} = 1.64$, $N = 161$), employment and job-related organizations (J, $\hat{b} = 1.16$, $N = 548$), and mutual membership benefit organizations (Y, $\hat{b} = 1.04$, $N = 352$) show substantial bunching. In contrast, environmental organizations (C, $\hat{b} = -0.79$), animal-related organizations (D, $\hat{b} = -1.05$), and medical research organizations (H, $\hat{b} = -0.27$) show negative estimates. This pattern suggests that organizations with more

Table 4: Robustness: Polynomial Order and Excluded-Region Sensitivity

Panel A: Polynomial Order			Panel B: Excluded Region		
Order	Excess Mass	$\hat{b}$	Width	Excess Mass	$\hat{b}$
5	12.4	0.024	\$10,000	18.3	0.036
6	-50.7	-0.098	\$15,000	-0.7	-0.001
7	-23.6	-0.046	\$20,000	-25.8	-0.05
8	34	0.066	\$25,000	-23.6	-0.046
9	17.7	0.034	\$30,000	-37.8	-0.073
			\$40,000	-165.4	-0.318

Notes: Panel A varies the polynomial order of the counterfactual density estimator from 5 to 9. Panel B varies the width of the excluded region around the \$500,000 threshold. The baseline specification uses polynomial order 7 and a \$25,000 excluded region. All estimates use organizations from states with a \$500,000 audit threshold.

sophisticated financial management or those facing less donor pressure for audited statements are more responsive to the threshold.

5.4 Robustness

Polynomial order. Table 4, Panel A shows that the pooled \$500,000 estimate is sensitive to the polynomial order of the counterfactual density. The bunching ratio ranges from -0.098 (degree 6) to 0.066 (degree 8), with all estimates economically small and statistically insignificant. This sensitivity is itself informative: when a bunching response is genuine and large, it is typically robust to polynomial order (Kleven and Waseem, 2013). The instability here supports the interpretation of a null result.

Excluded region. Panel B of Table 4 varies the width of the excluded region. With a narrow \$10,000 exclusion, the estimate is 0.036; with a wide \$40,000 exclusion, it is -0.318. The baseline (\$25,000) yields -0.046. Again, the sign instability and small magnitudes across specifications confirm that there is no robust bunching signal at the \$500,000 threshold.

Placebo thresholds. Table 5 reports bunching estimates at round-number thresholds in states without audit mandates. The placebo bunching ratios are: 0.227 at \$300,000, 0.080 at \$400,000, 0.085 at \$600,000, and 0.188 at \$900,000. Critically, the placebo estimate at \$500,000 itself in no-mandate states is $\hat{b} = -0.019$—essentially zero—confirming that the modest density patterns observed at this level in mandate states are not driven by round-number clustering. Several placebo estimates at other round numbers exceed the absolute value of the actual \$500,000 estimate, confirming that round-number clustering is a pervasive feature of nonprofit revenue reporting at various levels. The difference-in-bunching estimator

Table 5: Placebo Bunching at Round-Number Thresholds in No-Mandate States

Placebo Threshold	Excess Mass	Bunching Ratio $\hat{b}$
$3e+05	74	0.227
$4e+05	23.2	0.08
$6e+05	21.4	0.085
$7e+05	4.8	0.02
$8e+05	-31.8	-0.142
$9e+05	40	0.188
$1e+06	-6	-0.03
$500,000 (no-mandate)	-5	-0.019
Actual $500K threshold	-23.6	-0.046
		(0.119)

Notes: Placebo bunching estimated at round-number thresholds using only organizations in states without charitable audit mandates (AK, AZ, DE, FL, IA, ID, IN, LA, MT, ND, NE, NV, SD, TX, VT, WA, WY). The actual $500,000 estimate (last row) uses organizations from audit-mandate states. Bootstrap standard error in parentheses.

benchmarks audit-threshold bunching against this background.

6. Discussion and Conclusion

The central finding of this paper is that we cannot reject the null hypothesis of zero bunching at state charitable audit thresholds. Despite imposing compliance costs of 2–20% of marginal revenue, these thresholds produce, on average, no detectable distortion in the cross-sectional distribution of nonprofit revenue. The 95% confidence interval rules out bunching ratios larger than 0.19, placing a meaningful upper bound on the behavioral response. Two caveats temper this conclusion: (i) the EO BMF's total revenue variable is a noisy proxy for the contributions-based thresholds used in some states, which would attenuate true bunching toward zero; and (ii) the cross-sectional design cannot detect dynamic responses such as organizations slowing fundraising growth as they approach the threshold.

Three explanations are consistent with this null. First, *optimization frictions*: most nonprofits lack the financial sophistication to precisely control reported revenue near a threshold. Unlike wage earners who can adjust hours or self-employed individuals who can shift income across years, nonprofit revenue depends on donor behavior, grant timing, and programmatic outcomes that are largely outside the organization's control (Chetty et al., 2011). Second, *offsetting incentives*: audited financials signal credibility to institutional donors and grantmakers, so the audit mandate may confer reputational benefits that offset the direct costs (Krishnan et al., 2006; Garven et al., 2018). Third, *private ordering*: many foundations and government grantmakers require audited financials regardless of state mandates, rendering the statutory threshold inframarginal for a large share of organizations.

The heterogeneity across states and sectors is informative about which of these explanations dominates. The finding that science and employment organizations show bunching while environmental and animal-welfare organizations do not is consistent with the optimization friction story: NTEE categories with more sophisticated financial management (and perhaps more revenue controllability) show greater responsiveness. The state-level heterogeneity—with Connecticut and Mississippi showing strong bunching but New Jersey showing negative estimates—may reflect differences in enforcement intensity, organizational composition, or counterfactual misspecification in smaller samples. Negative bunching (excess mass *above* the threshold) is theoretically possible if audited financials serve as a quality signal that attracts donations, but more likely reflects sampling variation in states with few organizations near the threshold.

For policy, these results are reassuring. States can raise or lower audit thresholds without expecting large behavioral responses from the nonprofit sector as a whole. The compliance cliff exists in the statute books but does not, on average, distort organizational growth. Policymakers concerned about audit costs should focus on the direct burden rather than worrying about threshold-avoidance behavior. However, the state-level heterogeneity cautions against blanket conclusions: in states with active enforcement and threshold-aware organizations, bunching may be a real phenomenon deserving targeted attention.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: https://github.com/SocialCatalystLab/ape-papers

Contributors: @olafdrw

First Contributor: https://github.com/olafdrw

References

American Institute of CPAs, "Audit Fees Survey," Technical Report, AICPA 2020.

APEP Autonomous Research, "The Price of Subsidy Limits: Multi-Cutoff Evidence from Help to Buy," *APEP Working Paper*, 2026, (0492).

Botero, Juan C., Simeon Djankov, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The Regulation of Labor," *Quarterly Journal of Economics*, 2004, *119* (4), 1339–1382.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics*, 2011, *126* (2), 749–804.

Crimm, Nina J., "The Global Effects of Fund-Raising Regulation on Philanthropy," *Tulane Law Review*, 2008, *82*, 1491.

Djankov, Simeon, Rafael La Porta, Florencio Lopez de Silanes, and Andrei Shleifer, "The Regulation of Entry," *Quarterly Journal of Economics*, 2002, *117* (1), 1–37.

Fishman, James J. and Stephen Schwarz, *Nonprofit Organizations: Cases and Materials*, 5th ed., Foundation Press, 2015.

Garven, Sarah A., Amanda W. Beck, and Linda M. Parsons, "Are Audit-Related Factors Associated with Financial Reporting Quality in Nonprofit Organizations?," *Auditing: A Journal of Practice & Theory*, 2018, *37* (1), 49–68.

Internal Revenue Service, "Instructions for Form 990," Technical Report, Department of the Treasury 2024.

Irvin, Renee A., "State Regulation of Nonprofit Organizations: Accountability Regardless of Outcome," *Nonprofit and Voluntary Sector Quarterly*, 2005, *34* (2), 161–178.

Kleven, Henrik Jacobsen, "Bunching," *Annual Review of Economics*, 2016, *8*, 435–464.

_ **and Mazhar Waseem**, "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics*, 2013, *128* (2), 669–723.

Krishnan, Ranjani, Michelle H. Yetman, and Robert J. Yetman, "Expense Misreporting in Nonprofit Organizations," *The Accounting Review*, 2006, *81* (2), 399–420.

National Association of State Charity Officials, "State Charity Registration and Filing Requirements," Technical Report, NASCO 2020.

Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2010, *2* (3), 180–212.

Yildirim, Huseyin et al., "Charitable Giving and Audit Thresholds: Evidence from New York," *Working Paper*, 2018.

Table 6: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Bunching ratio ($\hat{b}$)	-23.6	61.4	234.6	-0.1	0.262	Moderate negative
Density discontinuity	134	—	234.6	0.571	—	Large positive
Diff.-in-bunching	0.00057	—	—	0.278	—	Large positive

Notes: **Country:** United States. **Research question:** Do state-mandated charitable audit thresholds distort nonprofit revenue reporting, and how large is the behavioral response to compliance cost discontinuities? **Policy mechanism:** Thirty-four US states require charitable nonprofits to submit independently audited financial statements when annual revenue exceeds a state-specific threshold (ranging from \$300,000 to \$2,000,000), creating a discontinuous jump in compliance costs of \$10,000–\$100,000 at the threshold. **Outcome definition:** Bunching ratio measures excess mass of organizations reporting revenue just below the threshold relative to counterfactual polynomial density; density discontinuity measures the ratio of bin counts in the \$25,000 window below vs. above the threshold. **Treatment:** Binary – organizations face audit mandate if revenue exceeds state-specific threshold (primary analysis at \$500,000, the modal threshold). **Data:** IRS Exempt Organizations Business Master File (2024 extract), all 501(c)(3) organizations with positive revenue across 50 states plus DC; approximately 555,000 organizations in audit-mandate states. **Method:** Polynomial bunching estimation following Kleven and Waseem (2013) with degree-7 counterfactual density, bootstrap standard errors (500 replications resampled within state), and difference-in-bunching comparing mandate vs. no-mandate states. **Sample:** Restricted to organizations with revenue within ±50% of their state's audit threshold; 34 states with audit mandates provide independent replications. SDE = $\hat{\beta}/\mathrm{SD}(Y)$ where SD(Y) is the cross-bin standard deviation of organization counts. Classification refers to magnitude, not statistical significance: Large (|SDE| > 0.15), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).

A. Standardized Effect Sizes