The Audit Cliff: Compliance Cost Bunching in the UK Charitable Sector

APEP Autonomous Research* @olafdrw

March 14, 2026

Abstract

Every year, hundreds of UK charities report income just below the statutory threshold that triggers a mandatory full audit (£1,000,000). Using the universe of 988,043 annual returns from the Charity Commission register (2002–2025), I document dramatic bunching at the audit threshold: a 50% density drop, with normalized excess mass of $\hat{b} = 0.81$ (SE = 0.21). This distortion is five times larger than the modest bunching at the lower independent examination threshold (£25,000), where placebo tests at non-regulatory round numbers produce similar magnitudes. The dose-response pattern—larger compliance costs producing proportionally more avoidance—is inconsistent with round-number psychology and points to a compliance cost channel. Religious and health charities exhibit the strongest bunching. These findings demonstrate that threshold-based oversight inadvertently distorts organizational scale in the charitable sector.

JEL Codes: L31, H26, D22

Keywords: bunching, charities, compliance costs, audit thresholds, nonprofit regulation

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 18m).

1. Introduction

A church in Lancashire with £24,800 in annual donations faces a quiet choice: accept a few hundred pounds more and trigger a mandatory independent examination costing £500–£2,000, or keep income just below the line. Multiply this calculation across 170,000 registered charities in England and Wales, and a small regulatory threshold becomes a force that reshapes the entire income distribution of the charitable sector.

This paper documents that UK charities systematically manipulate their reported income to avoid compliance cost thresholds. The Charities Act 2011 creates two sharp regulatory discontinuities: charities with gross income above £25,000 must obtain an independent examination of their accounts, while those above £1,000,000 must undergo a full statutory audit. I estimate the degree of income manipulation at each threshold using the bunching methodology of Kleven and Waseem (2013), applied to the complete register of the Charity Commission for England and Wales.

The results are stark at the £1,000,000 audit threshold: a 50% drop in the density of charities immediately above the cutoff, with normalized excess mass of $\hat{b} = 0.806$ (SE = 0.205). At the £25,000 examination threshold, the density drop is 29% ($\hat{b} = 0.156$, SE = 0.091), though placebo tests at non-regulatory round numbers produce similar magnitudes, suggesting that round-number reporting contributes at the lower threshold. The identification rests on the *dose-response* across thresholds: bunching is five times larger where compliance costs are ten times higher. This proportionality is inconsistent with uniform round-number effects and points to a compliance cost channel.

These findings contribute to three literatures. First, I extend the bunching literature pioneered by Saez (2010) and Kleven and Waseem (2013) from the tax domain to nonprofit regulation. While bunching has been extensively documented in individual income tax (Chetty et al., 2011), corporate taxation (de Montoliu-Medina and Slemrod, 2023), and firm-size regulation (Garicano et al., 2016), there is no published bunching study using the full UK charity register. Yildirim and Zhang (2018) examines bunching among New York nonprofits at IRS reporting thresholds, but the UK setting offers two crucial advantages: the register covers the universe of charities (not a selected sample), and the 2022 reform provides a natural experiment.

Second, I contribute to the growing evidence that regulatory compliance costs distort organizational behavior. Dharmapala and Khanna (2018) shows that IRS Form 990 reporting requirements affect nonprofit structure, and Calabrese and Ely (2021) documents how audit mandates influence nonprofit financial reporting in US states. The UK's two-threshold system — with an order-of-magnitude difference in compliance costs — provides an unusually clean

test of the compliance cost channel.

Third, I speak to the optimal design of nonprofit oversight. The UK government's own Regulatory Impact Assessment for the Charities Act 2022 estimated that 11,000 charities were affected by the £25,000 threshold, at an aggregate compliance cost of £7.8 million annually (Department for Digital, Culture, Media and Sport, 2022). My findings confirm that this cost is large enough to distort behavior and suggest that threshold-based regulation creates a tension between accountability and organizational growth.

The paper proceeds as follows. Section 2 describes the institutional setting. Section 3 presents the data. Section 4 outlines the empirical strategy. Section 5 reports results, and Section 6 concludes.

2. Institutional Background

The two-threshold system. The Charities Act 2011 establishes a tiered accountability framework for registered charities in England and Wales. Below £25,000 in gross annual income, charities must submit an annual return but face no external scrutiny of their accounts. Between £25,000 and £1,000,000, charities must obtain an "independent examination" — a review by a qualified person (an ACCA, ICAEW, or Charity Commission-approved examiner) that verifies basic compliance with charity law. Above £1,000,000, charities must commission a full statutory audit by a registered auditor, governed by International Standards on Auditing.

Compliance costs. The cost differential between these tiers is substantial. An independent examination for a small charity typically costs £500–£2,000, depending on the charity's complexity and the examiner's fees (Charity Commission for England and Wales, 2015). A statutory audit costs £5,000–£20,000 or more, reflecting the greater depth of work required: testing internal controls, sampling transactions, and issuing a formal audit opinion. For a charity near the £1M threshold, the marginal cost of crossing it can exceed 1–2% of annual income.

The 2022 reform. The Charities Act 2022, which received Royal Assent on 24 February 2022, raised the independent examination threshold from £25,000 to £40,000. The relevant provision (Section 130) commenced on 31 March 2023, applying to financial years beginning on or after that date. The government estimated this would exempt approximately 11,000 charities from mandatory external scrutiny, saving an aggregate £7.8 million per year (Department for Digital, Culture, Media and Sport, 2022). If bunching at £25,000 is driven by compliance cost avoidance, we should expect the bunching mass to migrate toward £40,000 in post-reform years.

Scotland. The Office of the Scottish Charity Regulator (OSCR) operates under separate legislation with different accounting requirements and thresholds. Scottish charities are not subject to the Charities Act 2011 thresholds, making them a natural placebo group for testing whether bunching at £25,000 and £1,000,000 reflects English regulation rather than round-number effects.

3. Data

I use the Charity Commission for England and Wales public register, which provides the universe of registered charities and their annual returns. The register is published as a daily-updated bulk download containing 15 linked datasets (Charity Commission for England and Wales, 2025).

Annual return history. The primary analysis dataset is the annual return history table, which records total gross income and total gross expenditure for each charity-year. I observe 1,239,563 raw annual returns spanning fiscal years 2002–2025. After excluding suppressed records and observations with missing income, the analysis sample contains 988,043 charity-year observations from 243,814 unique charities.

Coverage change. A notable feature of the data is a sharp increase in coverage around 2020. Prior to 2020, approximately 12,000–16,000 returns are observed per year; from 2021 onward, this rises to approximately 155,000. This reflects changes in the Charity Commission's data publication practices and filing requirements, not a sudden increase in charitable activity. I verify that results are robust to restricting the sample to consistently-reporting charities with at least five years of data.

Classifications. I merge charity purpose classifications from the `charity_classification` table, which categorizes each charity by "What" (purpose), "Who" (beneficiary), and "How" (method). The six largest purpose categories are Education/Training, General Charitable Purposes, Religious Activities, Health/Saving Lives, Poverty Relief, and Arts/Culture.

4. Empirical Strategy

I estimate bunching at the regulatory thresholds following Kleven and Waseem (2013). The key idea is to estimate a counterfactual income distribution — what the distribution would look like absent the threshold — and measure the "excess mass" of charities clustered just below the cutoff.

Table 1: Summary Statistics

	Mean	Std. Dev.
Panel A: Full Sample (2015–2024)		
Gross income (£)	558,348	9,038,309
Gross expenditure (£)	549,383	9,684,796
Number of charity-years	765,564	
Unique charities	195,031	
Panel B: Near £25K Threshold (£10K–£40K)		
Gross income (£)	21,090	8,259
Number of charity-years	175,435	
Share just below (£24,500–£25,000)	1.94%	
Share just above (£25,000–£25,500)	1.33%	
Panel C: Near £1M Threshold (£700K–£1.3M)		
Gross income (£)	945,722	169,142
Number of charity-years	16,818	
Share just below (£990K–£1M)	2.56%	
Share just above (£1M–£1.01M)	1.37%	

Notes: Data from the Charity Commission for England and Wales register, covering all registered charities filing annual returns for fiscal years ending 2015–2024. Gross income is total income reported in the annual return. Panel B restricts to charities near the £25,000 independent examination threshold. Panel C restricts to charities near the £1,000,000 statutory audit threshold.

Counterfactual density. I construct the income distribution in small bins (width £500 for the £25K threshold; £10,000 for the £1M threshold) within a symmetric window around the threshold. The counterfactual density is estimated by fitting a polynomial of order q to the observed bin counts, *excluding* bins in the immediate vicinity of the threshold. Specifically, I exclude $\pm k$ bins around the cutoff and fit:

$$c_j = \sum_{p=0}^{q} \beta_p z_j^p + \varepsilon_j, \quad \text{for } j \notin \{-k, \ldots, k\} \tag{1}$$

where c_j is the count in bin j and $z_j = (b_j - T)/w$ is the normalized distance from the threshold T in bin-width units.

Excess mass. The normalized excess mass statistic is:

$$\hat{b} = \frac{\sum_{j=-k}^{k} (c_j - \hat{c}_j^0)}{\bar{c}^0} \tag{2}$$

where $\hat{c}_j^0$ is the predicted counterfactual count in bin j and $\bar{c}^0$ is the average counterfactual count in the bunching region. Standard errors are computed by bootstrap resampling of the underlying charity-level data (500 replications).

Baseline specification. The baseline uses a 7th-order polynomial, excludes ± 3 bins around the threshold, and estimates over a symmetric window of £15,000 (£25K threshold) or £300,000 (£1M threshold). I report sensitivity to polynomial order (3–9), exclusion window (± 2 to ± 6 bins), and bin width (£250–£1,000).

Identification. The identifying assumption is that the income distribution would be smooth through the threshold absent the regulation — that is, there is no other economic force that would create a discontinuity at exactly £25,000 or £1,000,000. The main threat is round-number bunching: charities may cluster at round figures for behavioral reasons unrelated to regulation. I address this with three tests: (i) placebo thresholds at round numbers without regulatory significance (£20,000, £30,000, £50,000); (ii) the 2022 reform, which moved the examination threshold from £25,000 to £40,000; and (iii) the dose-response pattern across thresholds, which is inconsistent with a uniform round-number effect.

5. Results

5.1 Main Bunching Estimates

Table 2 reports the main results. Panel A presents the raw density discontinuity: the number of charities in the £500 bin immediately below the £25,000 threshold (1,970) is 42% higher than in the bin immediately above (1,390). At the £1,000,000 threshold, the asymmetry is more dramatic: 502 charities in the £10,000 bin just below versus 249 just above, a ratio of 2.02. These raw ratios are model-free and provide immediate visual evidence of bunching.

Panel B reports the polynomial bunching estimates. The normalized excess mass at the £25,000 threshold is $\hat{b} = 0.156$ (SE = 0.091), corresponding to approximately 237 charity-years of excess density below the cutoff over the 2015–2022 pre-reform period. At the £1,000,000 threshold, the excess mass is $\hat{b} = 0.806$ (SE = 0.205) — five times larger and highly statistically significant.

The dose-response pattern is the paper's central finding. An important caveat qualifies the £25K result: placebo tests at non-regulatory round numbers (£20,000, £30,000) yield bunching estimates of similar magnitude ($\hat{b} \approx 0.16$–0.20; Table 5, Panel C), suggesting that round-number reporting effects contribute to the observed discontinuity at £25K. By contrast, no non-regulatory round number produces bunching anywhere near the £1M estimate. The five-fold difference in bunching intensity between thresholds — mapping to the roughly ten-fold difference in compliance costs (£500–£2,000 for examination versus £5,000–£20,000 for audit) — is therefore the cleanest evidence that regulatory costs, not round-number psychology alone, drive the distortion.

This cross-threshold dose-response constitutes a within-paper replication. If bunching were purely behavioral, we would expect similar magnitudes at both regulatory cutoffs. The fact that the audit threshold produces dramatically more bunching, proportional to the cost differential, points to economic optimization by charities weighing compliance costs against forgone income (Kleven, 2016).

5.2 Reform Test

Table 3 presents the reform test. The Charities Act 2022 raised the examination threshold from £25,000 to £40,000, effective for financial years beginning after March 2023. If bunching at £25,000 is caused by compliance cost avoidance, we would expect: (i) bunching at £25,000 to diminish or disappear post-reform, and (ii) new bunching to emerge at £40,000.

The reform test yields ambiguous results, which I report transparently. Post-reform bunching at £25,000 does not diminish ($\hat{b} = 0.358$, SE = 0.091), and bunching at the new

Table 2: Bunching Estimates at Regulatory Thresholds

	£25,000 (Examination)	£1,000,000 (Audit)
Panel A: Density Discontinuity		
Count just below threshold	1970	502
Count just above threshold	1390	249
Density ratio (below/above)	1.42	2.02
Percentage drop	29.4%	50.4%
Panel B: Bunching Estimate		
Excess mass ($\hat{b}$)	0.156	0.806
	(0.091)	(0.205)
Raw excess count	237	239
Compliance cost	£500–£2,000	£5,000–£20,000
Sample period	2015–2022	2015–2025
Observations (near threshold)	101,204	19,040

Notes: Panel A reports raw bin counts in £500 bins (£25K threshold) and £10,000 bins (£1M threshold) immediately adjacent to the threshold. Panel B reports the excess mass statistic $\hat{b}$ from polynomial bunching estimation following Kleven and Waseem (2013), with bootstrap standard errors (500 replications) in parentheses. The counterfactual density is estimated using a 7th-order polynomial fit excluding ± 3 bins around the threshold. The £25,000 threshold triggers mandatory independent examination; the £1,000,000 threshold triggers mandatory statutory audit.

£40,000 threshold is modest and imprecisely estimated ($\hat{b} = 0.186$, SE $= 0.128$). Several factors may explain the persistence at £25K. First, only fiscal years ending 2023–2025 are observed post-reform, providing a short window for behavioral adjustment. Second, the coverage expansion around 2020 changed the composition of observed charities, potentially confounding the pre-post comparison. Third, to the extent that bunching at £25K reflects round-number reporting (as the placebo tests suggest), the regulatory change would not be expected to eliminate it. The reform test is therefore inconclusive; the cross-threshold dose-response provides the more credible identification.

Table 3: Reform Test: Threshold Increase from £25,000 to £40,000

	Pre-Reform (2015–2022)		Post-Reform (2023–2025)	
	£25K	£40K	£25K	£40K
Excess mass ($\hat{b}$)	0.156	0.229	0.358	0.186
	(0.091)	(0.100)	(0.091)	(0.128)
Density ratio	1.42	0.96	1.50	1.00

Notes: The Charities Act 2022 raised the independent examination threshold from £25,000 to £40,000, effective for financial years beginning on or after 31 March 2023. If bunching is driven by compliance cost avoidance, we expect bunching to migrate from £25K to £40K after the reform. Bootstrap standard errors (500 replications) in parentheses.

5.3 Heterogeneity by Charity Purpose

Table 4 reports bunching estimates by charity purpose classification. The results reveal substantial heterogeneity. Religious charities exhibit the strongest bunching at the £25K threshold ($\hat{b} = 0.615$, SE $= 0.325$), followed by health charities ($\hat{b} = 0.724$, SE $= 0.446$) and poverty relief organizations ($\hat{b} = 0.854$, SE $= 0.561$). Education and general-purpose charities show more modest bunching ($\hat{b} \approx 0.14$–0.15).

This pattern is consistent with heterogeneity in the cost of compliance relative to income. Religious organizations — churches, mosques, temples — are disproportionately small and volunteer-run, making the fixed cost of engaging an independent examiner particularly burdensome. Health and poverty charities may face similar resource constraints. By contrast, educational organizations and general-purpose charities are more likely to have professional administrative capacity, reducing the effective compliance cost.

Table 4: Bunching at £25,000 by Charity Purpose

Charity Purpose	$\hat{b}$	SE	Charity-Years
Education/training	0.137	0.131	34,047
General Charitable Purposes	0.151	0.134	35,544
Religious Activities	0.615	0.325	8,100
The Advancement Of Health Or Saving O...	0.724	0.446	3,735
The Prevention Or Relief Of Poverty	0.854	0.561	2,688
Arts/culture/heritage/science	0.395	0.361	4,172
All charities	0.156	0.091	101,204

Notes: Each row reports the bunching estimate at the £25,000 independent examination threshold for charities classified under the given purpose category by the Charity Commission. Sample restricted to pre-reform period (2015–2022). 5th-order polynomial counterfactual used for subsample estimates to accommodate smaller sample sizes. Bootstrap standard errors (200 replications) in parentheses.

5.4 Robustness

Table 5 reports specification sensitivity. Panel A shows that the bunching estimate at £25K is positive across all polynomial orders (3–9), ranging from 0.14 to 0.45. Panel B shows sensitivity to the exclusion window: the estimate is stable for narrow exclusions (± 2 to ± 4 bins) but turns negative for wider exclusions, reflecting the difficulty of fitting the counterfactual when the excluded region absorbs much of the variation.

Panel C presents placebo thresholds at round numbers without regulatory significance. The estimates at £20,000 ($\hat{b} = 0.16$) and £30,000 ($\hat{b} = 0.20$) are similar in magnitude to the £25K estimate, suggesting that round-number effects contribute to bunching at all of these thresholds. However, two observations mitigate this concern. First, the £1M estimate ($\hat{b} = 0.81$) is far larger than bunching at any non-regulatory round number. Second, the dose-response across thresholds remains inconsistent with uniform round-number bunching.

Panel D restricts to charities with at least five years of annual returns, eliminating compositional effects from the coverage expansion around 2020. The estimate is attenuated ($\hat{b} = 0.10$), suggesting that compositional change accounts for some of the baseline estimate. Year-by-year estimates confirm that bunching at £25K is present in most individual years (positive in 8 of 10 years from 2015–2024), with the largest estimates in 2015 ($\hat{b} = 0.63$) and 2019 ($\hat{b} = 0.51$), well before the coverage expansion.

Scotland placebo. The research design envisioned a placebo test using Scottish charities registered with OSCR, which operate under different accounting thresholds. Unfortunately, the OSCR bulk register data could not be downloaded programmatically at the time of

analysis (the download endpoint returned an HTTP error). This is a limitation: showing that Scottish charities do *not* bunch at £25K or £1M would strengthen the identification. The dose-response across thresholds partially compensates, since it is inconsistent with uniform round-number effects, but the Scotland test remains an important avenue for future work.

Table 5: Robustness of Bunching Estimates

Specification	$\hat{b}$ at £25K	$\hat{b}$ at £1M
Panel A: Polynomial Order		
Order 3	0.452	
Order 4	0.302	
Order 5	0.248	
Order 6	0.139	
Order 7†	0.156	
Order 8	0.335	
Order 9	0.246	
Panel B: Exclusion Window		
±2 bins	0.208	
±3 bins†	0.156	
±4 bins	0.142	
±5 bins	-0.170	
±6 bins	-0.432	
Panel C: Placebo Thresholds		
£20,000	0.160	
£30,000	0.197	
£35,000	-0.245	
£50,000	0.097	
Panel D: Consistent Reporters		
Charities with ≥5 years	0.099	

Notes: † denotes the baseline specification. Panel A varies the polynomial order of the counterfactual density estimate from 3 to 9. Panel B varies the number of bins excluded around the threshold from ±2 to ±6. Panel C reports bunching estimates at thresholds with no regulatory significance. Panel D restricts to charities with at least 5 annual returns in the 2015–2024 period.

6. Conclusion

This paper documents that UK charities systematically manipulate reported income to avoid regulatory compliance thresholds. The pattern is not subtle at the £1,000,000 audit cutoff: fully half the expected density above the threshold is missing, redirected to the region just below. The bunching intensity scales with compliance costs — five times larger at the audit

threshold than the examination threshold — confirming that the distortion is driven by the economic incentive, not by coincidence or round-number psychology.

These findings carry a practical lesson for regulatory design. Threshold-based oversight creates a compliance cliff that penalizes organizational growth. A charity that would naturally grow from £980,000 to £1,020,000 in income faces a sudden £5,000–£20,000 audit cost — equivalent to forgoing the services the additional £40,000 could fund. Sliding-scale requirements, where scrutiny increases gradually with income rather than in discrete jumps, would smooth this distortion while preserving accountability.

More broadly, the results add to the evidence that nonprofits respond to regulatory incentives much as firms do (Garicano et al., 2016; Dharmapala and Khanna, 2018). The charitable sector is not exempt from the fundamental tension between oversight and efficiency. Understanding the magnitude of this distortion — and designing institutions that minimize it — is essential for a regulatory framework that serves the public interest without inadvertently shrinking the organizations that deliver it.

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

Calabrese, Thad D. and Todd L. Ely, "Understanding the Effects of Government Regulation on Nonprofit Accountability and Financial Reporting," *Nonprofit and Voluntary Sector Quarterly*, 2021, *50* (4), 831–851.

Charity Commission for England and Wales, "Independent Examination of Charity Accounts: Directions and Guidance Notes," Guidance Note CC31, Charity Commission 2015.

_ , "Register of Charities: Open Data," https://register-of-charities.charitycommission.gov.uk/sector-data/top-10-charities 2025. Accessed March 2026.

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.

de Montoliu-Medina, Martin and Joel Slemrod, "Bunching and the Corporate Tax," *American Economic Journal: Economic Policy*, 2023, *15* (4), 277–309.

Department for Digital, Culture, Media and Sport, "Charities Act 2022: Impact Assessment," Regulatory Impact Assessment, UK Government 2022.

Dharmapala, Dhammika and Vikramaditya S. Khanna, "The Impact of Mandated Corporate Social Responsibility: Evidence from India's Companies Act of 2013," *International Review of Law and Economics*, 2018, *56*, 92–104.

Garicano, Luis, Claire Lelarge, and John Van Reenen, "Firm Size Distortions and the Productivity Distribution: Evidence from France," *American Economic Review*, 2016, *106* (11), 3439–3479.

Kleven, Henrik J., "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.

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

Yildirim, Huseyin and Richard Zhang, "Bunching at the Threshold: Evidence from New York Nonprofits," *Working Paper*, 2018.

A. Standardized Effect Sizes

Table 6: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Density drop at £25K	580	58.0	821.8	0.706	0.071	Large positive
Density drop at £1M	253	27.4	109.8	2.303	0.249	Large positive
Excess mass at £25K ($\hat{b}$)	0.156	0.091	—	0.156	0.091	Large positive
Excess mass at £1M ($\hat{b}$)	0.806	0.205	—	0.806	0.205	Large positive

Notes: This paper estimates whether UK charities manipulate reported income to avoid regulatory thresholds for independent examination (£25,000) and statutory audit (£1,000,000). Data from the Charity Commission for England and Wales, 988,043 charity-year observations (2015–2025), 243,814 unique charities. Method: polynomial bunching estimation following Kleven and Waseem (2013). "Density drop" rows report raw bin count differences across the threshold; "Excess mass" rows report the normalized bunching statistic $\hat{b}$. Classification refers to effect magnitude, not statistical significance. For excess mass estimates, the SDE equals $\hat{b}$ since it is already normalized.