How Do Nonprofits Respond to Regulatory Thresholds: Evidence From New York's Audit Requirements

Travis St. Clair

Abstract

Nonprofits in the United States must comply with various state and federal regulations to maintain their tax-exempt status. Despite persistent calls to increase accountability in the nonprofit sector, there is little research examining the burden imposed by existing regulatory requirements, especially at the state level. This paper uses a bunching design to estimate the avoidance behavior exhibited by tax-exempt charities in response to New York State's audit requirements. There is clear evidence of bunching in response to the requirement that nonprofits above certain revenue thresholds file financial statements reviewed by or audited by an independent certified public accountant. Measuring the extent of bunching around the revenue notches yields estimates of the average revenue that nonprofits either forego or fail to report in avoidance of the requirements. Results from dynamic estimation show that charities near the threshold for a review engagement report approximately $1,300 less revenue than otherwise predicted by a counterfactual; charities near the threshold for a full audit report approximately $1,400 less. The results have implications for the optimal design of state-level financial regulations.

INTRODUCTION

In exchange for tax-exempt status, nonprofits in the United States are subject to numerous federal and state regulations. In recent years, as the nonprofit sector has expanded, and as financial regulations in the for-profit sector have grown more stringent in the wake of Sarbanes-Oxley and the Enron scandal, demands for financial transparency on the part of nonprofit organizations have increased (Calabrese, 2011). These requirements, while raising the level of accountability required of the sector as a whole, come with compliance costs that may pose an especially large burden to smaller organizations with little back-office infrastructure or expertise.

The regulatory requirements that tax-exempt nonprofits face include the following: the IRS Form 990, state and federal initial registration forms, state registration renewal forms and annual reports, processing fees, and additional registration and reporting requirements related to fund-raising and solicitation activities (charitable solicitation registration). Requirements differ considerably by state, such that many have bemoaned the patchwork of regulations as inefficient (Irvin, 2005).

Despite this robust regulatory framework, there is little research examining the burden that regulations impose on nonprofits. The few articles that exist focus primarily on the Form 990, the annual federal information return, in part because of the abundance of data it affords (Keating & Frumkin, 2003; Marx, 2015). This

Journal of Policy Analysis and Management, Vol. 35, No. 4, 772–790 (2016)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
Supporting Information is available in the online issue at wileyonlinelibrary.com.
DOI:10.1002/pam.21931

paper focuses instead on a state-level rule: the requirement that nonprofits produce financial statements reviewed or audited by an independent certified public accountant (CPA). The federal government does not require nonprofits to produce audited financial statements; only nonprofits that expend federal funds in excess of $750,000 face requirements under the Single Audit Act (National Council of Nonprofits, 2013).[1] Organizations earning more than $200,000 in gross receipts must file the Form 990, while organizations earning less than $200,000 (more than 55 percent of public charities in 2012)[2] need file only the Form 990 EZ or Form 990-N, simpler forms requiring much less detailed information. Hence, for smaller organizations, which make up the vast majority of tax-exempt nonprofits, the state requirement to file audited financial statements can impose significant costs and is likely to be the most onerous regulation that they face.

To gain insight into the burden imposed by financial regulations, this paper uses the National Center for Charitable Statistics' (NCCS) Core Financial Files for public charities to estimate the effect of New York State's audit requirements on the reported revenues of public charities near the audit thresholds. Prior to 2014 New York required that all charitable organizations soliciting contributions and reporting annual revenue over $250,000 file an audited financial statement prepared by an independent CPA.[3] Organizations reporting less than $250,000 in total revenues but more than $100,000 were required to file statements reviewed by an independent CPA. There were no audit requirements for organizations reporting less than $100,000 (New York State Office of the Attorney General, 2012).[4]

New York's requirements are of interest for two reasons. During the period of study, they were the strictest in the country. Only 26 states require nonprofits to file audited financial statements, and none have revenue thresholds for review engagements as low as $100,000 (National Council of Nonprofits, 2013).[5] The strictness of these requirements makes it easier to isolate the impact of the treatment of interest; most larger nonprofits are likely to produce audited financial statements for other reasons, such as grant funding requirements, thereby frustrating attempts to identify the causal effects of audit requirements alone. The existence of two different thresholds also enables a comparison of the two types of responses and sheds light on the question of which imposes more costs: a full audit or the more basic requirement to produce financial statements.

In order to estimate the effect of audit requirements on reported revenues, this paper uses a bunching design. Bunching designs have been used frequently in the public economics literature in order to analyze kinks or notches along some policy-relevant parameter. Saez (2010) uses a bunching design to examine kink points in the U.S. income tax schedule and estimate the compensated elasticity of reported

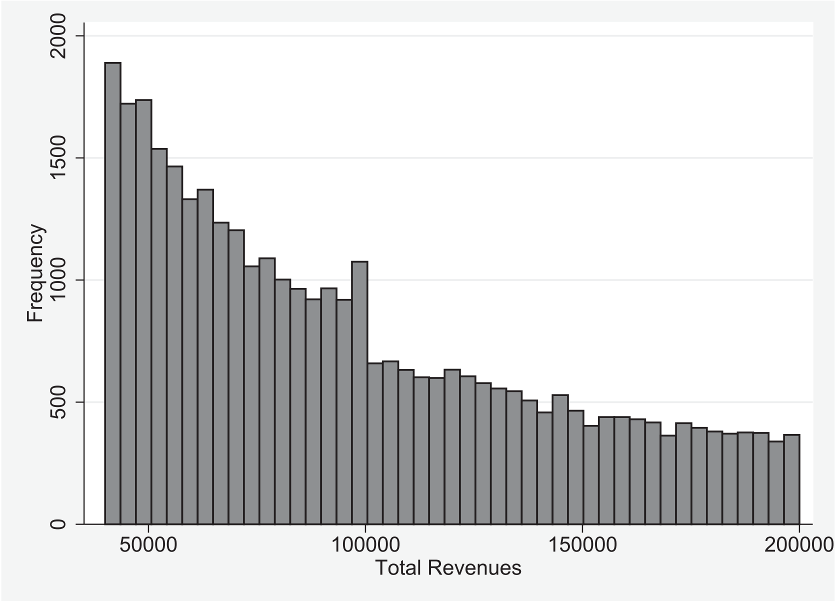
[1] The threshold was $500,000 prior to 2015. Some private foundations may also require grant recipients to conduct audits.

[2] This figure does not include charities that filed the Form 990-N, required of organizations whose gross receipts are ordinarily $50,000 or less, and so represents a lower bound.

[3] Although the statute governing nonprofit registration in New York (N.Y. U.C.C. Law 7A § 172) uses the term "gross revenues and support," in its instructions the state's Office of the Attorney General clarifies that the relevant metric is total revenues as listed on the IRS Form 990 (New York State Office of the Attorney General, 2012). The Form 990 reporting thresholds are based on gross receipts, which include money received from all sources without subtracting expenses. This differs from total revenues since nonprofits are allowed to deduct certain expenses in their calculation of total revenues (Internal Revenue Service, 2014).

[4] Organizations employing professional fundraisers also had to file, regardless of gross revenues. The state raised its filing thresholds to $250,000 and $500,000 effective July 1, 2014 (New York State Office of the Attorney General, 2012).

[5] Louisiana requires a compilation if annual revenue is at least $50,000 but only if the organization receives state funds (National Council of Nonprofits, 2013).



Note: The figure depicts bunching at the $100,000 threshold due to the state's requirement that charitable organizations with $100,000 or more in total revenues file financial statements reviewed by an independent CPA.

Figure 1. Number of Nonprofits in New York by Total Revenues, 2008 to 2012.

income with respect to the marginal tax rate. Onji (2009) documents the clustering of Japanese corporations just below that country's value-added tax threshold. Ramnath (2013) examines bunching of income in response to the saver's tax credit. Since the amount of bunching is proportional to the compliance costs associated with a regulation, and analysis of the frequency distribution can yield estimates of the average income that organizations forego or fail to report as a result of the requirement, bunching designs represent a particularly useful tool for analyzing the burden imposed by regulations.

Figure 1 illustrates the intuition of the design. The frequency distribution of nonprofits in New York is generally smooth by total revenues, with the exception of bunching just below the $100,000 threshold. The typical bunching design involves the construction of a counterfactual distribution that is smooth over the revenue range, thereby allowing for estimation of the extent of bunching around the threshold. To complement that analysis, this paper also takes advantage of the fact that a natural comparison group is available—nonprofits from all other states. Comparing the frequency distribution of nonprofits in New York to a weighted distribution of nonprofits from other states permits an additional test of the extent of bunching and a further check on the identification assumptions in the standard analysis.

This paper builds in part on Marx (2015), who estimates the bunching of public charities in response to the 990 reporting requirements. Unlike the Form 990, however, which has two assignment variables—gross receipts and assets—New York's audit requirements are based only on total revenues. Moreover, whereas the Form 990 is a reporting requirement, audits and review engagements require engaging the services of an independent CPA (though organizations may use a CPA for 990 preparation as well) and thus are likely to impose greater compliance costs. Finally, while there has been some research looking at federal requirements, there is a dearth of research examining the impact of *state*-level regulations, especially those that may affect smaller organizations, making this paper somewhat unique in the nonprofit finance literature.

This paper proceeds as follows. The next two sections summarize the literature on nonprofit regulations and audit requirements and present descriptive findings based on the density distribution of nonprofits. The following section outlines the bunching methodology and discusses the difference between static and dynamic estimates. The final sections present the results and discuss implications.

NONPROFIT REGULATIONS AND AUDIT REQUIREMENTS

The benefits to an organization of incorporating as a 501(c)3 are significant. The federal benefits include exemption from federal income tax and the ability to collect tax-deductible contributions. While state laws vary, they typically offer qualifying organizations exemption from sales, property, and income taxes (Cordes & Weisbrod, 1998). In exchange for these benefits, tax-exempt organizations must comply with the series of state and federal regulatory requirements outlined above. Blumenthal and Kalambokidis (2006) estimate that the total compliance costs nonprofits faced in the year 2000 to maintain their tax-exempt status totaled \$3.2 billion, or 0.4 percent of total revenues.

In theory, the regulation of nonprofits is intended to enforce the contract with the state that grants nonprofits certain benefits in return for producing public goods. Regulation also protects donors to nonprofits, beneficiaries of charity, and other nonprofit and for-profit organizations who may do business with a fraudulent entity. Irvin (2005) outlines other possible incentives for regulators, including revenue generation (fees may be a source of revenues rather than merely cost recovery) and political motivations (there are political benefits to cracking down on organizations that abuse the public trust).

Just as regulation of publicly traded corporations has grown in the wake of Sarbanes-Oxley and the Enron scandal, so too have calls for accountability within the nonprofit sector increased (Calabrese, 2011; Irvin, 2005). These demands have coincided not only with a sharp rise in the number of public charities, but also in the level of government funding; as public charities have stepped in to provide crucial services in some public service categories, they have come to rely on government funding for the majority of their revenues (Smith, 2010). Often these calls for accountability have been realized in the form of increased financial oversight. For most nonprofits, audit requirements are dictated by state law. However, nonprofits receiving federal money are subject to more stringent auditing requirements under the Single Audit Act, which supplements traditional CPA audits with additional procedures. These requirements occur on top of the requirement to file the Form 990, which requires much of the same information as financial statements but follows different accounting standards, with the 990 following IRS guidance, and financial reporting following Generally Accepted Accounting Principles (GAAP) set by the Financial Accounting Standards Board (FASB).[6] Though the Form 990 was redesigned in 2008 for the first time in more than 15 years, these differences in

[6] 990 forms prepared according to IRS guidelines and financial statements prepared using GAAP use largely similar accounting practices, but do differ in several key ways. Certain types of revenue, such as special events, are netted against related expenses on the Form 990, as opposed to being reported separately under GAAP. Donated services are not counted as contributions under IRS guidelines, whereas GAAP recognizes certain types of donations. GAAP requires investments in equity and debt securities to be reported at fair value, while the Form 990 recognizes only realized gains and losses (Gordon et al., 2007).

accounting practices persist and have led to calls for greater uniformity such that organizations can more easily fulfill the two requirements simultaneously (Keating & Frumkin, 2003).

Despite the increasing calls for enhanced accountability and financial transparency, little is known about the public benefits of audit requirements or their cost to nonprofit organizations. Calabrese (2011) finds that increased public oversight, including public audit requirements, influences organizations to switch from cash to accrual accounting, suggesting that regulations can positively impact nonprofit organizations' financial management practices. However, Neely (2011) finds that California's Nonprofit Integrity Act (2004), which required California charities with at least $2 million in gross revenues to file audited financial statements and establish audit committees, had little effect on financial reporting quality and instead simply led to an increase in accounting fees.

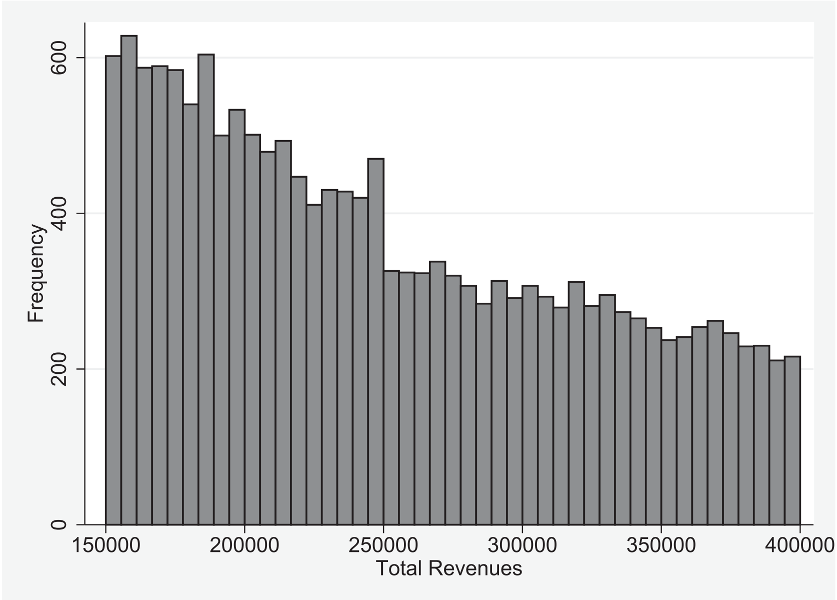
Part of the challenge in studying financial regulations is that there are several different types of reporting requirements, which differ according to the cost and level of scrutiny. The most basic requirement is a compilation, in which a CPA assists in preparing financial statements without undertaking any review or analysis of the underlying transactions. The accountant gives no assurance that the financial statements are without misstatements. In a review, the CPA reviews financial statements prepared by the organization and conducts analytical procedures to ensure that the information contained in the statements is presented accurately, but the CPA does not perform sufficient testing to be able to assess the level of fraud risk or gain a full understanding of the organization's internal control. In an audit, the most comprehensive type of engagement, the auditor gives an opinion as to whether the financial statements are presented accurately and provide a fair portrait of the organization's financial position (Finkler et al., 2013).

Of the 26 states with financial reporting requirements, nine specify only that nonprofits above a revenue threshold file audited financial statements prepared by an independent CPA. A typical example is Arkansas, where charitable organizations with gross revenue over $500,000 must file with the state an audit report of a CPA (Ark. Code § 4-28-403[b]). Many states, including Florida, Illinois, and Michigan, set thresholds around the amount of annual *contributions* rather than revenue, which, if set at the same level, affect far fewer organizations (National Council of Nonprofits, 2013).

New York's requirements are of particular interest for two reasons. As noted above, the thresholds are particularly low compared to other states. The most common threshold for an audit is $500,000 (National Council of Nonprofits, 2013). Since most nonprofits of this size are likely to produce audited financial statements for other reasons, such as grant funding requirements, it is more difficult to isolate the causal impact of audit requirements alone. Furthermore, the fact that New York has two distinct thresholds—for review engagements and full audits—allows for a comparison of the behavioral response to the two thresholds.

DATA AND DESCRIPTIVE FINDINGS

To analyze the effect of audit requirements on nonprofits, this paper uses the NCCS 2005 to 2012 Core Financial Files for public charities. The Core Files are based on the IRS' annual Return Transaction Files, which contain data on all 501(c)(3) organizations required to file the Form 990 or 990-EZ. Before making the data available, the NCCS performs various cleaning procedures to ensure data integrity (NCCS, 2013). Though researchers have identified discrepancies in the past between the NCCS' 990 data and audited financials (Lampkin & Boris, 2002), it remains the most comprehensive database on nonprofit finances.



Note: The figure depicts bunching at the $250,000 threshold due to the state's requirement that charitable organizations with $250,000 or more in total revenues file financial statements audited by an independent CPA.

Figure 2. Bunching at $250,000 Revenue Threshold, 2005 to 2009.

As noted above, New York requires independent accountants' reports only from charities that solicit contributions. Hence, the sample excludes organizations that report zero contributions (approximately 10 percent of charities). The final sample (2005 to 2012) includes data on 30,841 public charities in New York (157,471 observations) as well as a nation-wide comparison group of 396,836 charities (1,974,891 observations) for a total sample of 427,677 organizations and 2,132,362 observations.

Figure 2 displays the bunching visible at the $250,000 revenue threshold due to New York's audit requirement. As noted above, New York-based nonprofits with total revenues less than $250,000 and greater than $100,000 in total revenues must have their financial statements *reviewed* by an independent CPA. Charitable organizations with total revenues greater than $250,000 must file financial statements *audited* by an independent CPA. Thus, the bunching at the higher threshold is due to the additional cost associated with a full audit relative to a review engagement. The bunching at the lower threshold is due to the cost associated with a review engagement; since organizations at that threshold are not otherwise responsible for producing financial statements and need only complete the Form 990 EZ, this cost is essentially the cost of producing financial statements (with or without help from an independent CPA) as well as the additional cost of a review engagement.

Unlike the Form 990, which uses both assets and gross receipts thresholds to determine who must file the longer form, New York's audit thresholds—as with those of most states—are based strictly on revenues, making the estimation of the behavioral response more straightforward. However, since the 990 reporting threshold has at times overlapped with New York's audit thresholds, it is necessary to limit the analyses to years during which the two did not overlap. Prior to 2008, the 990 gross receipts threshold was $100,000; after a transition period (2008 to 2009), the threshold was raised to $200,000 in 2010. So as not to confound the effect of the two different requirements, the analysis of the $100,000 threshold for a review is limited to the years 2008 to 2012, while the analysis of the $250,000 audit threshold is limited to the years 2005 to 2009.

BUNCHING METHODOLOGY

Static Estimation

A bunching design exploits the excess mass in the density distribution around a policy-relevant threshold (or "notch" or "kink") to obtain estimates of the behavioral response to the threshold.[7] So long as the costs associated with a regulation exceed the expected amount of income that a tax filer would have to forego or delay to avoid the requirement (including the risks associated with misreporting), the rational response is to reduce income to an amount below the threshold. The design is similar to other quasi-experimental designs, such as regression discontinuity designs, that exploit a policy-relevant discontinuity, with the difference being that bunching designs measure the extent of avoidance in response to the threshold, whereas regression discontinuity designs assume that there is no manipulation around the cutoff. Like the regression discontinuity design, bunching designs estimate a local average treatment effect at the discontinuity rather than an average treatment effect taken over the entire sample.

In a widely cited paper, Saez (2010) showed how to obtain estimates for the compensated elasticity of reported taxable income by measuring the amount of bunching around the kink points of the U.S. income tax schedule. Since then, a number of other researchers have applied similar methods to related issues in public finance. Onji (2009) examines the distribution of small businesses to measure the extent to which Japan's value-added tax threshold influences firms to separately incorporate segments of the business. Ramnath (2013) documents how U.S. taxpayers manipulate their income to remain below the notch created by the Saver's Credit. Kleven and Waseem (2013) use administrative data to examine tax notches in Pakistan where each bracket is associated with a fixed average tax rate rather than a marginal tax rate.

The design involves estimating the extent of bunching to one or both sides of a discontinuity. This requires modeling the density distribution by constructing revenue "bins" of a certain width, as in Figures 1 and 2, and counting the number of organizations that fall within each bin. By modeling the distribution of, in this case, total revenues, one can measure the excess mass on one side of the notch—along with the reduced mass on the other side—by comparing the distribution around the threshold to a counterfactual distribution in which no such bunching occurs. The identifying assumption is that the counterfactual distribution is smooth across the discontinuity of interest in the absence of bunching and can be correctly modeled.

To complement the standard bunching analysis, this paper also makes use of the fact that the distribution of nonprofits in states other than New York should be similar in shape to the distribution of nonprofits in New York and that, if appropriately weighted, may provide a further counterfactual against which to test the extent of

[7] Whereas kinks are defined as discontinuities in the slope of a choice set, as in graduated income tax schedules that feature changes in marginal tax rates, notches create discontinuous jumps, such as when certain regulations apply only to firms above a certain size (Slemrod, 2010) or where, as in the case studied by Kleven and Waseem (2013), tax brackets are associated with fixed average tax rates. The discontinuities in audit requirements studied here are thus notches, though I will continue to also use the term "threshold."

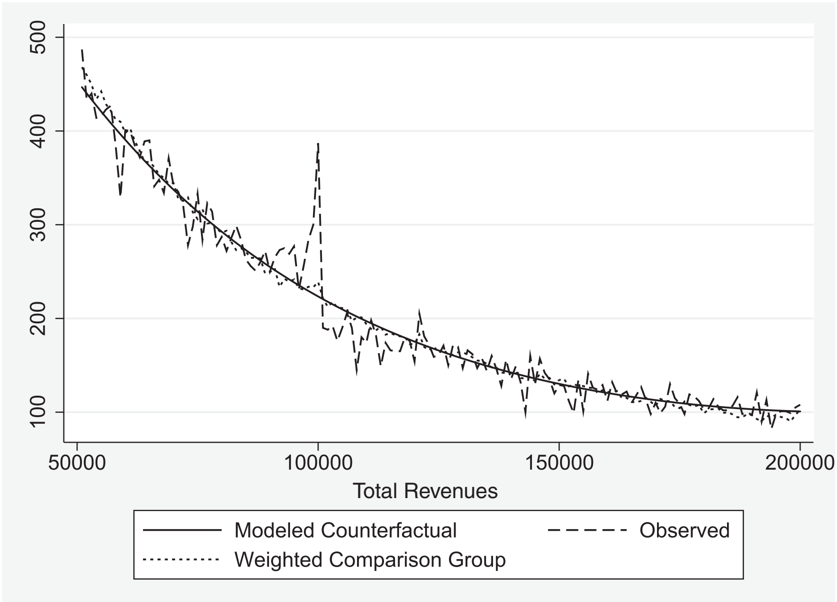


Figure 3. Deviation From Counterfactual at $100,000 Revenue Threshold.

bunching. The comparison group design works similarly to the counterfactual analysis except that the number of organizations within each bin near the notch can simply be counted; no modeling of the distribution is necessary. This comes with a different set of assumptions. On the one hand, it is not necessary to model the counterfactual using regression. On the other hand, the comparison group may not be a good fit with the observed distribution as a result of differences between nonprofits in New York and the rest of the country, thus biasing the extent of bunching.

Figures 3 and 4 compare the observed distribution near the thresholds to these two counterfactual distributions: the "modeled counterfactual" and the counterfactual based on the weighted comparison group. The modeled counterfactuals are estimated by applying a log transformation to the count data and then fitting a polynomial (degree three in the case of the lower threshold, degree four in the higher) to the revenue distribution. The Akaike Information Criterion (AIC) can help to decide the degree of the polynomial, but visual inspection also provides a useful check for determining whether the counterfactual fits the data well on either side of the notch.[8] The weighted counterfactual is produced by equating the total size of the comparison group to the size of the New York distribution within the range of interest.

For the most part, the modeled counterfactual fits the observed data very well until it reaches the area of bunching, where there is excess mass observed to the left of the notch and reduced mass to the right. The comparison groups show that there is no equivalent distortion of the frequency distribution in other states, confirming that the bunching in New York is due to the audit requirements. In Figure 3, the weighted comparison group almost perfectly matches the modeled counterfactual.

[8] In this case, the difference between the AIC values is a matter of decimals, and visual inspection indicates that higher-order polynomials may lead to over-fitting. Hence, the degree of the polynomials differs slightly from what would be determined solely by reliance on the AIC.

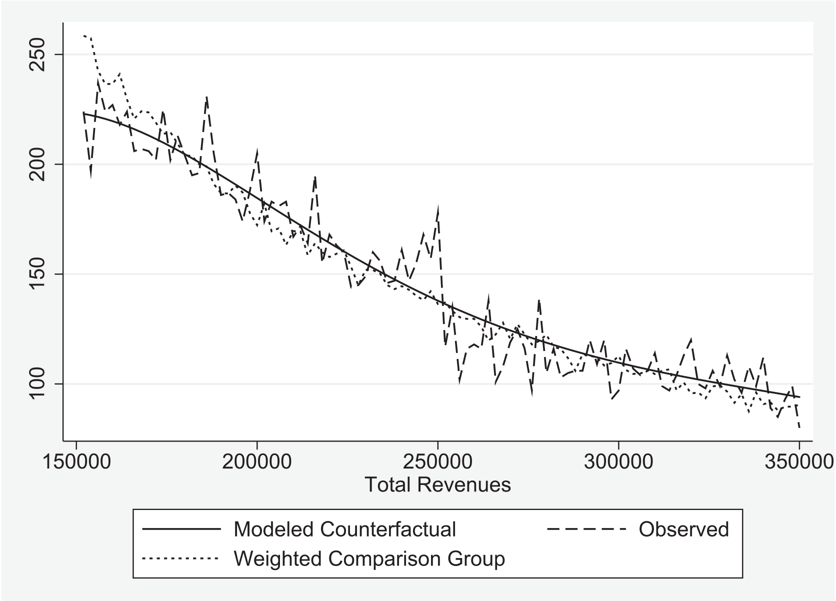


Figure 4. Deviation From Counterfactual at $250,000 Revenue Threshold.

In Figure 4, where the data are more sparse, the fit is not as close, particularly at the left end of the distribution. As with other empirical analyses of tax bunching (Kleven & Waseem, 2013), the reduced mass above the notch does not exhibit the same sharp spike that the excess mass does but instead extends somewhat farther into the distribution, implying that proper estimation of the reduced mass requires examining a broader range of the distribution than estimation of the excess mass.

This highlights one of the challenges in properly estimating the extent of bunching: selecting the region around the discontinuity where bunching occurs. If one estimates bunching over too narrow a stretch of the distribution, then the extent of bunching will be underestimated. On the other hand, if one selects a bunching region that is too large, it may compromise the estimation of the counterfactual in the proximity of the notch. In this case, the range used for estimation is based on visual inspection; the distribution appears smooth until approximately $90,000 and regains its trajectory at approximately $120,000, and thus these points are used as the "omitted range" over which bunching is measured. The range for the full audit (at $250,000) is $230,000 to $290,000. Because estimation of the modeled counterfactual is imprecise at very high and low incomes, it is also useful to restrict the range of the counterfactual so as to more efficiently estimate the distribution in the region near the notch. (A similar logic applies to the choice of weights for the comparison group.) Counts are based on a bin width of $1,000. To explore the sensitivity of these assumptions, Appendix tables will present a range of estimates based on alternative choices for the bin width and range of estimation.[9] The robustness checks reveal that the estimates of bunching are fairly insensitive to the choice of either parameter.

[9] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Standard errors in bunching designs can be obtained in two ways: by using the delta method, a technique of estimating the variance of a function of a random variable by applying Taylor series expansions, or by resampling the revenue distribution. This paper uses the latter method, drawing 500 bootstrap samples and using the distribution of estimates across the 500 draws to construct confidence intervals.

Dynamic Estimation

While providing a measure of behavioral response that lends itself to graphical description, static estimates of bunching suffer from certain shortcomings. Marx (2015) shows how static estimation is subject to bias in settings with serial correlation or extensive margin responses. Panel data offer opportunities to extend the standard methodology, especially in settings where organizations face a threshold repeatedly as opposed to a single time. The basic bunching design also does not establish whether the response to the notch results in a permanent loss in income or simply represents intertemporal shifting; bunching organizations that merely shift income from one period to the next may not ultimately report less income due to the requirement, a response that is not captured by static estimation. le Maire and Schjerning (2013) show how the Danish self-employed defer taxes in response to tax kinks by retaining earnings in the firm.

To address these shortcomings, this paper uses two dynamic approaches based on Marx (2015). First, to obtain estimates conditional on current revenue, a dynamic bunching design is employed that examines bunching in *growth rates* rather than in levels of revenue by considering the distribution of growth rates for separate bands of revenue and then summing over the entire revenue distribution. That is, rather than considering

$$c(y_i) = f(y_i) + \sum_{j=r_L}^{r_U} \beta_j \cdot 1\,[j = r_i] + \epsilon_i \tag{1}$$

as in the static bunching design, the extent of bunching is instead measured by

$$c(g_i) = \sum_{k=1}^{K} \left(f_k(g_i) + \sum_{j=r_L}^{r_U} \left(\beta_{kj} \cdot 1\,[j = r_i]\right) \right) + \epsilon_i \tag{2}$$

where $c(y_i)$is the log count of the number of charities as a function of total revenues, $f(y_i)$ is a polynomial representing the counterfactual distribution, and $r_U - r_L$ is the "omitted" revenue range. The beta coefficients within the summation indicate the extent that each bin in the omitted range departs from the modeled counterfactual. $c(g_i)$ is the log count of the number of charities as a function of the growth of revenues, K is the number of revenue bands, and $f_k(g_i)$ is a polynomial representing the counterfactual distribution of revenue growth. The key difference in equation (2) is that the polynomial expressing the counterfactual is determined separately for each revenue band. Standard errors are again obtained by drawing 500 bootstrapped samples, clustered at the organization level. As in the static analysis, the comparison group design works the same way as the standard analysis except that the number of organizations within each bin near the notch can simply be counted, and no modeling of the distribution is necessary.

Figure 5 illustrates how the dynamic estimator works. The figure shows bunching over one income band: log revenues between 10.825 and 10.850 (approximately $50,000 to $51,500). The vertical lines delineate the omitted range—growth between 0.56 and 0.87 log revenues. (0.87 represents the growth that an organization at the

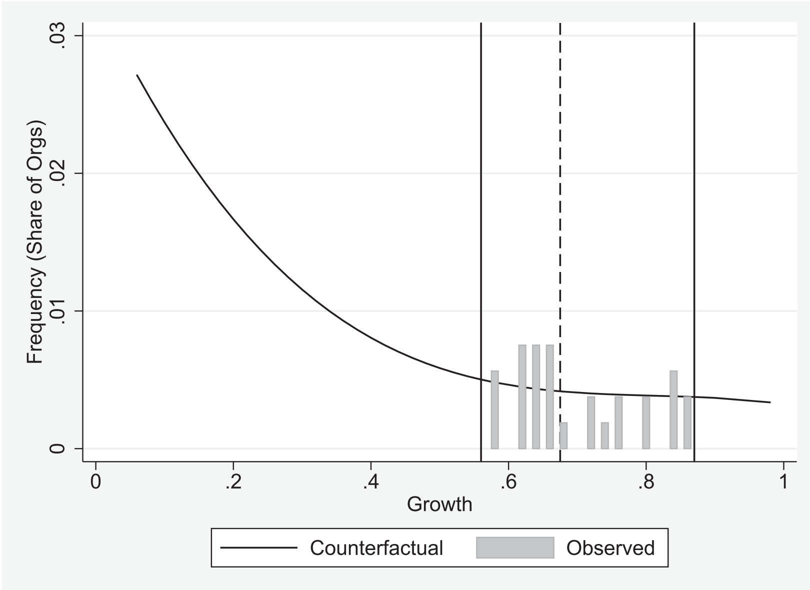


Figure 5. Illustration of Bunching Over Income Growth Within One Income Band: Log Revenues between 10.825 and 10.850 (Approximately $50,000 to $51,500).

bottom end of the income band [approximately $50,000] needs to achieve to reach the upper end of the omitted range [$120,000]. 0.56 is the growth required of an organization at the top of the income band [approximately $51,500] to reach the lower end of the range [$90,000]). The dashed vertical line represents the notch, with the excess mass appearing in the left and the reduced mass appearing on the right. While the bunching estimates from any one income band will be noisy due to the small number of observations, as shown in the figure, the estimator gains power by drawing estimates over all the income bands. (In practice, the number of income bands can be limited to those that might realistically be expected to grow or shrink to a size close to the notch).

There is one drawback to this approach. Since the distance from the threshold varies by income, the bunching range and the reduced range are not as well defined as they are in static estimation. Within each income band there are “overlapping bins” where it is unclear whether charities fall above or below the notch. For example, a growth rate of 0.675 brings a charity with log revenues of 10.85 (approximately $51,500) just above the notch (approximately $101,000), but brings a charity with log revenues of 10.825 (approximately $50,000) just below the notch (approximately $99,000). The simplest approach to deal with this problem is to treat the notch as though it falls in the middle of the overlapping bins, bringing with it the possibility that some organizations falling very close to the notch are counted as bunchers when they should not be and vice versa. An alternative approach is to exclude the income-growth bins that fall into this range and adjust the scaling of the bunching ratio accordingly, though this too brings with it the danger of underestimating bunching if there are many organizations falling very close to the notch. The results of both approaches are presented below.

In addition to a dynamic estimator, further tests can help to reveal whether or not the responses to the notches represent real responses or simply the intertemporal shifting of income. One such test is to compare the growth rates of organizations just above and below the notches. If organizations avoid the audit requirement by

Table 1. Static results.

	$100,000 threshold		$250,000 threshold	
	Modeled counterfactual	Comparison group	Modeled counterfactual	Comparison group
Excess mass	0.0043**	0.0043**	0.0018**	0.0015**
	(0.0007)	(0.0007)	(0.0006)	(0.0005)
Reduced mass	0.0038**	0.0035**	0.0014	0.0017**
	(0.0008)	(0.0006)	(0.0010)	(0.0005)
Average missing revenue (using excess mass)	$1,983	$1,982	$2,590	$2,052
Years	2008–2012		2005–2009	
N	102,860	1,397,182	93,578	1,270,787

Note: The excess and reduced masses are measured as a share of the number of total organizations. Standard errors in parentheses are block-bootstrapped at the organization level using 500 bootstrap samples. Estimates are based on a bin width of $1,000 and omitted ranges of $90,000 to $120,000 and $230,000 to $290,000. Appendix Tables A1 and A2 show results based on different bin width and estimation ranges. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
*$p < 0.05$; **$p < 0.01$.

simply timing revenue so that it does not hit until the following year, then the growth rates of those organizations should be artificially high the following year. If the loss in (reported) revenues due to the notch is permanent, then one would expect to see no difference in the growth rates.

RESULTS

Table 1 reports the results of static estimation. The first two columns report results for the $100,000 threshold using counterfactual modeling and the comparison group, respectively. The next two columns report results for the $250,000 threshold. The polynomials are the same as those presented in Figures 3 and 4. Bunching is measured as the amount of excess mass in the revenue distribution as a share of the total number of charities in the sample. This describes the extent to which bunching distorts the frequency distribution and facilitates comparison with behavioral responses in other contexts. The second row also reports the amount of reduced mass above the threshold. In theory, the amount of reduced mass should equal the amount of excess mass since the excess mass represents charities that would otherwise be found above the threshold in the counterfactual distribution. However, this may not be the case if organizations drop out of the sample or if the counterfactual is imperfectly specified.

In addition to measures of excess and reduced mass, Table 1 also reports the amount of revenue "missing" from the average charity in the vicinity of the threshold. This is calculated by determining how far the area represented by the excess mass would extend into the distribution above the threshold. In the case of static estimation, this is given by

$$\left(\sum_{j=r_L}^{r^*} \beta_j \cdot 1\left[j = r_i\right]\right) \cdot \rho = f(p) \cdot \delta \tag{3}$$

where the summation represents the number of excess organizations, ρ is the bin width ($1,000 here), $f(p)$ is the height of the counterfactual density distribution at

Table 2. Dynamic results.

	\$100,000 threshold		\$250,000 threshold	
	Modeled counterfactual	Comparison group	Modeled counterfactual	Comparison group
Excess mass	0.0028**	0.0026**	0.0010**	0.0009
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Reduced mass	0.0028**	0.0029**	0.0011**	0.0011**
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Average missing revenue (using excess mass)	\$1,276	\$1,181	\$1,374	\$1,205
Years	2008–2012		2005–2009	
N	98,770	1,343,960	72,782	982,529

Note: The excess and reduced masses are measured as a share of the number of total organizations. Standard errors in parentheses are block-bootstrapped at the organization level using 500 bootstrap samples. Estimates are based on a growth bin width of 0.01 log points, a revenue bin width of 0.025 log points, and omitted ranges of \$90,000 to \$120,000 and \$230,000 to \$290,000. Appendix Tables A1 and A2 show results based on different bin width and estimation ranges.
$^*p < 0.05$; $^{**}p < 0.01$.

the notch, and δ is the parameter of interest, the implied amount of missing revenue due to the notch. The estimates of δ are necessarily somewhat imprecise as they do not take into account the curvature of the density distribution at the notch. If charities exhibit heterogeneity in their response to the threshold, then the estimate of δ gives the *average* revenue response (Kleven & Waseem, 2013).

The amount of excess mass below the \$100,000 threshold is statistically significant and equivalent to 0.43 percent of the total number of charities. This is more than twice the size of the static coefficients Marx (2015) reports as the amount of bunching due to the 990 threshold, which makes sense given the more onerous nature of audit requirements.[10] This implies that the average charity reports \$1,983 less revenue than they would in the absence of the requirement, either because they have foregone revenue or simply failed to report it. The estimate of reduced mass above the threshold is quite comparable—0.38 percent of charities—and is also significant at the 1 percent level. At the \$250,000 threshold, the amount of bunching is equivalent to 0.18 percent of charities, and the coefficient is significant at the 1 percent level. This implies that the average charity reports \$2,590 less revenue than they would were there no requirement for a full audit.

The results from the comparison group are extremely close to those from modeling the counterfactual at the \$100,000 threshold. There is a slight difference in the estimates at the \$250,000 threshold, but this is not surprising given that the fit between the distributions is not as close in Figure 4 as it is in Figure 3; nevertheless, the findings are quite similar across the two approaches.

Table 2 presents the results from dynamic estimation based on a growth bin width of 0.01 log points and a revenue bin width of 0.025 log points.[11] As discussed above,

[10] For comparability with his dynamic estimates, Marx limits his sample to those organizations that appeared in the prior year. Limiting the sample in a similar way produces estimates of \$2,157 and \$2,973 for the average missing revenue at the two thresholds, slightly higher than the modeled counterfactual estimates presented in Table 1.
[11] The analytic choices for dynamic estimation are based on visual inspection of the growth distribution for those revenue ranges most likely to bunch.

Table 3. Growth rates of charities just below the threshold relative to charities just above the threshold.

	\$100,000 threshold	\$250,000 threshold
Indicator: below threshold in year t	0.036	−0.011
	(0.027)	(0.034)
N	2,970	1,120
Years	2008–2012	2005–2009

Note: If charities delay revenue, then they should show greater reported revenue in the year after bunching; the growth rate would be positive and statistically significant. The dependent variable is the change in log total revenues from year t to year $t + 1$. The comparison is between charities reporting revenues in year $t \pm \$10{,}000$ of the relevant threshold.

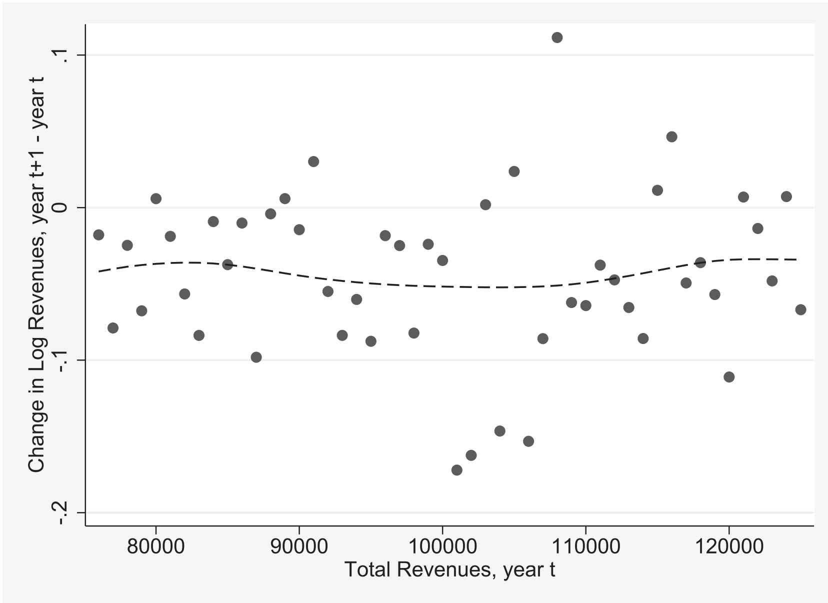
dynamic estimation is not susceptible to the same problems of serial correlation and attrition that static estimation is. The bunching estimates in Table 2 have coefficients that are 40 percent lower on average than in Table 1: an excess mass and a reduced mass equivalent to 0.28 percent of charities at the \$100,000 threshold (for the modeled counterfactual) and 0.10 percent and 0.11 percent at the \$250,000 threshold. These estimates suggest that the average charity reports \$1,276 less revenue as a result of the requirement for a review engagement and \$1,374 as a result of the requirement for a full audit.

It is not surprising that the amount of excess mass is smaller at the higher threshold. As Figures 1 and 2 show, the density distribution is much higher at lower levels of revenues; in other words, there are many more charities with revenues close to \$100,000 than there are charities with revenues of \$250,000 and thus more organizations will be affected by the notch at the lower threshold. On the other hand, a full audit is more expensive than a review engagement, and larger organizations may be more willing to forego what is a relatively smaller proportion of their income. This explains why the estimates of missing revenue are slightly larger at the \$250,000 threshold despite the smaller amount of bunching. This is expressed mathematically in equation (3); the smaller amount of excess mass is balanced by the smaller value of $f(p)$.

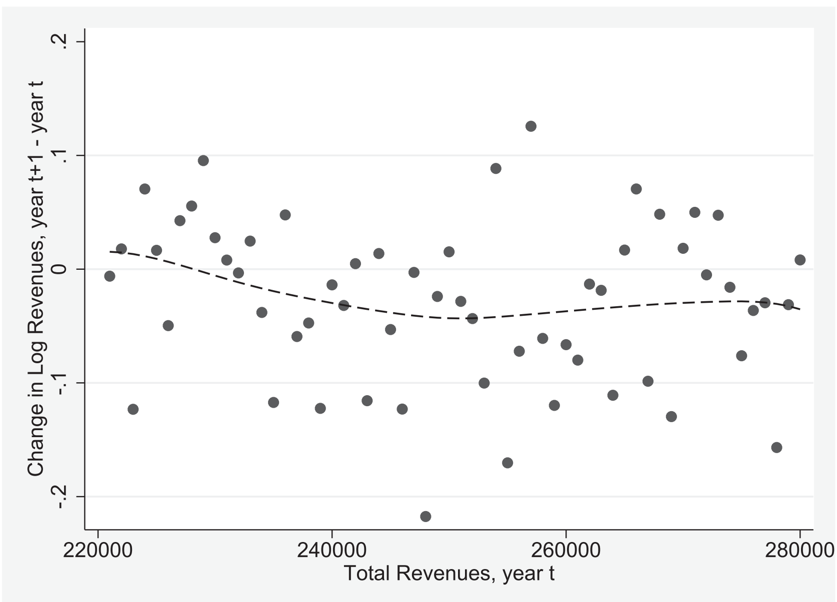
Table 3 presents the results from the test of income shifting. The test compares the growth rates of charities just below and above the threshold in the following year. If a bunching organization reduces revenue to just below the cutoff by re-timing income, then this would result in higher growth rates the following year. Assuming that organizations just above the cutoff would otherwise have comparable growth rates, a comparison of the two should reveal the extent to which the response to the threshold represents a real response rather than simply a shifting of income.

The first column again reports results for the \$100,000 threshold, while the second column reports results for the \$250,000 threshold. In both cases, the coefficients are not significant. In the case of the second threshold, the coefficient is actually negative. That is, there is no statistically significant difference in the growth rates of charities that are just above and below the thresholds, and on average charities that are just below the \$250,000 threshold see lower growth rates, not higher. This would suggest that they are not simply re-timing their income in response to the audit threshold.

Figure 6a and b provides graphical evidence of the test. The change in log revenues between year t and year $t + 1$ is nonparametrically regressed on total revenues in year t, using bins \$1,000 in size. There is no obvious evidence in either figure of a spike in growth rates just below the notch, as one would expect if organizations grew relatively more quickly the year after facing a notch. This test is not without limitations. As noted above, rather than reporting that certain income appeared in



(a)



(b)

Note: The circles represent the average change in log growth for each $1,000 bin. Figure 6a includes observations from 2008 to 2011. Figure 6b includes observations from 2005 to 2009.

Figure 6. (a) Growth Rates on Either Side of $100,000 Threshold. (b) Growth Rates on Either Side of $250,000 Threshold.

Table 4. The effect of audit requirements on attrition rates.

	\$100,000 threshold	\$250,000 threshold
Difference in attrition rate	0.010*	0.002
	(0.005)	(0.009)
N	619,830	418,558
Years	2008–2012	2005–2009

Note: If charities respond to audit thresholds by dropping out of the sample, then the attrition rate will be higher among organizations that face the threshold in the following year, all else being equal. The table presents the results of a difference-in-difference analysis where the difference in attrition rates for organizations above and below the thresholds in New York is compared to the corresponding difference in rates for organizations in the comparison group. The comparisons are between organizations most likely to grow to a range just above the threshold based on comparison group data (\$73k ± \$10k, \$183k ± \$10k) and organizations above both thresholds where differences in attrition rates are not attributable to the audit requirements (>\$290k).
*$p < 0.05$.

the year $t + 1$ rather than year t, some organizations may simply never report it. Moreover, if charities bunch repeatedly in response to the threshold, it will skew the average growth rates of charities below the notch. Nevertheless, there is no clear evidence that charities are re-timing income.

One final concern with bunching, particularly as it concerns estimating behavioral responses to *notches* as opposed to kinks, is that organizations above the threshold may drop out of the sample (an extensive margin response) rather than lower their revenues (an intensive margin response; Kleven & Waseem, 2013). In this case, dropping from the sample would mean that an organization failed to file their 990 with the IRS in a given year. This is relevant for two reasons. If charities are more likely to go missing from the sample when their revenues exceed the thresholds, then OLS may not provide consistent estimates (Marx, 2015). But if charities respond to the thresholds, not by bunching, but by leaving the sample (even if only temporarily), then it may help to explain the difference between the amount of excess mass and the amount of reduced mass as well as the difference between the static and dynamic estimates.

A comparison of attrition rates shows that New York charities below the notches are more likely to disappear from the sample than charities above the notches, though the average difference is somewhat sensitive to the range over which the rates are calculated. However, these differential rates of attrition cannot be directly attributed to the notches, as attrition may be negatively correlated with income, i.e., smaller organizations may simply be more likely to cease operations than larger organizations. The existence of a comparison group presents a straightforward way of calculating the effect of the notches on attrition. The attrition rates of New York charities below the notches can be compared with the rates of charities in other states in a similar income range.

Table 4 presents the results of such a comparison. To make the comparison more robust, the test compares the difference in attrition rates above and below the notch in New York with the difference in attrition rates above and below the notch in the comparison group. This difference-in-difference analysis accounts for the possibility that charities in New York may demonstrate higher attrition at all income levels. The results show that organizations below the \$100,000 threshold in New York are 1.0 percentage points more likely to disappear from the sample, a result that is significant at the 5 percent level. Thus, the notches may not only cause organizations to report less revenue, but they may also cause charities to disappear from the sample at higher rates. In contrast, while there is some evidence of extensive margin responses at the \$250,000 threshold, the effect is small and not significant.

Appendix Tables A1 to A3 present static and dynamic bunching estimates under different sets of assumptions.[12] Appendix Tables A1 and A2 present estimates using different values for the size of the reduced range, the width of the revenue bins, and the range over which the counterfactual was estimated. These sensitivity tests reveal that the bunching estimates are fairly robust to slight changes in assumptions. The static estimates of the missing revenue at the lower threshold range from $1,924 to $2,058, compared to the baseline estimate of $1,983, while the dynamic estimates range from $992 to $1,224, compared to the baseline estimate of $1,276. Results at the higher threshold are slightly more sensitive; static estimates range from $1,630 to $2,716, compared to a baseline of $2,590, while the dynamic estimates range from $1,018 to $1,565 compared to a baseline of $1,374. The estimate that appears most sensitive to analytic choices is the static estimate at the higher threshold; changing the range over which the counterfactual is estimated lowers the point estimate by $960. However, the results for the comparison group analysis show much less sensitivity, likely because fewer modeling assumptions are required. Appendix Table A3 presents dynamic estimates that exclude the "overlapping" bins in order to account for the fact that the location of the threshold is not as well defined in dynamic estimation.[13] The results for the missing revenue are similar to the baseline estimates, suggesting that bunching estimates are not significantly biased by large numbers of organizations that fall right at the notch.

DISCUSSION

Bunching designs represent a useful approach to estimating the behavioral response to a policy threshold. This paper examines the amount of bunching that occurs in the reported revenues of nonprofits in New York in response to that state's audit requirements, estimating the percentage of organizations that report less revenue to avoid either the requirements for a review engagement at revenue levels of $100,000 or above or a full audit at revenue levels of $250,000 or above. While static estimates coincide with graphical depictions of bunching, dynamic estimates are preferred in this case because of how they incorporate information about the counterfactual and the susceptibility of the static estimates to serial correlation and attrition. Dynamic estimation suggests that the average nonprofit reports $1,200 to $1,300 less revenue in response to the $100,000 threshold, while the average nonprofit in range of the $250,000 threshold reports $1,200 to $1,400 less. A comparison of growth rates between organizations just above and below the thresholds suggests that these responses do not represent intertemporal shifting.

How do these estimates compare to the cost of hiring a CPA to perform audit functions? According to one accounting firm in the New York area that specializes in not-for-profit and tax-exempt organizations (www.jvcpapllc.com/pricing-fees/), the cost of an audit is in the range of $8,000 to $20,000. The cost of a review starts at $4,500. Form 990 preparation alone starts at $2,500. (Form 990 preparation is also included in a full audit or review engagement.) Thus, the estimates of the average avoidance costs are considerably below the pricing of audit services.

How do the responses to the two different thresholds compare? Both static and dynamic estimates suggest that bunching occurs with much more frequency at the lower threshold. In other words, many more organizations lower their income in response to the lower threshold. The implied amount of average missing revenue is

[12] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.
[13] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

slightly higher in absolute terms at the higher threshold but lower when measured as a percentage of income for charities in that range. This is consistent with Blumenthal and Kalambokidis' (2006) finding that nonprofits exhibit economies of scale in regard to compliance costs; large organizations bear proportionately smaller costs than smaller organizations.

What do these estimates suggest about the burden imposed on nonprofits in New York by the state's audit requirements? In the absence of comparative data on other regulations, it is difficult to draw conclusions about the severity of New York's requirements. Nevertheless, a few observations can be made. The requirement to complete financial statements and undergo a review engagement appears to be more burdensome to nonprofits than filing the form 990, as a comparison with Marx's (2015) estimates shows. Moreover, the comparison group data clearly show that New York nonprofits in the vicinity of the notches were at a disadvantage relative to other states, lowering their reported revenues by 0.5 to 1.3 percentage points; this may explain why the state has since raised the level of its audit thresholds. The number of organizations affected also appears to be relatively large; according to the static estimates, as much as 0.43 percent of nonprofit charities in New York, representing approximately 440 organizations, exhibit bunching at the lower threshold alone. And while the average amount of missing revenue appears to be well below the price of audit services, it is difficult to state definitely what this means for the average buncher, given that responses to the notch are likely to be heterogeneous.

Of course a proper benefit-cost analysis of New York's regulations will also depend on the size of the benefits, which are not measured here. These benefits include the prevention of fraud and accounting irregularities and abuse of the public trust. Unfortunately, little is known about the public benefits of financial transparency, despite frequent calls for more detailed reporting. There are some indications that the public benefits may be overstated (Neely, 2011), though it should also be noted that nonprofits themselves may benefit from increased transparency; 21 percent of the survey respondents in Blumenthal and Kalambokidis' (2006) study reported deriving benefits from reporting requirements, including the opportunity to review the organization's finances and enhanced ability to qualify for external funding. Nevertheless, this paper highlights the need to consider the costs and avoidance behavior associated with regulations as well as the benefits, particularly for smaller organizations.

As calls for financial transparency increase, public officials must ensure that regulations aimed at improving financial oversight do not represent an undue burden on nonprofit organizations that have the potential to serve an essential role in the delivery of public services. This paper examines the behavioral response to a regulatory notch in order to estimate the avoidance behavior associated with nonprofit regulatory requirements. As states move to bring their regulatory requirements into greater uniformity, further research on both the costs and benefits of financial regulations will help to better inform the efficient design of state-level fiscal institutions.

TRAVIS ST. CLAIR is an Assistant Professor in the School of Public Policy at the University of Maryland, 2101 Van Munching Hall, College Park, MD 20742 (e-mail: tstclair@umd.edu).

ACKNOWLEDGMENTS

I am grateful to Ben Marx, Lesley Turner, Peter Reuter, Angela Bies, Bob Grimm, and seminar participants at the 2014 Association for Budgeting and Financial Management conference for their feedback on earlier drafts. Any errors are my own.

REFERENCES

Blumenthal, M., & Kalambokidis, L. (2006). The compliance costs of maintaining tax exempt status. National Tax Journal, 56, 235–252.

Calabrese, T. (2011). Public mandates, market monitoring, and nonprofit financial disclosures. Journal of Accounting and Public Policy, 30, 71–88.

Cordes, J. T., & Weisbrod, B. A. (1998). Differential taxation of nonprofits and the commercialization of nonprofit revenues. Journal of Policy Analysis and Management, 17, 195–214.

Finkler, S. A., Purtell, R. M., Calabrese, T. D., & Smith, D. L. (2013). Financial management for public, health, and not-for-profit organizations. Upper Saddle River, NJ: Pearson Education.

Gordon, T., Khumawala, S. B., Kraut, M. A., & Meade, J. A. (2007). The quality and reliability of form 990 data: Are users being misled? Academy of Accounting and Financial Studies Journal, 11, 27–49.

Internal Revenue Service. (2014). 2014 instructions for Form 990 return of organization exempt from income tax. Retrieved July 10, 2015, from http://www.irs.gov/pub/irs-pdf/i990.pdf.

Irvin, R. (2005). State regulation of nonprofit organizations: Accountability regardless of outcome. Nonprofit and Voluntary Sector Quarterly, 34, 161–178.

Keating, E. K., & Frumkin, P. (2003). Reengineering nonprofit financial accountability: Toward a more reliable foundation for regulation. Public Administration Review, 63, 3–15.

Kleven, H. J., & Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Evidence from Pakistan. Quarterly Journal of Economics, 128, 669–723.

Lampkin, L. M., & Boris, E. T. (2002). Nonprofit organization data. American Behavioral Scientist, 45, 1675–1715.

le Maire, D., & Schjerning, B. (2013). Tax bunching, income shifting and self-employment. Journal of Public Economics, 107, 1–18.

Marx, B. (2015). Dynamic bunching estimation and the cost of regulatory hurdles for charities. Working Paper, March 2015.

National Center for Charitable Statistics (NCCS). (2013). Guide to using NCCS data. Retrieved July 10, 2014, from http://nccsdataweb.urban.org/NCCS/extracts/NCCSDataGuide2013.pdf.

National Council of Nonprofits. (2013). State law nonprofit audit requirements. Retrieved June 1, 2015, from www.councilofnonprofits.org/nonprofit-audit-guide/state-law-audit-requirements.

Neely, D. G. (2011). The impact of regulation on the U.S. nonprofit sector: Initial evidence from the nonprofit integrity act of 2004. Accounting Horizons, 25, 107–125.

New York State Office of the Attorney General. (2012). Instructions for form CHAR500: Annual filing for charitable organizations. Retrieved August 2015, from http://www.charitiesnys.com/pdfs/char500i.pdf.

Onji, K. (2009). The response of firms to eligibility thresholds: Evidence from the Japanese value-added tax. Journal of Public Economics, 93, 766–775.

Ramnath, S. (2013). Taxpayers' responses to tax-based incentives for retirement savings: Evidence from the Saver's Credit notch. Journal of Public Economics, 101, 77–93.

Saez, E. (2010). Do taxpayers bunch at kink points? American Economic Journal, 2, 180–212.

Slemrod, J. (2010). Buenas notches: Lines and notches in tax system design. University of Michigan, Mimeo, September 2010.

Smith, S. R. (2010). Nonprofits and government: Implications for policy and practice. Journal of Policy Analysis and Management, 29, 621–625.

APPENDIX

Table A1. Bunching estimates at the $100,000 threshold based on different bin widths and omitted ranges.

a. Static estimates

	Modeled counterfactual				Comparison group			
	Baseline estimate				Baseline estimate			
Excess mass	0.0043^{**}	0.0043^{**}	0.0041^{**}	0.0042^{**}	0.0043^{**}	0.0043^{**}	0.0042^{**}	0.0044^{**}
	(0.0007)	(0.0007)	(0.0010)	(0.0008)	(0.0007)	(0.0006)	(0.0008)	(0.0006)
Reduced mass	0.0038^{**}	0.0037^{**}	0.0034^{**}	0.0044^{**}	0.0035^{**}	0.0034^{**}	0.0033^{**}	0.0033^{**}
	(0.0008)	(0.0008)	(0.0010)	(0.0010)	(0.0006)	(0.0006)	(0.0008)	(0.0006)
Avg. missing revenue	$1,983	$1,981	$1,924	$1,928	$1,982	$1,999	$1,933	$2,058
Bin width	$1,000	$500	$1,000	$1,000	$1,000	$500	$1,000	$1,000
Omitted range	$90–120k	$90–120k	$85–125k	$90–120k	$90–120k	$90–120k	$85–125k	$90–120k
Counterfactual range	$50–200k	$50–200k	$50–200k	$50–150k	$50–200k	$50–200k	$50–200k	$50–150k
N	102,860	102,860	102,860	102,860	1,397,182	1,397,182	1,397,182	1,397,182

b. Dynamic estimates

	Modeled counterfactual				Comparison group			
	Baseline estimate				Baseline estimate			
Excess mass	0.0028^{**}	0.0024^{**}	0.0027^{**}	0.0026^{**}	0.0026^{**}	0.0021^{**}	0.0023^{**}	0.0023^{**}
	(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0005)
Reduced mass	0.0028^{**}	0.0025^{**}	0.0030^{**}	0.0030^{**}	0.0029^{**}	0.0028^{**}	0.0031^{**}	0.0028^{**}
	(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0006)	(0.0005)
Avg. missing revenue	$1,276	$992	$1,233	$1,224	$1,181	$992	$1,075	$1,080
Bin width	0.01 log	0.02 log	0.01 log	0.01 log	0.01 log	0.02 log	0.01 log	0.01 log
Omitted range	$90–120k	$90–120k	$85–125k	$90–120k	$90–120k	$90–120k	$85–125k	$92–115k
Counterfactual range	–1.5–1.5	–1.5–1.5	–1.5–1.5	–1.75–1.75	–1.5–1.5	–1.5–1.5	–1.5–1.5	–1.5–1.5
N	98,770	98,770	98,770	98,770	1,343,960	1,343,960	1,343,960	1,343,960

Note: Bootstrapped standard errors in parentheses.
$^{*}p < 0.05$; $^{**}p < 0.01$.

Table A2. Bunching estimates at the $250,000 threshold based on different bin widths and omitted ranges.

a. Static estimates

	Modeled counterfactual				Comparison group			
	Baseline estimate				Baseline estimate			
Excess mass	0.0018**	0.0019**	0.0016**	0.0012*	0.0015**	0.0015+	0.0014**	0.0013**
	(0.0006)	(0.0007)	(0.0005)	(0.0006)	(0.0005)	(0.0004)	(0.0005)	(0.0005)
Reduced mass	0.0014	0.0013	0.0012	0.0022*	0.0017**	0.0017**	0.0017**	0.0020**
	(0.0010)	(0.0010)	(0.0008)	(0.0009)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Avg. missing revenue	$2,590	$2,716	$2,252	$1,630	$2,052	$2,033	$1,872	$1,845
Bin width	$1,000	$2,000	$1,000	$1,000	$1,000	$2,000	$1,000	$1,000
Omitted range	$230–290k	$230–290k	$235–285k	$230–290k	$230–290k	$230–290k	$235–285k	$230–290k
Counterfactual range	$150–350k	$150–350k	$150–350k	$150–400k	$150–350k	$150–350k	$150–350k	$150–400k
N	93,578	93,578	93,578	93,578	1,270,787	1,270,787	1,270,787	1,270,787

b. Dynamic estimates

	Modeled counterfactual				Comparison group			
	Baseline estimate				Baseline estimate			
Excess mass	0.0010*	0.0009*	0.0008*	0.0011*	0.0009*	0.0007	0.0009*	0.0008
	(0.0005)	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0005)	(0.0004)	(0.0005)
Reduced mass	0.0011*	0.0007	0.0009	0.0009	0.0011*	0.0009	0.0008	0.0012*
	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Avg. missing revenue	$1,374	$1,293	$1,331	$1,565	$1,205	$1,018	$1,182	$1,031
Bin width	0.01 log	0.02 log	0.01 log	0.01 log	0.01 log	0.02 log	0.01 log	0.01 log
Omitted range	$230–290k	$230–290k	$235–285k	$230–290k	$230–290k	$230–290k	$235–285k	$225–295k
Counterfactual range	–1.5 to 1.5	–1.5 to 1.5	–1.5 to 1.5	–1.25 to 1.25	–1.5 to 1.5	–1.5 to 1.5	–1.5 to 1.5	–1.5 to 1.5
N	72,787	72,787	72,787	72,787	982,529	982,529	982,529	982,529

Note: Bootstrapped standard errors in parentheses.
**p* < 0.05; ***p* < 0.01.

Table A3. Dynamic estimates that ignore overlapping bins.

	$100,000 threshold		$250,000 threshold	
	Modeled counterfactual	Comparison group	Modeled counterfactual	Comparison group
Excess mass	0.0024**	0.0020**	0.0011*	0.0008
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Reduced mass	0.0023**	0.0029**	0.0010*	0.0014**
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Average missing revenue	$1,123	$907	$1,503	$1,029
Years	2008–2012		2005–2009	
N	98,770	1,343,960	72,782	982,529

Note: Bootstrapped standard errors in parentheses.
**p* < 0.05; ***p* < 0.01.