Dynamic Bunching Estimation with Panel Data

Marx, Benjamin M.

University of Illinois at Urbana-Champaign

23 August 2018

Dynamic Bunching Estimation with Panel Data

Benjamin M. Marx*

August 2018

Abstract

An increasingly common technique for studying behavioral elasticities uses bunching estimation of the excess mass in a distribution around a price or policy threshold. This paper shows how serial dependence of the choice variable and extensive-margin responses may bias these estimates. It then proposes new bunching designs that take greater advantage of panel data to avoid these biases and estimate new parameters. Standard methods over-reject in simulations using household income data and over-estimate bunching in an application with charities. Designs exploiting panel data provide unbiased bunching estimates, improved heterogeneity analysis, and the ability to estimate extensive-margin responses and long-run effects.

JEL: C23, D92, H21, H26.

* Department of Economics, University of Illinois at Urbana-Champaign, 214 David Kinley Hall, 1407 W. Gregory, Urbana, IL 61801. Correspondence should be sent to benmarx@illinois.edu. I am especially grateful to Bruce Kogut, Wojciech Kopczuk, Robert McMillan, Brendan O'Flaherty, Bernard Salanié, and Miguel Urquiola for invaluable advice and discussions. I also thank Raj Chetty, Julie Berry Cullen, Jonathan Dingel, Francois Gerard, Jessie Handbury, Todd Kumler, Ilyana Kuziemko, Corinne Low, David Munroe, Giovanni Paci, Abigail Payne, Petra Persson, Jonah Rockoff, Maya Rossin-Slater, Ugo Troiano, Lesley Turner, Eric Verhoogen, Reed Walker, Caroline Weber, and seminar participants at meetings of the International Institute for Public Finance, the Office of the U.S. Department of the Treasury Office of Tax Analysis, the National Tax Association Annual Conference, the Association for Public Economic Theory, the Midwestern Economics Association, and the University of Illinois and other universities. I gratefully acknowledge the financial support of a Nonprofit Research Fellowship from the National Bureau of Economic Research and the Outstanding Dissertation Award from the National Tax Association.

1 Introduction

Taxes, income eligibility limits, and other policies can change agents' incentives at threshold values of a choice variable. Some policies create a kink in the cost function at the threshold, while others create discontinuities, or "notches" (Slemrod, 2010). If costs increase above the threshold, the incentive to locate at or below the threshold can create "bunching" in the distribution of the choice variable. Saez (2010a) launched a bunching-estimation literature by showing how the extent of bunching around a kink in the tax schedule can be used to obtain an estimate of the tax price elasticity of income. The theory and technique of bunching estimation have been developed by Kleven and Waseem (2013), and bunching has been estimated at many kinks and notches in tax and regulatory schedules.[1]

This paper proposes and evaluates new "dynamic bunching estimation" designs for panel data. In the extant literature, researchers with access to panel datasets have treated them as repeated cross sections by pooling years. The methods proposed in this paper take greater advantage of the panel structure, and they offer improved identification and the ability to quantify parameters that have not been estimated with standard techniques. I examine two threats to identification for the standard design and show that the dynamic alternatives rely on assumptions that are arguably more plausible, particularly in settings where agents face a notch repeatedly. These designs can identify new parameters including long-run effects of a notch, which provide evidence on whether bunching is driven by retiming or misreporting, and the extent of extensive-margin responses to a notch. In addition, they offer new capabilities for analyzing heterogeneity in the bunching response and concurrent responses in other variables.

I first study two identification issues for the standard, static approach. One issue is extensive-margin responses, which have been discussed in the bunching literature but cannot be quantified with existing methods. Another issue that has not been discussed as an identification concern arises from persistent notches and serial dependence of the choice variable.[2] To see the issue for identification, suppose there is a tax notch that imposes a fixed cost when household income exceeds a threshold, and suppose that income exhibits positive serial dependence, i.e. that reducing income in the present will also reduce income in the future.[3] When the notch is introduced, we would expect some households with potential incomes just above

[1]Recent papers estimating bunching have studied firms' responses to taxes and regulations (Sallee and Slemrod 2012, Devereux et al. 2014, Best et al. 2015, Sallee and Ito 2018), individuals' responses to taxes and benefits (Chetty et al. 2013a, Burman et al. 2014, Kleven and Schultz 2014, Kleven et al. 2014, Kline and Tartari 2016, Rees-Jones 2018), consumer responses to prices (Ito 2014, Grubb and Osborne 2015, Einav et al. 2017), saving (Ramnath 2013, Weber 2016), teacher responses to testing incentives (Brehm et al. 2017, Dee et al. 2015), and corruption (Camacho and Conover 2014). Kleven (2016) provides an extensive review of bunching research.

[2]Kleven and Waseem (2013) note that serial dependence should reduce bunching elasticities if agents are forward-looking because they will not want to reduce future income, whereas I highlight the implication for identification.

[3]Serial dependence may be mechanical or formulaic, such as when the choice variable is assets (e.g., Seim 2017, Brülhart et al. 2016). Other variables like income may also exhibit serial dependence, such as if agents receive raises that are a percentage of the prior year's wage.

it to bunch below it, but households far above the notch would not be willing to reduce their income all the way to the notch. This motivates an identifying assumption in bunching estimation, which is that the notch will not affect the number of agents at values far from the threshold, and hence the density at these values can be used to construct a counterfactual near the notch. However, agents in many settings face the same notch repeatedly. If the running variable exhibits positive serial dependence, then reducing its value today in order to bunch will lower its value tomorrow. This is true regardless of where that value lies in tomorrow's distribution, implying the potential for distortion far from the notch and excess curvature of the true counterfactual distribution within the omitted region. When a notch is faced more than once and income is serially dependent, the standard methodology describes accumulated effects but does not recover the behavioral parameters governing these effects.

Dynamic bunching estimation follows the intuition of the static approach but extends it to dynamic processes. In light of the concerns about the identifying assumption that the notch only affects the distribution locally, I propose three variants that exploit panel data in differing ways. I present these methods in order of increasing complexity. The first method is ideally suited to a notch that agents face only once, which avoids concerns about serial dependence and repeated bunching. This method simply involves widening bins to create one unselected "treatment" bin around the notch, and with panel data it provides estimates of short- and long-run effects and of serial dependence. The second and third methods condition on income in the year prior to the year that an agent approaches the notch, implying that the localness assumption is applied not to the overall income distribution but only to the distribution of one-year growth rates. This identification strategy can be applied even when agents face the same notch repeatedly. I first implement it with bins and OLS, which provides transparent evidence of manipulation, improved heterogeneity analysis, and tests of long-run effects. The OLS implementation is easy to execute and has already been employed, based on an earlier draft of this paper, by St.Clair (2016). The final method uses maximum likelihood estimation to implement the same identification strategy. Though more complicated, the MLE implementation retains the flexibility of the binning approach while offering improved precision and the first estimates of extensive-margin responses in the literature. R code is provided with this paper to facilitate adoption.

Simulations provide evidence on the performance of the static and dynamic estimation strategies. These simulations use draws of household income profiles from the Panel Study of Income Dynamics (PSID). I first provide graphical intuition for the less-studied issue of serial dependence of income. With serial dependence, bunching in one year causes distortions throughout the distribution of incomes in later years, increasing the mass far below the threshold and decreasing the mass far above. I then document the biases in static estimates caused by each of (a) extensive-margin responses and (b) serial dependence in income. Both issues cause the static estimates to reject a true null of no bunching in considerably more than 5 percent of draws.

In contrast, simulations with dynamic estimators produce bias close to zero and coverage rates close to 0.05.

Benefits of each of the proposed methods are demonstrated by application to the bunching of charities at a reporting threshold. Charities account for about 9 percent of all U.S. wages and salaries (Roeger et al., 2012). Though exempt from taxes, these organizations must file an annual information return with the IRS. I find that charities bunch below the income eligibility threshold for filing a simplified “EZ” form. A more detailed analysis of this policy and its welfare implications appears in Marx (2018); here, I focus on how the application exemplifies the benefits of each of the three proposed methods. Temporary-notch estimation shows that bunching permanently reduces income, indicating positive serial dependence, an identification concern for standard bunching estimation. Dynamic OLS estimation reveals significant heterogeneity in the bunching response and shows that the notch permanently reduces the growth of some charities. Dynamic MLE estimation finds significant extensive-margin responses (most likely non-filing or late filing) that appear to be at least as important as the bunching response. Static estimates capture none of these phenomena and overstate bunching by a factor of 2 or 3.

A few papers have studied dynamic aspects of bunching. For example, thresholds in time may induce bunching in intertemporal decisions such as the choice of when to claim retirement (Manoli and Weber, 2016) or purchase a vehicle (Sallee, 2011). Others have documented how bunching at an income threshold varies over time; Chetty et al. (2013b) and Mortenson and Whitten (2018) use income tax microdata to study kinks in the tax code, and both studies find that the amount of bunching has increased several-fold since 1996. Gelber et al. (2013) examine whether bunching persisted after elimination of the Social Security earnings test, and Kleven and Waseem (2013) use panel data to estimate the share of taxpayers remaining just above or below tax notches. The approach of le Maire and Schjerning (2013) augments the estimating equation with estimated deviations from mean income, which a model of income-smoothing predicts will distinguish income retiming from real income responses. In contrast to the existing research, I exploit panel data to avoid selection bias even in settings where agents face repeatedly. Incorporating income dynamics requires neither structural assumptions nor identification based on comparisons between agents who bunch and agents who do not. The intuition is similar to that for the discrete-distribution application of Schivardi and Torrini (2007), who estimate the effect of a 15-employee notch in Italian labor law on employment growth by solving for the steady state of a counterfactual one-year transition matrix between employment levels. The designs in this paper enable analysis of notches in continuous variables such as income.

The paper proceeds as follows. Section 2 provides information about the household and charity datasets. In Section 3, I describe the standard estimation approach in the literature, discuss identification challenges, and evaluate the standard approach with simulations. The next three sections each present a dynamic estimation strategy that exploits panel data, evaluate the strategy with simulations, and show its benefits in application. The strategy of Section 4 exploits a temporary notch, that in Section 5 generalizes this to an OLS regression that can be estimated for temporary or persistent notches, and Section 6 presents the MLE implementation. Section 7 provides concluding remarks, including a summary list of diagnostics that researchers can check when estimating bunching.

2 Data

I use two panel datasets to illustrate and evaluate each bunching estimation method. Unlike the administrative records that are required to estimate bunching in many settings, both of the datasets used in this paper are publicly available, which will facilitate replication and extension of the proposed methods. One of the datasets is the widely familiar Panel Study of Income Dynamics. The other dataset covers charitable organizations, enabling an application of the methods to a federal reporting requirement for these charities. Marx (2018) studies this application in greater detail, relating this notch to the broader question of optimal regulation of information provision, and providing a general theoretical framework for evaluating welfare effects of regulatory notches. Here, the setting provides a particularly useful application for the proposed methods because there is publicly available panel data, variation in the location of the notch, and interesting patterns in all of the dimensions in which dynamic bunching estimation provides new information.

The dataset for the charities application is provided by the National Center for Charitable Statistics (NCCS), an initiative of the Urban Institute. This dataset is the union of all annual "Core Files," which provide digital records of the information returns that tax-exempt organizations are required to file with the IRS. Large organizations must file IRS Form 990, while those with both total assets and gross receipts below certain thresholds may file Form 990-EZ. Bunching is visible at the gross receipts threshold, which was held at \$100,000 from before the beginning of the panel in 1991 until 2008. Further details are provided by Marx (2018). The variables used here are organization (Employer Identification Number), year, and gross receipts.

To conduct simulations, I use a panel of household taxable income from the Panel Study of Income Dynamics. Use of real-world data from the PSID shows that results are not driven by a particular distribution of known function form. I restrict attention to the nationally-representative sample within the PSID and exclude the over-sample of low-income families. The sample covers years from 1967 through 2013, and I inflate all amounts to 2013 dollars using the Consumer Price Index. The variable of interest is the household's taxable income.[4] Where $T \geq 2$ years of data are required for the analysis, I use all combinations in which household income is observed T times in succession. For most of the analysis I need only two years of

[4]According to the PSID Data Custom Codebook generated with the data extract, "This variable is the sum of Head's labor income, Wife's labor income, asset part of income from farm, business, roomers, etc., rental, interest and dividend income, and Wife's income from assets."

consecutive observations of household income, which I label the "base-year" income and "next-year" income.

Comparisons of bunching estimators requires the ability to create or eliminate bunching. I eliminate bunching by estimating a smoothed version of the PSID distribution of consecutive-year incomes. To do so I define r_{it} as the log of realized base-year income and $g_i = r_{it+1} - r_{it}$ as the growth rate from base-year income to next-year income. I first bin the inverse cdf of g_i and estimate it as a flexible function with range $[0,1]$ so that I can generate random numbers u_i from a uniform distribution and convert them to values of g_i.[5] I then bin the joint distribution of r_{it} and g_i to estimate the condition cdf $F^{-1}(r_{it}|g_i)$ as a flexible function with range $[0,1]$ so that I can generate random numbers v_i from a uniform distribution and convert them to values of r_{it}.[6] This allows me to generate a sample of any size according to the smoothed joint distribution of income and growth.

I generate bunching over a heterogeneous distribution of abilities and elasticities. Saez (2010b) studies bunching at kinks for individuals with ability n, consumption c, earnings z, earnings elasticity ϵ, and marginal tax rate t, the quasi-linear utility function $u(c,z) = c - \frac{n}{1+1/e}\left(\frac{z}{n}\right)^{1+1/\epsilon}$ and linear budget constraint $c = (1-t)z + R$. To this framework we can easily introduce a notch by assuming a discrete cost ϕ is incurred when income exceeds a threshold ρ. Then the budget constraint becomes $c = (1-t)z + R - \phi \cdot \mathbf{1}_{\{z>\rho\}}$.[7] I create a hypothetical notch with $\rho = \$40,000$ and $\phi = \$1000$. I assume that $e = 0$ for half of all households (e.g., due to optimization frictions), and for the other half of households $e \sim Unif(0,1)$.

3 Benchmark Static Bunching Estimation

Before presenting the dynamic bunching research design I follow the static approach used in the literature. I describe the technique, discuss some concerns for identification, and evaluate these with simulations. Throughout this section and the rest of the paper I will refer to bunching as occurring in "income," as it does in both datasets I analyze, as short-hand for "the variable in which costs change discretely at a threshold value." I will also minimize repetition by describing the notch as inducing a cost that increases when income lies above the threshold, leaving implicit the straightforward adjustments for settings in which costs decrease at the threshold.

[5] The flexible function includes the tangent of w_i (transformed to have domain $[0,1]$), which has the approximate shape of the inverse cdf. I include a cubic function of $\tan(u_i)$ as well as the interaction of this cubic function with a trend break where $g_i = 0$, which allows the pdf of g_i to have a non-differentiable peak at zero, as appears in the raw data.

[6] The flexible function includes the tangent of v_i (transformed to have domain $[0,1]$), the level of v_i, the level of g_i, and the log of g_i. I include a quintic function of each v_i variable, a cubic function of each g_i variable, and all interactions of v_i variables with g_i variables.

[7] Unless bunching, an individual chooses income of $z^* = n(1-t)^\epsilon$. The utility obtained from choosing $z = z^*$ is $u(c(z^*), z^*) = n(1-t)^{1+\epsilon} + R - \phi \cdot \mathbf{1}_{\{z^*>\rho\}} - \frac{n}{1+1/\epsilon}(1-t)^{1+\epsilon}$. The utility obtained from choosing $z = \rho$ is $u(c(\rho), \rho) = (1-t)\rho + R - \frac{n}{1+1/\epsilon}\left(\frac{\rho}{n}\right)^{1+1/\epsilon}$. Individuals will bunch if $z^* > \rho$ and $u(c(z^*), z^*) > u(c(\rho), \rho)$.

3.1 Description of Method

Bunching empirics can exploit distortions in distributions around thresholds at which income or prices change discretely. By estimating the excess mass around a threshold one can obtain reduced-form estimates of policy-relevant behavioral elasticities. Saez (2010a) introduced this insight by showing how kinks in marginal tax rates produce a pattern of bunching in the income distribution that reveals the taxable income elasticity without the need to specify a particular utility function. Individuals with incomes above a kink that raises the marginal tax rate have an incentive to reduce reported income, and the greater the income elasticity, the more bunching will be observed in the distribution around the kink. Similarly, those facing a notch at which total expenses increase discontinuously have an incentive to stay below the notch. Bunching estimation, both at kinks and at notches, quantifies the extent of bunching by comparing the observed distribution to an estimate of the smooth counterfactual that would be expected in the absence of the threshold.

The key to bunching estimation is to construct the counterfactual distribution of income. Static bunching estimates use parts of the density above and below a threshold to construct a counterfactual for the amount of mass that should be at the threshold. Most studies approach the distribution as a histogram, constructing bins and plotting the count of observations in each bin. The number of observations within some number of bins of the threshold is compared to a counterfactual constructed using bins further away from the threshold. That is, the researcher estimates the counterfactual density by omitting a certain number of bins around the threshold (the "omitted region") and then estimating a smooth function through the values of the other bins. For a notch at which costs increase, there should be excess mass in the "bunching range" below the notch and reduced mass in the "reduced range" above the notch.[8] Both the excess mass and reduced mass can be estimated by comparing the observed density to the counterfactual.[9]

Consider a notch that imposes a discrete cost on agents with income greater than ρ. Let the data on agents i with income z_i be collapsed into income bins of width ω and maximum value bin_b, giving bin counts $c_b = \sum_i \mathbf{1}\left[z_i \in (bin_b - \omega, bin_b]\right]$. The standard estimating equation for bunching at a notch is

$$c_b = \sum_{h=0}^{n_E - 1} \beta_{E,j} \cdot \mathbf{1}\left[bin_b = \rho - \omega h\right] + \sum_{j=1}^{n_R} \beta_{R,j} \cdot \mathbf{1}\left[bin_b = \rho + \omega h\right] + \sum_{k=0}^{K} \alpha_k bin_b^k + e_i \tag{1}$$

where n_E and n_R are the numbers of bins on either side of the threshold that are omitted from estimating the counterfactual. The counterfactual is a polynomial of order K with parameters α_k. This equation provides

[8] In the literature this is typically referred to as the "missing range." I use the term "reduced" to distinguish between this response and the mass that is missing because observations leave the data for reasons potentially including extensive-margin responses to the notch.

[9] In the case of a kink, agents may bunch on either side of the threshold, so the bunching estimate is the sum of all excess mass observed in the omitted region. Estimation of bunching at kinks requires the further step of raising or lowering the counterfactual distribution on the side of the kink where prices are affected, but this step is not necessary for a notch at which fixed costs are imposed because other than income effects there should be no responses far from the threshold.

two measures of bunching: $\sum_{h=0}^{n_E-1} \beta_{E,j}$ gives the excess mass of bunchers just below the notch, and $\sum_{j=1}^{n_R} \beta_{R,j}$ gives the reduction of mass just above the notch.

Visual examples of the standard approach are provided in Figure 2 for the charity and PSID datasets, respectively. In both figures, a polynomial is fit to the counts of income bin to estimate the counterfactual distribution, and mass is elevated in the bunching range just below the notch and reduced in the reduced range above the notch. Assuming no extensive-margin responses, the excess mass in the bunching range should equal the reduction of mass in the reduced range, and hence either of these can be used as the measure of bunching. Often this mass is divided by the counterfactual level of the density at the notch, and this bunching ratio gives an estimate of the amount by which the average buncher reduces income (Kleven and Waseem 2013).

3.2 Identification Challenges

Extensive-margin responses present one concern for static bunching estimation. The standard approach relies on the identifying assumption that in the absence of the bunching response there would be a smooth, continuous distribution of income in the neighborhood of the threshold of interest. If agents who would be above a notch are instead missing from the data, then this will reduce the mass found above the notch even if there is no bunching.

Some bunching papers have considered the nature of extensive-margin responses and how they might be detected or addressed. A point made by Kleven and Waseem (2013) and generalized by Best and Kleven (2018) is that there should be no such responses just above the notch when agents can simply bunch at very little cost relative to their preferred income level. This condition may not hold if frictions prevent bunching, as studied in much of the literature, or if exceeding the threshold causes attrition from the sample for reasons outside of the agents' control, such as an increased probability of being audited and removed from the sample. Kopczuk and Munroe (2015) note that even when the result holds in the limit, extensive-margin responses may still occur close enough to the notch to affect estimation, and that it is not possible (with the standard bunching design) to measure extensive-margin responses without making strong assumptions. They suggest using only the data below the notch to estimate bunching and testing for the existence of extensive-margin responses by comparing the excess mass below to the reduced mass above. In this paper I measure the degree to which extensive-margin responses affect estimation, and I propose the first tool for estimating their extent.

An issue that has not received attention in the literature is serial dependence in income. If income in one year depends positively on income in the prior year, then bunching in one year lowers an agent's income in later years. This distortion can affect incomes far from a notch and therefore violate the identifying assumptions of static bunching estimation. In Appendix A, I mathematically demonstrate the issue for discrete and discrete-continuous distributions. Here I provide a visual example using the first ten years of earnings from each household in the PSID.[10] To capture a situation without serially-dependent income I simply impose bunching at the hypothetical $40,000 notch in each year. To capture serial dependence I construct another version of the panel in which I impose bunching in each year before applying the growth rates in the data to obtain potential income for the next year.

Figure 1 displays the distribution of income at different points in time. Panel A shows incomes at the 10-year horizon with and without serial dependence. With no serial dependence the distribution is, by construction, consistent with the expected features of bunching: there is a spike below the threshold, a trough above, and no distortion further from the threshold. The distribution with serial dependence also appears to have these features, indicating that it may not be obvious whether or not an empirical distribution has been influenced by serial dependence. By comparing this distribution to that without serial dependence, however, we can see that the mass is relatively high everywhere below the threshold and relatively low above it. This suggests the first diagnostic for researchers to evaluate, which is a "donut-RD" regression discontinuity design that excludes the bunching region and tests whether the estimated counterfactuals meet at the threshold. In the charity setting, annual RD estimates (not shown) reveal that the discontinuity in the distribution of income at the notch has grown steadily over most of the sample period, and annual bunching estimates fail to uncover this pattern. In Figure 1, Panel B shows how the difference between the two distributions emerges over time, with large differences emerging within five years in this example. Going beyond this example, the nature of this evolution will vary with the underlying dynamics.

For clarity, I will abstract from a variety of other potential complications with bunching estimation. For example, this paper focuses on empirics and will not discuss the challenges of interpreting bunching estimates in terms of the parameters of decision models (Einav et al. 2017, Blomquist and Newey 2017, Bertanha et al. 2018). For the sake of clarity, I have eschewed some adjustments that have been made to basic bunching estimation. For one, researchers have sometimes attempted to address complications such as income effects by estimating the counterfactual separately on either side of the threshold or adjusting the counterfactual above the threshold so as to equalize the excess- and reduced-mass estimates (e.g., Chetty et al. 2011). For another, the mass in a strictly dominated region just above the notch can be used to estimate the number of agents facing frictions that prevent bunching (Kleven and Waseem 2013). Optimization frictions have

[10] I make 300 copies of each household to increase the sample size and multiply initial income by a random number to smooth its distribution. The random number was drawn from a gamma distribution with shape parameter equal to 0.3 and scale parameter equal to 20. This gamma distribution has positive support and gives a base-year median household income of $45,070.04, which is close to the U.S. median for 2013. I retained the original growth rates and multiplied these by the smoothed base-year income to obtain income levels in later years.

garnered increasing attention in the literature, and the methods proposed in this paper might find useful application in studying the nature of such frictions in both the short and long run. These adjustments can easily be incorporated into the dynamic designs in this paper, but I present versions without them to focus attention on the identification issues addressed by dynamic methods and the new parameters that these methods allow researchers to estimate.

3.3 Simulation

To examine the severity of this problem posed by extensive-margin responses I take 10,000 draws of 100,000 observations from the smoothed distribution of incomes in the PSID. I create a hypothetical notch at income of \$40,000 and impose that no households bunch but that a fraction whose incomes fall above the notch are not observed. These missing observations represent responses on the extensive margin, and I vary the share of responding households between 0 and 10 percent. I estimate the average income foregone by bunchers using the standard, static approach. The response can be calculated using the estimates of either the excess mass below the threshold or the reduction in mass above the threshold, which should be equal absent extensive-margin responses, and I report the estimates from both approaches. Because the true amount of bunching is zero, the average estimate gives the bias, and the percentage of statistically significant results gives the coverage. I also report the root mean squared error (root-MSE).

Table 1 shows the effect of extensive-margin responses on static bunching estimates. Estimates using the reduced mass above the notch show small biases, as well as coverage rates close to 0.05, when 2 percent of the sample or less responds on the extensive margin. When these responses are more prevalent, the reduced-mass estimates falsely attribute them to bunching and reject too frequently. The pattern is similar for the excess-mass estimates, though these show inflated rejection rates even for small shares of extensive-margin responders. This may be surprising, given that the extensive-margin response occurs entirely above the notch and not in the bunching range. Even so, these estimates are biased because the reduced mass above the omitted range lowers the counterfactual mass within the omitted range, including below the notch.

Table 2 quantifies the bias in the static estimates arising from serial dependence of income. I impose bunching in the base year but estimate bunching in the next year when there is no bunching. I vary the degree of serial dependence by applying empirical growth rates to a weighted average of the preceding potential income and actual income. When the weight η on potential income is zero there is perfect serial dependence, and when the weight is one there is no serial dependence. I also consider weights larger than one, which could arise if bunching simply represents retiming of income into the future. When income exhibits no serial dependence ($\eta = 1$), static estimates have coverage rates close to 0.05 and relatively small bias. With positive serial dependence ($\eta < 0$), both estimates show a positive bias, and coverage rates rise rapidly. With negative serial dependence ($\eta > 0$), the bias in the excess-mass estimates becomes negative, while coverage rates are again high for both estimates. Based on the graphical evidence in Figure 1, these biases would likely grow over time.

3.4 Application

Static bunching estimates for charities appear in Table 3.[11] I use the sample of charities in years up to 2007 (before the notch was moved) that also appear in the prior year (for maximum comparability with the dynamic estimates that follow). The first row of the table shows estimates of excess mass below the notch. An estimate of .1 would indicate .1 percent of all charities in that year's sample are below the notch and should be above it. The results from the basic specification indicate that the share of charities appearing below the notch is .148 percentage points greater than predicted by the counterfactual. In the second row this number is divided by the value of the density at the notch to give the bunching ratio, which can be interpreted as the average amount by which agents are willing to reduce income to bunch.[12] The bunching ratio is reported with the density in log scale, and it indicates that the number of bunching charities is roughly equal to the number of charities that should be above the notch by up to $600 (=$100,000*.00592). If all income responses are real (and not simply reporting responses), then this estimate would imply the average charity is willing to pay $600 to file Form 990-EZ instead of Form 990. The third row displays the estimated reduction of mass above the notch, which is nearly 70 percent larger than the estimate of the excess mass below.

The estimates in Table 3 raise the question of why the reduction in the number of charities above the notch is significantly larger than the increase in number below the notch. The "Basic" specification suggests the excess is only about 60 percent of the reduction, and the size of the reduction suggests charities may be willing to pay as much as $1000 to avoid Form 990. The second and third columns present the results of more flexible specifications motivated by regression discontinuity designs, but these do not reconcile the two results. The ("Discontinuous") specification in the second column allows for a discontinuity at the notch. This reduces the estimate on both sides by a very small amount, leaving the asymmetry in the estimates. The ("Two-Sided") specification in the third column estimates separate polynomials on each side of the notch. This gives a negative point estimate for the reduced mass, which would imply bunching in the range where mass should be reduced, and again failing to reconcile excess- and reduced-mass estimates.

[11] I use charities with receipts of $50-200,000 and a polynomial of degree 3, which minimizes the Akaike information criterion.

[12] With a homogeneous elasticity, there will be no mass in the missing range. With no extensive-margin responses, the reduced mass in this "hole" will equal the excess mass below the notch. With a constant density, the width of the hole will equal its mass divided by its height, which is the bunching ratio. Kleven and Waseem (2013) show that this interpretation generalizes to cases with heterogeneous elasticities.

These findings suggest the comparison of excess-mass estimates and reduced-mass estimates as a diagnostic. As noted above, Kopczuk and Munroe (2015) use this diagnostic to test for extensive-margin responses, then focus on the one-sided estimate of the excess mass. In the charity setting, the excess-mass estimate is more robust to the choice of specification in Table 3. However, as the simulations have shown, it may still be biased by serial dependence. I will provide further evidence of this by comparing the static estimate to the dynamic MLE estimate in Section 6. In settings for which a lack of panel data would make dynamic methods impossible, researchers can still compare the static excess- and reduced-mass estimates. Researchers should also be cautious about constraining estimation so as to require that the excess mass equal the reduced mass, as this equality need not hold in the presence of extensive-margin responses or serial dependence.

It should also be noted that the literature has done little to exploit panel data for heterogeneity analysis. It is straightforward to split a sample by pre-determined categorical variables, such as gender, and obtain static estimates for each subsample. It is less clear how to incorporate continuous and time-varying covariates into the static design. Some researchers have estimated the share of agents who remain in a bin over time, but slower growth by those in bunching bins could reflect either negative selection into current bunching or a greater likelihood of future bunching. A more dynamic approach is required to assess causation and to more fully exploit panel data to learn about heterogeneity in responses to the notch.

4 Dynamic Bunching Estimation: Temporary Notches

I now propose and evaluate designs that exploit panel data. I present these in order of abstraction from the static design. The design in this section makes a small adjustment that provides an alternative estimate of responses in the year of bunching. With panel data and a temporary notch, it also yields information about the dynamic effects of bunching. The design offers a test for serial dependence, which may provide evidence about whether bunching is driven by real income responses or misreporting.

4.1 Description of Method

The small adjustment I propose to the standard bunching estimate is to widen the bins to identify an unselected "treatment group" of agents with incomes *near* the notch. If the response to the notch is only local, then it will be possible to identify a range around the notch that includes all responders, i.e. that includes both the bunching range and reduced range. If one takes the entire omitted range as a bin, rather than using smaller bins on either side of the notch, then the sample within the bin is not selected because it contains both agents who respond and agents who don't.

Here, the observation i is now an individual agent, rather than a bin. Agent i has realized income r_i

measured in logs, so that estimates can be interpreted as percentage changes.[13] Let the bunching region and reduced region have respective widths ω_E and ω_R, and bins will now have width $\omega_E + \omega_R$. To simplify notation, rather than denoting the maximum value in the bin, bin_i will now be defined by the bin straddling the notch, with $r_i \in (\rho - \omega_E, \rho + \omega_R] \Leftrightarrow bin_i = \rho \Leftrightarrow NearNotch_i = 1$. The following equation can then be estimated with either cross-sectional or panel data.

$$r_i = \beta \cdot \mathbf{1}\left[bin_i = \rho\right] + \sum_{k=0}^{K} \alpha_k bin_i^k + e_i \tag{2}$$

As in Equation 1, the parameters α_k describe the counterfactual distribution using a polynomial of order K. The parameter β gives the average deviation from this counterfactual of the observations in the bin that surrounds the notch.[14]

Identification is similar but not identical to that in the standard bunching design. Both rely on an identifying assumption that a polynomial can approximate a counterfactual function of bin level. Here, that function describes conditional mean income of agents in the bin, rather than the count of agents in the bin. Neither design's assumption implies that of the other.[15] Thus, researchers may wish to test robustness by estimating bunching using both Equation 1 and Equation 2. Moreover, this approach becomes particularly useful with panel data and with a temporary notch that only exists in one period. Panel data make it possible to replace the outcome of Equation 2 with later years' values of income (or other variables). A temporary notch simplifies interpretation because these later outcomes should not be affected by whether notch-year income was close to the notch unless the induced bunching in that year had persistent effects on these later outcomes. That is, if there is only a notch in the base year, then there should be no active bunching in the next year. A temporary notch also makes identification more credible because the treated group of agents with income near the notch should not be a selected sample, whereas a long-standing notch might have accumulated a selected sample of long-term bunchers. With panel data, researchers can test exogeneity of $1\left[bin_i = \rho\right]$ by estimating "effects" on outcomes determined before the year of the notch.

Figure 3 demonstrates how the approach provides an alternative estimate of bunching. In each panel, observations are binned by log income relative to the threshold in the base year, when the notch is imposed. As with the standard approach, I use data generated from the smoothed distribution of incomes in the PSID

[13] Static estimation can be performed with log income; this transformation does not account for the differences in results between static and dynamic estimates.

[14] Diamond and Persson (2016) use a similar design to estimate long-run effects of test score manipulation by teachers. They allow for a discontinuity in the counterfactual density; for simplicity and comparability with the standard estimator I assume a continuous counterfactual density.

[15] As a counterexample in one direction, suppose that income is uniformly distributed in all bins and that the bin count takes one value below the notch and another value above it. As a counterexample in the other direction, suppose the the bin count is constant but that below the notch all agents are at the minimum value of their bins and above the notch they all take the maximum value of their bins.

and generate bunching at a notch at an income threshold of $40,000. Unlike the figures employed for the standard approach, which plot the counts in each of the bins, this figure displays conditional means. Panel A displays mean log income relative to a "bin threshold" that corresponds to the location of the notch in the treatment bin. Here the bin width is 0.2 and width of the bunching range is 0.05, so the bin threshold for each bin is a log income of 0.05 above the minimum log income in the bin.[16] Panel A shows that bunching in the (red) treatment bin has reduced its mean log income relative to the counterfactual mean predicted by surrounding bins. Changing the outcome of the regression to an indicator for having income above the bin threshold provides an estimate of the share of bunchers in the bin. The ratio of these two estimates should equal the average amount by which bunchers reduce income, the usual quantity of interest in bunching estimation.

The great advantage of this approach lies in the ability to estimate effects on future outcomes occurring after the notch is removed. Panel B of Figure 3 plots income in the year after the notch. Importantly, outcomes are again examined as a function of base-year income, which should not be influenced by the notch except in the treatment bin. For this illustration I have assumed perfect serial dependence, which implies that base-year bunchers' incomes should be reduced one-for-one in the next year. Panel B of Figure 3 shows that this is the case, with next-year incomes reduced in the treatment bin by about the same amount that they were in the base year. Because having income *near* the notch in the base year was not endogenous to the notch, these figures show the *causal* effect of bunching on later income. Moreover, the treatment bin should have no effect on next-year income except through base-year income and can therefore be used as an instrument to estimate the degree of serial dependence in income. Additional outcomes can include functions of next-year income, including indicator variables for having income above the base-year bin threshold, above either of the minimum or maximum income levels in the bin, or in between the two.

4.2 Simulation

I perform two types of simulation to evaluate the temporary-notch design. First, I estimate bunching in base-year distributions in which there is no bunching. Second, I generate bunching in base-year distributions and estimate next-year outcomes, which I do for several degrees of serial dependence. For comparability with the corresponding static estimates I again use the smoothed distribution of annual taxable incomes from the PSID and draw 10,000 samples of size 100,000 for each of these simulations.

Table 4 quantifies the bias in the temporary-notch base-year estimates. The truth in this simulation is that there is no bunching, and hence all estimates should be close to zero. The results are as desired;

[16]The normalization of log income by subtracting the bin threshold level is irrelevant for estimation but makes the figure easier to read; one could alternatively plot mean log income relative to the minimum value in the bin.

on average, I estimate an income reduction of 0.017 percent, and the share of households in the treatment bin that bunch is 0.00033. The bunching ratio, which is expressed in thousands, indicates that the average household in the treatment bin reduces income by $6, in contrast with the static approach estimates in Table 2 that approach $1000 on average in some simulations. Coverage rates are also appropriate with roughly five percent of the estimates reject the true null of no bunching.

Table 5 displays the accuracy of the estimator under various degrees of serial dependence. Each cell of the table gives the average of the estimates for a particular outcome and simulation. The standard error in parentheses is the standard deviation of the estimates. As in Table 2 for the standard approach, I vary the weight of potential income relative to observed income. A weight of 0 indicates perfect serial dependence, 1 indicates no serial dependence, and 2 indicates that bunching households' income rebounds in the next-year, as it would if base-year income was retimed. This weight is unrelated to the amount of bunching in the base year, and hence all base-year estimates of the reduction in income and share of households above the notch are similar. Moving to next-year outcomes, there should be no significant effects if there is no serial dependence, and this is indeed what I find when the weight on potential income is 1. With positive serial dependence (weight on potential income less than one), bunching has negative effects on next-year income. Conversely, with negative serial dependence (weight on potential income greater than one), bunching has positive effects on next-year income. Using the treatment-bin dummy as an instrument for base-year income provides an estimate of its effect on next-year income, and one minus this coefficient gives an estimate of the weight on potential income. The estimated weights are all slightly greater than the true weights, but they are all within one standard error of the true values.

The simulations provide evidence that the temporary-notch estimator has good properties. This approach can be used as an alternative to the standard approach in the cross-section, and with panel data it can be used to estimate dynamic effects.

4.3 Application

Variation in the IRS reporting notch for public charities makes it possible to estimate effects of a temporary notch. After decades with a notch at a nominal value of $100,000, the notch was moved to $1,000,000 for 2008, $500,000 for 2009, and $200,000 thereafter. The 2008 and 2009 notches were therefore only temporary (one-year) phenomena that should not have induced either repeated bunching or manipulation at incomes far from the level of the notch. I exploit this temporary nature, focusing on the 2009 notch, which fell at an income level affecting many more charities than the 2008 notch.

I estimate Equation 2 using 2009 log receipts as the binning variable.[17] I examine three different functions

[17]The results depicted in this section are obtained with a simple quadratic function of bin level, a range starting at receipts

of income as outcomes. The first is simply income (log gross receipts), which is expected to be negatively affected in 2009 by the opportunity to bunch. The second, labeled "Cross 2009 Threshold," is an indicator for whether an organization's current receipts are above the level in its bin that correspond to the notch.[18] Bunching should also have a negative effect on the probability that treated observations cross the 2009 threshold in 2009. Finally, I construct an indicator for current receipts that lie within the same bin that the charity occupies in 2009. This dummy takes the value of one for all observations in 2009 but provides another useful measure of pre-trends or long-term effects. Marx (2018) presents figures akin to Figure 3, and these provide visual support for the continuity assumption.

Results of the temporary-notch regressions for charities appear in Table 6. Each cell of the table reports the estimate of the parameter β for a different outcome and year. In the first row the outcome is log receipts, which is negatively impacted in 2009, as expected. The coefficient for this year implies that the average charity near the notch reduces log income by .003, i.e. reducing income by roughly \$1500 (\$500,000*.003). Estimated deviations from expectation are negative for subsequent years, but standard errors are large because in this setting (as will also be seen for the PSID), the distribution of year-over-year income growth has fat tails. The alternative income measures, which focus on more central moments, can therefore offer evidence that is more precisely estimated. In the second row, one can see that the probability of being above the notch in 2009 is reduced by 10.1 percentage points among the treatment group. These charities are not significantly different in this regard prior to 2009, but bunching in 2009 appears to cause a permanent 2 percentage-point reduction in the probability that treated charities ever achieve receipts greater than \$500,000. Similarly, the third row shows that while these charities were no more or less likely to be in their year-2009 income bin in years before 2009, the probability that they remain in this bin (rather than growing out of it) is permanently increased by at least .4 percentage points.

Permanent effects of a temporary notch provide evidence of positive serial dependence of income. Marx (2018) discusses the apparent implication that manipulation in the charity setting was not entirely carried out by misreporting income. Here I will emphasize the implication, taking into account the simulation results in Section 3.3, that there is likely to be bias in static estimates of bunching at the repeated notch that was in place for all sample years before 2008. That the static estimates are biased is consistent with the evidence that follows.

---

greater than \$100,000 to avoid lingering effects of the original notch, and log receipts bins of width .155, the width of the smallest omitted range that provides robust static bunching estimates. This approach is still potentially susceptible to extensive-margin responses, but for the temporary notch of 2009 this does appear to be a problem: the static estimate of excess mass below the notch is not statistically different from the static reduced-mass estimate or from the dynamic estimate obtained from the methodology described in Section 6, and the dynamic estimate of extensive-margin responses is not significantly different from zero.

[18]This dummy variable indicates that the charity has receipts greater than .065 plus the level of the minimum income in the bin that it occupies in 2009, since the minimum income in the treatment bin is .065 log points below the notch.

5 Dynamic Bunching Estimation: Ordinary Least Squares Estimation

This section describes another easily-implemented dynamic extension of the ordinary-least-squares binning approach to bunching estimation. This extension can be applied to settings in which a notch is faced repeatedly. Serial dependence is addressed by conditioning on past income.

5.1 Description of Method

Consider agents observed in a base year and the next year, with a notch existing in the next year and possibly the base year as well. Figure 4 depicts such a situation using the smoothed PSID distribution. Panel A of the figure shows the distribution of log income in the next year for two illustrative levels of base-year income. Because modal growth is roughly zero, each conditional distribution of next-year income is centered around the level of current income. For each group, the distribution of next-year income is distorted around the notch as one would expect, with excess mass just to the left and reduced mass just to the right. Panel B shows the distribution of growth rates for each group. Because income is logged, the growth rates are a simple translation of the group's next-year income that subtracts base-year income.

Panel B of Figure 4 conveys the intuition for a dynamic bunching estimation identification strategy. Households with different levels of base-year income have similarly-shaped growth distributions, except that each has a bunching distortion at growth rates that would take the household close to the notch. Because the depicted groups are starting from different income levels, each approaches the notch at a different level of growth, and hence the distortions lie in different parts of the two groups' growth distributions. The extent of the distortions can therefore be estimated by comparing the shape of one group's growth distribution around its notch to the corresponding, undistorted section of the growth distribution among households in the other group. This insight can then be extended from two levels of base-year income to all levels, with each providing a counterfactual distribution for the rest.

The dynamic bunching identification strategy can be implemented in a variety of ways. In theory, one could estimate the entire multivariate density of income in all available years, but allowing for such generality would be computationally expensive. I propose two implementations that focus on pairs of base-year log income (r_{it}) and growth to next-year income ($g_{it} = r_{it+1} - r_{it}$). Section 6 presents an implementation using MLE, while this Section adheres to a faster and more familiar approach of binning and using OLS. The binning approach provides transparent evidence of whether agents manipulate income when approaching a notch. If these income responses are local to the notch, then it is possible to construct treatment groups and control groups for a permanent notch in much the same way as for a temporary notch, as described in Section 4. Also like the temporary-notch approach, outcomes of interest will be means of income or growth conditional on falling within each bin.

To develop intuition before considering the full OLS implementation, consider restricting a dataset to agents' whose growth rate from any base year to the next year falls within a particular range. As an illustration using the empirical application, Figure 5 plots the mean growth rate by bins of base-year income. The sample includes all observations of charities growing log income by .1 to .2 between the base year and the next year. For some bins of base-year income, this range of growth implies next-year income that is near the notch but mostly or entirely to one side of it; these bins may therefore include a selected sample of bunchers or non-bunchers, and hence they are depicted with light gray markers and excluded from estimation. However, for the bin with base-year income between -0.18 and $-.013$, growth of .1 to .2 translates to a range of future income that straddles the notch. Thus, agents in this bin should provide a sample that is not selected based on whether an agent would bunch. This bin can therefore be labeled as the group treated with the opportunity to bunch, as in the temporary-notch estimation. It is represented in the figure by the filled circle with standard error bands. Empty circles display the growth rates of charities with higher or lower base-year income, and the curve with standard error bands illustrates a quadratic fit to these control bins. The interpolated counterfactual implies that average growth, conditional on growth in the range of .1 to .2, should be nearly .147 for charities nearing the notch, but the estimate for this group is instead less than .145. The difference provides an estimate of the effect of nearing the notch on observed income.

A more general design incorporates multiple ranges of growth rates by stacking estimates for each of γ different growth-rate bins of width ω_g and maximum value labeled $gbin_{it}$ of equal width. Again, denote by r_{it} agent i's log income in base year t, and label the growth rate to the next year's income $g_{it} = r_{it+1} - r_{it}$. Bins of base-year income can be selected with width ω_r and maximum value labeled $rbin_{it}$. To estimate the effect of nearing the notch on outcome Y_{it+1}, I propose estimating equations of the form

$$Y_{it+1} = \beta \cdot NearNotch_{it} + \sum_{k=0}^{K} \sum_{\gamma} \alpha_{k\gamma} r_{it}^{k} \cdot \mathbf{1}\left[gbin_{it} = \gamma\right]. \qquad (3)$$

Here, the "treatment" variable is $NearNotch_{it} = \mathbf{1}\left[rbin_{it} - \omega_r + gbin_{it} - \omega_g < \rho_{t+1} < rbin_{it} + gbin_{it}\right]$, which indicates pairs of income and growth-rate bins that produce a range of next-year income that straddles the notch ρ_{t+1}. As in previous methods, the α coefficients describe the counterfactual, which now varies in two dimensions by allowing for a separate polynomial of base-year income for each bin of growth. As in temporary-notch estimation, β gives the difference between the conditional mean of the outcome and what is predicted by the counterfactual. If the outcome is either g_{it} or r_{it+1}, then bunching would imply that

$\beta < 0$. Note that base-year characteristics should not be affected by nearing the notch in the next year, and hence interactions with these characteristics can be used to describe heterogeneity in the bunching response, even if these characteristics are continuous variables.

Appendix B provides additional details. As Figure 5 indicates, some care is required in constructing the treatment bins and choosing omitted bins. In the Appendix B propose the use of bin count as an outcome to test the exogeneity of the treatment bin. I also describe the construction of a useful outcome variable that provides a direct estimate of the share of agents who bunch.

5.2 Simulation

I evaluate the dynamic OLS approach with the same type of simulations used for the static and temporary-notch approaches. Using the smoothed PSID income distribution, I generate 10,000 random samples of 100,000 observations in each. I induce bunching at the hypothetical notch at base-year income of $40,000, then apply an observation's growth rate to its (potentially bunching) base-year income to obtain next-year income. I then estimate bunching in next-year income, for which the true value is zero, and examine the bias, mean squared error, and coverage rates.[19]

Table 7 presents the results of the dynamic OLS simulation. The first row presents results for log income, which would be reduced in the treated bins approaching the notch if households in these bins chose to bunch. The coverage rate is close to 0.05, and the bias and root-MSE are very small. These magnitudes can be evaluated more easily for levels than for logs because the static estimates offer a comparison in levels. The second row of the table transforms the estimates to levels by exponentiating the direct estimates, which changes the bias and root-MSE but not the coverage rage. The relevant comparison among the simulations of static estimation is the top row of Table 2, in which the setup is identical because serial dependence is perfect. Compared to the dynamic OLS estimates, the excess-mass static estimates (which performed better than the reduced-mass estimates) have a coverage rate this is three times larger, root-MSE that is nearly 8 times larger, and absolute bias that is 2000 times larger. Moreover, the dynamic approach provides a direct estimate of the effect of approaching the threshold on the probability of crossing it. The last row of Table 7 shows that for this outcome the absolute bias is half of one percentage point, and the coverage rate is less than 0.06.

[19] The estimation uses base-year log-income bins of width 0.05, growth rate bins of width 0.1, and an omitted range of base-year income of 10.55 to 10.65 (around the notch at $\log(40,000) \approx 10.6$).

5.3 Application

Dynamic OLS estimation can be performed with nearly the same ease as static estimation, and it can offer several improvements. For one, it can address the identification issue posed by serial dependence, as shown by the simulation. Another benefit is the opportunity to study heterogeneity in agents' responsiveness to the notch, particularly when this heterogeneity relates to time-varying characteristics. Such characteristics can be made time-invariant, such as by defining them as having their value in the base year, and then interacting them with the treatment-group dummy and potentially other variables. Using this approach, Marx (2018) estimates that the responses of charities to the reporting notch are related to both size and staffing, providing evidence describing the nature of the compliance cost and induced avoidance behavior.

Here I focus on the effect of the notch on long-run growth, which also cannot be estimated with the standard design. The last row of Table 7 showed simulation results for a binary outcome indicating growth exceeded that required to be above the notch at time $t+1$. Corresponding indicators can be defined for any horizon h to estimate effects of approaching the notch at time $t+1$ on crossing it at time $t+h$. [20] Table 8 displays the results of regressions with h varying from 1 to 12.[21] For charities in the bins approaching the notch in year $t+1$, the estimated counterfactual is that 40 percent should have income greater than the notch at $t+1$, and 75 percent should have income above the notch in year $t+10$. The estimates in the table suggest that bunching reduces these probabilities in both the short- and long-run. I find the largest effect, a 5.3 percentage point reduction, in the year that the notch is first approached. I then find effects of around 1.5 to 2 percentage points at all other horizons, implying that the notch permanently reduced the growth of a significant number of charities.

6 Dynamic Bunching Estimation: Maximum Likelihood Estimation

Dynamic OLS estimation is straightforward and provides a number of potential advantages over static OLS estimation. However, as in static bunching estimation, this design will not account for extensive-margin responses without additional assumptions. Moreover, the OLS estimates are unlikely to be efficient because binning the data treats observations within a bin as equivalent, and throughout the literature, the choices of bin widths and locations have been ad-hoc. To address these limitations, I now propose a maximum

[20] For the group with $NearNotch_{it} = 1$, this outcome indicates exceeding the level of growth that would be required to cross the notch at $t+1$. For those with $NearNotch_{it} = 0$ it indicates a corresponding growth rate depending on position in the bin of base-year income.

[21] Counterfactual estimated as a quadratic function of current receipts, controlling for growth-rate bins of width .1. The sample includes charities within one log point of the notch in any starting year from 1990 to 1997 and growing by 0 to 1 log points.

likelihood bunching estimator.

6.1 Description of Method

Dynamic bunching estimation with MLE retains the flexibility of OLS estimation. This is because identification is obtained by comparing parts of the joint distribution of income in multiple years, as was described in Section 5 and depicted in Figure 4. OLS estimation involves choosing a functional form (the polynomial) and then adding parameters as needed to flexibly fit the data. MLE estimation proceeds similarly, starting from a functional form that reflects the fact that the researcher is estimating a probability distribution.[22] As before, r_{it} is a agent i's log income in a base-year t, and growth to the next year is $g_{it} = r_{it+1} - r_{it}$. A researcher can first choose a functional form with parameters Θ for $F^* (g_{it}|r_{it}, \Theta)$, the latent cdf that would be observed if no observations were bunching or going missing from the data. Bunching and extensive-margin parameters Ω can then be incorporated to construct the distribution $F(g_{it}|r_{it}, \Theta, \Omega)$ that is fit to the data.

Laplace (or "double exponential") distributions offer a natural choice for the basic form of $F^* (g_{it}|r_{it}, \Theta)$. These distributions have been used extensively to model financial data and are described by Kotz et al. (2001) as "rapidly becoming distributions of first choice whenever 'something' with heavier than Gaussian tails is observed in the data."[23] I estimate a flexible version of the Laplace cdf by allowing for flexible functions $P_l(g, r, \Theta)$ and $P_u(g, r, \Theta)$ to enter the lower and upper pieces of the distribution:[24]

$$F^* (g_{it}|r_{it}, \Theta) = \begin{cases} \exp\left(P_l\left(g_{it}, r_{it}, \Theta\right)\right) & g_{it} < \theta \\ 1 - \exp\left(P_u\left(g_{it}, r_{it}, \Theta\right)\right) & g_{it} \geq \theta \end{cases}, \tag{4}$$

where θ is a location parameter that I set equal to zero based on the shapes of the distributions in both datasets. Appendix B motivates the focus on the cumulative distribution function and provides details of the specification and implementation, including derivation of parameter restrictions that for flexible choices of $P_l(g, r, \Theta)$ and $P_u(g, r, \Theta)$ while ensuring that the resulting function satisfies the properties of a probability distribution. R code for implementation is available online.

Figure 6 provides a visual example using draws from the smoothed PSID distribution. The figure focuses on illustrative base-year log income level of 0.2 less than the hypothetical notch at $\log(\$40,000)$. The empirical density of growth rates conditional on this level of receipts (and others) is non-Normal, with a sharper peak around zero growth and fatter tales, motivating the use of a Laplace distribution, as in

[22] Kopczuk and Munroe (2015) use MLE to estimate bunching within the static framework.

[23] See Kozubowski and Nadarajah (2010) for other applications.

[24] The symmetric Laplace distribution with location parameter θ and scale parameter σ has this form with $P_l(g, r, \Theta) = P_u(g, r, \Theta) = \left|\frac{g-\theta}{\sigma}\right| - \log(2)$.

Equation 4. It is also not well approximated by the basic, two-parameter Laplace distribution, requiring higher-order functions for $P_l(g_{it}, r_{it}, \Theta)$ and $P_u(g_{it}, r_{it}, \Theta)$. The estimated functions, listed with other details in the following Subsection on simulation, provide a counterfactual (dark curve) that approximates the PSID distribution (circular markers). The notch for this level of base-year income occurs at a growth rate of 0.2, and the omitted range is drawn around this. Using two other illustrative levels, the figure also displays how the counterfactual varies with base-year income.

Responses to the notch are estimated by maximizing the likelihood of the observed data according to a censored version, $F(g_{it}|r_{it}, \Theta, \Omega)$, of the latent distribution $F^*(g_{it}|r_{it}, \Theta)$. The added parameters Ω describe bunching and attrition. The censoring occurs in the omitted region, which varies with r_{it}, as depicted for one level of r_{it} in Figure 6. $F^*(g_{it}|r_{it}, \Theta)$ determines the probabilities, for a given level of r_{it}, that an observation will appear in its bunching region or reduced region. Outside of this region, one can use the pdf derived from $F^*(g_{it}|r_{it}, \Theta)$. Attritors can be assigned a specific, fill-in value of g_{it}, and then attrition can also be estimated with censoring. For example, in the simulation I treat all attritors as having $r_{it+1} = \log(10,000)$, with corresponding g_{it}, and then estimate a probability mass for this level of g_{it} that can vary with r_{it} and g_{it}. In this way, I capture extensive-margin responses of households whose next-year income would exceed the notch-level growth ρ_{it} if they were observed. I do so by incorporating a parameter that shifts mass from $(1 - F(\rho_{it}|r_{it}, \Theta))$ to $Pr(r_{it+1} = \log(10,000))$.

6.2 Simulation

The simulation of MLE dynamic bunching estimation follows the earlier simulations. Again, observations are generated from the bivariate PSID distribution. I draw 10,000 random samples of 10,000 observations, a smaller sample size because the time required to obtain each MLE estimate is many times that for the OLS estimates. As in previous simulations, I impose bunching in the base year but no bunching or notch-related extensive-margin responses in the next year. I randomly select observations to attrit at a rate of about four percent, matching the attrition rate among high-income households in the raw data. I estimate extensive-margin responses of households that should cross the notch (as in the application below). I also allow for attrition that is a flexible function of base-year income.[25]

Table 9 presents results from the simulation. I report estimates describing bunching and extensive-margin responses, both of which have a true value of zero. The top row describes estimates of the share of bunchers among the total number of households that should move to the reduced region. The second row reflects the

[25]Estimation uses base-year log incomes within 0.5 of the notch at $\log(40,000) \approx 10.6$, with omitted range of next-year log incomes of 10.5 to 10.9. The flexible functions $P_l(g_{it}, r_{it}, \Theta)$ and $P_u(g_{it}, r_{it}, \Theta)$ include a number of subfunctions of g_{it}, each multiplied by a quadratic function of base-year log income. These subfunctions are quintic, exponential, and the arc-tangents of each of 1, 2, and 3 times g_{it}. Attrition is estimated as a quartic and logarithmic function of r_{it}.

estimated effect on the income of these households, and the third row represents estimates of the percentage of those that should cross the notch that instead exit the sample. Despite the imperfect specification of the counterfactual, as seen in Figure 6, all of these estimates exhibit a bias close to zero and a coverage rate close to 5 percent.

6.3 Application

For the charity application, I estimate functions similar to those used for the simulation.[26] The latent distributions are similar to the PSID distributions in Figure 6, but attrition is more common in the charity data. Attrition could be due to late filing, earning income below the level at which filing is required, shutting down, merging, or simply failing to comply with the reporting requirement. I estimate types of attrition that vary with r_{it} and may or may not be systematically related to the notch.[27]

Table 10 displays maximum likelihood dynamic bunching estimates. The first column provides static bunching estimates for comparison with the dynamic bunching estimates in the second column. The top panel of Table 10 shows estimates of parameters governing bunching and systematic attrition. The first parameter gives the bunching propensity among charities that have base-year receipts below the notch. An estimated 2.6 percent of the charities that should enter the reduced range from below are instead bunching. The second row shows this bunching propensity is lower for those with current receipts above the notch (who have already filed Form 990 in the base year).

The third and fourth rows in the table show that attrition is systematically related to the notch. The estimated parameters indicate the share of charities that should be crossing the notch from below but instead go missing from the sample. This is estimated separately for those that would have entered the reduced range and those that would have grown to a point further above the notch. These estimates turn out to be similar and highly significant, revealing extensive-margin responses by 8 or 9 percent of charities that should cross the notch. Comparing the attrition and bunching propensities, a combined 10.6 percent of charities avoid filing when first crossing the notch to the reduced range, and the number of charities doing so by bunching is dwarfed by the number responding on the extensive margin. The static approach provides no such estimates of these extensive-margin responses.

The lower panel of Table 10 reveals the estimated excess share of charities below the notch and reduction

[26] The charity notch occurs at $\rho = \log(100,000) \approx 11.51$, and I include in the sample all observations with $r_{it} < 14$. The omitted range of next-year log incomes is $(\log(95,000), \log(130,000))$. The flexible functions $P_l(g_{it}, r_{it}, \Theta)$ and $P_u(g_{it}, r_{it}, \Theta)$ again include a number of subfunctions of g_{it}, each multiplied by a quadratic function of base-year log income. These subfunctions are linear, exponential, the exponent of the square, and the arc-tangent. The argument of the arctangent is also multiplied by a quadratic function of base-year log income. I adjust $F(g_{it}|r_{it}, \Theta, \Omega)$ to account for truncation of the sample due to the fact that agents with income below $25,000 do not have to report.

[27] Baseline attrition is estimated as a quadratic function of r_{it}. I allow this attrition rate to have additional slope and intercept terms for $r_{it} > \rho$. As explained in the description subsection for this method, I also estimate extensive-margin responses of charities that should cross from below the notch to either the reduced range or to even higher levels of income.

in the share above it. The excess and reduction are found by aggregating the bunching and attrition propensities across all observations in the base year according to their counterfactual probability of moving to the reduced region in the next year. The dynamic estimates of the share of charities that appear in the bunching range exceeds by .103 percentage points (or roughly 200 charities per year) the quantity that would have been found in the absence of a response. Consistent with the difference in estimated magnitudes of the propensities to bunch or leave the sample, the reduction of mass in the range just above the notch is significantly greater at .366 percentage points. Applying the static approach gives estimates of the excess and reduced mass that are much closer to each other, and both are significantly different from the dynamic estimates. Finally, the bunching ratio gives the ratio of excess bunching mass to the counterfactual density at the notch, here reported for the density in levels so that the ratio can be interpreted as a dollar amount. This ratio gives an estimate of average willingness to forego income to bunch, and as with the non-normalized excess-mass estimates, the result of static estimation appears to be biased upward by a factor of nearly two.

Finally, Figure 7 provides a depiction of the difference between the two estimation approaches. The figure plots the density of receipts in the next year and estimated counterfactual densities. The MLE counterfactual is what Kleven and Waseem (2013) refer to as "a 'partial' counterfactual stripped of intensive-margin responses only." Due to serial dependence of income and extensive-margin responses, the counterfactual is not smooth at the notch. That is, even if charities did not bunch in these years, the density would not be a smooth function of income as assumed in the static approach. The static approach estimates a smooth curve connecting the discontinuous pieces of the density, and as can be seen in the figure, this will underestimate the density below the notch and therefore overestimate the excess mass in the bunching region. Also, because the density is reduced everywhere above the notch, treating it as unaffected by the notch assumes away some of the true reduction above the reduced range and so underestimates the reduction in the reduced range. The figure shows that static bunching estimation conflates multiple responses, including repeated bunching and extensive-margin responses, and is not directly informative of the underlying behavior.

7 Conclusion

This paper proposes new tools for analyzing bunching and provides evidence on the behavior of charities. Dynamic bunching estimation offers an identification strategy based on assumptions that are likely to hold in more general settings than the assumptions required for static estimates. Simulations demonstrate the robustness of dynamic estimators to factors that bias the static estimator. The differences in estimates can be quite large, as seen in an application to bunching at a reporting threshold for charities. Dynamic estimation also provides new opportunities to describe behavior by estimating extensive-margin responses, preference heterogeneity, long-run effects of approaching a notch, and the effect of bunching in one year on income in subsequent years.

Data analysis suggests a number of diagnostics that researchers can use to test for potential bias in static bunching estimates. Researchers studying notches can, even with only a single cross section of data, perform a statistical test of whether the excess mass to one side of the notch equals the reduction in mass on the other side. This test can be generalized by estimating the counterfactual distributions separately on each side of the notch (Kopczuk and Munroe 2015). With repeated cross sections, the researcher can examine whether bunching or donut-RD estimates vary by year, and in particular whether mass accumulates or discontinuities grow over time. With panel data, the researcher can compare the static and dynamic OLS estimates to assess whether estimation with dynamic MLE appears warranted. Finally, with a temporary notch, the researcher can estimate the degree of serial dependence in income.

Dynamic bunching estimation can be used to learn from agents' responses to a variety of thresholds. Early bunching papers focused on tax and labor supply elasticities, but the design is being applied in a widening array of settings. The tools presented in this paper apply to notches, which may induce responses in such decisions as income choices around the eligibility limits of social welfare programs, housing choices around school district and political boundaries, commerce around tolls and security checkpoints, product pricing with quantity discounts, or charitable fundraising strategies that reward gifts above discrete size thresholds. Future work could extend these methods to estimating bunching at kinks. Across settings, dynamic bunching designs will offer researchers tools to take greater advantage of panel data.

References

Bertanha, Marinho, Andrew McCallum, and Nathan Seegert, "Better Bunching, Nicer Notching," 2018.

Best, Michael and Henrik Kleven, "Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK," *Review of Economic Studies*, 2018, *85*, 157–193.

Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem, "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan," *Journal of Political Economy*, 2015, *123* (6), 1311–55.

Blomquist, Sören and Whitney Newey, "The Bunching Estimator Cannot Identify the Taxable Income Elasticity," *NBER Working Paper No. 24136*, 2017.

Brehm, Margaret, Scott A. Imberman, and Michael F. Lovenheim, "Achievement Effects of Individual Performance Incentives in a Teacher Merit Pay Tournament," *Labour Economics*, September 2017, *44* (C), 133–150.

Brülhart, Marius, Jonathan Gruber, Matthias Krapf, and Kurt Schmidheiny, "Taxing Wealth: Evidence from Switzerland," *NBER Working Paper No. 22376*, 2016.

Burman, Leonard E., Norma B. Coe, Kevin Pierce, and Liu Tian, "The Effects of the Taxation of Social Security Benefits on Older Workers' Income and Claiming Decisions," *National Tax Journal*, 2014, *67(2)*, 459–486.

Camacho, Adriana and Emily Conover, "Manipulation of Social Program Eligibility," *American Economic Journal: Economic Policy*, 2014, *3(2)*, 41–65.

Chetty, Raj, John N. Friedman, and Emmanuel Saez, "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, 2013, *103(7)*, 2683–2721.

_ , _ , and _ , "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," *American Economic Review*, 2013, *103(7)*, 2683–2721.

_ , _ , Tore Olsen, and Luigi Pistaferri, "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *The Quarterly Journal of Economics*, 2011, *126*, 749–804.

Dee, Thomas S., Brian A. Jacob, Justin McCrary, and Jonah Rockoff, "Rules and Discretion in the Evaluation of Students and Schools: The Case of the New York Regents Examinations," 2015. Mimeo, Columbia Business School.

Devereux, Michael P., Li Liu, and Simon Loretz, "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records," *American Economic Journal: Economic Policy*, 2014, *6(2)*, 19–53.

Diamond, Rebecca and Petra Persson, "The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests," *NBER Working Papers No. 22207*, 2016.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf, "Bunching at the kink: Implications for spending responses to health insurance contracts," *Journal of Public Economics*, 2017, *146*, 27–40.

Gelber, Alexander, Damon Jones, and Dan Sacks, "Earnings Adjustment Frictions: Evidence from the Social Security Earnings Test," *NBER Working Paper No. 19491*, 2013. Wharton School of the University of Pennsylvania, mimeo.

Grubb, Michael D. and Matthew Osborne, "Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock," *American Economic Review*, 2015, *105(1)*, 234–271.

Ito, Koichiro, "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing," *American Economic Review*, 2014, *104(2)*, 537–563.

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

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

Kleven, Henrik Jacobsen and Esben Anton Schultz, "Estimating Taxable Income Responses Using Danish Tax Reforms," *American Economic Journal: Economic Policy*, 2014, *6(4)*, 271–301.

_ , Camille Landais, Emmanuel Saez, and Esben Schultz, "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark," *The Quarterly Journal of Economics*, 2014, *129(1)*, 333–378.

Kline, Patrick and Melissa Tartari, "Bounding the Labor Supply Responses to a Randomized Welfare Experiment: A Revealed Preference Approach," *American Economic Review*, January 2016, *106* (4), 972–1014.

Kopczuk, Wojciech and David Munroe, “Mansion Tax: The Effect of Transfer Taxes On the Residential Real Estate Market,” *American Economic Journal: Economic Policy*, 2015, *7* (2), 214–57.

Kotz, Samuel, Tomasz J. Kozubowski, and Krzysztof Podgórski, *The Laplace Distribution and Generalizations: A Revisit With Applications To Communications, Economics, Engineering, and Finance*, Birkhauser, 2001.

Kozubowski, Tomasz J. and Saralees Nadarajah, “Multitude of Laplace Distributions,” *Statistical Papers*, 2010, *51(1)*, 127–48.

le Maire, Daniel and Bertel Schjerning, “Tax Bunching, Income Shifting and Self-Employment,” *Journal of Public Economics*, 2013, *107*, 1–18.

Manoli, Dayanand and Andrea Weber, “Nonparametric Evidence of the Effects of Financial Incentives on Retirement Decisions,” *American Economic Journal: Economic Policy*, 2016, *8* (4), 160–182.

Marx, Benjamin M., “The Cost of Requiring Charities to Report Financial Information,” 2018. University of Illinois at Urbana-Champaign, mimeo.

Mortenson, Jacob and Andrew Whitten, “Bunching to Maximize Tax Credits: Evidence from Kinks in the U.S. Tax Schedule,” 2018.

Ramnath, Shanthi, “Taxpayers’ Response to Notches: Evidence from the Saver’s Credit,” *Journal of Public Economics*, 2013, *101*, 77–93.

Rees-Jones, Alex, “Quantifying Loss-Averse Tax Manipulation,” *Review of Economic Studies*, 2018, *85* (2), 1251–1278.

Roeger, Katie L., Amy Blackwood, and Sarah L. Pettijohn, “The Nonprofit Almanac 2012,” Technical Report, Urban Institute Press 2012.

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

_ , “Do Taxpayers Bunch at Kink Points?,” *American Economic Journal: Economic Policy*, 2010, *2(3)*, 180–212.

Sallee, James M., “The Surprising Incidence of Tax Credits for the Toyota Prius,” *American Economic Journal: Economic Policy*, 2011, *3(2)*, 189–219.

— **and Joel Slemrod**, "Car Notches: Strategic Automaker Responses to Fuel Economy Policy," *Journal of Public Economics*, 2012, *96(11)*, 981–999.

— **and Koichiro Ito**, "The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel Economy Standards," *Review of Economics and Statistics*, 2018, *100* (2), 319–36.

Schivardi, Fabiano and Roberto Torrini, "Identifying the Effects of Firing Restrictions Through Size-Contingent Differences in Regulation," *Labour Economics*, 2007, *15*, 482–511.

Seim, David, "Behavioral Responses to Wealth Taxes: Evidence from Sweden," *American Economic Journal: Economic Policy*, 2017.

Slemrod, Joel, "Buenas Notches: Lines and Notches in Tax System Design," 2010. University of Michigan, mimeo.

St.Clair, Travis, "How Do Nonprofits Respond to Regulatory Thresholds: Evidence From New York's Audit Requirements," *Journal of Policy Analysis and Management*, 2016, *35*, 772–790.

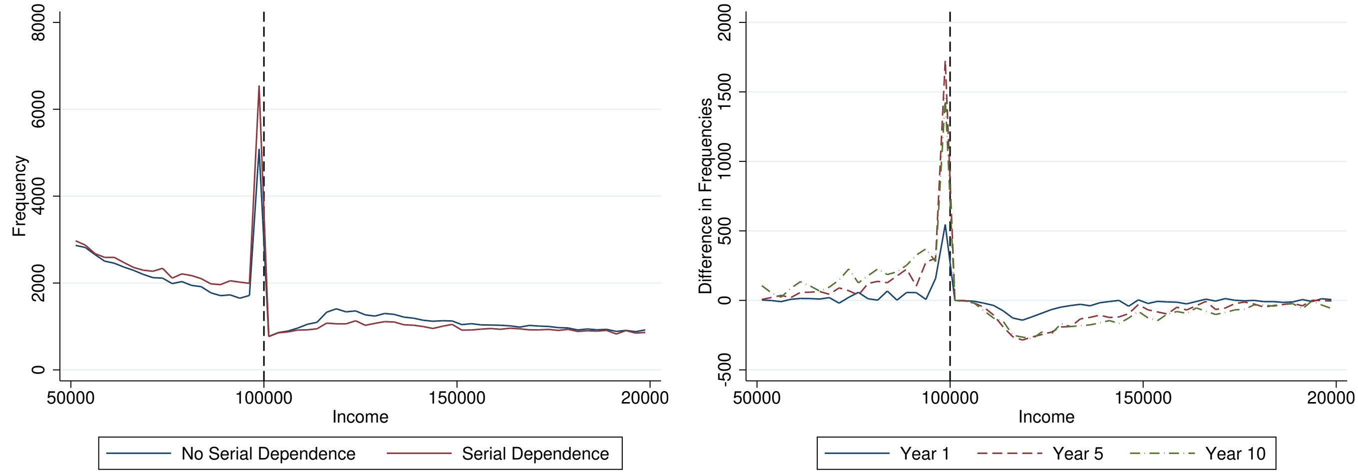
Weber, Caroline E., "Does the Earned Income Tax Credit Reduce Saving by Low-Income Households?," *National Tax Journal*, April 2016, *69* (1), 41–76.

Figures and Tables

Figure 1: Bunching After Introduction of Notch with Serially-Dependent Incomes

A. 10 years later, w/ and w/o serial dependence

B. Differences, w/ vs. w/o serial dependence

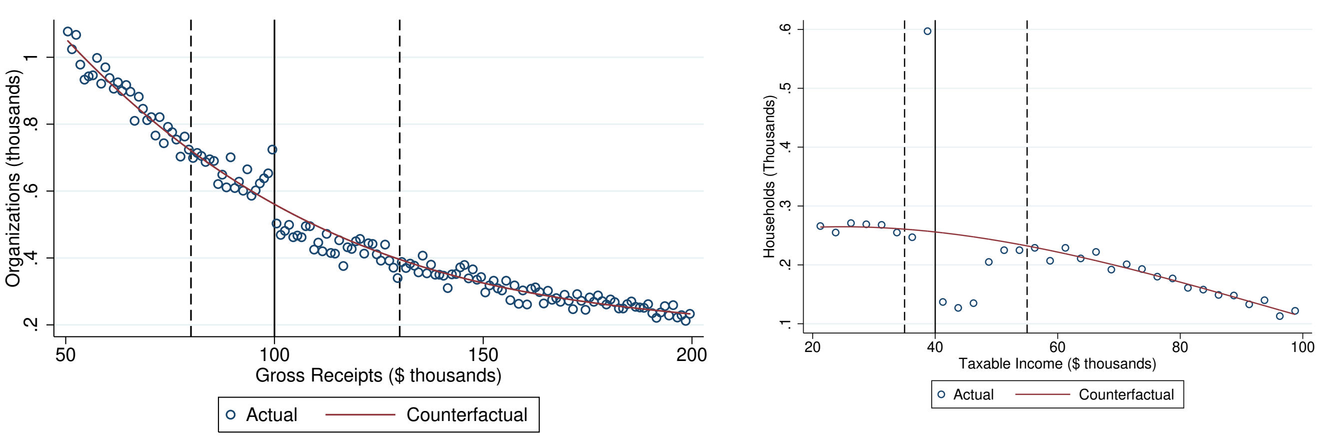


Notes: The figure depicts distributions of income in years after the introduction of a notch that induces bunching. Income growth follows all 10-year income profiles drawn from the PSID, with observations copied and base-year incomes smoothed as described in Section 3.3. Income in years after the base year is either serially dependent, meaning that it grows from the observed income in the prior year, or not, meaning that it grows from the potential income that would have been earned if a household had not bunched. Panel A shows that the distribution of income 10 years after introduction of the notch depends on whether income is serially dependent. Panel B shows the difference that serial dependence makes in the evolution of income over time.

Figure 2: Standard bunching estimation

A. Charities

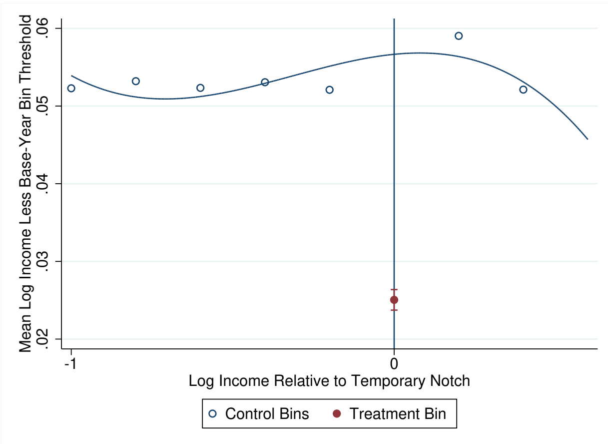
B. Households



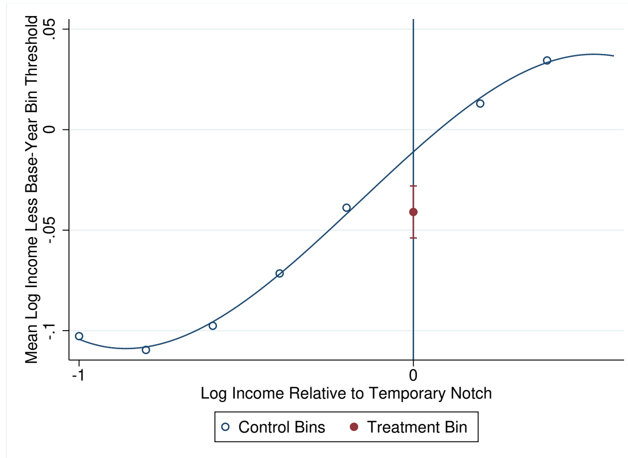
Notes: The figure depicts standard bunching estimation with charity data (Panel A) and the smoothed PSID income distribution described in Section 3.3 (Panel B). In both panels, the underlying data are represented by a histogram in blue circles, and each bin is treated as an observation. Bin counts are regressed on a polynomial, which estimates the counterfactual distribution depicted by the red curve, and a dummy variable for each bin in the omitted range indicated by the dashed lines. Excess "bunching" mass is calculated as the sum of coefficients on dummy variables for each bin in the bunching region between the first dashed line and the solid line at the notch. Similarly, the estimated reduction in mass above the notch is the sum of coefficients on dummies for the bins between the solid line and second dashed line. For Panel A, the polynomial has degree 3, the omitted range is $80-130,000, bin width=$1000, and N=264,770. For Panel B, the polynomial has degree 5, the omitted range of $35-55,000, bin width=$2500, and N=100,000.

Figure 3: Effect of a Temporary Notch on Income in Notch Year

A. Base-Year Income

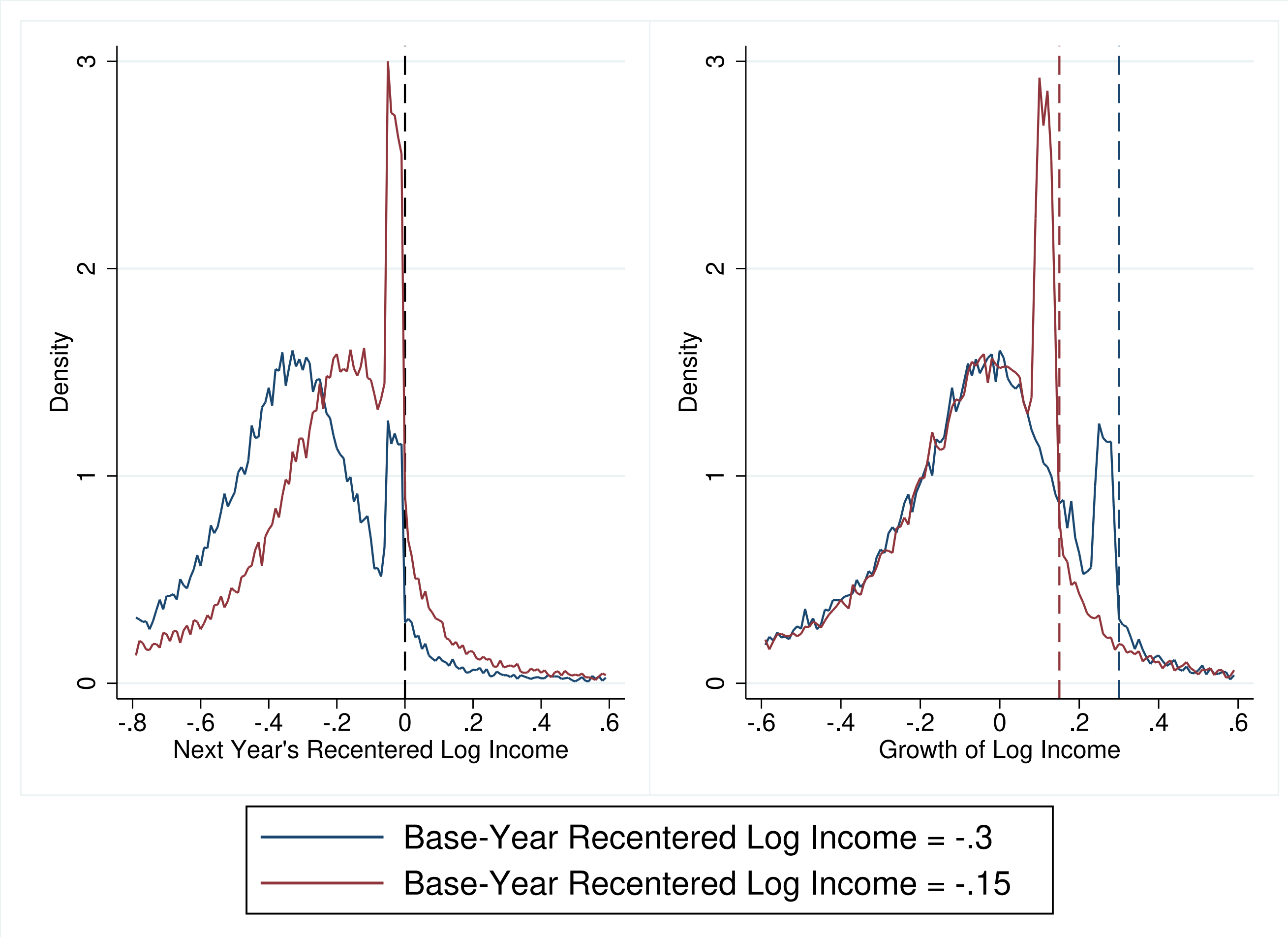


B. Next-Year Income



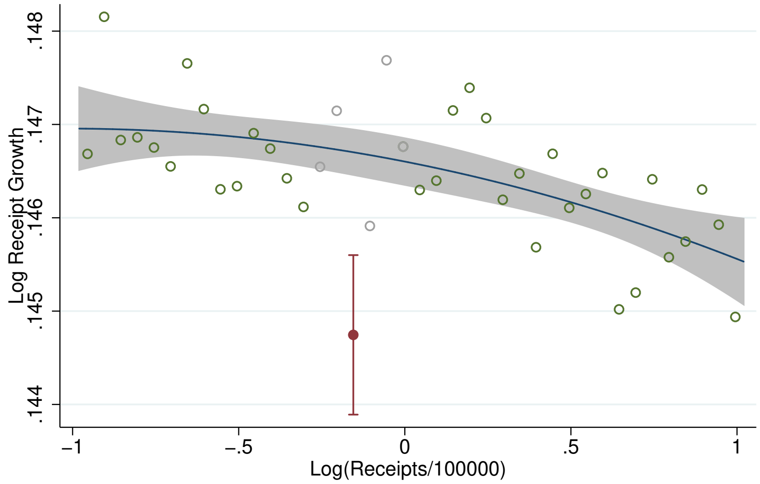
Notes: The figure shows results of temporary-notch dynamic bunching estimation performed on data in the year of bunching. The sample consists of 100,000 observations generated from the smoothed PSID income distribution described in Section 3.3. Bins have width .2, and the width of the bunching range is .05, meaning that the notch lies .05 from the minimum income in the range defining the treatment bin, and the bin threshold for each bin is its minimum plus .05. The outcome in Panel A is the difference between log income and the bin threshold. The outcome in Panel B is a dummy variable for having income above the bin threshold. The curve in each figure is a counterfactual estimated as a cubic polynomial through non-treatment bins. Due to bunching, the average observation in the treatment bin surrounding the threshold has a lower income and lower probability of being above the notch than it would have in the absence of bunching.

Figure 4: Illustration of Identification in Dynamic Bunching Estimation



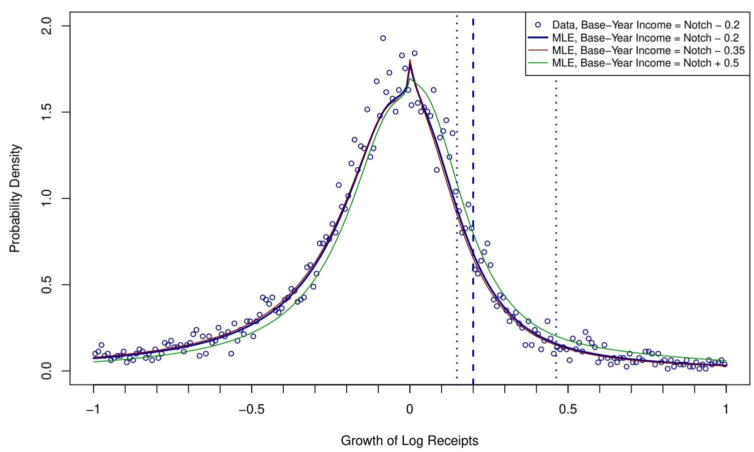
Notes: The figure shows the distribution of next-year log income (Panel A) and growth from base-year income to next-year income (Panel B) for households in two illustrative levels of base-year income. The sample is generated from the smoothed PSID income distribution described in Section 3.3. Due to bunching, the distributions for each group exhibit a spike just below the notch and a depression just above it. The growth distribution of each group is similar except around the notch, which appears in a different part of each distribution. Because the growth distribution does not vary too much with current income, growth distributions for varying levels of base-year income can be compared to estimate the extent of distortion to the rates of growth that bring households close to the notch.

Figure 5: Growth of Treated Charities vs. Counterfactual for an Illustrative Bin of Growth Rates



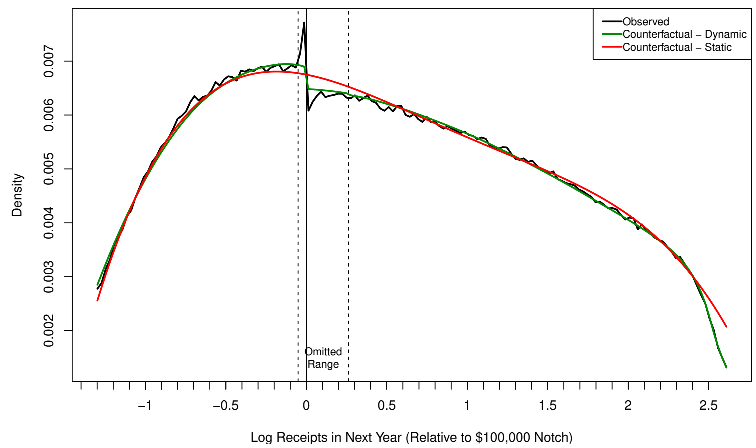
Notes: The figure shows growth of income from the current year to the next year as a function of current income (recentered around the reporting notch at $100,000). The figure sample consists of organizations in an illustrative growth bin that includes are organizations with growth between .1 and .2 log points. The marker with a 95-percent confidence interval represents the bin (defining the "$near\ notch_{it}$" dummy described in the text) for which growth of .1 to .2 implies that future receipts lie in the "omitted range" straddling the notch. The conditional average growth rate of these charities is just below .145, which is significantly less than the counterfactual growth rate interpolated from charities with higher and lower current incomes. The difference is interpreted as a measure bunching; some charities that approach the notch reduce their income to stay below it, and therefore conditional average growth is less than predicted. N=152,191. Omitted range is -.08 to .07. Bin width = .05.

Figure 6: Estimation of the Distribution of Household Income Growth Rates



Notes: The figure shows the density of growth in log income conditional on current income for an illustrative group of households. The curve depicts the estimated counterfactual that has been fit to draws from the smoothed empirical density depicted by circular markers, which was generated from the smoothed PSID income distribution described in Section 3.3. The figure sample contains 30,000 households starting from a range of income levels 0.2 to 0.3 log points below the hypothetical notch at $40,000. The notch for each household therefore lies at a growth rate of 0.2 to 0.3, and if households bunched there would be excess mass in or below this range.

Figure 7: Non-smooth Counterfactual Distribution of Charities' Income



Notes: The dynamic bunching estimation fits growth rates from each base year of data to the next year, and the figure shows the density of log gross receipts in the next year. Details of the dynamic estimates are provided in Section 6 and Appendix B. Both the dynamic and static estimates of the counterfactual fit the observed distribution fairly closely away from the notch. The counterfactual estimated using the dynamic MLE strategy is not smooth around the notch because it allows for extensive margin responses. The two estimation strategies imply different counterfactual distributions within the omitted range around the notch and therefore give different estimates of the amount of bunching. N=2,815,026.

Table 1: Bias of Static Estimates in Simulation with Extensive-Margin Responses

Responder	Reduced-Mass Estimates			Excess-Mass Estimates		
Share	Bias	Coverage	Root-MSE	Bias	Coverage	Root-MSE
0	-90	0.051	516	72	0.068	213
.01	2	0.049	506	103	0.078	227
.02	103	0.055	529	128	0.095	243
.05	438	0.132	690	207	0.166	295
.10	1023	0.450	1161	362	0.374	423

Notes: The table shows results of static bunching estimation performed on data with *no* bunching. Thus, the estimate for each outcome should equal zero, and the coverage rate should be close to 0.05. Each row presents results for 10,000 random samples of 100,000 observations generated from the smoothed PSID income distribution described in Section 3.3. The outcome is the common bunching-ratio estimate of the average number of dollars of income foregone by bunchers: either the reduced or excess number of observations, depending on column, divided by bin width and counterfactual number of observations. The Responder Share is the percentage of observations with incomes above the threshold that exhibit extensive-margin responses, i.e. are dropped from the data. Responders are drawn at random. Hypothetical notch at income of $40,000. Estimation using 5th-order polynomial with omitted range of $35,000 to $55,000.

Table 2: Bias of Static Estimates in Simulation with Serially-Dependent Income

Weight on Base-Year Potential Income	Reduced-Mass Estimates Bias	Coverage	Root-MSE	Excess-Mass Estimates Bias	Coverage	Root-MSE
0	879	0.358	1040	210	0.161	306
.5	251	0.074	602	150	0.103	269
1.0	51	0.050	549	-7	0.051	222
1.5	250	0.074	606	-177	0.125	285
2	623	0.199	837	-315	0.289	387

Notes: The table shows results of static bunching estimation performed on data for the year *after* bunching. Thus, the estimate for each outcome should equal zero, and the coverage rate should be close to 0.05. Each row presents results for 10,000 random samples of 100,000 observations generated from the smoothed PSID income distribution described in Section 3.3. The outcome is the common bunching-ratio estimate of the average number of dollars of income foregone by bunchers: either the reduced or excess number of observations, depending on column, divided by bin width and counterfactual number of observations. The weight on base-year potential income captures serial dependence; if the weight is ω and, in the base-year, observed income is I_o and potential income is I_p, then next-year income depends on $\omega I_p + (1 - \omega) I_o$. Thus, $\omega = 1$ implies no serial dependence on observed income, $\omega = 0$ implies that *only* observed income matters, and $\omega = 2$ implies that agents who bunch in the base year grow by *more* in the next year (e.g. if income is retimed from the base year to the next year). Hypothetical notch at income of $40,000. Estimation using 5th-order polynomial with omitted range of $35,000 to $55,000.

Table 3: Static Bunching Estimates for Charities

	Basic	Discontinuous	Two-Sided
Excess mass below the notch (*100)	.148***	.135***	.152
	(.020)	(.029)	(.093)
Bunching ratio (*100)	.592***	.537***	.608*
	(.096)	(.123)	(.312)
Reduction in mass above the notch (*100)	.250***	.223***	-.055
	(.026)	(.049)	(.066)

Notes: The table shows deviations of the binned income distribution from a counterfactual estimated in the range of $50-200,000. In the Basic specification, the counterfactual is a cubic in gross receipts. The Discontinuous specification allows for a discontinuity at the notch, and the Two-Sided specification allows for a separate quadratic on each side of the notch. The excess mass shows the estimated extra share of charities with incomes below the notch relative to the counterfactual, the bunching ratio is the ratio of the excess mass to the counterfactual density at the notch, and the reduction above the notch is the difference between the counterfactual and actual share above. The Basic specification indicates that .148 percent of charities appear below the notch when they shouldn't, which is roughly equal to the number of charities that should be above the notch by up to $600 (=$100,000*.00592 because the bunching ratio is reported with the density in log scale). The reduction in the number of charities above the notch is significantly larger than the addition below the notch, suggesting either misspecification or missing observations, and the flexible specifications do not reconcile the two results. The sample includes observations in years up to 2007 for charities also appear in the prior year (for comparability with the dynamic estimates). Bin width = $250. N = 969,842 in the range used for estimation and 2,907,476 total. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Simulation of Temporary Notch: No Bias with No Bunching

Variable	Bias	Coverage	Root-MSE
Log Income	-0.00015	0.07	0.001
Income	-5.85	0.07	31.46
Cross	-0.00005	0.06	0.003

Notes: The table shows results of temporary-notch dynamic bunching estimation performed on data with *no* bunching. Thus, the estimate for each outcome should equal zero, and the coverage rate should be close to 0.05. 10,000 random samples of 100,000 observations were generated from the smoothed PSID income distribution described in Section 3.3. Each row presents results for an outcome: "Income" is log taxable income, "Cross" is a dummy for having income above the level of the bin corresponding to the notch, and the "Bunching Ratio" is the product of the Income effect and the counterfactual average income level. Hypothetical notch at income of $40,000. Estimation using 3rd-order polynomial in log income with treatment and other bins of width 0.2.

Table 5: Temporary-Notch Simulations by Outcome and Degree of Serial Dependence of Income

Weight on Potential Income	Estimated Weight	Base-Year Log Income	Base-Year Cross	Next-Year Log Income	Next-Year Cross
0	0.06	-0.038	-0.419	-0.036	-0.419
	(0.195)	(0.001)	(0.006)	(0.008)	(0.006)
0.5	0.605	-0.038	-0.419	-0.015	-0.004
	(0.196)	(0.001)	(0.006)	(0.008)	(0.003)
1	1.116	-0.038	-0.419	0.004	-0.004
	(0.197)	(0.001)	(0.006)	(0.008)	(0.003)
1.5	1.598	-0.038	-0.419	0.023	-0.004
	(0.199)	(0.001)	(0.006)	(0.008)	(0.003)
2	2.049	-0.038	-0.419	0.04	-0.004
	(0.202)	(0.001)	(0.006)	(0.008)	(0.003)

Notes: The table shows results of temporary-notch dynamic bunching estimation performed on data in the year with bunching ("Base-Year") and the subsequent year ("Next-Year"). Each row presents results of 10,000 random samples of 100,000 observations generated from the smoothed PSID income distribution described in Section 3.3, with weights determining serial dependence of income as described in the notes of Table 2. Outcomes across columns include the estimate of the weight on base-year potential income, log income, and the "Cross" dummy for having income above the notch. Each cell contains the average estimate of the effect of being in the treated range around the notch, and the standard deviation of the estimates appears in parentheses. Bunching reduces income and crossing in Year 1, and if potential income does not have weight equal to 1 then this affects income and crossing in year 2. Hypothetical notch at income of $40,000. Estimation using 3rd-order polynomial in log income with treatment and other bins of width 0.2.

Table 6: Temporary-Notch Estimates for Charities

	2007	2008	2009	2010	2011	2012
Log Receipts	-0.001	0.009	-0.003	-0.007	-0.013	-0.018
	(0.010)	(0.008)	(0.001)***	(0.008)	(0.009)	(0.012)
Cross 2009 Threshold	0.003	0.006	-0.104	-0.022	-0.020	-0.018
	(0.008)	(0.008)	(0.007)***	(0.008)***	(0.008)**	(0.010)*
Same Receipts as in 2009	-0.000	0.001		0.005	0.004	0.007
	(0.001)	(0.002)		(0.002)**	(0.002)*	(0.003)***
Observations	48,716	107,579	105,160	127,855	115,601	104,760

Notes: The table shows the results of regressions of three different variables on a quadratic function of binned log receipts in 2009 and a "Near Notch" dummy for the bin that straddles the $500,000 notch for that year. The table shows the estimate of the coefficient on the "Near Notch" dummy, which represents the causal effect of having income near the notch in 2009. The first row shows that log receipts of charities near the threshold in 2009 are significantly lower than expected in that year, as expected due to bunching. Point estimates remain negative in subsequent years but standard errors are large. The outcome in the second row is a dummy for crossing the level of growth corresponding to the notch ("Cross" as defined in the text). The coefficients indicate charities experience a significant, permanent reduction of at least one percentage point in the probability of having income over $500,000 in any year after 2009. The outcome for the third row is an indicator equal to one if the charity is in the same log receipts bin that it is in 2009, and the results indicate that charities are significantly less likely to have grown out of their bin in 2009. In years before 2009 there are no significant differences between the treated charities and the interpolated counterfactual. Robust Huber-White standard errors are displayed. Bins have width .155 and extend from 1.615 log points below the notch to 3.19 log points above it (roughly $100,000 to $12 mil). These parameters give 35 control bins in addition to the treatment bin, and bunching estimates are robust to changes in these parameter choices. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Dynamic OLS Estimates in Simulation with Serially-Dependent Income

Variable	Bias	Coverage	Root-MSE
Log Income	-0.000003	0.053	0.001
Income	-0.105	0.053	38.46
Cross	-0.005	0.059	0.017

Notes: The table shows results of OLS dynamic bunching estimation performed on data with bunching occurring in the base year but *no* bunching and *no* extensive margin responses in the next. Thus, the estimate for each outcome should equal zero, and the coverage rate should be close to 0.05. 10,000 random samples of 100,000 observations were generated from the smoothed PSID income distribution described in Section 3.3. "Log Income" is the estimated effect of moving to the omitted range, which will be negative if households bunch. "Cross" is a dummy for having income growth putting the household above the notch or its equivalent for the household's starting income and growth bin. Hypothetical notch at income of $40,000. Estimation using log income, omitted base-year range of 10.55 to 10.65, and growth rate bins of width 0.1.

Table 8: Dynamic OLS Estimates for Charities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Near Notch	-0.053***	-0.021***	-0.018**	-0.015**	-0.016**	-0.017***	-0.016**	-0.015**	-0.017***	-0.022***	-0.012*	-0.014***
	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.008)	(0.006)	(0.005)	(0.006)	(0.007)	(0.005)
N	307,526	260,209	261,771	256,548	252,669	247,364	245,228	240,193	234,728	231,303	225,570	221,296

Notes: The table shows the results of regressing a dummy for crossing the level of growth corresponding to the notch ("Cross" as defined in the text) (t) years in the future on the "Near Notch" dummy for bins that straddle the notch in the next year, controlling for bins of growth rate (of width .1) and a quadratic function of current receipts. The coefficients show charities a significant reduction of at least one percentage point in the probability of crossing the notch at all horizons. The sample includes charities within one log point of the notch in any starting year from 1990 to 1997 and growing by 0 to 1 log points. Standard errors are clustered by state. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Dynamic MLE Estimates in Simulation with Serially-Dependent Income

Variable	Bias	Coverage	Root-MSE
Share Bunching	-0.0003	0.049	0.008
Income Reduction	-9.20	0.049	241.1
Share Exiting	0.0039	0.052	0.025

Notes: The table shows results of MLE dynamic bunching estimation performed on with bunching occurring in the base year but *no* bunching and *no* extensive margin responses. Thus, the estimate for each outcome should equal zero, and the coverage rate should be close to 0.05. 10,000 random samples of 10,000 observations were generated from the smoothed PSID income distribution described in Section 3.3. "Bunching" is the percentage of households, of those that would have moved into the reduced range, who instead bunch below the notch. "Extensive Margin" is the percentage of households, of those that would have crossed from below the notch to above, who instead exit the sample. Hypothetical notch at income of $40,000. Estimation using log income and omitted range of 10.5 to 10.9.

Table 10: MLE Estimates for Charities

	Static 1991-2007	Dynamic 1991-2007
Share bunching from below notch		0.026***
		(.003)
Share bunching from above notch		0.005***
		(.001)
Attrition of those crossing to reduced range		0.080***
		(.004)
Attrition of those crossing to higher incomes		0.093***
		(.005)
Excess mass just below the notch (*100)	.194***	0.103***
	(.017)	(.007)
Reduction in mass in reduced range (*100)	.293***	0.354***
	(.019)	(.012)
Bunching ratio	753.21***	404.92***
	(67.04)	(65.14)
N	2,196,564	2,815,026

Notes: The table shows the results of maximum likelihood dynamic bunching estimation, along with static estimates on similar sample for comparison. The top panel of the figure provides estimates from the dynamic design that cannot be obtained from the static approach. The top two parameter estimates indicate that charities that approach the notch from below are significantly more likely to manipulate receipts to remain below the notch in the next year. The next two parameter estimates imply that significant share of the charities with current income below the notch should grow to an income level above the notch but instead exit from the sample. The lower panel shows that the static approach overestimates the excess number of organizations just below the notch and underestimates the number that should be just above it. All regressions allow for attrition that can vary with current income as described in the text. The sample size for static estimation is smaller than that for dynamic because the latter includes all charities appearing in the base year while the former excludes charities that were missing or far above the notch in the next year, but static estimates are rescaled to have the same denominator as the dynamic estimates for comparability. Standard errors for dynamic estimates are calculated numerically. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Appendix A - Examples of Serial Dependence

First, consider a simple discrete income distribution $d = \begin{bmatrix} d_1 & d_2 & d_3 & d_4 \end{bmatrix}$ that is left-multiplied by a transition matrix

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & .05 & .95 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrix T provides a discrete-case example with perfect serial dependence and a stochastic compliance cost that causes some bunching in the form of shifting of mass from d_3 to d_2. If the initial distribution is $\begin{bmatrix} .4 & .3 & .2 & .1 \end{bmatrix}$, then the distribution after one period is $\begin{bmatrix} .4 & .31 & .29 & .1 \end{bmatrix}$. A counterfactual distribution estimated from a linear combination of d_1 and d_4 will correctly recover the initial values of d_2 and d_3 and the amount of bunching that occurred. The limit distribution, however, is $\begin{bmatrix} .4 & .5 & 0 & .1 \end{bmatrix}$. The researcher estimating the same linear counterfactual would obtain a bunching estimate of .2. Such an estimate reflects the full accumulation of excess mass over numerous periods' choices rather than a description of the choice itself. If one were to generalize T to include positive probabilities of transition between each pair of states then the empirical challenge would be even greater because today's bunching would affect tomorrow's values of d_1 and d_4, and bunching estimation relies on the assumption that the distribution is unaffected at levels sufficiently far from the notch. Static estimation would not identify the share of agents that will bunch or the amount by which they will be willing to reduce income (respectively b and $\bar{\delta}$ in the model of Section 3) but rather some unknown function of these parameters and the rates of transition between different income levels.

Alternatively, suppose the conditional distribution of income in the next period is discrete-continuous with strictly positive mass at today's income level. Say that income has observed distribution $f_t(y_t)$ in the current year, and the pre-bunching (counterfactual) cumulative distribution function (cdf) for the following year is given by $F_{t+1}(y_{t+1}) = \int G_{t+1}(y_{t+1}|y_t) f_t(y_t)\, dy$, with $G_t(y_{t+1}|y_t) = \alpha \cdot 1_{\{y_{t+1} \geq y_t\}} + (1-\alpha) H_t(y_{t+1}|y_t)$ for some constant $\alpha \in (0,1)$ and continuous cdf $H_t(y_{t+1}|y_t)$. Say there is a notch at $y_t = n$ and bunching at the notch in current year. Then $\lim\limits_{y_{t+1}\to n^+} F_{t+1}(y_{t+1}) - \lim\limits_{y_{t+1}\to n^-} F_{t+1}(y_{t+1}) = \lim\limits_{y_{t+1}\to n^+} \int \left[\alpha \cdot 1_{\{y_{t+1}\geq y_t\}} + (1-\alpha) H_t(y_{t+1}|y_t)\right] f_t(y_t)\, dy - \lim\limits_{y_{t+1}\to n^-} \int \left[\alpha \cdot 1_{\{y_{t+1}\geq y_t\}} + (1-\alpha) H_t(y_{t+1}|y_t)\right] f_t(y_t)\, dy = \alpha \left[\lim\limits_{y_{t+1}\to n^+} F_t(y_{t+1}) - \lim\limits_{y_{t+1}\to n^-} F_t(y_{t+1})\right] \neq 0.$

Appendix B - Details of Dynamic Bunching Designs

1. Ordinary Least Squares

This appendix describes a useful outcome in binning designs and then bin selection. The useful outcome provides a direct estimate of the share of agents that bunch. Consider a growth rate range/bin of $[\gamma, \gamma + .1)$, a notch ρ_{it+1}, a bunching (excess-mass) range of width ω_E, and bins of current income r_{it} with minimum values denoted by $rmin_{it}$. The treatment bin of current income has minimum value $rmin_{it} = \rho_{it+1} - \omega_E - \gamma$. Agents with income at the minimum of this bin will cross the notch if they grow by $\gamma + \omega_E$. Other agents in the bin will cross the notch if growth is greater than $\gamma + \omega_E - (r_{it} - rmin_{it})$. Let $cross_{\gamma it} = \mathbf{1}\left[g_{it} > \gamma + \omega_E - (r_{it} - rmin_{it})\right]$. This same outcome is relevant for all observations, regardless of bin, and indicates whether the agent achieves growth at a rate that *would* correspond to the notch if the observation were in the treatment bin. Because $cross_{\gamma it}$ only has significance for the treatment bin, the probability that $cross_{\gamma it} = 1$ should be reduced in the bin of interest by the share of agents that bunch, and it should not be affected for other bins. This outcome for growth bin γ can then be generalized to all levels of growth by defining $cross_{it} = \sum_{\gamma} \left(cross_{\gamma it} * \mathbf{1}\left[\gamma \leq growth_{it} < \gamma + .1\right]\right)$.

One issue to be considered for any binning approach to bunching estimation is bin selection. In the dynamic binning designs, the treatment bins with $NearNotch_{it} = 1$ should be constructed so as not selected on bunching status. Thus, for each growth bin γ, the treated bin (denoted here by indicator $NearNotch_{\gamma it}$) should not differ in the number of agents from what would be predicted by a counterfactual constructed from other bins. This suggests a test that the choice of bins is reasonable: estimate equation 3 separately for each growth rate range, using the base-year bin count as the "outcome," and then test the hypothesis that $\forall\gamma, NearNotch_{\gamma it}$ has no effect on bin count. In the charity application, this test fails to reject, with p-value .1361. Marx (2018) provides a figure depicting this test for charities with growth of log receipts in the range of $[0.1, 0.2)$, which appears similar to Figure 5 here but shows that the count in the treatment bin is in line with the counterfactual. The choice of bin width and location remains ad-hoc, as it has been throughout the bunching literature; an econometric procedure for optimal bin construction remains a topic for future research.

2. Maximum Likelihood Estimation

It is possible to perform maximum likelihood estimation by estimating a flexible function for the pdf and constraining it to integrate to unity, but starting from the cdf offers several advantages. First, it is desirable to estimate excess attrition among those who cross the notch or points of sample truncation, and the cdf gives the probabilities of these occurrences. Second, the cdf makes it straightforward to constrain the reduced mass to equal the bunching mass (except for differences due to systematic attrition). Third, truncation requires integration of the likelihood between limits that vary with the level of current receipts, a practical issue for programs performing multidimensional integration. A disadvantage of specifying the cdf is the need for functions that appear more arbitrary than their derivatives. For example, I include inverse tangents to allow for curvature at growth rates close to zero because the derivative of $\arctan(x)$ is $\frac{1}{1+x^2}$.

As noted in equation 4, the latent cdf of conditional growth for the Laplace family of distributions can be written as

$$F^*(g_{it}|r_{it},\Theta) = \begin{cases} \exp\left(P_l\left(g_{it}, r_{it}, \Theta\right)\right) & g_{it} < \theta \\ 1 - \exp\left(P_u\left(g_{it}, r_{it}, \Theta\right)\right) & g_{it} \geq \theta \end{cases}.$$

To simplify notation slightly, I will hereafter omit the subscripts on g_{it} and r_{it}. I am able to obtain a reasonable fit for the both the PSID and charity data with slight variants of the following structure.

$$P_l(g,r,\theta) = \pi^l_{0,0} + \pi^l_{0,1}r + \pi^l_{0,2}r^2 + \left(\pi^l_{1,0} + \pi^l_{1,1}r + \pi^l_{1,2}r^2\right)(g-\theta) + \left(\pi^l_{2,0} + \pi^l_{2,1}r + \pi^l_{2,2}r^2\right)\left[\exp(g-\theta) - 1\right]$$

$$+\left(\pi^l_{3,0} + \pi^l_{3,1}r + \pi^l_{3,2}r^2\right)\left[\exp\left(-(g-\theta)^2\right) - 1\right] + \left(\pi^l_{4,0} + \pi^l_{4,1}r + \pi^l_{4,2}r^2\right)\arctan\left(\left(\pi^l_{5,0} + \pi^l_{5,1}r + \pi^l_{5,2}r^2\right)(g-\theta)\right)$$

$$P_u(g,r,\theta) = h(r) + \left(\pi^u_{1,0} + \pi^u_{1,1}r + \pi^u_{1,2}r^2\right)(g-\theta) + \left(\pi^u_{2,0} + \pi^u_{2,1}r + \pi^u_{2,2}r^2\right)\left[\exp(-(g-\theta)) - 1\right]$$

$$+\left(\pi^u_{3,0} + \pi^u_{3,1}r + \pi^u_{3,2}r^2\right)\left[\exp\left(-(g-\theta)^2\right) - 1\right] + \left(\pi^u_{4,0} + \pi^u_{4,1}r + \pi^u_{4,2}r^2\right)\arctan\left(\left(\pi^u_{5,0} + \pi^u_{5,1}r + \pi^u_{5,2}r^2\right)(g-\theta)\right)$$

I now list, and impose as needed, conditions to constrain $F^*(g|r,\Theta)$ to have properties of a cumulative distribution function. First, the function must have infimum 0 and supremum 1. The appropriate limits can be achieved by two sets of restrictions on the parameters:

1. $\left(\pi^l_{1,0} + \pi^l_{1,1}r + \pi^l_{1,2}r^2\right) < 0 \Rightarrow \lim_{g\to-\infty} P_l(g,r,\Theta) = -\infty \Leftrightarrow \lim_{g\to-\infty} F^*(g|r,\Theta) = 0$
2. $\left(\pi^u_{1,0} + \pi^u_{1,1}r + \pi^u_{1,2}r^2\right) < 0 \Rightarrow \lim_{g\to\infty} P_u(g,r,\Theta) = -\infty \Leftrightarrow \lim_{g\to\infty} F^*(g|r,\Theta) = 1$

Both constraints are easily implemented by using exponentiated coefficients in the numerical maximization.

Second, $F^*(g|r,\Theta)$ must be nondecreasing. Because the posited functional form has one point of non-differentiability at $g=\theta$, the nondecreasing property requires $\lim_{g\to\theta^-} F^*(g|r,\Theta) \leq \lim_{g\to\theta^+} F^*(g|r,\Theta)$. I require this relation to hold with equality, giving continuity of the cdf and ruling out point mass at a growth rate of θ. This gives

$$\exp\left(P_l(\theta,r,\Theta)\right) = 1-\exp\left(P_u(\theta,r,\Theta)\right)$$

$$\exp\left(\pi^l_{0,0}+\pi^l_{0,1}r+\pi^l_{0,2}r^2\right) = 1-\exp\left(h(r)\right)$$

$$3.\, h(r) = \log\left(1-\exp\left(\pi^l_{0,0}+\pi^l_{0,1}r+\pi^l_{0,2}r^2\right)\right)$$

The implied latent density is

$$f^*(g_{it}|r_{it},\Theta) = \begin{cases} P'_l(g,r,\Theta)\exp\left(P_l(g,r,\Theta)\right) & g<\theta \\ -P'_u(g,r,\Theta)\exp\left(P_u(g,r,\Theta)\right) & g\geq\theta \end{cases}$$

where $P'_l(g,r,\Theta) =$ and $P'_u(g,r,\Theta)$ are derivatives with respect to g. These derivatives can be assured of the correct sign by exponentiating each of the relevant coefficients, but this would impose more than is required because nonnegativity of the density does not necessitate that *all* the coefficients have the same sign. In practice, I instead impose a prohibitive penalty on the value of the likelihood function if the pdf is negative for any observations. Similarly, I do not impose conditions 1 and 2, which arise naturally during the optimization, but I do impose condition 3, which has the added benefit of reducing the number of parameters to be estimated. I set the location parameter $\theta=0$ for the PSID data, and doing so for the charity application gives very similar results to using a nonparametric estimate of the mode, as detailed by Marx (2018). One could also exclude observations with current receipts in the omitted region or allow the density to be discontinuous in r at the threshold for the base-year notch.

This completes the specification of the latent distribution. The observed distribution $F(g^*|r,\Theta,\Omega)$ involves modifications for bunching and attrition through parameters Ω. I describe these modifications in steps and then present the comprehensive function for $F(g^*|r,\Theta,\Omega)$.

To measure bunching I estimate b, the share of mass from the reduced region that instead appears in the bunching region. I estimate two parameters for b, allowing the bunching propensity to depend on whether base-year income is above the notch, but require the bunching mass to equal the reduced mass regardless. I define a notch ρ and allow agents to shift income from a reduced range of width ω_R to a region of width ω_E. Thus, there is excess mass B in the bunching region $g+r\in[\rho-\omega_E,\rho)$ that would otherwise lie in the reduced region $g+r\in[\rho,\rho+\omega_R)$. Combining these ranges gives an omitted region of $g+r\in[\rho-\omega_E,\rho+\omega_R)$.

Agents moving to the omitted region are excluded from identification of the latent distribution. However, these observations should be incorporated into the observed distribution to estimate bunching and attrition parameters. To do this I generate a variable $g*$ equal to $(\rho + \omega_R - r)$ for agents moving to the reduced range, $(\rho - r)$ for charities moving to the bunching range, and g for other charities. The fact that $g*$ is assigned as such is then incorporated into the likelihood function. Missing and bunching observations could be assigned to any value of g^*; identification uses the count of missing and the count of omitted and not the location of either.

Next, one must model attrition. I allow for 3 channels through which agents observed in the base year go unobserved in the next year. First, I truncate the samples at a lower-bound growth rate ℓ because these take agents far from the notch and because charities do not file any information return if log receipts are below $\log(25,000)$. I set the share of truncated observations equal to the value taken by the latent conditional cdf at $\ell - r$. Second, some share $\lambda(r)$ of current filers will not appear in the next year's data file regardless of income. Third, I allow that an additional share $\delta(r,g)$ go missing when crossing ρ. In all three cases, growth is unobserved, so for these observations I set the value of $g*$ equal to the minimum observable growth $(\ell - r)$. The share exiting instead of crossing the notch, $\delta(r,g)$ is allowed to take different values depending on whether growth takes the observation to the reduced range, where bunching offers an alternative to going missing, or to income levels above the reduced range.

The observed conditional cdf is then

$$F(g^*|r,\Theta,\Omega) = \begin{cases} 0 \\ \lambda(r) + (1-\lambda(r))\,F(\ell - r|r) + \delta(r,g)\,(1 - F(\rho - r|r)) \\ \lambda(r) + (1-\lambda(r))\,F(g^*|r) + \delta(r,g)\,(1 - F(\rho - r|r)) \\ \lambda(r) + (1-\lambda(r))\,F(\rho - \omega_E - r|r) + \delta(r,g)\,(1 - F(\rho - r|r)) \\ (1-\lambda(r))\,[F(\rho - r + \omega_R|r) - F(\rho - r - \omega_E|r)] \\ \qquad + b\,(1 - \lambda(r) - \delta(r,g))\,[F(\rho - r + \omega_R|r) - F(\rho - r|r)] \\ (1-b)\,(1 - \lambda(r) - \delta(r,g))\,[F(\rho - r + \omega_R|r) - F(\rho - r|r)] \\ \lambda(r) + \delta(r,g) + (1 - \lambda(r) - \delta(r,g))\,(F(g^*|r)) \end{cases}$$

$$
\begin{cases}
 & for \quad g* < \ell - r \\
 & for \quad g* = \ell - r \\
 & for \quad \ell - r < g* < \rho - r - \omega_E \\
 & for \quad \rho - r - \omega_E \leq g* < \rho - r \\
 & for \quad g* = \rho - r \\
 & \\
 & for \quad g* = \rho - r + \omega_R \\
 & for \quad \rho - r + \omega_R < g*
\end{cases}
$$

and maximizing the likelihood function $\sum_{i=1}^{N} \log\left[f^*\left(g_i^*|r_i\right)\right]$, where $f^*\left(g_i^*|r_i\right)$ is the discrete-continuous implementation of the conditional likelihood implied by $F\left(g^*|r, \Theta, \Omega\right)$, gives an estimate of the value of each parameter. For any value of r one can then obtain counterfactual growth estimates by plugging the desired value(s) of g into the estimated distribution function(s). Integrating over r gives the total counterfactual mass for the next year.