Optimization Theory Behind the `bunching` Package

Panos Mavrokonstantis

2022-08-16

This vignette introduces the optimization theory underlying the `bunching` package. For an explanation of the package's estimation methods and examples, see the companion vignette `bunching_examples`.

Introduction

The `bunching` package implements the bunching estimator, as developed (in different flavors and applications) by Saez (2010), Chetty et al. (2011) and Kleven et al. (2013). The original aim of the estimator was to tackle questions within the fields of labor and public economics (and taxation in particular), but can be applied to any setting where a constrained optimization problem involves a discontinuity in a constraint, which can be related to a discontinuity in the observed density of the decision variable. It naturally extends to any setting that involves prices, since taxes are simply instruments that scale or shift prices. Indeed, it has since been applied in numerous studies outside of economics, including empirical questions in finance (Dagostino 2019), psychology (Allen et al. 2017) and law (Dharmapala 2019). Due to its generality, it can be readily used in any commercial setting involving non-linearities in constraints, such as non-linear pricing schemes, non-linear incentive contracts in the workplace, etc.

The goal of the estimator is to credibly estimate the resulting bunching (i.e. excess mass) at a particular level of a choice set that features a discontinuity, over and above what would have been observed in the absence of the discontinuity in the given constraint. The estimator solves a challenge of both causal inference and prediction, since we would like to (causally) attribute the excess mass to a particular discontinuity in a constraint, but cannot typically observe the counterfactual density in the absence of that constraint. The estimator's novelty is to provide a prediction method for this counterfactual by combining reduced-form methods with micro-economically founded optimization theory, and relies on minimal assumptions.

The vignette will explain the intuition using the original setting for which bunching was developed: analyzing behavioral responses to income taxes, and will closely follow Mavrokonstantis and Seibold (2022). For a longer review, see Kleven (2016). Two main types of discontinuities are considered: kinks, i.e. changes in the slopes of choice sets, and notches, i.e. changes in the levels of choice sets. Both are supported by the `bunching` package.

Theoretical foundations: Why would densities exhibit bunching?

The case of kinks

Let's consider the setting of income taxes. In most countries, income taxes follow what is called a progressive tax schedule, which features several brackets, with the marginal tax rate increasing as incomes enter higher brackets. This creates kinks in choice sets at bracket thresholds because only income falling in a particular bracket is subject to its marginal tax rate. The other case is a notch, where all income is subject to the tax rate of the final bracket it falls in, including the amounts falling in lower brackets.

From a taxpayer's optimization perspective, we want to understand how the presence of a kink affects incentives, and thereby predict what behavioral responses tax rate reforms can elicit. To do this, we start by

first considering an optimization setting where no kinks are in place, and then compare these to optimal behavior in the presence of a kink.

We will consider the following simple model. There is a population of individuals, each having preferences over consumption c and taxable income z, who differ only by ability n. Utility $U(c, z; n)$ is increasing and concave in consumption ($U_c > 0$, $U_{cc} \leq 0$), and decreasing and convex in earnings ($U_z < 0$, $U_{zz} > 0$), capturing the idea that the effort associated with labor supply is costly. Tax on income z is denoted by the function $T(z)$, which is non-linear in the presence of kinks and notches. $T(z)$ can be linearized by using "virtual" income $R = z - T(z) - [1 - T'(z)]z$ and marginal tax rate t, which simplifies the optimization problem to:

$$\max_{c,z} \quad U(c, z; n)$$

$$s.t. \quad c = (1 - t)z + R$$

The first-order condition $(1 - t)u_c + u_z = 0$ implies an optimal level of earnings given by $z = z((1 - t); n)$, where the only source of heterogeneity is ability n. If this ability distribution is smooth and the tax schedule is linear ($T(z) = t_0 \ \forall z$), then the model predicts that the distribution of earnings will also be smooth, with cdf $H_0(z)$ and pdf $h_0(z)$.

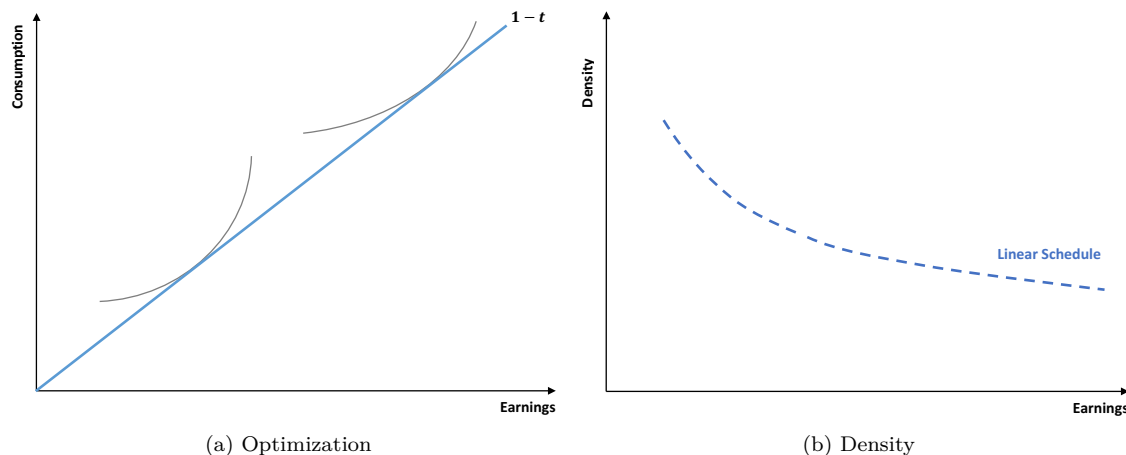(a) Optimization   (b) Density

Figure 1: Optimization with Linear Constraints

Figure 1 provides a visualization of the constraint and density in this situation. Sub-figure (a) depicts typical indifference curves and a linear budget constraint. It shows that under a linear tax schedule, individuals optimize by locating at the tangency points between their indifference curves and their budget set (the constraint). Importantly, they distribute themselves along this constraint smoothly, according to their heterogeneous ability n. Those with higher ability will optimize at higher combinations of consumption and earnings. This is because the opportunity cost of leisure, given by ability n, is higher for these individuals. Sub-figure (b) shows the associated density, which as a result is also smooth.

Next consider the creation of a second bracket, increasing the marginal tax rate from t_0 to t_1 for earnings above z^*. This situation is shown in Figure 2. Sub-figure (a) shows how this non-linear constraint features a convex kink. Recall that optimal earnings are given by $z = z((1 - t); n)$, which has now changed for those above z^*, who will now re-optimize by decreasing their earnings. Sub-figure (b) shows the effect of this on the distribution of earnings. Individuals in a range $z \in (z^*, z^* + \Delta z^*]$ optimize by moving to the kink at z^*, which creates the bunching mass. Not everyone above z^* will find it optimal to reduce earnings to the kink. The bunching range can be identified by a marginal buncher, i.e. the individual with the highest earnings under the linear tax schedule who chooses to bunch under the non-linear schedule. This is the individual tangent both at $z^* + \Delta z^*$ under the linear schedule, and the upper part of the non-linear budget set at z^*. Hence, the bunching mass is given by:
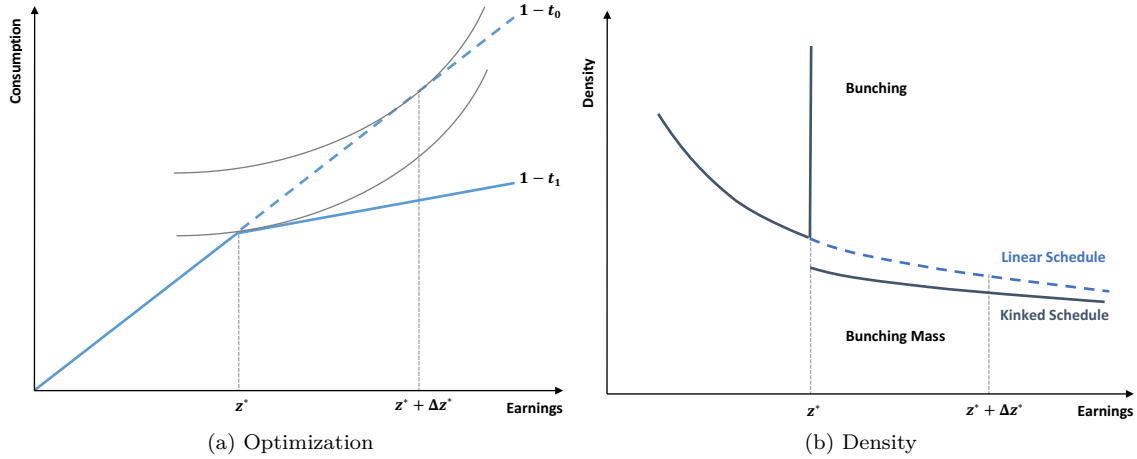
Figure 2: Optimization with Kinked Constraints

$$B = \int_{z^*}^{z^* + \Delta z^*} h_0(z) dz$$

A central question is how to empirically estimate this credibly, and is covered in the companion vignette `bunching_examples`.

From bunching at kinks to elasticities

The key point of estimating the bunching mass is to recover an economic parameter of interest. Since individuals are responding to a change in their choice set, it is feasible to estimate an elasticity. In our setting, we can back out the elasticity of earnings e with respect to the marginal tax rate by relating the response of the marginal buncher to the change in the marginal tax rates. If the kink's size $\Delta t = t_1 - t_0$ is small, then a non-parametric estimate of the (observed) elasticity of earnings can be obtained using:

$$e = \frac{\Delta z^* / z^*}{\Delta t / (1 - t_0)}$$

We can back this out from our estimate of the bunching mass by observing that:

$$B = \int_{z^*}^{z^* + \Delta z^*} h_0(z) dz \approx h_0(z^*) \Delta z^*$$

Then, all we need in order to estimate the elasticity is an estimate of the bunching mass B and the height of the counterfactual density at the kink $h_0(z^*)$, both of which are estimable. The elasticity estimate is then:

$$e = \frac{B / h_0(z^*) z^*}{\Delta t / (1 - t)}$$

This approximation will perform poorly if the kink is large. In this case, we can make progress by parametrizing the utility function. The usual approach is to specify a quasi-linear and iso-elastic function of the form:

$$U(c, z; n) = c - \frac{n}{1 + \frac{1}{e}} \left(\frac{z}{n} \right)^{1 + \frac{1}{e}}$$

where $c = z - T(z)$. The optimal earnings supply function is now $z = n(1-t_1)^e$. Denoting the ability level of the marginal buncher by n^*, his optimal earnings are $z^* + \Delta z^* = n^*(1-t_0)^e$ under the linear schedule and $z^* = n^*(1-t_1)^e$ under the kinked schedule. Combining these gives:

$$e = -\frac{ln\left(1 + \frac{\Delta z^*}{z^*}\right)}{ln\left(\frac{1-t_0}{1-t_1}\right)}$$

or using our estimable quantities:

$$e = -\frac{ln\left(1 + \frac{B}{h_0(z^*)z^*}\right)}{ln\left(\frac{1-t_0}{1-t_1}\right)}$$

Note that this can be interpreted as the "observed", but not necessarily the "structural" elasticity. The latter is a deep preference parameter that does not depend on the policy environment. Whether the observed is a good approximation of the structural elasticity will depend on whether there are any optimization frictions affecting (and here, attenuating) the observed bunching responses. To recover structural estimates in the presence of frictions, the observed bunching mass must be combined with a structural model that includes frictional parameters. For an example of such an approach, see Gelber et al. (2020) and Mavrokonstantis and Seibold (2022).

The case of notches

Whereas a kink creates a discrete change in the slope of the constraint, a notch creates a discrete change in its level. A typical example is a taxation system where above a given threshold z^*, the tax rate increases from t_0 to t_1 but applies to all income, not just that falling in the new bracket. In other words, instead of the marginal tax rate changing, the average tax rate changes. This can be analyzed using the same optimization setup as with kinks, the only difference being that the tax function is now given by $T(z) = z(t_0 + (t_1 - t_0)\mathbb{1}(z > z^*))$. Figure 3 depicts such a constraint, with its associated density.
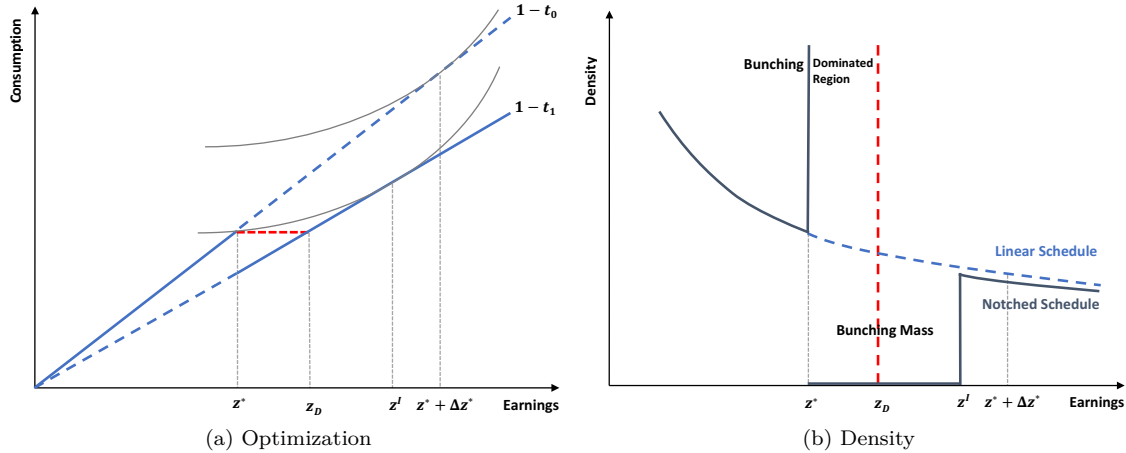

Figure 3: Optimization with Notched Constraints

Sub-figure (a) shows how a notch creates a drop in the budget constraint as an individual crosses z^*, essentially rotating the budget constraint through the origin. From our optimization perspective, this again leads to bunching from those with pre-notch earnings $z \in (z^*, z^* + \Delta z^*]$. The marginal buncher is now the individual tangent to both the linear tax schedule at $z^* + \Delta z^*$ and to the interior of the notched schedule at some lower level z_I, while at the same time being indifferent between z_I and z^*.

4

Compared to the kink case, notches are more involved in two ways. First, they lead to a hole in the density between z^* and z_I, as depicted in sub-figure (b). While everyone with pre-notch earnings above the marginal buncher will also respond to the higher average tax rate (by finding their tangency point along the notched schedule), this can only be above z_I since that was the tangency point of the marginal buncher. Second, notches also create a dominated region $z \in (z^*, z_D]$ (marked by the red dashed line) where no individual wants to locate. At any point in this range, consumption is lower and effort cost is higher than at z^*, placing individuals at lower indifference curves and making them strictly worse off. Hence, it is never optimal to locate there. Note that the level of z_D is assumption-free as it only depends on the size of z^*, t_0 and t_1.

Frictions attenuating bunching

In reality, individuals may locate in the dominated region due to optimization frictions stopping them from moving to z^*. This will be evident from a visual inspection of the density. In this case, the observed bunching mass is:

$$B = \int_{z^*}^{z^* + \Delta z^*} (1 - \alpha) h_0(z) dz \approx (1 - \alpha) h_0(z^*) \Delta z^*$$

where α is the proportion of individuals who would optimize by bunching at z^* but cannot due to frictions (adjustment costs that must be incurred to change their level of z). The estimation of the marginal buncher's earnings level under the counterfactual linear constraint scenario is therefore modified to take this into account.

From bunching at notches to elasticities

As with kinks, we can estimate both a reduced-form approximation of the elasticity, or rely on functional form assumptions on preferences. The reduced-form approximation is more involved than in the case of kinks, because we now need to convert the average tax to an implicit marginal tax rate that would have generated an equivalent bunching response. Denoting the marginal tax rate in this hypothetical scenario by t^*, and using the trapezoid rule, we get the following approximation:

$$t^* \approx t + (2 + \Delta z^*/z^*) \frac{z^* \Delta t}{\Delta z^*}$$

which results in the following reduced form elasticity:

$$e = \frac{\Delta z^*/z^*}{\Delta t^*/(1 - t^*)} \approx \frac{1}{2 + \Delta z^*/z^*} \frac{(\Delta z^*/z^*)^2}{\Delta t/(1 - t)}$$

For further details on this reduced-form approach, see Kleven's (2018) note.

The second approach is to employ a parametric functional form for utilities, and use the condition that the marginal buncher is indifferent between locating at z^* and z_I and hence equate the two utilities. As before, we can parametrize the utility function using the usual quasi-linear and iso-elastic form, which leads to the following utility levels at z^* and z_I:

$$U_{z^*} = (1 - t_0)z^* - \frac{n^*}{1 + \frac{1}{e}} \left(\frac{z^*}{n^*} \right)^{1 + \frac{1}{e}}$$

$$U_{z_I} = \left(\frac{1}{1 + e} \right) n^* (1 - t_1)^{1 + e}$$

Equating these and using the first-order condition $z = n^*(1 - t)^e$, we get the following condition:

$$\frac{1}{1 + \Delta z^*/z^*} - \frac{1}{1 + 1/e}\left(\frac{1}{1 + \Delta z^*/z^*}\right)^{1+1/e} - \frac{1}{1+e}\left(1 - \frac{t_1 - t_0}{1 - t_0}\right)^{1+e} = 0$$

This can then be solved numerically to recover the elasticity.

Estimating z_I One benefit of the parametric approach is that we can use it to also estimate z_I, i.e. the location of the indifference condition for the marginal buncher. Using again the first-order-condition and denoting the marginal buncher's ability by n_m, we have that $z^* + \Delta z^* = n_m(1 - t_0)^e$ and $z_I = n_m(1 - t_1)^e$. Thus,

$$z_I = (z^* + \Delta z^*)\left(\frac{1 - t_1}{1 - t_0}\right)^e$$

which is estimable using our estimates of e and Δz^*.

References

Allen, E.J, Dechow, P.M., Pope, D.G. and Wu, G. (2017) https://doi.org/10.1287/mnsc.2015.2417

Dharmapala, D. (2019) http://dx.doi.org/10.2139/ssrn.2817151

Dagostino, R. (2019) https://www.brookings.edu/wp-content/uploads/2019/07/Dagostino2019_muni.pdf

Chetty, R., Friedman, J.N., Olsen, T. and Pistafferi, L. (2011) https://doi.org/10.1093/qje/qjr013

Gelber, A.M., D. Jones and Sacks, D.W. (2020) https://doi.org/10.1257/app.20170717

Kleven, H. and Waseem, M. (2013) https://doi.org/10.1093/qje/qjt004

Kleven, H. (2018) https://www.henrikkleven.com/uploads/3/7/3/1/37310663/notchelasticity_kleven_2018.pdf

Mavrokonstantis, P. and Seibold, A. (2022) http://dx.doi.org/10.2139/ssrn.4127660

Saez, E. (2010) https://doi.org/10.1257/pol.2.3.180