ALTERNATIVE PARAMETRIC BUNCHING ESTIMATORS OF THE ETI

THOMAS ARONSSON, KATHARINA JENDERNY, AND GAUTHIER LANOT

ABSTRACT. We propose a maximum likelihood (ML) based method to improve the bunching approach of measuring the elasticity of taxable income (ETI), and derive the estimator for several model settings that are prevalent in the literature, such as perfect bunching, bunching with optimization frictions, notches, and heterogeneity in the ETI. We show that the ML estimator is more precise and likely less biased than ad-hoc bunching estimators that are typically used in the literature. In the case of optimization frictions in the form of random shocks to earnings, the ML estimation requires a prior of the average size of such shocks. The results obtained in the presence of a notch can differ substantially from those obtained using ad-hoc approaches. If there is heterogeneity in the ETI, the elasticity of the individuals who bunch exceeds the average elasticity in the population.

JEL Classification: C51; H24; H31;
Keywords: Bunching Estimators, Elasticity of Taxable Income, Income Tax.

1. INTRODUCTION

A large literature addresses the problem of estimating the elasticity of taxable income (ETI). The early literature focused on changes in income levels over time, eventually leading to instrumental variables (IV) regression approaches that essentially regress changes in taxable income on changes in the net-of-tax rate (Feldstein, 1995; Gruber and Saez, 2002; Weber, 2014). Recently, bunching approaches to measuring the ETI have evolved as an alternative to regression methods, following the seminal paper of Saez, (2010). Saez argues that if the taxable income responds to marginal taxation, then the observed income distribution depends on the net-of-tax rate. Therefore, the change in the net-of-tax rate at a kink point keeps some individuals from earning income above that kink. These individuals will create an excess mass (bunching) at the kink point, the size of which can identify the ETI.

The appeal of the bunching approach is that it circumvents endogeneity and weak instrument problems that are typical for the IV regression methods.[1] Instead, data containing a single cross-section with at least one change in the marginal net-of-tax

---

DEPARTMENT OF ECONOMICS, UMEÅ UNIVERSITY, SWEDEN

E-mail addresses: thomas.aronsson@umu.se, katharina.jenderny@umu.se, gauthier.lanot@umu.se.

[1]Aronsson et al., (2017) address some of these problems based on Monte Carlo simulations and show that IV regression methods to estimate the ETI can be heavily biased and/or imprecise.

rate is sufficient to identify the ETI in principle. Yet, the bunching approach faces practical problems. In particular, individuals may not bunch exactly at the kink, due to optimization frictions or income shocks that are outside their control. In that case, in order to estimate the ETI, it is necessary to compare the observed income distribution to a counter-factual distribution in the absence of the kink point. The literature relies on non-parametric methods such as polynomial smoothing based on histograms, that are typically not best in a non-parametric sense. This is because they rely on the visual identification of the bunching range, because grouped data can lead to a biased estimate using such methods, and because the non-parametric estimate is based on the observed income density with exception of the bunching range, which makes that range an out-of-sample prediction that is not necessarily fitted well.

We argue that parametric methods, the assumptions of which are fully described, are preferable to non-parametric methods, whose statistical consequences are opaque. This paper therefore provides a structural approach to the measurement of the ETI based on the bunching method. We propose a maximum likelihood estimation (MLE) based method to improve currently used approaches (but see Bertanha et al., (2018) for an alternative take on the same issues). Specifically, the proposed method is based on the log linear labor supply model with log-normal unobserved components which can be applied to the case of data with perfect bunching, as well as extended to account for imperfect bunching, or to account for the possibilities of notches, i.e., noncontinuous jumps in the tax due. The MLE approach has four advantages over procedures based on polynomial smoothing of the distribution around the kink.[2] First, it is transparent in terms of the underlying model of income formation, as well as in terms of functional form and distributional assumptions. Second, measurement error/ optimization frictions can be modeled explicitly, which allows for the estimation of their size as opposed to the visual determination of the bunching interval. As indicated above, this aspect is potentially very important, since individuals do not bunch exactly at the kink. Third, the ML estimator can be extended to include covariates and can thus control for specific characteristics of different types of taxpayers, while still assuming a common distribution of the unobserved income component or optimization frictions. Fourth, the approach is in principle more flexible, as the precise model can be adjusted and the number of behavioral margins of the underlying model can be increased. It is also applicable to non-convex budget sets, which can be exemplified by the presence of notches (see, e.g., Pudney, 1989, chapter 5, and Hausman, 1980. See also Kleven and Waseem, 2013 for a recent application).

We obtain four main results: First, the MLE estimator is more precise and likely less biased than the ad-hoc bunching estimators that are typically used in the literature. Second, in the case of bunching at kink points, the MLE estimation fits the data of several published papers very well, and produces results in a similar order of

---

[2]As advocated for example by Chetty et al., (2011) and Kleven and Waseem, (2013).

magnitude. Third, results obtained in the presence of a notch can differ substantially from those obtained using ad-hoc approaches. Last, if there is heterogeneity in the ETI parameter, the elasticity of the individuals who bunch exceeds the average elasticity in the population. Yet, as many empirical results for the ETI are very small, the estimated heterogeneity of the ETI is typically also small.

After presenting Saez's basic intuition of the bunching estimator of the ETI, and some modifications used in the literature (section 2) we discuss how our parametric estimator can improve the measurement in terms of efficiency. Then, we define the parametric estimators for the ETI that are adapted to the theoretical benchmark case of perfect bunching (Section 3), and several alternative environments such as and imperfect bunching (Section 4), bunching in the presence of a notch (Section 5), and heterogeneity in the ETI (Section 6). For each environment, we apply the method to the data of published studies in the field. Section 7 concludes.

2. Basic Intuition

We base our parametric estimator on a model used by Saez, (2010) which is the seminal paper on the bunching estimator, and similar to other models used in earlier research on the ETI. By using the same model, we can compare our results to those obtained in the literature. Furthermore, we can use Saez' (2010) original rationale to illustrate the intuition of the bunching estimator.

Suppose that the preferences over consumption, c, and work hours, h, can be represented by the quasi-linear utility function

$$u(c,h) = c - \frac{\eta}{1 + 1/\alpha}\left(\frac{h}{\eta}\right)^{1+\frac{1}{\alpha}}, \tag{1}$$

which yields the log linear labor supply function given the before-tax wage rate w, and the marginal tax rate, τ,

$$\ln h^*(w, \eta) = \alpha \ln w(1 - \tau) + \ln \eta. \tag{2}$$

In Equation (2), $\alpha > 0$ is the wage elasticity of the labor supply with respect to the marginal net wage rate, $w(1 - \tau)$, and $(-\eta) < 0$ is the disutility of work. Assuming that we cannot observe the wage rate but only the taxable earnings, we are interested in the optimal earnings function

$$\ln wh^* = \alpha \ln \tau^c + \ln \omega \equiv \ln z, \tag{3}$$

where $\tau^c \equiv (1 - \tau)$ is the marginal net-of-tax rate. Given our assumptions, $\omega = w^{(\alpha+1)}\eta$ describes an unobserved component of the individual's income.

As the preferences are quasi-linear and the labor supply is isoelastic, the behavioral response of earnings, z, to a marginal change in the net-of-tax rate is then defined as: $\frac{d\ln z}{d\ln \tau^c} = \alpha$. As earnings are the only source of income, α is the elasticity of taxable income (ETI) w.r.t. the net-of-tax rate.

Given the preferences in Equation (1), the optimal level of earnings z is an increasing function of both the unobserved income component ω and the net-of-tax rate τ^c. In a tax environment with one kink at the earnings level k, where the net-of-tax rate τ_2^c to the right of k is less than the net-of-tax rate τ_1^c to the left of k, the earnings distributions below and above the kink will follow different distributions: $z(\tau_1^c, \omega)$ below the kink and $z(\tau_2^c, \omega)$ above the kink. For a particular value of the unobserved income component $\hat{\omega}$, $z(\tau_1^c, \hat{\omega})$ exceeds $z(\tau_2^c, \hat{\omega})$, because the individual would choose a higher income level if she can keep more at the margin. Analogously, for a particular income level $\hat{z}$, there are two values of ω, ω_1 and ω_2, such that $\hat{z} = z(\tau_1^c, \omega_1) = z(\tau_2^c, \omega_2)$. For both income functions $z(\tau_1^c, \omega)|(z < k)$ and $z(\tau_2^c, \omega)|(z > k)$, there is an exact correspondence between income, z, and ω, as the marginal net-of-tax rate τ^c is well-defined. At the kink, however, a range of values of the unobserved income component are possible: define $\underline{\omega}$ such that $z(\tau_1^c, \underline{\omega}) \equiv k$ and $\bar{\omega}$ such that $z(\tau_2^c, \bar{\omega}) \equiv k$. Then, Saez, (2010) argues that individuals with an unobserved income component $\omega \in [\underline{\omega}, \bar{\omega}]$ choose $z = k$. If the density function of ω is smooth, then the observed income distribution will spike at the kink income. Furthermore, if the densities of the income distribution to the left and to the right of the kink income are known, then the size of the spike (the amount of bunching) identifies the ETI.

Let us formalize the corresponding bunching estimator of the ETI, still following Saez exactly. Denote the probability to be observed at the kink by B, and let $f(\omega)$ denote the density function for the unobserved income component. As the income level and the level of the unobserved component identify each other uniquely given the marginal net-of-tax rate, B can be expressed in terms of either the income level or the unobserved income component:[3]

$$
\begin{aligned}
B = \mathbb{P}[z = k] &= \mathbb{P}[z(\tau_1^c, \omega) \leq k \leq z(\tau_2^c, \omega)] \\
&= \mathbb{P}[\underline{\omega} \leq \omega \leq \bar{\omega}] \\
&= \int_{\underline{\omega}}^{\bar{\omega}} f(\omega) d\omega.
\end{aligned}
\tag{4}
$$

In order to estimate B, we would prefer to express B in terms of observables instead of unobservables, that is in terms of the income distributions $\tilde{f}(z; \tau_1^c)|(z < k)$ and $\tilde{f}(z; \tau_2^c)|(z > k)$, instead of the distribution $f(\omega)$. To this aim, it is instructive to note that in the absence of a kink (if the net-of-tax rate was τ_1^c for all levels of income), the tax payers who bunch would instead realize earnings in the interval $z \in [z(\tau_1^c, \underline{\omega}), z(\tau_1^c, \bar{\omega})] \equiv [k, (k+d)]$ of length d. We can then express B in terms of the density of the income distribution under the net-of-tax rate τ_1^c (i.e., the hypothetical income distribution in the absence of the kink):

---

[3]We can also express ω as the inverse function of the income level and the tax rate, such that $z(\tau^c, z^{-1}(\tau^c, k)) = k$.

FIGURE 1. Bunching
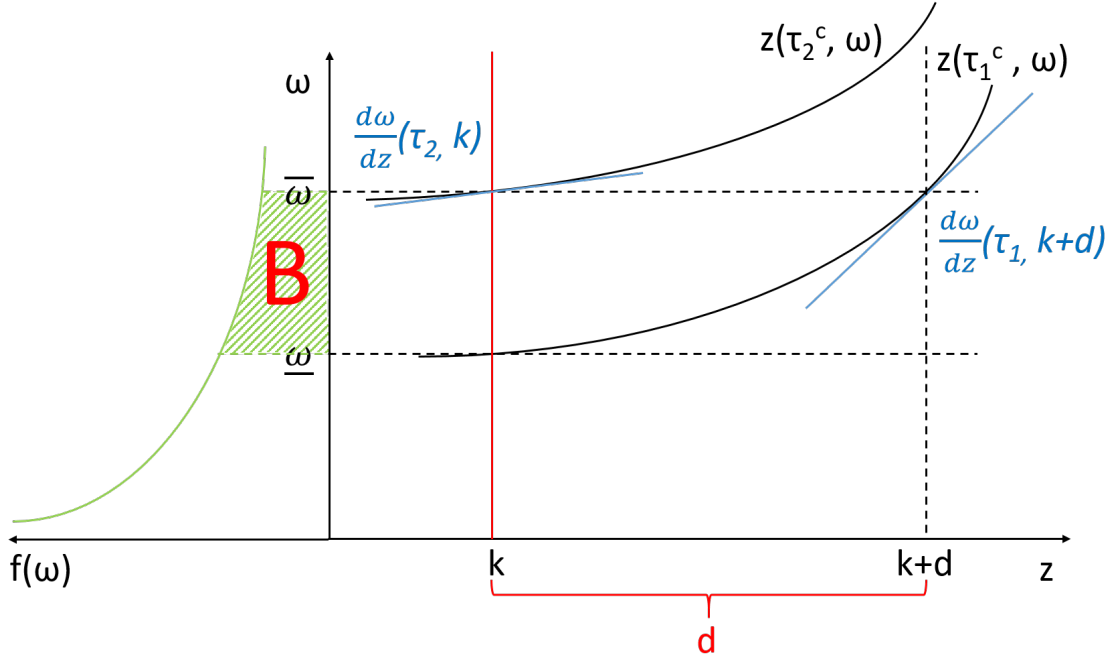
Note: ω describes the unobserved income component. $z(\tau_1^c, \omega)$ and $z(\tau_2^c, \omega)$ describe optimal earnings functions given the respective net-of-tax rate τ^c and the unobserved component ω. $\frac{d\omega}{dz}(\tau, z)$ describes the derivative of $\omega(\tau, z)$, i.e., the inverse earnings function, with respect to optimal earnings, evaluated at the earnings level z and under the tax rate τ.

$$B = \int_k^{k+d} \tilde{f}(z; \tau_1^c) dz, \qquad (5)$$

where $(k + d) = z(\tau_1^c, \bar{\omega})$ is the optimal level of earnings which the individual with the smallest disutility of work among those bunching at k would choose under the uniform net-of-tax rate τ_1^c.[4] Figure 1 illustrates the relation between the unobserved component ω and the income level z given the net-of-tax rate τ^c, and depicts B in terms of Equation (5). Saez' approach approximates Equation (5) (area $\hat{B}$ in Figure 1) using a trapezoidal approximation

$$\begin{aligned} B &= \int_k^{k+d} \tilde{f}(z; \tau_1^c) dz \\ &\approx d\big(\tilde{f}(k; \tau_1^c) + \tilde{f}(k + d; \tau_1^c)\big)\frac{1}{2}. \end{aligned} \qquad (6)$$

---

[4]The individual would prefer that quantity to k. Under the piecewise linear tax system, this level of earnings is not available, however, at the tax rate before the kink.

In the approximation in Equation (6), $\tilde{f}(k; \tau_1^c)$ is in principle observable, while d and $\tilde{f}(k + d; \tau_1^c)$ are not. To derive $\tilde{f}(k + d; \tau_1^c)$, note that the density function of the unobserved component $f(\omega)$ can be transformed to the density function of the income distribution $\tilde{f}(z)$ by using the inverse earnings function $\omega(\tau^c, z) = z^{-1}(\tau^c, z)$ such that

$$\tilde{f}(z; \tau^c) = f(\omega)\frac{d\omega}{dz}(\tau^c, z).$$

Therefore, the densities of the earnings distributions at a given level of the unobserved income component $\hat{\omega}$ are in a strict relation to each other:

$$\tilde{f}(z; \tau_1^c) = \tilde{f}(z; \tau_2^c)\frac{\frac{d\omega}{dz}(\tau_1^c, z)}{\frac{d\omega}{dz}(\tau_2^c, z)} \equiv \tilde{f}(z; \tau_2^c)\gamma(\tau_1^c, \tau_2^c, z).$$

The unobservable term $\tilde{f}(k+d; \tau_1^c)$ can thus be replaced by $\tilde{f}(k; \tau_2^c)\gamma(\tau_1^c, \tau_2^c, k)$: Saez' approximation can now be expressed in terms of the densities of the observed income distributions before and after the kink:

$$B \approx d\big(\tilde{f}(k; \tau_2^c)\gamma(\tau_1^c, \tau_2^c, k) + \tilde{f}(k; \tau_1^c)\big)\frac{1}{2}. \tag{7}$$

The sizes of d and γ can be constructed from observables using the model structure, in particular the optimal earnings level in Equation (3), which after substitution in Equation (7) gives Saez's Equation (5) exactly.[5]

Saez bases his estimate for α on the approximation in Equation (7) and replaces the unknown components such as the two densities by empirical analogs. Furthermore, the empirical results in Saez, (2010) are based on the assumption that bunching is imperfect, i.e., that it occurs in an interval around the kink rather than precisely at the kink, as displayed in Figure 2. Imperfect bunching complicates the estimation of B, because the bunching population is now spread out over an interval that contains also individuals that do not bunch. The bunching probability B (the green striped area in Figure 2) then equals the probability to be in the interval (both shaded areas), minus the probability to be one of the individuals in the interval that did not intend to be at k (the yellow shaded area). B is therefore referred to as excess bunching. This correction requires an assumption about the counter-factual density in the interval around k in the absence of a kink point. Saez, (2010) assumes that the counter-factual density on each side of the kink equals the density next to the bunching interval.

While the intuition of the bunching estimator is clear and provides a simple and theory-based method to estimate the ETI, it is empirically challenging. The estimator in Equation (7) demands the measurement of several quantities which

---

[5]with $z^{-1}(\tau^c, e) = \tau^{c\alpha}/e$; $z(\tau_1^c, z^{-1}(\tau_2^c, k)) - k = ((\frac{\tau_1^c}{\tau_2^c})^\alpha - 1)k$; $z_\zeta^{-1}(\tau^c, \zeta) = \frac{\tau^{c\alpha}}{\zeta^2}$; and $\gamma(\tau_1^c, \tau_2^c, k) = (\frac{\tau_1^c}{\tau_2^c})^{-\alpha}$

FIGURE 2. Excess Bunching
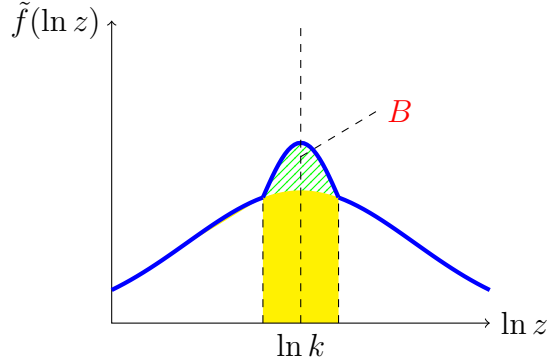
Note: The graph shows the density of observed earnings in the case of imperfect bunching. Shaded areas depict the bunching interval. The green striped area depicts excess bunching B.

may not be straightforward to obtain from a sample. If bunching is imperfect (which virtually all studies in the literature assume), the measurement of B requires both the identification of the bunching interval and an assumption about the counter-factual density around the kink. The bunching interval is typically determined visually, which is not necessarily best. The densities $\tilde{f}(k; \tau_2^c)$ and $\tilde{f}(k; \tau_1^c)$ are even less obvious to measure. The estimation problem is made harder since in the simple model we discuss above, we can only estimate these quantities from the observations to the left or to the right of the kink. In the case of imperfect bunching, the estimation must even rely on observations outside the bunching interval.

Several authors have proposed further refinements of Saez's first approach which are mostly concerned with the questions of estimating the counter-factual earnings distribution if bunching is imperfect, and estimating the earnings density on either side of the kink, i.e., $\tilde{f}(k; \tau_1^c)$ and $\tilde{f}(k; \tau_2^c)$. In particular, Chetty et al., (2009, 2011), fit a higher-order polynomial through the observed income density with exception of the bunching interval, which has been adopted my many other studies (for example Bastani and Selin, 2014 and Kleven and Waseem, 2013).[6] While this approach is more complicated than Saez' assumption of a constant density on either side of the bunching interval, it is not obvious to us that it leads to more reliable results, as the statistical consequences of the underlying assumptions are not entirely described. Furthermore, it relies on methodological choices that are not necessarily best in a non-parametric sense.

To demonstrate this, let us take a closer look at the polynomial approach of Chetty et al., (2009, 2011). The earnings data is first collected into relatively narrow

---

[6]Some studies use the income distribution from other years or groups of taxpayers that are not subject to the kink or notch as counter-factual income distribution.

bins to construct a histogram of the distribution of earnings before and after the kink. Excluding observations near the kink on either side, these observation counts become the dependent variable of a regression of a polynomial function of earnings (on the basis of simulations, Chetty et al., 2009, 2011 fit a polynomial of order 7). The prediction of the model is used to redistribute the proportion of the population around the kink and create a hypothetical earnings distribution that would apply in the absence of the kink. This produces a new counter-factual distribution on which they apply the previous procedure. This process is repeated until the parameters which determine the polynomial have converged. Bootstrapping is then used to measure the precision of the parameters. The final parameter estimates are eventually used to calculate the excess mass at, or around the kink, relative to the constructed counter-factual distribution which yields the estimates of the ETI. The procedure produces simultaneously a smoothed estimator of the histogram (because of the fit of a polynomial function of earnings) and an estimator of the counter-factual distribution of earnings in the absence of the tax kink.

This procedure is not based on any theoretical result together with the necessary technical conditions which would apply to it. Yet, it is not likely that it is best in a non-parametric sense because the initial binning of the data may have a cost in terms of integrated squared error over the non-excluded range of the earnings data relative to a method based on a kernel density estimator with a smoothing parameter (Bosq and Lecoutre (1987), Theorem 3.2 and remark 3.3). Whether this cost is large or not in practice is difficult to assess, but in the conventional cases and in large samples it can be sizable (the exact values depend on various constants which depend themselves on the details of the methods and on the unknown density. The order of magnitude ESI for large sample sizes clearly favors the kernel estimates). The costs of estimation related to the construction of the counter-factual distribution based on the smoothed histogram are beyond our discussion here. However, the method clearly does embody a particular trade off between the bias and the variance of the estimated earnings density, which in this instance may favor a reduction of the variance (i.e., favor smoothness) at a cost in terms of bias. Finally the bias of the estimated counter-factual density around the kink is likely to increase further since the data near the kink is essentially ignored and does not contribute to the construction of the counter-factual.

The bunching framework has also been adjusted to the case of notches, which create non-convexities in the individual's budget set. At a notch, the marginal tax rate exceeds 100%, such that there is a certain range of taxable income above the notch in which net income is lower than net income at the notch. In this range, both leisure and net income would be lower than at the notch point, which is why it is typically assumed that no rational agent would choose to be in that range. In the presence of a notch, the tax system thus creates incentives for individuals to avoid an income range (k, e^+). Because of the notch however, the net-of-tax rate after the notch depends on the exact location of the earning level that is considered.

This means that we can not use the analogy between the statistical model and the economic model directly as we did in the case of a kink.

Kleven and Waseem, (2013) develop this point nicely. They assume that the tax schedule around the kink takes the following form of a pure proportional notch:

$$T(z) = \tau_1 z + \Delta\tau z \mathbf{1}_{[z>k]},$$

for $\Delta\tau > 0$. For increases in earnings beyond the kink k, the implied approximate marginal tax rate after the kink varies as well:

$$\tau_2(\Delta z) \equiv \frac{T(k+\Delta z) - T(k)}{\Delta z} = \tau_1 + \Delta\tau + \Delta\tau \frac{k}{\Delta z},$$

for $\Delta z > 0$. For relatively small values of Δz this rate can be significantly larger than 1. This arises since the notch leads to large marginal tax rate for small increases in earnings. The approximate net-of-tax rate beyond the notch takes the form:

$$\tau_2^c(\Delta z) \equiv 1 - \tau_2(\Delta z) = \tau_1^c - \Delta\tau - \Delta\tau \frac{k}{\Delta z}, \tag{8}$$

In the case of a notch, it is thus crucial to determine the distance Δz empirically. Kleven and Waseem, (2013) determine Δz by the point where the excess mass before the notch equals the "missing mass" after the notch (compared with a counter-factual income that is determined using a polynomial as described above). For notches, the derivation of the counter-factual income density is thus even more crucial than in the case of a kink in a convex budget set.

There are thus few arguments why the non-parametric methods that are typically used in the bunching literature should produce reliable results. In particular, because they rely on a combination of visual determination of the bunching interval and a potentially large bias in the measurement of the counter-factual earnings density around the kink. We argue that a parametric estimation of the model outlined above is preferable to the non-parametric approach for at least three reasons: because the underlying assumptions about both the earnings process and the distribution of the unobserved income component are transparent; because these assumptions can be modified (in particular, the model can be extended to allow for several margins of income generation); and because additional parameters, such as the size of random earnings shocks in the case of imperfect bunching, can be estimated (rather than assumed).

In order to improve the measurement, we therefore propose a parametric estimator of the two densities and the bunching probability which can be estimated by the maximum likelihood method. We start by deriving the estimator for the case of perfect bunching, and then adapt the basic case to the case of imperfect bunching and notches. We aim here at keeping the baseline income generating process simple, in order to demonstrate the general methodology. Yet, the approach is suitable to entail more complicated income generating processes.

3. A Statistical Model of Perfect Bunching

Based on the intuition in part 2, we derive a parametric model of the distribution of earnings in an interval that includes the kink, which allows for the estimation of the probability to be at the kink as well as for the estimation of the densities around the kink. Using this model, we show that in the case of perfect bunching, the Saez estimator is biased if the tax units at the kink are identified within an interval rather than at a precise income level (which is typical for the literature and can be a restriction of the available data), and that the bias can be substantial. We further show that using the suggested parametric model, it is possible to correct for that bias. This enables the researcher to use the perfect bunching formula even if only grouped data are available. We also show simulation results that suggest that using the parametric model considerably reduces the variance of the Saez estimator.

A log-normal specification of the model. We consider a model based on the preferences in Equation (1), where the unobserved income component is log normally distributed, such that $\ln \omega \sim \mathcal{N}(\mu, \sigma_\omega^2)$.[7] Using this distributional assumption as well as the inverse earnings function $\omega = z^{-1}(\tau^c, z)$, we can evaluate expression (4) precisely:

$$
\begin{aligned}
B &= \mathbb{P}[\ln z^{-1}(\tau_1^c, k) < \ln \omega < \ln z^{-1}(\tau_2^c, k)] \\
&= \int_{(\alpha \ln \tau_1^c - \ln k - \mu)/\sigma_\omega}^{(\alpha \ln \tau_2^c - \ln k - \mu)/\sigma_\omega} \phi(u) du, \\
&= \Phi[(\alpha \ln \tau_2^c - \ln k - \mu)/\sigma_\omega] - \Phi[(\alpha \ln \tau_1^c - \ln k - \mu)/\sigma_\omega] \\
&= \Phi[(\ln k - \alpha \ln \tau_1^c + \mu)/\sigma_\omega] - \Phi[(\ln k - \alpha \ln \tau_2^c + \mu)/\sigma_\omega],
\end{aligned}
\tag{9}
$$

where $\phi(\omega)$ is the standard normal density function and $\Phi(\omega)$ the standard normal distribution function. The last expression follows from the properties of the normal distribution. The approximation comparable to (7) now takes the simpler form (using the distribution of the logarithm of earnings instead of earnings directly):

$$
B \approx \alpha(\ln \tau_1^c - \ln \tau_2^c)\left(\phi(\frac{\alpha \ln \tau_1^c - \ln k - \mu}{\sigma_\omega}) + \phi(\frac{\alpha \ln \tau_2^c - \ln k - \mu}{\sigma_\omega})\right)\frac{1}{2\sigma_\omega}.
$$

Hence, provided we can measure the quantities B, $\phi(\frac{\alpha \ln \tau_1^c - \ln k - \mu}{\sigma_\omega})/\sigma_\omega$ and $\phi(\frac{\alpha \ln \tau_2^c - \ln k - \mu}{\sigma_\omega})/\sigma_\omega$, we obtain a direct estimate of α using Saez's approach:

$$
\hat{\alpha} = 2\frac{1}{\ln \tau_1^c - \ln \tau_2^c}\frac{\sigma_\omega B}{\phi(\frac{\alpha \ln \tau_1^c - \ln k - \mu}{\sigma_\omega}) + \phi(\frac{\alpha \ln \tau_2^c - \ln k - \mu}{\sigma_\omega})}.
\tag{10}
$$

We observe first that the estimation problem is local in nature, i.e., it concerns only the observations near or at a particular kink. The parametric model we provide

---

[7]The method described below is, of course, applicable also for other functional form assumptions.

here captures this feature. The model is based on the truncated normal distribution over a range of earnings $[\underline{z}, \bar{z}]$ which includes the kink k.

Since bunching is perfect, it will take place at k exactly. Yet, in practice, it may only be feasible to measure earnings to a relative precision of δ, where δ is a small and positive number.[8] We assume here that an observation is identified at the kink if it lies between $ke^{-\delta}$ and k. Then, over the range $[\underline{z}, \bar{z}]$, log earnings are distributed as follows:[9]

$$\begin{cases} \text{if } z \in [\underline{z}, ke^{-\delta}), \text{then } \tilde{f}_1(\ln z) = s\phi(s\ln z - \lambda_1)/\mathcal{P}[\underline{z}, \bar{z}], \\ \text{if } z \in [ke^{-\delta}, k], \text{then } B(\ln k) = (\Phi(s\ln k - \lambda_2) - \Phi(s(\ln k - \delta) - \lambda_1))/\mathcal{P}[\underline{z}, \bar{z}], \\ \text{if } z \in (k, \bar{z}], \text{then } \tilde{f}_2(\ln z) = s\phi(s\ln z - \lambda_2)/\mathcal{P}[\underline{z}, \bar{z}], \end{cases}$$

(11)

where $\mathcal{P}[\underline{z}, \bar{z}] \equiv \Phi(s\ln \bar{z} - \lambda_2) - \Phi(s\ln \underline{z} - \lambda_1)$ is the probability of being in the earnings range $[\underline{z}, \bar{z}]$, while s, λ_1, and λ_2 are parameters with $s > 0$. In the context of a normal distribution s is one over the standard deviation of the unobserved component σ_ω, and the the parameters λ_i are the ratios of the mean to the standard errors to the left and right of the kink. In the case of the isoelastic model of earnings we have $\lambda_i \equiv s\alpha \ln \tau_i^c - s\mu$ and $s \equiv 1/\sigma_\omega$.

If $\lambda_1 \geq \lambda_2$, the model allows for bunching at k for any positive δ. This condition must be true if taxable income responds negatively to taxation and if the net-of-tax rate to the left of the kink exceeds the net-of-tax rate to the right of the kink.[10]

The parameters s, λ_1, and λ_2 can be estimated by maximum likelihood given a sample in the interval $[\underline{z}, \bar{z}]$.

For a sample of n individual observations, z_i, in the interval $[\underline{z}, \bar{z}]$ and for a positive number δ given, the log-likelihood takes the form:

$$\begin{aligned} \ln \mathcal{L}_{B,n} = &\sum_{i \in I_-} \left(\ln s - \frac{1}{2}s^2(s\ln z_i - \lambda_1)^2\right) + \\ &\sum_{i \in I_+} \left(\ln s - \frac{1}{2}s^2(s\ln z_i - \lambda_2)^2\right) + \\ &n_k \ln(\Phi(s\ln k - \lambda_2) - \Phi(s(\ln k - \delta) - \lambda_1)) - \\ &n \ln(\Phi(s\ln \bar{z} - \lambda_2) - \Phi(s\ln \underline{z} - \lambda_1)), \end{aligned}$$

(12)

---

[8]For example, Bastani and Selin, (2014) and Kleven and Waseem, (2013) base their estimations on small earnings intervals with roughly $\delta = 0.5$. The model allows for δ to be zero, of course.

[9]In the case of a distribution like the log-normal distribution, the ratio of the distribution of z and of the distribution of its logarithm, $\ln z$, depends on z only and not on the parameters of the distribution. To clarify the presentation we describe the distribution of $\ln z$ only.

[10]If $\lambda_1 = \lambda_2$ the bunching is approximately equal to $s\phi(s\ln k - \lambda_2)/\mathcal{P}[\underline{e}, \bar{e}]$ whenever δ is small enough. Finally if $s\delta + \lambda_1 < \lambda_2$ (which requires that $\lambda_1 < \lambda_2$), the model above does not describe a probability distribution over the interval, i.e. the expression of the probability at the kink is negative.

where I_- and I_+ collect the observations such that $\ln z_i < \ln k - \delta$ and $\ln z_i > \ln k$ respectively. n_k denotes the number of observations such that earnings are between $\ln k - \delta$ and $\ln k$.

The maximum likelihood estimates of the parameters, $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{s}$, can be obtained together with an estimate of their (asymptotic) precision. We can then use the model structure[11] and identify α directly from the parameter estimates by:

$$\hat{\alpha}_{norm} = \frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\ln(\tau_1^c) - \ln(\tau_2^c)} \frac{1}{\hat{s}} \qquad (13)$$

Observe that this expression is simple: it suggests that we consider the difference between the latent (latent because it is conditional on being below or above the kink) mean log earning before the kink and the latent mean log earning after the kink and divide by the difference in the log of net-of-tax rates.

We have shown in related work that the precision of the ML estimator exceeds the precision of the original Saez estimator in a simulated environment (see our related Monte Carlo study Aronsson et al., 2017 for details). Figure 3 illustrates the increased precision that the ML estimator provides.

Does δ matter for the baseline Saez approximation? Instead of relying on the model structure, we can reproduce the Saez approximation from the maximum likelihood estimates by deducing estimates of the bunching probability B and the two earnings densities, $\tilde{f}_1(\ln k - \delta)$ and $\tilde{f}_2(k)$. If δ is known, we can of course correct for the imprecision caused by δ given the parameter estimates. Without a parametric estimation, however, the researcher would not be able to correct for δ. We ask here what error we would expect from an imprecise identification of B. In that case, the bunching probability is estimated by

$$\hat{B} = \frac{\Phi(\hat{s}\ln k - \hat{\lambda}_2) - \Phi(\hat{s}(\ln k - \delta) - \hat{\lambda}_1)}{\Phi(\hat{s}\ln \bar{z} - \hat{\lambda}_2) - \Phi(\hat{s}\ln \underline{z} - \hat{\lambda}_1)},$$

and of the densities at the edge of the bunching interval:

$$\hat{\tilde{f}}_1(\ln k - \delta) = \hat{s}\frac{\phi(\hat{s}(\ln k - \delta) - \hat{\lambda}_1)}{\Phi(\hat{s}\ln \bar{z} - \hat{\lambda}_2) - \Phi(\hat{s}\ln \underline{z} - \hat{\lambda}_1)},$$

$$\hat{\tilde{f}}_2(k) = \hat{s}\frac{\phi(\hat{s}\ln k - \hat{\lambda}_2)}{\Phi(\hat{s}\ln \bar{z} - \hat{\lambda}_2) - \Phi(\hat{s}\ln \underline{z} - \hat{\lambda}_1)}.$$

We can use these estimates to replace the theoretical expressions in Equation (10) to obtain an estimate of α based on the methodology Saez proposes

$$\hat{\alpha}_{SaezN} = \frac{2}{\hat{s}}\frac{1}{\ln \tau_1^c - \ln \tau_2^c}\frac{\Phi(\hat{s}\ln k - \hat{\lambda}_2) - \Phi(\hat{s}(\ln k - \delta) - \hat{\lambda}_1)}{\phi(\hat{s}(\ln k - \delta) - \hat{\lambda}_1) + \phi(\hat{s}\ln k - \hat{\lambda}_2)} \qquad (14)$$

---

[11]That is, recognize that the isoelastic model of earnings implies that $\lambda_i = s\alpha\ln \tau_i^c + s\mu$

FIGURE 3. Saez and lognormal bunching estimators
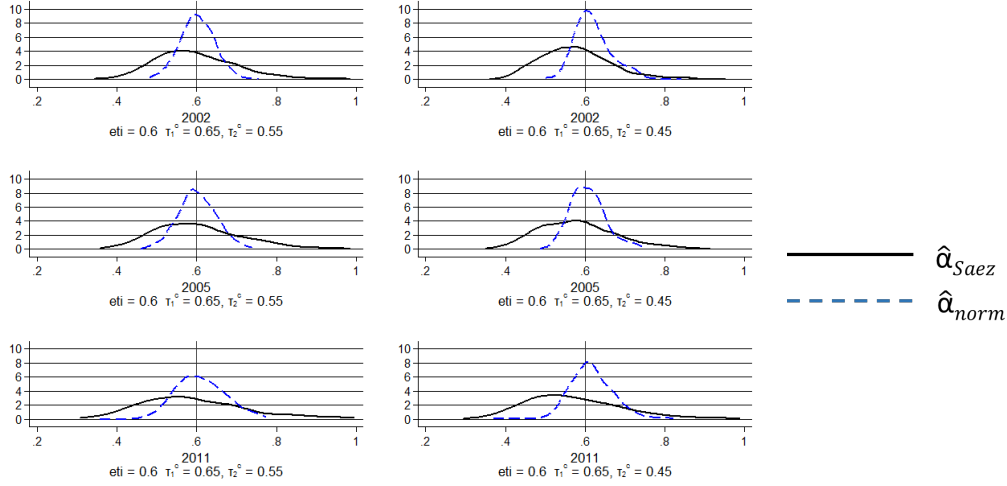
Note: Saez and log-normal bunching estimators. 1,000 replications with 10,000 individual observations over 12 years each, facing two tax environments: small kink: $\tau_1^c = 0.65$, $\tau_2^c = 0.55$; large kink: $\tau_1^c = 0.65$, $\tau_2^c = 0.45$. See Aronsson et al., (2017) for details and precise figures. Estimators: $\hat{alpha}_{Saez}$: Saez approximation (Equation 7);$\hat{alpha}_{norm}$: MLE based ETI bunching estimator (Equation 13).

The difference between the structural estimator $\hat{\alpha}_{norm}$ and the (uncorrected) Saez approximation $\hat{\alpha}_{SaezN}$ can be deduced if $(\hat{\lambda}_1 + \hat{s}\delta - \hat{\lambda}_2)$ is small:

$$\Phi(\hat{s}\ln k - \hat{\lambda}_2) - \Phi(\hat{s}(\ln k - \delta) - \hat{\lambda}_1) \approx$$
$$(\hat{\lambda}_1 + \hat{s}\delta - \hat{\lambda}_2)\frac{1}{2}\big(\phi(\hat{s}(\ln k - \delta) - \hat{\lambda}_1) + \phi(\hat{s}\ln k - \hat{\lambda}_2)\big).$$

Then we have

$$\hat{\alpha}_{SaezN} \approx \hat{\alpha}_{norm} + \frac{\delta}{\ln(\tau_1^c) - \ln(\tau_2^c)}, \tag{15}$$

which suggests that the two estimators differ little if δ is small and the difference in the tax rates is large. Yet, if the kink is small, such that τ_1^c and τ_2^c are close in size, small imprecisions can lead to large differences in the estimate of α. For example, if the tax rate at the kink increases from 20% to 22%, a small imprecision such as 0.5% of the kink income leads to an increase of roughly 0.2 in the estimate of α. While the bias can be substantial, the second term in the expression above does not depend on any unknown parameter. Hence it is possible to deduce the value of the structural estimator $\hat{\alpha}_{norm}$ from the $\hat{\alpha}_{SaezN}$. In the case of perfect bunching, a researcher using

the Saez approximation can thus correct for the bias created by binned data even without the use of ML estimation.

As an illustration of the method, we estimate the model of perfect bunching using the observed earnings distribution of self-employed individuals in Pakistan that Kleven and Waseem, (2013) have analyzed. Their data displays very precise bunching, which is thus suitable for the perfect bunching model framework. As the tax system in Pakistan features notches[12] instead of kinks, this application is illustrative at this point. We come back to a modified version of our model that explicitly accounts for both the notch and frictions in section 5.

We focus on on the kink at 400k Pakistani Rupees (PR).[13] Figure 4 illustrates the fit of the model in the earnings range around the 400kPR kink, for which the authors estimate an ETI between 0.06 and 0.19[14], using a polynomial approximation to estimate a counter-factual income distribution in the absence of a notch. The data in Figure 4 is grouped in small bins of width 2.5kPR. We argue that our parametric model reproduces the main features of the data, although the observed density is more variable than the model would suggest.

Assuming that $\delta = 0.5\%$, which roughly corresponds to the size of the income intervals the authors use relative to the notch income (2.5kPR/400kPR), we estimate the parameters of the distribution of earnings around the kink and obtain the estimated value $\hat{\alpha}_{norm} = 0.55$ (0.02). The difference between the estimates $\hat{\alpha}_{SaezN}$ and $\hat{\alpha}_{norm}$, i.e. the correction $\delta/(\ln \tau_1^c - \ln \tau_2^c)$, for this application is equal to 0.18, so that $\hat{\alpha}_{SaezN} \approx 0.55 + 0.18 = 0.73$; a direct calculation of $\hat{\alpha}_{SaezN}$ yields an almost identical value. Recall that as we do not account for the notch here, it is not surprising that our estimates are much larger than the estimates that Kleven and Waseem, (2013) provide, because we underestimate the incentive to undercut the notch.

4. Imperfect Bunching and Optimization Frictions

As discussed in Section 2, the literature typically assumes that bunching is imperfect. Individuals may not be able to aim perfectly at the kink and their earnings may vary in ways that they do not control. Yet, optimization errors are typically not explicitly modeled in the literature. Instead, it is assumed that bunching occurs in an interval around the kink. As discussed above, the methods used in the literature are problematic both because the bunching interval has to be determined visually,

---

[12]From the evidence Kleven and Waseem, (2013) present in Figure A.5, we can retrieve a measurement of their bunched data. This involves using a ruler to measure the distance from the horizontal axis and scaling that distance to the vertical axis. . We rescale the data to be densities instead of the number of tax payers, and can then fit the theoretical density we describe earlier to the observed one.

[13]We refer to their paper for further detail on the particularities of the income tax schedule in Pakistan.

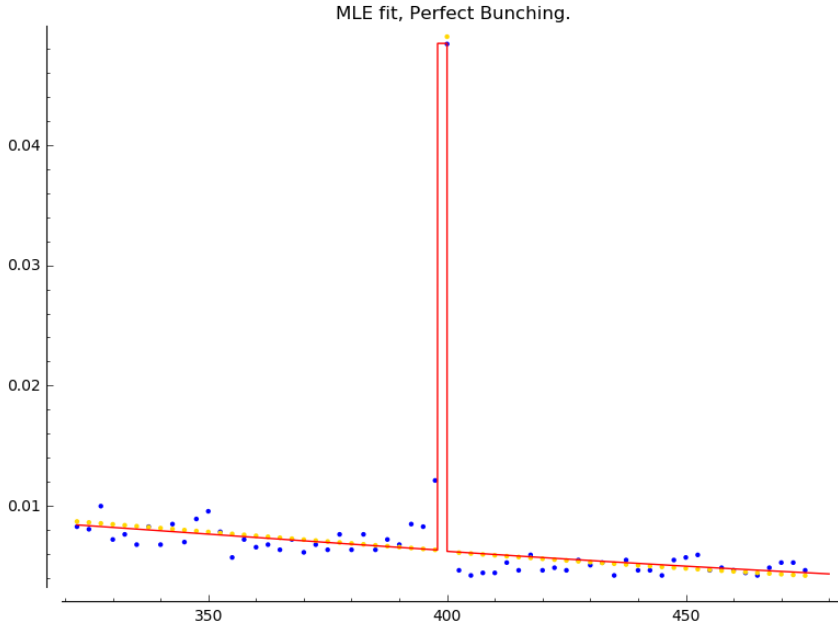[14]See Figure VI and Table II in Kleven and Waseem, (2013)

FIGURE 4. Kleven and Waseem (2013), $K=400$kPR, $\hat{\alpha}_{norm}=0.55(0.02)$.

Note: The data is not identical to Kleven and Waseem, (2013). The figure shows the observed data (blue dots) and the prediction of the distribution of earnings in the interval given the best fit parameters. In the case of perfect bunching, the height at the kink is a probability whereas elsewhere the height measures the density.

The maximum likelihood estimator for the ETI (see Equation (13), is 0.55 (0.02), while the estimator following Saez's approach assuming log normality is about 0.73. Kleven and Waseem, (2013) estimate a structural elasticity between 0.02 and 0.04.

and because the methods of determining a counter-factual density of earnings in the absence of a kink are not best in a non-parametric sense.

We therefore suggest a modification of the ML estimator of the ETI that directly models optimization frictions, such that observed earnings are a mixture of planned earnings and some (log-normal) noise. This allows us to estimate both the size and variance of the shock, and the size of the bunching interval using maximum likelihood.[15]

We start again from the model based on the preferences in Equation (1) which determines the optimal level of earnings as well as the relevant net-of-tax rate. In

_____

[15]In addition to imperfect bunching, the literature has modeled optimization frictions by assuming that a certain fraction of taxpayers (so-called non-responders) has a (short-term) elasticity of zero (Aronsson et al., 2017; Kleven and Waseem, 2013). This is particularly common in the case of notches, as a notch always creates an income range that a rational agent who values both consumption and spare time would not choose. This range can be used to identify the non-responders.

the single-kink case we have:

$$\begin{cases} \quad \text{if } \alpha \ln \tau_1^c + \ln \omega < \ln k : \ \ln z = \alpha \ln \tau_1^c + \ln \omega, \\ \quad \text{if } \alpha \ln \tau_2^c + \ln \omega > \ln k : \ \ln z = \alpha \ln \tau_2^c + \ln \omega, \\ \quad \text{if } \alpha \ln \tau_1^c + \ln \omega > \ln k \text{ and } \alpha \ln \tau_2^c + \ln \omega < \ln k : \ \ln z = \ln k. \end{cases} \tag{16}$$

The observed z^o earnings that the individual experiences in the end is determined by the planned earnings z and a shock ϵ such that

$$\ln z^o = \ln z + \ln \epsilon. \tag{17}$$

We assume that $\ln \omega$ and $\ln \epsilon$ are independently normally distributed. As before, $\ln \omega$ has mean μ and variance σ_ω^2. ϵ is a multiplicative shock to the planned earnings z, and is log normally distributed with mean 1, which implies that the logarithm of ϵ is distributed normally with $\ln \epsilon \sim \mathcal{N}(-\frac{\sigma_\epsilon^2}{2}, \sigma_\epsilon^2)$. We denote $\varsigma \equiv 1/\sigma_\epsilon$ and $\nu \equiv -\sigma_\epsilon/2 = -1/(2\varsigma)$.

In order to obtain maximum likelihood estimates for this modified model, we need to derive the density function of observed earnings. To that aim, we first derive the distribution function of observed earnings. We describe the distribution of observed earnings $\ln z^o$ given our assumptions concerning the distribution of planned earnings $\ln z$ (which depend on the unobserved component ω, the elasticity of taxable income α, and the net-of-tax rate τ^c) and the distribution of $\ln \epsilon$.

The distribution function of observed earnings at some level t equals the probability that observed earnings $z^o = z\epsilon$ are less than t, or, equivalently, that planned earnings z are less than $\frac{t}{\epsilon}$:[16]

$$H(t) \equiv \mathbb{P}[z\epsilon < t] = \mathbb{E}_\epsilon[\mathbb{P}[z\epsilon < t|\epsilon]] = \mathbb{E}_\epsilon[\mathbb{P}[z < \frac{t}{\epsilon}|\epsilon]]$$
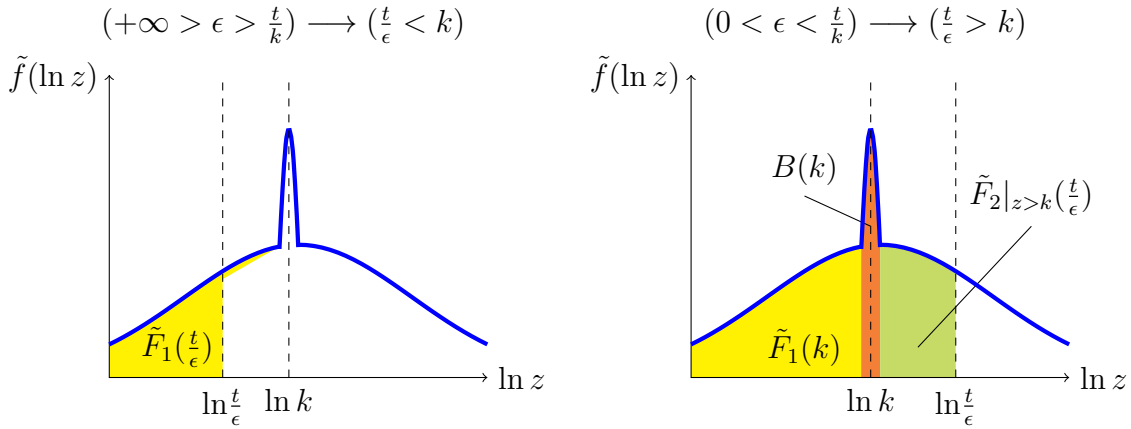
Given the distribution of the unobserved component of planned earnings ω, the distribution of planned earnings z depends on the net-of-tax rate, and is therefore different to the left and to the right of the kink. For ϵ larger than $\frac{t}{k}$, the distribution function of $\frac{t}{\epsilon}$ corresponds to the distribution function of planned earnings given the net-of-tax rate to the left of the kink. For ϵ smaller than $\frac{t}{k}$, the distribution function of $\frac{t}{\epsilon}$ corresponds to the distribution function of planned earnings given the net-of-tax

---

[16]Note that any positive level of observed earnings is consistent with any positive level of planned earnings, as the absolute value of ϵ is unbounded.

rate to the right of the kink.

$$H(t) = \int_0^{\frac{t}{k}} \mathbb{P}[z < \frac{t}{\epsilon}]g(\epsilon)d\epsilon + \int_{\frac{t}{k}}^{+\infty} \mathbb{P}[z < \frac{t}{\epsilon}]g(\epsilon)d\epsilon =$$
$$\underbrace{\int_0^{\frac{t}{k}} \tilde{F}_2(\frac{t}{\epsilon})g(\epsilon)d\epsilon}_{\frac{t}{\epsilon} \text{ above kink}} + \underbrace{\int_{\frac{t}{k}}^{+\infty} \tilde{F}_1(\frac{t}{\epsilon})g(\epsilon)d\epsilon}_{\frac{t}{\epsilon} \text{ below kink}}, \tag{18}$$

FIGURE 5. Distribution Function of Observed Earnings

Note: The graphs show the densities of planned earnings z. Shaded areas depict the probability that observed earnings are below a certain level t, given the size of the shock ϵ and the distribution of planned earnings z.

where $g(\epsilon)$ represents the probability density function of the shock ϵ, $\tilde{F}_1(x) = \mathbb{P}[z < x]$ is the distribution function of planned earnings under the net-of-tax rate τ_1^c, and $\tilde{F}_2(x) = \mathbb{P}[z < x]$ is the distribution function of planned earnings under the net-of-tax rate τ_2^c. We can decompose $\tilde{F}_2(x)$ using $\tilde{F}_1(x)$, the probability that planned earnings are at the kink $B(k) = \mathbb{P}[z = k]$, and the distribution function of planned earnings above the kink $\tilde{F}_2(x)|_{z>k} = \mathbb{P}[k < z < x]$. This is done in Equation (19). Figure 5 depicts that decomposition.

$$H(t) = \underbrace{\int_0^{\frac{t}{k}} \left(\tilde{F}_1(k) + B(k) + \tilde{F}_2|_{z>k}(\frac{t}{\epsilon})\right)g(\epsilon)d\epsilon}_{\frac{t}{\epsilon} \text{ above kink}} + \underbrace{\int_{\frac{t}{k}}^{+\infty} \tilde{F}_1(\frac{t}{\epsilon})g(\epsilon)d\epsilon}_{\frac{t}{\epsilon} \text{ below kink}} \tag{19}$$

The density of observed earnings $h(t)$ is such that $h(t) \equiv \dfrac{dH(t)}{dt}$. Some algebra yields:

$$h(t) = \frac{1}{k}B(k)\, g(\frac{t}{k}) + \int_0^{\frac{t}{k}} \tilde{f}_2(\frac{t}{\epsilon})g(\epsilon)\frac{d\epsilon}{\epsilon} + \int_{\frac{t}{k}}^{+\infty} \tilde{f}_1(\frac{t}{\epsilon})g(\epsilon)\frac{d\epsilon}{\epsilon}, \qquad (20)$$

where $\tilde{f}_1$ and $\tilde{f}_2$ describe the density functions of planned earnings given the net-of-tax rates τ_1^c and τ_2^c. Given the normality assumptions we have made, the density of observed earnings with imperfect bunching takes the form:[17]

$$\begin{aligned}
h(t) =& \frac{\varsigma}{t}\phi(\varsigma \ln t - \varsigma \ln k - \nu)B(k) + \\
& \frac{s\varsigma}{St}\phi(\frac{1}{S}(\varsigma s \ln t - (\lambda_2\varsigma + \nu s)))\Phi[\frac{1}{S}(\varsigma^2 \ln t - S^2 \ln k + \varsigma\nu - \lambda_2 s)] + \\
& \frac{s\varsigma}{St}\phi(\frac{1}{S}(\varsigma s \ln t - (\lambda_1\varsigma + \nu s)))\Phi[\frac{1}{S}(S^2 \ln k - \varsigma^2 \ln t + \lambda_1 s - \varsigma\nu)],
\end{aligned} \qquad (21)$$

where $S^2 = s^2 + \varsigma^2$, and $B(k) = \Phi[s \ln k - \lambda_2] - \Phi[s \ln k - \lambda_1]$ so that we require $\lambda_1 - \lambda_2 > 0$ to insure some bunching.

Figure 6 illustrates the effect of imperfect bunching on the distribution of earnings in a simple case where the variance of the shock to planned earnings is substantial.

We can now evaluate the amount of bunching around the kink following Saez's intuition. The imperfect nature of the bunching means that the earnings of individuals who aimed for the kink are now distributed around the kink, as shown in Figure 7. This is captured by the first term in Equation (21): $\frac{\varsigma}{t}\phi(\varsigma \ln t - \varsigma \ln k - \nu)F(k)$. Hence, the log normal model of the bunching error suggests that we consider an interval of earnings values which covers a large percentage of the realizations of the bunching error around the kink.

In terms of the first term in Equation (21), we may require that $-3 < \varsigma \ln z - \varsigma \ln k - \nu < 3$ to ensure the 99% of the bunching errors have been observed.[18] We should therefore consider all observed earnings in the range $\underline{z}_k \equiv \exp(-\frac{3-\nu}{\varsigma})k < z < \exp(\frac{3+\nu}{\varsigma})k \equiv \overline{z}_k$. Whatever the precise interval, as long as it contains approximately

---

[17]We derive the density in Appendix A

[18]This requires that the analyst has some ideas about the likely size of the variance of the shocks σ_ϵ, since it determines both ς and ν.

FIGURE 6. Perfect vs Imperfect Bunching
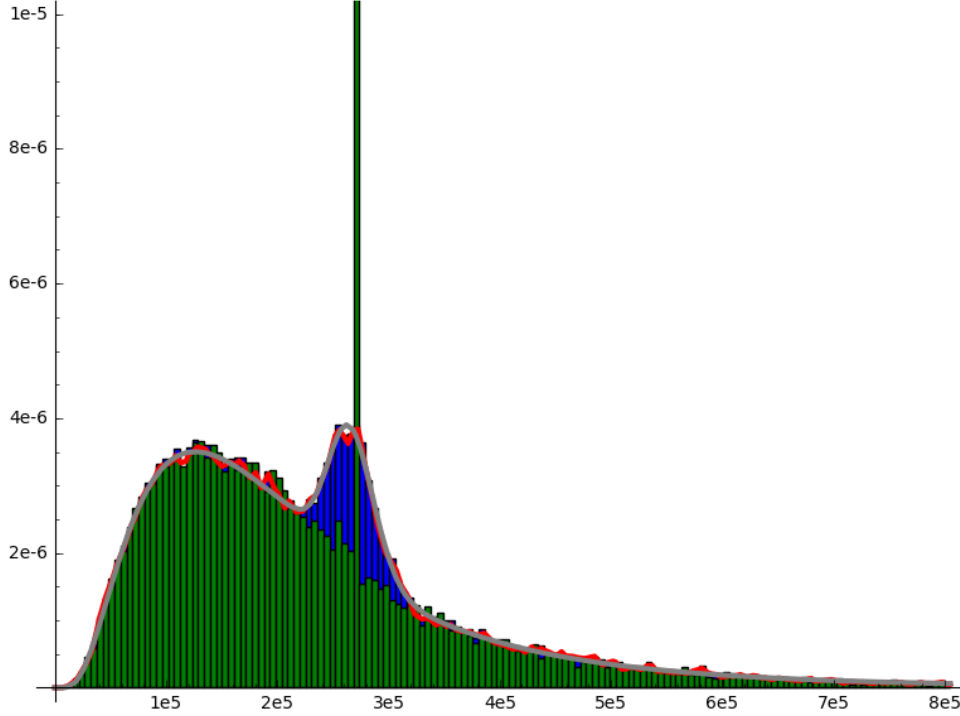
Note: The figure presents histograms from simulated data in the case of perfect bunching (in green) and imperfect bunching (behind the first histogram, in blue). The red line is the kernel density estimator when the data is generated under imperfect bunching while the grey line corresponds to the theoretical density.

all tax payers who aimed at the kink, i.e., $\int_{\underline{z}_k/k}^{\overline{z}_k/k} g(u)du \approx 1$ we find:

$$
\begin{aligned}
\mathbb{P}[\underline{z}_k < z < \overline{z}_k] &= \int_{\underline{z}_k}^{\overline{z}_k} h(t)dt \\
&\approx F(k) + \int_{\underline{z}_k}^{\overline{z}_k} \int_0^{\frac{t}{k}} f_2(\frac{t}{u})g(u)\frac{du}{u}dt + \int_{\underline{z}_k}^{\overline{z}_k} \int_{\frac{t}{k}}^{+\infty} f_1(\frac{t}{u})g(u)\frac{du}{u}dt \\
&= F(k) + I_2 + I_1,
\end{aligned}
\tag{22}
$$

where I_2 and I_1 are shorthand for the last two terms in the previous expression. We can understand I_2 as the proportion of observations such that desired earnings are beyond the kink and such that the bunching error is consistent with observed earnings in the observation range around the kink. A similar interpretation can be given for I_1.

FIGURE 7. Decomposing the observed density $h(t)$
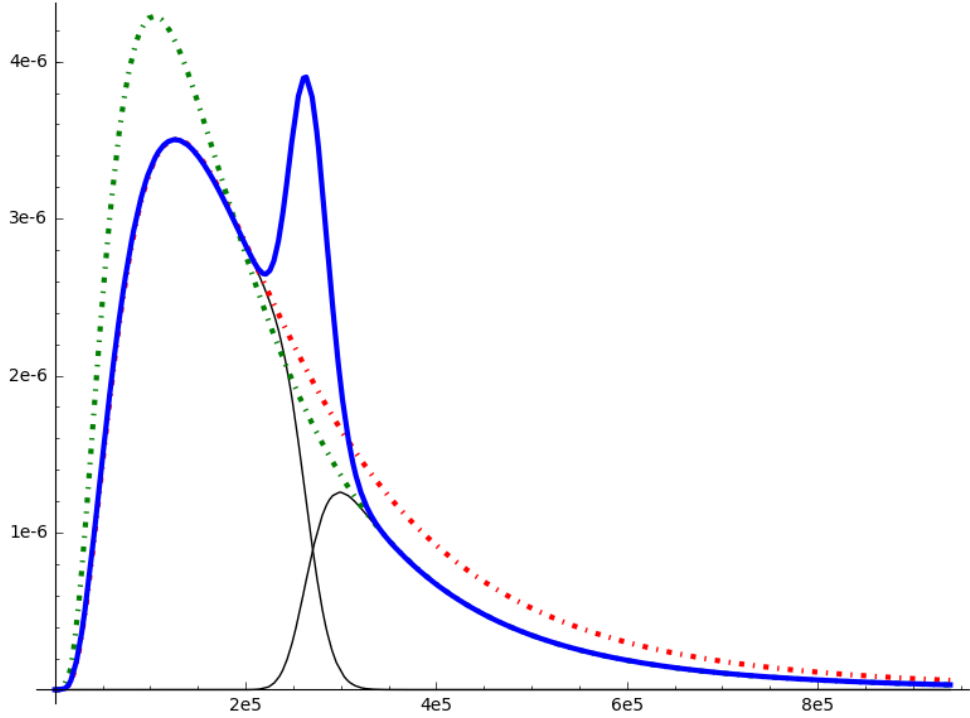
Note: The figure shows the theoretical density $h(t)$ in dark blue. The dashed and dotted lines correspond to the density of desired earnings assuming that the net-of-tax rates are τ_1^c (in red) or τ_2^c (in red). The second and third term in the expression of $h(t)$ are represented in black.

We can then understand Saez's expression for excess bunching since approximately:

$$F(k) \approx \mathbb{P}[\underline{z}_k < z < \overline{z}_k] - I_2 - I_1, \qquad (23)$$

where the RHS measures the amount of "net bunching". The approximation we used in Equation (4) can be used here too to approximate $B(k)$:

$$B(k) \approx \big(z(\tau_1^c, z^{-1}(\tau_2^c, k)) - k\big)\big(\tilde{f}(k; \tau_2^c)\gamma(\tau_1^c, \tau_2^c, k) + \tilde{f}(k; \tau_1^c)\big)\frac{1}{2} \qquad (24)$$

which provides a link back to Saez's suggested measurement procedure where an estimate of α is obtained by solving:

$$\alpha\big(z(\tau_1^c, z^{-1}(\tau_2^c, k)) - k\big)\big(\tilde{f}(k; \tau_2^c)\gamma(\tau_1^c, \tau_2^c, k) + \tilde{f}(k; \tau_1^c)\big)\frac{1}{2} =$$
$$\mathbb{P}[\underline{z}_k < z < \overline{z}_k] - I_2 - I_1. \qquad (25)$$

Saez proposes to estimate I_2 and I_1 from their empirical analog above or below the kink. Yet, this is difficult to rationalize in the context of imperfect bunching, since

the extent of the bunching error will be difficult to assess ex-ante. The estimation of the values of the density functions that appear on the LHS of Equation (25) seem even less obvious. In principle, the density of the bunching shock will make the estimation of the density of desired earnings next to the kink difficult, as this is the place where the density of "observed" earnings is the least likely to approach the density of "desired" earnings. It follows that the typically applied method of determining the bunching interval visually may not be efficient.

Estimation of the model with imperfect bunching. In the presence of imperfect bunching as described above, the solution is to estimate the parameters of the model using maximum likelihood over a subsample around the kink. With imperfect bunching, all earnings values have a positive density, and the likelihood is expressed in terms of $h(t)$ only. The probability to observe a given earnings value in the range around the kink, i.e., in some interval $[\underline{z}, \bar{z}]$ such that $\underline{z} < k < \bar{z}$, takes the form:

$$\mathbb{P}[\underline{z} < z < \bar{z}] = \int_{\underline{z}}^{\bar{z}} h(t)dt.$$

In general, the log-likelihood for a sample of n observations of individual earnings, z_i^o (in the interval $[\underline{z}, \bar{z}]$) is then simply:

$$\ln \mathcal{L}_{\text{IB},n} = \sum_{i=1}^{n} \ln h(z_i^o) - n \ln \mathbb{P}[\underline{z} < z^o < \bar{z}]. \tag{26}$$

The maximization of the likelihood above relative to its parameters λ_1, λ_2, s and ς will provide the MLE estimates $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{s}$.[19] Using the link between the statistical model and the economic structure, the estimator for the ETI will take the form:

$$\hat{\alpha}_{\text{IB},norm} = \frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\ln(\tau_1^c) - \ln(\tau_2^c)} \frac{1}{\hat{s}}. \tag{27}$$

In this case, we can estimate all the parameters of the model, in particular that of the variance of the shock to optimal earnings, and fully control for the effects of the shock ϵ when estimating the ETI. The ML estimation method is in principle extendable to settings with more kinks, which would allow the researcher to combine information of the whole observed income distribution and the tax system to improve the estimate of α.

---

[19] The likelihood in (26) is more difficult to evaluate than the likelihood with perfect bunching (see Equation 12). In addition to increased model complexity, Equation (26) also necessitates the estimation of an additional parameter, σ_ϵ. The latter necessitates, in turn, some density in an interval around the kink. However, even though the computation of the likelihood in (26) is less straightforward than in the perfect bunching case, they can nevertheless be implemented and maximum likelihood estimators for the parameters of the model can be obtained.
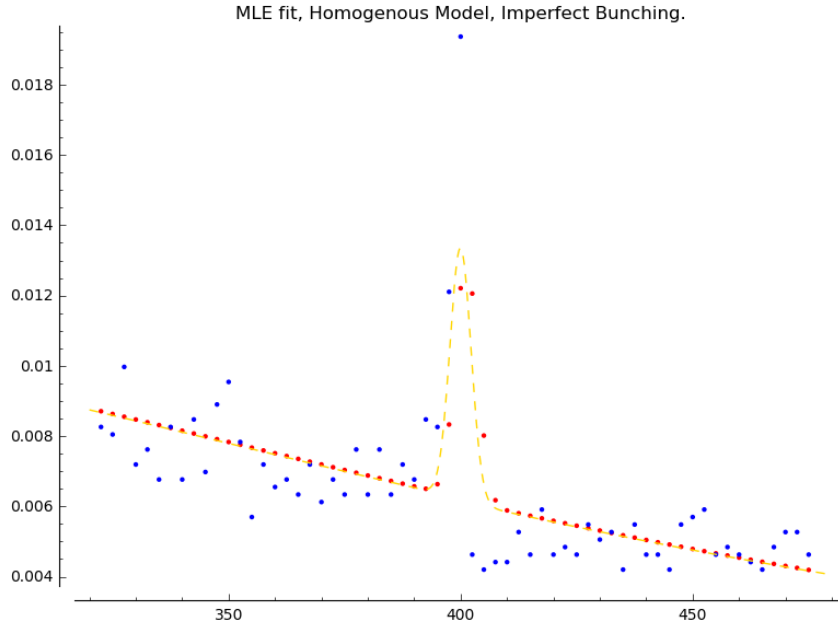
MLE fit, Homogenous Model, Imperfect Bunching.

FIGURE 8. Revisiting Kleven and Waseem (2013), K=400kPR, $\hat{\alpha} = 0.6(0.03)$. $\hat{\sigma}_\epsilon = 0.006(0.0003)$

Note: The figure shows the observed data (blue dots) and the prediction of the distribution of earnings in the interval given the best fit parameters (red dots). The dashed line draws the theoretical density at the MLE. Although the original data is identical, the scale of this graph is different from the one in Figure 4, since in the imperfect bunching case the height measures a density everywhere.

As in Section 3, we illustrate the method, by applying the model of imperfect bunching to (published) binned data of three studies. We first repeat the example from Section 3 using the data of Kleven and Waseem, (2013) (still not accounting for the notch), displayed in Figure 8. Then, we use the data of two studies that analyze imperfect bunching at a kink, and whose results are thus comparable to ours: Bastani and Selin, (2014), displayed in Figure 9, and Chetty et al., (2009, 2011), displayed in Figure 10. In all cases, we reproduce the estimate of $\hat{\alpha}$, using the same estimation interval $[\underline{z}, \bar{z}]$ around the kink as the authors, and apply the ML estimator with imperfect bunching defined in Equation (27) instead of the authors' original method based on visual detection of the bunching interval and polynomial smoothing. In the (illustrative) case of the data used by Kleven and Waseem, (2013), allowing for a shock increases the MLE estimate $\hat{\alpha}$ from 0.55(0.02) to 0.67(0.03). We estimate the standard deviation of the shock $\sigma_\epsilon = 0.006(0.003)$, and at the tax threshold of 400 this yields a standard deviation of the shock on earning of about 2.4kPR.

The ML estimates for the two studies that estimate bunching at kink points are in both cases in the same order of magnitude, but slightly higher than the

FIGURE 9. Imperfect bunching application 1
Bastani and Selin, (2014)

MLE fit, Homogenous Model, Imperfect Bunching.
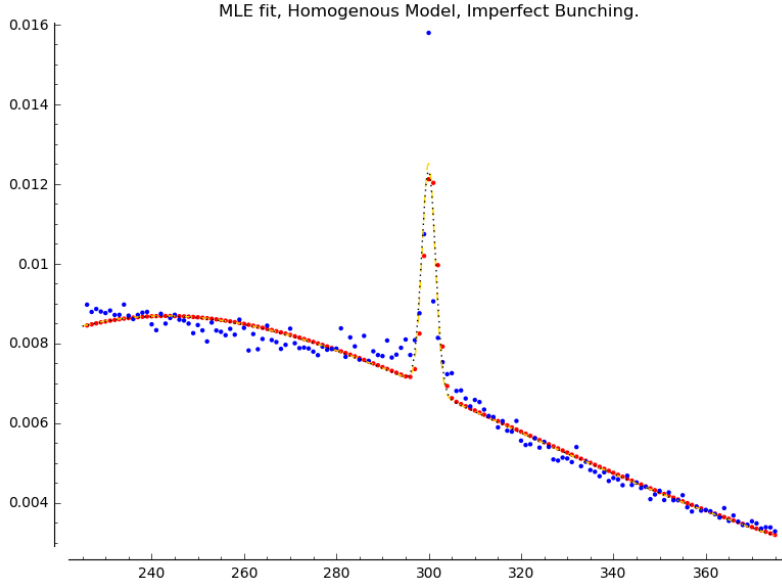
$K = 325\text{kSEK} \ (\approx 35\text{k} \ €), \ \hat{\alpha} = 0.028(0.0003), \ \sigma_\epsilon = 0.005(0.00007)$
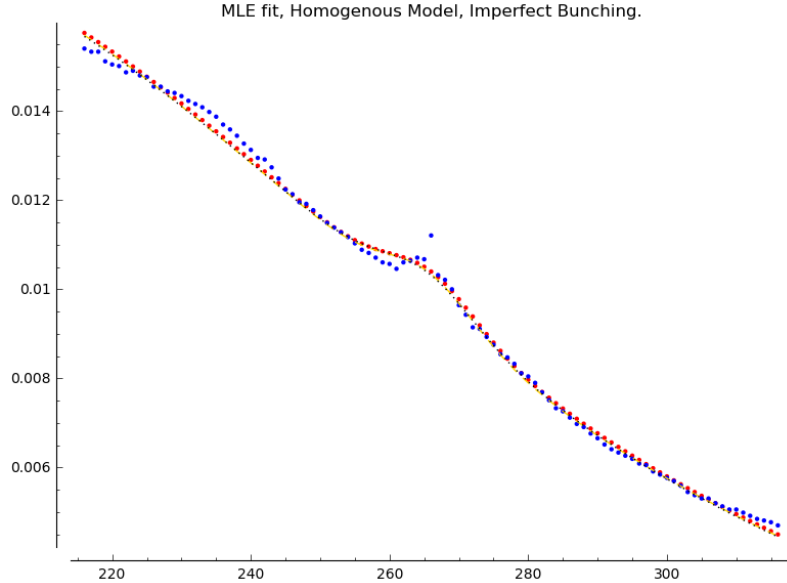
Note: The figure shows the observed data displayed in Bastani and Selin, (2014), figure 6a (blue dots) and the prediction of the distribution of earnings in the interval given the best fit parameters (red dots) based on the ML estimator in Equation (27). The dashed line draws the theoretical density at the MLE. The population are all self-employed tax payers in Sweden between 2000 and 2008, whose taxable income is in a range of 75k SEK around the first government tax kink point (at 325kSEK on average), at which the marginal tax rate increases by 20 percentage points. The data is grouped in income intervals of $1k$. The elasticity estimate in Bastani and Selin, (2014) is 0.024.

original estimate. In the case of Bastani and Selin, (2014), the MLE estimate is $\hat{\alpha} = 0.028(0.0003)$, which exceeds the original estimate of 0.024 by 16%. In the case of Chetty et al., (2009, 2011), the MLE estimate is $\hat{\alpha} = 0.017(0.0002)$, which exceeds the original estimate of 0.01 by 73%. Our estimates based on Bastani and Selin, (2014) data imply a standard deviation of the earnings shock at the tax threshold of about 1.5kSEK, while for Chetty et al., (2009, 2011) the standard deviation of the shock is 6.27kDKK. In this last case, Figure 10 illustrates the smaller concentration around the kink.

5. A STATISTICAL MODEL OF BUNCHING IN THE PRESENCE OF A NOTCH

The presence of a notch, i.e., a discontinuity of the budget constraint such that the tax due changes suddenly, inducing a spike in the marginal net-of-tax rate that exceeds 100%, gives an additional feature to bunching: it may create a "hole" in the

FIGURE 10. Imperfect Bunching Application 2
Chetty et al., (2009, 2011)

MLE fit, Homogenous Model, Imperfect Bunching.

$$K = 267.6\text{kDKK } (\approx 36\text{k€}), \hat{\alpha} = 0.017(0.0002), \sigma_\epsilon = 0.023$$.

Note: The figure shows the observed data displayed in Chetty, (2009), (blue dots) and the prediction of the distribution of earnings in the interval given the best fit parameters (red dots) based on the ML estimator in Equation (27). The dashed line draws the theoretical density based on the maximum likelihood estimation. The population are all wage earners in Denmark between 1994 and 2001, whose taxable income is in a range of 50k SEK around the largest tax kink point (at 267.6kDKK on average), at which the marginal tax rate increases by roughly 13 percentage points. The data is grouped in income intervals of 1k DKK. The elasticity estimate in Chetty et al., (2009, 2011) is 0.01.

density of observed earnings. A simple example of a notch is the existence of fixed cost of work associated with the decision to participate in the labor market. Under such a condition, no individual would be willing to start working if her earnings were less than the fixed costs associated with working. In fact, because individuals in general dislike work, we expect that the level of earnings required to start working is larger than the amount exactly equal to the fixed cost. This suggests that if all individuals face the same fixed costs, then the lower bound of support of the earnings distribution is larger than the fixed costs. Below the value of the fixed cost, the earnings density is equal to 0. The tax and benefit system potentially creates similar features of the budget constraint although for larger levels of earnings.

Perfect Bunching in the Presence of a Notch. The statistical model of bunching we present now can be understood as an extension of the model of perfect bunching we presented in Equation (11). We rely again on a parametric model (based on

the normal distribution) to describe the distribution of earnings in an interval of the range of observed earnings $[\underline{e}, \bar{e}]$. We assume that both the kink k and the region from k to some value e^+ with zero earnings density belong to this interval as well. Furthermore, the level e^+ becomes a parameter of the model which needs to be estimated. Hence we have: $\underline{e} < k \leq e^+ < \bar{e}$.

$$
\begin{cases}
\text{if } z \in [\underline{e}, ke^{-\delta}), \text{then } f_1(\ln z) = s\phi(s\ln z - \lambda_1)/\mathcal{P}[\underline{e}, \bar{e}], \\
\text{if } z \in [ke^{-\delta}, k], \text{then } F(\ln k) = (\Phi(s\ln e^+ - \lambda_2) - \Phi(s(\ln k - \delta) - \lambda_1))/\mathcal{P}[\underline{e}, \bar{e}], \\
\text{if } z \in (k, e^+], \text{then } f_0(\ln z) = 0, \\
\text{if } z \in (e^+, \bar{e}], \text{then } f_2(\ln z) = s\phi(s\ln z - \lambda_2)/\mathcal{P}[\underline{e}, \bar{e}].
\end{cases}
$$
(28)

It is easy to verify that the model defined this way is coherent (the probabilities are positive and sum to 1 over the range) as long as $\lambda_1 - \lambda_2 > s(\ln k - \delta - e^+)$. If $\tau_1^c \geq \tau_2^c$ then e^+ can take any values greater than k (but less than $\bar{e}$); if instead $\tau_1^c < \tau_2^c$ (i.e., the net-of-tax rate to the right of the kink is larger than the rate to the left of the kink), the coherency condition requires that $\ln k - \delta - \alpha \ln(\tau_1^c/\tau_2^c) < \ln e^+ < \bar{e}$. We can interpret $\ln e^+ - \ln k + \delta$ as the the "width of the hole" to the right of the kink and, when $\tau_1^c < \tau_2^c$, the condition above requires it to be larger than $-\alpha \ln(\tau_1^c/\tau_2^c) > 0$.

The parameters of this model may be difficult to identify, since we do not typically observe a range with zero density. For example, if the data is such that the smallest earnings observed above the kink is near the kink, then the maximum likelihood estimator of e^+ will be that value (or one slightly smaller), and the fitted model will be nearly identical to the model of perfect bunching in the absence of a notch.[20]

The model yields a likelihood which shares a similar structure to the likelihood we describe in Equation (12). The parameters of the model and the ETI in particular can be estimated using ML, following the same procedure as before.

Imperfect Bunching in the Presence of a Notch: random error. Following the same approach as we used to analyze imperfect bunching around a kink in Section 4, we can analytically derive the density of observed earnings in the presence of a notch assuming that the optimization friction takes the log normal form. We proceed as in the previous section and derive the density of observed earnings from first principles. Here the definition of the quantities changes to accommodate the hole in the desired distribution of earnings between k and e^+. Define $F_1(t) = \mathbb{P}[z < t]$ if $t < k$; $F(k) = \mathbb{P}[z = k]$ and $F_2(t) = \mathbb{P}[e^+ < z < t]$ if $t > e^+$ and let $F_2(e^+) = 0$. Finally note that $k < e^+ \Leftrightarrow \frac{t}{k} > \frac{t}{e^+}$.

The probability that observed earnings $z\epsilon$ are less than some level, say t, takes the

---

[20]It must be the case that any candidate value for e^+ to the right of the smallest value of earnings that is larger than k yields a likelihood that is infinitely smaller than than the likelihood evaluated for e^+ at or below the smallest value of earnings larger than k.

form:

$$H(t) \equiv \mathbb{P}[z\epsilon < t] = \mathop{\mathbb{E}}_{\epsilon}[\mathbb{P}[z\epsilon < t|\epsilon]] = \mathop{\mathbb{E}}_{\epsilon}[\mathbb{P}[z < \frac{t}{\epsilon}|\epsilon]] =$$

$$\int_0^{\frac{t}{e^+}} \left(\tilde{F}_1(k) + B(k) + \tilde{F}_2(\frac{t}{u}) \right) g(u)du+ \tag{29}$$

$$\int_{\frac{t}{e^+}}^{\frac{t}{k}} \left(\tilde{F}_1(k) + F(k) \right) g(u)du + \int_{\frac{t}{k}}^{+\infty} \tilde{F}_1(\frac{t}{u}) g(u)du.$$

Equation (29) relies on the fact that there are no observations between k and e^+ under perfect bunching, and therefore $\tilde{F}_2(e^+) = 0$. We then deduce the density of observed earnings in this case:[21]

$$h(t) = \frac{1}{k} B(k) g(\frac{t}{k}) + \int_0^{\frac{t}{e^+}} \tilde{f}_2(\frac{t}{u}) g(u) \frac{du}{u} + \int_{\frac{t}{k}}^{+\infty} \tilde{f}_1(\frac{t}{u}) g(u) \frac{du}{u}. \tag{30}$$

Our earlier analysis of the imperfect bunching case with log-normal distributions carries over to this context with a notch, and yields :

$$h(t) = \frac{\varsigma}{t} \phi(\varsigma \ln t - \varsigma \ln k - \nu) B(k)+$$
$$\frac{s\varsigma}{S_2 t} \phi(\frac{1}{S}(\varsigma s \ln t - (\lambda_2 \varsigma + \nu s_2))) \Phi[\frac{1}{S}(\varsigma^2 \ln t - S^2 \ln e^+ + \varsigma \nu - \lambda_2 s)]+ \tag{31}$$
$$\frac{s\sigma}{St} \phi(\frac{1}{S_1}(\varsigma s_1 \ln t - (\lambda_1 \varsigma + \nu s))) \Phi[\frac{1}{S}(S^2 \ln k - \varsigma^2 \ln t + \lambda_1 s - \varsigma \nu)],$$

where $B(k) = \Phi[s \ln e^+ - \lambda_2] - \Phi[s \ln k - \lambda_1]$ and we now require $s \ln k - s \ln e^+ < \lambda_1 - \lambda_2$ so that bunching arises with positive probability. Figure 11 illustrates the difference the notch would create relative to the earnings distribution in the absence of a notch (all else constant).

Again, let us apply the model to the data used by Kleven and Waseem, (2013), this time accounting for both imperfect bunching and the notch, and thus producing comparable estimates to theirs. We present the fit of the homogeneous model with imperfect bunching in Figure 11. The model clearly attempts to capture the lower density of earnings to the right of the kink. Our estimate of the first level of earnings which is chosen after the notch is 406kPR. Kleven and Waseem, (2013) in their paper calculate the dominated range of earnings to extend up to 410kPR. In our estimation, while we do not impose that restriction, the model discovers a minimum level of desired earnings which would be consistent with it. The distribution of the shock, and in particular its variance σ_ϵ^2, plays a substantial role here. It explains the imperfect bunching around 400kPR and the reduced density after the kink and around

---

[21]Note that the existence of a "hole" in the support of the density translates into distinct bounds of integration

FIGURE 11. Imperfect Bunching Density with or without a Notch
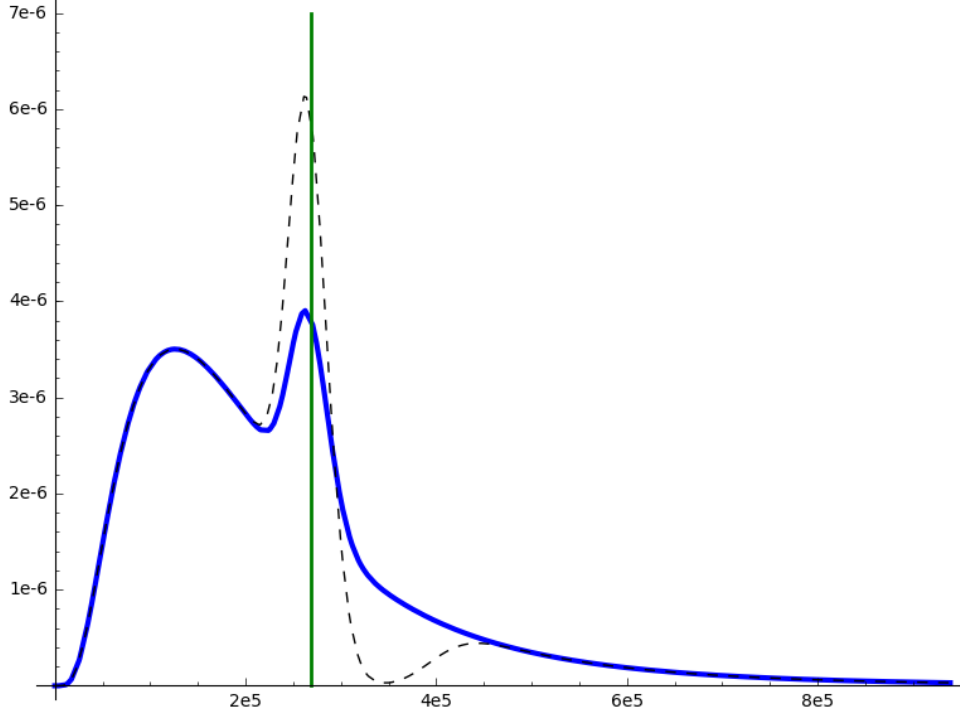
Note: The figure shows in dark blue the theoretical density $h(t)$ in the absence of a notch but with imperfect bunching. The dashed black line corresponds to the same density with a notch which creates a "hole" in the density without optimization friction between k and $1.5k$. The green vertical line indicates the position of the kink (or the discontinuity) at k. The variance of the optimization friction is the same for both densities. In this particular example, the length of the interval of the support with zero density is not determined by an optimizing agent (i.e. it is not consistent with a possible value of α derived from the comparison between λ_1 and λ_2). The log normal model of optimization friction means that the bunching reaches a local mode to the left of the position of the kink.

e^+ (the trough of the density to the right of 400kPR). The normality assumption on the distribution of ϵ requires that it is symmetric and invariant to the location. This specifically determines how the density increases before the kink and after e^+. Accounting for the possibility of a notch reduces the estimated value of the ETI to $0.41(0.03)$ which is substantially smaller than our previous estimate, yet more than twice as large as the authors' upper bound estimate of 0.194. As a consequence of the normality assumption, the size of the variance of the earnings shock σ_ϵ depends on the width of the bunching interval (which is narrow in this case). The model predicts that some density of the population would be observed directly after the kink. The
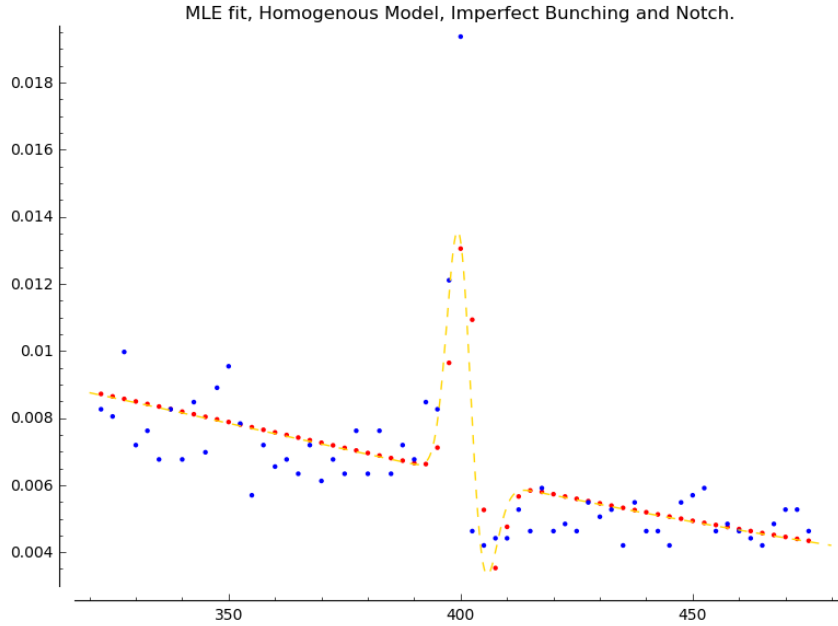
FIGURE 12. Revisiting Kleven and Waseem (2013), K=400kPR, $\hat{\alpha} = 0.41(0.03)$, $\hat{\sigma}_\epsilon = 0.00(0.0003)$, $\hat{e}^+ = 406(0.022)$.

Note: The figure shows the observed data (blue dots) and the prediction of the distribution of earnings in the interval given the best fit parameters (red dots). The dashed line draws the theoretical density at the MLE. Although the original data is identical, the scale of this graph is different from the one in Figure 4, since in the imperfect bunching case the height measures a density everywhere.

observed density directly after the kink is therefore interpreted as tax payers who planned to be at the kink, thereby increasing our elasticity estimate as compared to Kleven and Waseem, (2013). Also note that our estimate of e^+ is lower than the original estimate, as our model does not put too much weight on the comparatively low observed density to the right of the kink, given that the data is quite noisy in the rest of the income distribution.

6. HETEROGENEITY IN THE LOG-NORMAL MODEL

Saez, (2010) argues that the bunching approach applies even when the population is characterized by both an heterogeneous disutility of work and a heterogeneous ETI. In this case the bunching estimator measures the average of the ETI among those who are bunching. The log-normal model can be extended to provide a specific example

of this case.[22] To that effect, we consider the log linear model of optimal earnings given in Equation (3). In addition, we assume now that the parameter α, the ETI, is independently (from $\ln \omega$) and identically normally distributed in the population, so that $\alpha \sim \mathcal{N}(\bar{\alpha}, \sigma_\alpha^2)$, for σ_α^2 "small" enough. That is we are assuming that α is distributed among individuals around a mean value $\bar{\alpha}$ but we are requiring that $\mathbb{P}[\alpha < 0]$ is close to zero. In what follows we will set $\alpha = \bar{\alpha} + \tilde{\alpha}$, with $\tilde{\alpha} \sim \mathcal{N}(0, \sigma_\alpha^2)$.

Intuitively, since the response to taxation is heterogeneous, more responsive individuals, i.e., individuals with values of α larger than the average, are likely to be located below the kink, while individuals with values of α less than the average are likely to be located above the kink.[23] The selection/sorting is not exact since the disutility of work and the individual wages introduce additional sources of randomness that compete to determine the observed level of earnings.

The variability of α modifies the components of the model, and in particular the expression of the probability to observe a tax payer at the kink, which becomes:

$$B = \mathbb{P}[z = k] = \mathbb{P}[z(\tau_2^c, \omega) < k \leq z(\tau_1^c, \omega)]$$
$$= \mathbb{P}[u_1 \geq \ln k - \bar{\alpha} \ln \tau_1^c - \mu, u_2 < \ln k - \bar{\alpha} \ln \tau_2^c - \mu],$$

where $u_1 \equiv \tilde{\alpha} \ln \tau_1^c - \ln \omega - \mu$ and $u_2 \equiv \tilde{\alpha} \ln \tau_2^c - \ln \omega - \mu$.

Allowing for some heterogeneity in the parameter α has several consequences: it generates heteroscedasticity since the variance of the log-earnings vary depending on whether the observation is to the left or the right of the kink. To the left of the kink, the variance of the log-earnings is equal to $\sigma_\alpha^2(\ln \tau_1^c)^2 + \sigma^2$ while it is $\sigma_\alpha^2(\ln \tau_2^c)^2 + \sigma^2$ to the right of the kink. Furthermore, for a given individual the covariance between (latent log-)earnings to the right and to the left of the kink is $\sigma_\alpha^2 \ln \tau_1^c \ln \tau_2^c + \sigma^2$, which implies that the correlation between log-earnings on either side of the kink is different from one.

The probability to be at the kink can be expressed in terms of the bivariate normal distribution.

$$\mathbb{P}[u_1 \geq \ln k - \bar{\alpha} \ln \tau_1^c - \mu, u_2 < \ln k - \bar{\alpha} \ln \tau_2^c - \mu] =$$
$$\Phi\left[\frac{\ln k - \bar{\alpha} \ln \tau_2^c - \mu}{\sqrt{\sigma_\alpha^2(\ln \tau_2^c)^2 + \sigma^2}}\right] - \Phi_2\left[\frac{\ln k - \bar{\alpha} \ln \tau_1^c - \mu}{\sqrt{\sigma_\alpha^2(\ln \tau_1^c)^2 + \sigma^2}}, \frac{\ln k - \bar{\alpha} \ln \tau_2^c - \mu}{\sqrt{\sigma_\alpha^2(\ln \tau_2^c)^2 + \sigma^2}}, \rho\right], \quad (32)$$

where Φ_2 is the distribution function of the bivariate normal distribution, and $\rho \equiv \frac{\sigma_\alpha^2 \ln \tau_1^c \ln \tau_2^c + \sigma^2}{\sqrt{\sigma_\alpha^2(\ln \tau_1^c)^2 + \sigma^2}\sqrt{\sigma_\alpha^2(\ln \tau_2^c)^2 + \sigma^2}}$. ρ is positive for all values of the parameters if $\ln \tau_1^c$ and $\ln \tau_2^c$ share the same sign. Finally observe that the probability given in Equation (32) is always positive.

---

[22]Blomquist and Newey, 2017 show that if α is heterogeneous, the bunching estimator may not be well-defined if the density of the distribution of unobserved characteristics $f(\omega)$ is not smooth. By assuming a log-normal distribution, we assume away that case here.

[23]For this intuition to be correct it must be the case that there are individuals on either side of the kink.

Thanks to the normality assumptions, much of the analysis we provide earlier on applies equally to this heteroscedastic model. The model with nearly perfect bunching developed in Section 3 is easily derived in this heteroscedastic case, allowing for variances of the unobserved components which vary on either side of the kink and such that the probability to be at the kink is given exactly by the expression in Equation (32). Furthermore, the derivation of the density of earnings in the imperfect bunching case applies directly as it is presented in Appendix A. The term $F(k)$ which appears in Equation (A.1) is equal to the expression given in (32) and the parameters λ_1, λ_2, s_1 and s_2 are such that:

$$
\begin{aligned}
\frac{1}{s_1} &\equiv \sqrt{\sigma_\alpha^2 (\ln \tau_1^c)^2 + \sigma^2}, \\
\frac{1}{s_2} &\equiv \sqrt{\sigma_\alpha^2 (\ln \tau_2^c)^2 + \sigma^2}, \\
\lambda_1 &\equiv s_1 \bar{\alpha} \ln \tau_1^c + s_1 \mu, \\
\lambda_2 &\equiv s_2 \bar{\alpha} \ln \tau_2^c + s_2 \mu.
\end{aligned}
\tag{33}
$$

Given our model assumptions we deduce that $s_1 > s_2$ and $\frac{s_2}{s_1} \leq \rho < \frac{s_1}{s_2}$. Hence given a set of parameter estimates for $\bar{\alpha}$, μ, σ_α^2 and σ^2, we can derive an estimate of the expected ETI for the individuals attempting to locate exactly at the kink. As such, we wish to evaluate $\mathbb{E}[\alpha | u_1 \geq \ln k - \bar{\alpha} \ln \tau_1^c - \mu, u_2 < \ln k - \bar{\alpha} \ln \tau_2^c - \mu]$. Here we recognize that α is normally distributed and it is correlated with the unobserved components u_1 and u_2 since both depend on $\tilde{\alpha}$. Let

$$
\begin{aligned}
a &\equiv \ln k - \bar{\alpha} \ln \tau_1^c - \mu \\
b &\equiv \ln k - \bar{\alpha} \ln \tau_2^c - \mu
\end{aligned}
$$

In the Appendix C, we show that the mean ETI among those at the kink, $\mathbb{E}[\alpha | u_1 \geq a, u_2 < b]$, satisfies the expression:

$$
\begin{aligned}
\left\{ \mathbb{E}[\alpha | u_1 \geq a, u_2 < b] - \bar{\alpha} \right\} \mathbb{P}[u_1 \geq a, u_2 < b] = \\
\frac{\theta}{\ln \tau_1^c / \tau_2^c} \left\{ (\frac{1}{s_1} - \frac{\rho}{s_2}) \phi[s_1 a] \Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1 - \rho^2}}] + (\frac{1}{s_2} - \frac{\rho}{s_1}) \phi[s_2 b] \bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1 - \rho^2}}] \right\},
\end{aligned}
\tag{34}
$$

with $\theta \equiv 1 + \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \ln \tau_1^c \ln \tau_2^c$. We also show that individuals who wish to locate at the kink are likely to exhibit a larger response than the average in the population, i.e. we show that $\mathbb{E}[\alpha | u_1 \geq a, u_2 < b] \geq \bar{\alpha}$. In the limit, if σ_α^2 is equal to 0, then $\rho = 1$ and $\mathbb{E}[\alpha | u_1 \geq a, u_2 < b] = \bar{\alpha}$ which conforms with the homogeneous case .

The two applications that we used to illustrate the analysis, see Figures 9 and 10 , do not leave much space for any sizeable variability of α in the population. Under the normal model we have developed, for α to take positive value with a large

probability, we must limit α's standard deviation so that σ_α is significantly smaller than $\bar{\alpha}$, for example if $\sigma_\alpha \leq \bar{\alpha}/2.5$ and $\bar{\alpha} > 0$ more than 99% of the values of α are positive.

As a further illustration, we estimate the model with heterogeneity using data from Bastani and Selin, (2014). Starting from the maximum likelihood parameters estimated for the homogenous model with imperfect bunching the likelihood increases until $\cdot\sigma_\alpha = 0.0076$ leaving the maximum likelihood estimates of all other parameters essentially unchanged from their estimated value under the assumption of homogeneity . Using these parameters values we can then calculate the expected ETI at the kink, as described in equation (34). We estimate $\bar{\alpha}$ at 0.028 and we calculate that the average of the ETI at the kink is $\mathbb{E}[\alpha|u_1 \geq \ln k - \bar{\alpha} \ln \tau_1^c - \mu, u_2 < \ln k - \bar{\alpha} \ln \tau_2^c - \mu] = 0.03$. Accounting for heterogeneity in this particular instance does not produce significant differences between the mean ETI and the mean ETI at the kink. Our findings suggest therefore that there is little evidence supporting substantial ETI heterogeneity in this case.

7. Summary and Discussion

In this paper, we have presented a structural, parametric alternative to the bunching approach of measuring the ETI, where the excess mass of observations close to a tax kink is used for identification. Although the bunching approach is convenient (as it avoids several difficult problems characterizing the IV-regression approach to the ETI), the literature to date relies on more or less ad-hoc procedures of measuring the bunching range and the counter-factual density. The statistical properties of the prevailing non-parametric methods of identifying the bunching interval are not fully described, and there is no clear distinction between unobserved behavioral components and measurement/optimization errors. The latter is particularly problematic since individuals bunch in a neighborhood around the kink (and not exactly at the kink), and this excess mass around the kink is used for purposes of identification. Our parametric alternative is related to the model used by Saez (2010) in his seminal contribution to the bunching estimator (and subsequently used by other researchers). We use this model to characterize the preferences as well as the nonlinear budget constraint underlying the formation of income, and show how the model can accommodate measurement/optimization errors as well as non-convexities in the budget set (such as those created by notches in the tax system). The parameters of the model (including the fixed preference parameters as well as the parameters of the assumed distributions of the unobserved components) can be estimated simultaneously by using the maximum likelihood method.

Our approach has several advantages compared to the prevailing methodology of identifying the ETI based on bunching. One is the clear relationship between the underlying theory of income formation and the statistical problem to be solved. As such, the behavioral assumptions, functional form assumptions, and distributional assumptions are clearly stated, and their contributions to the statistical problem are

easy to describe and understand. This means, among other things, that account for the presence of measurement/optimization errors when estimating the size of the bunching interval. Since people seem to bunch in an interval around the tax threshold, it is key to distinguish between measurement/optimization errors and random components of the utility function. Furthermore, our parametric approach appears flexible enough and can be extended to more comprehensive models of income formation with several behavioral margins.

The take-home message of the paper is that the parametric approach provides a transparent alternative to the existing non-parametric methods, since the statistical model is derived from the underlying model of income formation. As such, the methodological changes proposed here are reminiscent of those in the labor supply literature in the late 1970s and, of course, motivated by the same desire to understand the various mechanisms behind income formation. Without such information, which requires a well-specified parametric model, very little can be said about the consequences of taxation for income formation and welfare. The role of the present paper is to provide a methodological route through which this is made possible. We would like to emphasize three results here. First, the maximum likelihood estimates are less biased and more precise than the bunching estimators based on ad-hoc non-parametric methods prevailing in the literature. Second, based on data from two published papers using the excess mass at a kink point for purposes of identification, our parametric estimates of the ETI are slightly higher than the estimates presented in these studies. However, by adding a notch to the model and applying the data presented in Kleven and Waseem (2013), our estimate of the ETI differs quite much from theirs. Third, by allowing for (unobserved) heterogeneity in the ETI, we find that individuals whose desired taxable income equals the income at the tax kink have a higher ETI than average.

Future research may take several directions, and we will briefly discuss two of them here. First, as mentioned above, it is interesting and arguably relevant to use the methodology proposed here to estimate more realistic models of income formation with several behavioral margins. This can be exemplified by a framework where the labor income and capital income are determined simultaneously through labor supply and saving behavior. Note that this issue is relevant regardless of whether labor income and capital income are taxed jointly or separately. Second, the methodological framework developed here is suitable for analyzing the consequences of complex tax reforms on income formation (as well as from the perspective of economic efficiency). In our view, this is one of the major advantages of a structural approach to income formation since a single measure of the ETI may not be a sufficient statistic for tax policy design.

References

Aronsson, T., K. Jenderny, and G. Lanot. *The Quality of the Estimators of the ETI.* Umeå School of Business and Economics, Umeå Economic Studies 955, 2017.

Bastani, S. and H. Selin. "Bunching and non-bunching at kink points of the Swedish tax schedule". In: *Journal of Public Economics* 109.C (2014), pp. 36–49.

Bertanha, M., A. H. McCallum, and N. Seegert. *Better Bunching, Nicer Notching.* mimeo. 2018.

Blomquist, S. and W. Newey. *The Bunching Estimator Cannot Identify the Taxable Income Elasticity.* Working Paper 24136. National Bureau of Economic Research, Dec. 2017.

Bosq, D. and J. Lecoutre. *Théorie de l'Estimation Fonctionnelle.* Ed. Economica, Paris. 1987.

Chetty, R. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods". In: *Annual Review of Economics* 1.1 (2009), pp. 451–488.

Chetty, R., J. Friedman, T. Olsen, and L. Pistaferri. *Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records.* Working Paper 15617. National Bureau of Economic Research, Dec. 2009.

— "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records". In: *Quarterly Journal of Economics* 126.2 (2011), pp. 749–804.

Feldstein, M. "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act". In: *Journal of Political Economy* 103.3 (1995), pp. 551–72.

Gouriéroux, C., A. Monfort, and Q. Vuong. *Statistics and Econometric Models.* Statistics and Econometric Models 2 volume set v. 1. Cambridge University Press, 1995.

Gruber, J. and E. Saez. "The elasticity of taxable income: evidence and implications". In: *Journal of Public Economics* 84.1 (2002), pp. 1–32.

Hausman, J. A. "The effect of wages, taxes, and fixed costs on women's labor force participation". In: *Journal of Public Economics* 14.2 (1980), pp. 161–194.

Kleven, H. J. and M. Waseem. "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan*". In: *The Quarterly Journal of Economics* (2013).

Pudney, S. *Modelling Individual Choice: The Econometrics of Corners, Kinks, and Holes.* Blackwell, Oxford and New York. 1989.

Saez, E. "Do Taxpayers Bunch at Kink Points?" In: *American Economic Journal: Economic Policy* 2.3 (2010), pp. 180–212.

Weber, C. E. "Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences". In: *Journal of Public Economics* 117.C (2014), pp. 90–103.

APPENDIX A. EARNINGS DENSITY WITH IMPERFECT BUNCHING

We develop here the general heteroscedastic distribution of the model with imperfect bunching. This means that the variance of (log-) earnings to the left of the kink is different from the variance to the right of the kink.

Starting point is the density of observed earnings $h(t)$:

$$h(t) = \frac{1}{k}F(k)g(\frac{t}{k}) + \int_0^{\frac{t}{k}} f_2(\frac{t}{u})g(u)\frac{du}{u} + \int_{\frac{t}{k}}^{+\infty} f_1(\frac{t}{u})g(u)\frac{du}{u}.$$

The first term corresponds to the distribution of observed earnings t given that desired earnings are located at the kink precisely. The other two terms correspond to observations such that desired earnings, i.e. t/u, are below or above the kink, but the shock is such that observed earnings are t exactly.

The last two terms take the general form:

$$\int_a^b f(\frac{t}{u})g(u)\frac{du}{u} = \int_{t/b}^{t/a} f(v)g(\frac{t}{v})\frac{dv}{v}$$

Given the normality assumptions, $f(v) = \frac{s}{v}\phi(s\ln v - \lambda)$ and $g(u) = \frac{\varsigma}{u}\phi(\varsigma\ln u - \nu)$ where $\phi()$ is the standard normal density function. In this case the integrand becomes:

$$f(v)g(\frac{t}{v})\frac{1}{v} = \frac{s\varsigma}{S}\frac{S}{vt}\phi(S\ln v - \frac{m}{S}))\phi(\frac{1}{S}(s\varsigma\ln t - (\lambda\varsigma + \nu s))),$$

where $S^2 = s^2 + \varsigma^2$, $m = \varsigma^2\ln t + \lambda s - \nu\varsigma$. Hence m depends on $\ln t$. Finally we can integrate:

$$\int_{t/b}^{t/a} f(v)g(\frac{t}{v})\frac{dv}{v} =$$
$$\frac{s\varsigma}{St}\phi(\frac{1}{S}(\varsigma s\ln t - (\lambda\varsigma + \nu s)))\big[\Phi[S\ln\frac{t}{a} - \frac{m}{S}] - \Phi[S\ln\frac{t}{b} - \frac{m}{S}]\big].$$

In terms of the standard normal p.d.f. and c.d.f., the second and third components in the earlier expression become:

$$\int_0^{\frac{t}{k}} f_2(\frac{t}{u})g(u)\frac{du}{u} = \int_k^{+\infty} f_2(v)g(\frac{t}{v})\frac{dv}{v} =$$
$$\frac{s_2\varsigma}{S_2 t}\phi(\frac{1}{S_2}(\varsigma s_2\ln t - (\lambda_2\varsigma + \nu s_2)))\big[1 - \Phi[S_2\ln k - \frac{m_2}{S_2}]\big],$$

and

$$\int_{\frac{t}{k}}^{+\infty} f_1(\frac{t}{u})g(u)\frac{du}{u} = \int_0^k f_1(v)g(\frac{t}{v})\frac{dv}{v} =$$
$$\frac{s_1\varsigma}{S_1 t}\phi(\frac{1}{S_1}(\varsigma s_1\ln t - (\lambda_1\varsigma + \nu s_1)))\Phi[S\ln k - \frac{m_1}{S_1}],$$

where $S_k^2 = s_k^2 + \varsigma^2$, and $m_k = \varsigma^2 \ln t + \lambda_k s_k - \nu \varsigma$ for $k = 1, 2$. This provides us with an analytical expression for the observed density of earnings given our model of imperfect bunching:

$$
\begin{aligned}
h(t) =& \frac{1}{k} F(k) g(\frac{t}{k}) + \int_0^{\frac{t}{k}} f_2(\frac{t}{u}) g(u) \frac{du}{u} + \int_{\frac{t}{k}}^{+\infty} f_1(\frac{t}{u}) g(u) \frac{du}{u} \\
=& \frac{\varsigma}{t} \phi(\varsigma \ln t - \varsigma \ln k - \nu) F(k) + \\
& \frac{s_2 \varsigma}{S_2 t} \phi(\frac{1}{S_2}(\varsigma s_2 \ln t - (\lambda_2 \varsigma + \nu s_2))) \Phi[\frac{1}{S_2}(\varsigma^2 \ln t - S_2^2 \ln k + \varsigma \nu - \lambda_2 s_2)] + \\
& \frac{s_1 \varsigma}{S_1 t} \phi(\frac{1}{S_1}(\varsigma s_1 \ln t - (\lambda_1 \varsigma + \nu s_1))) \Phi[\frac{1}{S_1}(S_1^2 \ln k - \varsigma^2 \ln t + \lambda_1 s_1 - \varsigma \nu)].
\end{aligned}
\tag{A.1}
$$

APPENDIX B. THE CONDITIONAL EXPECTATION OF A TRUNCATED NORMAL DISTRIBUTION

This is a technical result that can be found elsewhere in the literature. We describe it here for completeness sake.

Assume that the component of the vector $\mathbf{u} = (u_1, u_2)$ are distributed according to a bivariate normal distribution with zero means, unit variances and correlation ρ. Σ is the matrix which collects all the variance and covariance terms. We wish to find the expression for the moments:

$$
\begin{aligned}
\mathbb{E}[u_1 | \{a_1 \leq u_1 < b_1\} \cap \{a_2 \leq u_2 < b_2\}], \\
\mathbb{E}[u_2 | \{a_1 \leq u_1 < b_1\} \cap \{a_2 \leq u_2 < b_2\}].
\end{aligned}
$$

Denote $\Psi_{\mathbf{u}}(\mathbf{t})$ the characteristic function of the conditional distribution of $\mathbf{u}$ given $\{a_1 \leq u_1 < b_1\} \cap \{a_2 \leq u_2 < b_2\}$ evaluated at $\mathbf{t} = (t_1, t_2)$. The definition of $\Psi_{\mathbf{u}}(\mathbf{t})$ yields:

$$
\mathbb{P}[\{a_1 \leq u_1 < b_1\} \cap \{a_2 \leq u_2 < b_2\}] \Psi_{\mathbf{u}}(\mathbf{t}) =
$$
$$
\exp(\mathbf{t}' \Sigma \mathbf{t}) \int_{a_1}^{b_1} \int_{a_2}^{b_2} \exp(-\frac{1}{2}(\mathbf{u} - i\Sigma \mathbf{t})' \Sigma^{-1} (\mathbf{u} - i\Sigma \mathbf{t})) du_2 du_1.
$$

A change in variable such that $\mathbf{v}(\mathbf{u}, \mathbf{t}) \equiv \mathbf{u} - i\Sigma \mathbf{t}$, yields a relatively simple expression in terms of the bivariate normal distribution $\Phi_2(w, z, \rho)$:

$$
\begin{aligned}
\Psi_{\mathbf{u}}(\mathbf{t}) \propto \exp(\mathbf{t}' \Sigma \mathbf{t}) \big\{ & \Phi_2(v_1(b_1, \mathbf{t}), v_2(b_2, \mathbf{t}), \rho) - \Phi_2(v_1(a_1, \mathbf{t}), v_2(b_2, \mathbf{t}), \rho) - \\
& \Phi_2(v_1(b_1, \mathbf{t}), v_2(a_2, \mathbf{t}), \rho) + \Phi_2(v_1(a_1, \mathbf{t}), v_2(a_2, \mathbf{t}), \rho) \big\}.
\end{aligned}
$$

The derivatives of the characteristic function determine the moments of the distribution since:

$$
\frac{\partial \Psi_{\mathbf{u}}(\mathbf{0})}{\partial \mathbf{t}} = i \, \mathbb{E}[\mathbf{u}].
$$

Applying this property to the expression for $\Psi_{\mathbf{u}}$, we find that:

$$\mathbb{P}[\{a_1 \leq u_1 < b_1\} \cap \{a_2 \leq u_2 < b_2\}] \, \mathbb{E}[u_1 | \{a_1 \leq u_1 < b_1\} \cap \{a_2 \leq u_2 < b_2\}] =$$
$$- \phi[b_1] \left\{ \Phi\left[\frac{b_2 - \rho b_1}{\sqrt{1 - \rho^2}}\right] - \Phi\left[\frac{a_2 - \rho b_1}{\sqrt{1 - \rho^2}}\right] \right\} + \phi[a_1] \left\{ \Phi\left[\frac{b_2 - \rho a_1}{\sqrt{1 - \rho^2}}\right] - \Phi\left[\frac{a_2 - \rho a_1}{\sqrt{1 - \rho^2}}\right] \right\}$$
$$- \rho \phi[b_2] \left\{ \Phi\left[\frac{b_1 - \rho b_2}{\sqrt{1 - \rho^2}}\right] - \Phi\left[\frac{a_1 - \rho b_2}{\sqrt{1 - \rho^2}}\right] \right\} + \rho \phi[a_2] \left\{ \Phi\left[\frac{b_1 - \rho a_2}{\sqrt{1 - \rho^2}}\right] - \Phi\left[\frac{a_1 - \rho a_2}{\sqrt{1 - \rho^2}}\right] \right\}.$$
$$\text{(B.1)}$$

The expression can then be specialised to satisfy the bounds on the truncation intervals given in the text.

Appendix C. Conditional Expectation of the ETI in the Heteroscedastic Case

In this appendix, we give some details about the calculation of the expected value of the random parameter α as it is given in Equation (34).

We need to describe the the conditional expectation of $\tilde{\alpha}$ given u_1 and u_2, where $u_1 \equiv \tilde{\alpha} \ln \tau_1^c + \ln \omega - \mu$ and $u_2 \equiv \tilde{\alpha} \ln \tau_2^c + \ln \omega - \mu$. We set the variance of $\ln \omega - \mu$ to σ^2.

Since $\tilde{\alpha}$, u_1 and u_2 are jointly normally distributed with zero means, the conditional mean of $\tilde{\alpha}$ given u_1 and u_2 takes the form:

$$\tilde{\alpha} = \delta_1 u_1 + \delta_2 u_2,$$

where the parameters δ_1 and δ_2 are deduced from the variance covariance matrix of $\tilde{\alpha}$, u_1 and u_2.

Given our model assumptions, the latter takes the form:

$$\sigma_\alpha^2 \begin{bmatrix} 1 & \ln \tau_1^c & \ln \tau_2^c \\ \ln \tau_1^c & (\ln \tau_1^c)^2 + \frac{\sigma^2}{\sigma_\alpha^2} & \ln \tau_1^c \ln \tau_2^c + \frac{\sigma^2}{\sigma_\alpha^2} \\ \ln \tau_2^c & \ln \tau_2^c + \frac{\sigma^2}{\sigma_\alpha^2} & (\ln \tau_2^c)^2 + \frac{\sigma^2}{\sigma_\alpha^2} \end{bmatrix}.$$

δ_1 and δ_2 are then such that [24]:

$$\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} (\ln \tau_1^c)^2 + \frac{\sigma^2}{\sigma_\alpha^2} & \ln \tau_1^c \ln \tau_2^c + \frac{\sigma^2}{\sigma_\alpha^2} \\ \ln \tau_2^c + \frac{\sigma^2}{\sigma_\alpha^2} & (\ln \tau_2^c)^2 + \frac{\sigma^2}{\sigma_\alpha^2} \end{bmatrix}^{-1} \begin{bmatrix} \ln \tau_1^c \\ \ln \tau_2^c \end{bmatrix},$$

which yields:

$$\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \frac{1}{\ln \tau_1^c - \ln \tau_2^c} \begin{bmatrix} \theta \\ -\theta \end{bmatrix}.$$

[24]see for example Property B.43 in Gouriéroux et al., (1995).

The result in section (B) of the appendix, allows us to determine that:

$$\mathbb{E}[s_1 u_1 | u_1 \geq a, u_2 < b] \, \mathbb{P}[u_1 \geq a, u_2 < b] =$$

$$\phi[s_1 a] \Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1 - \rho^2}}] - \rho \phi[s_2 b] \bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1 - \rho^2}}],$$

and

$$\mathbb{E}[s_2 u_2 | u_1 \geq a, u_2 < b] \, \mathbb{P}[u_1 \geq a, u_2 < b] =$$

$$- \phi[s_2 b] \bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1 - \rho^2}}] + \rho \phi[s_1 a] \Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1 - \rho^2}}],$$

where in our context $a \equiv \ln k - \bar{\alpha} \ln \tau_1^c - \mu$, $b \equiv \ln k - \bar{\alpha} \ln \tau_2^c - \mu$, with $b > a$, and $\bar{\Phi}[x] \equiv 1 - \Phi[x]$. These two results together lead to

$$\mathbb{P}[u_1 \geq a, u_2 < b] \, \mathbb{E}[\alpha | u_1 \geq a, u_2 < b] = \mathbb{P}[u_1 \geq a, u_2 < b] \, \bar{\alpha} +$$

$$\frac{1}{\ln \tau_1^c / \tau_2^c} \frac{\theta}{s_1} \left(\phi[s_1 a] \Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1 - \rho^2}}] - \rho \phi[s_2 b] \bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1 - \rho^2}}] \right) - \qquad \text{(C.1)}$$

$$\frac{1}{\ln \tau_1^c / \tau_2^c} \frac{\theta}{s_2} \left(\rho \phi[s_1 a] \Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1 - \rho^2}}] - \phi[s_2 b] \bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1 - \rho^2}}] \right),$$

with $\theta \equiv 1 + \frac{\sigma_\alpha^2}{\sigma_\epsilon^2} \ln \tau_1^c \ln \tau_2^c$. This expression then simplifies to the expression given in the text in equation (34).

Finally we determine the sign of the correction for the selection among individuals at the kink. From equation (34) it is clear that the following expression determines the sign of this selection:

$$\mathcal{C} \equiv (\frac{1}{s_1^2} - \frac{\rho}{s_2 s_1}) s_1 \phi[s_1 a] \Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1 - \rho^2}}] + (\frac{1}{s_2^2} - \frac{\rho}{s_1 s_2}) s_2 \phi[s_2 b] \bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1 - \rho^2}}].$$

By construction, we know that $A \equiv \frac{1}{s_2^2} - \frac{\rho}{s_1 s_2}$ and $B \equiv \frac{1}{s_1^2} - \frac{\rho}{s_2 s_1}$ are such that: $A > 0 > B$ and $A + B \geq 0$ and therefore $A > |B|$. Denote $\kappa_1 \equiv s_1 \phi[s_1 a] \Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1 - \rho^2}}]$ and $\kappa_2 \equiv s_2 \phi[s_2 b] \bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1 - \rho^2}}]$ and both quantities are non negative. Hence we can rewrite $\mathcal{C} = B \kappa_1 + A \kappa_2$.

Observe furthermore that:

$$\frac{dF_k}{d \ln k} = \kappa_2 - \kappa_1,$$

which describe how the proportion of individuals at the kink varies with the position of the (log) kink, $\ln k$.

If $\kappa_2 > \kappa_1$, i.e. $\frac{dF_k}{d \ln k} > 0$, then $\mathcal{C}$ is positive.

If instead $\kappa_2 \leq \kappa_1$ and this is equivalent to:

$$\frac{s_2}{s_1} > \frac{\phi[s_1 a]\Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1-\rho^2}}]}{\phi[s_2 b]\bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1-\rho^2}}]},$$

and this allows us to find a lower bound for $\mathcal{C}$. Indeed we can express $\mathcal{C}$ as:

$$\mathcal{C} = \phi[s_2 b]\bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1-\rho^2}}]\left(Bs_1 \frac{\phi[s_1 a]\Phi[\frac{s_2 b - \rho s_1 a}{\sqrt{1-\rho^2}}]}{\phi[s_2 b]\bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1-\rho^2}}]} + As_2\right)$$

$$\geq \phi[s_2 b]\bar{\Phi}[\frac{s_1 a - \rho s_2 b}{\sqrt{1-\rho^2}}]\left(Bs_1 \frac{s_2}{s_1} + As_2\right)$$

$$= \kappa_2(A + B) \geq 0.$$