WILEY ═APPLIED ECONOMETRICS═

A maximum likelihood bunching estimator of the elasticity of taxable income

Thomas Aronsson | Katharina Jenderny | Gauthier Lanot

Department of Economics, Umeå School of Business, Economics and Statistics, Umeå University, Umeå, Sweden

Correspondence

Thomas Aronsson, Department of Economics, Umeå School of Business, Economics and Statistics, Umeå University, Umeå, Sweden.
Email: thomas.aronsson@umu.se

Summary

This paper develops a maximum likelihood (ML) bunching estimator of the elasticity of taxable income (ETI). Our structural approach provides a natural framework to simultaneously account for unobserved preference heterogeneity and optimization errors and for measuring their relative importance. We characterize the conditions under which the parameters of the model are identified and show that the ML estimator performs well in terms of bias and precision. The paper also contains an empirical application using Swedish data, showing that both the ETI and the standard deviation of the optimization friction are precisely estimated, albeit relatively small.

KEYWORDS

bunching, elasticity of taxable income, maximum likelihood

1 | INTRODUCTION

A large literature addresses the problem of estimating the elasticity of taxable income (ETI) with respect to the net-of-tax rate, which is a central parameter for tax policy design. The early ETI literature focused on changes in income levels over time, eventually leading to instrumental variables (IV) approaches that regress changes in taxable income on changes in the net-of-tax rate (Feldstein, 1995; Gruber & Saez, 2002; Weber, 2014). More recently, bunching approaches to measuring the ETI have evolved as an alternative to regression methods, following the seminal contribution of Saez (2010). The idea is that behavioral responses to marginal taxation lead to an excess mass (bunching) in the earnings distribution at or around a given tax kink, the size of which can be used to identify the ETI under certain assumptions. The purpose of the present paper is to develop a structural maximum likelihood (ML) method for the bunching approach to measuring the ETI, which has a number of advantages compared with other methodological approaches. This will be described in greater detail below.

The bunching approach is appealing, since it circumvents endogeneity and weak instrument problems that are typical for the IV methods. However, the bunching approach faces practical challenges. In particular, individuals may not bunch exactly at the kink, but in an interval around it, 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 that would have applied in the absence of the tax kink. The literature largely relies on non-parametric methods such as polynomial smoothing based on histograms to calculate this counter-factual distribution (e.g., Chetty et al., 2011; Kleven & Waseem, 2013). In a Monte Carlo study, Aronsson et al. (2022) show that the polynomial bunching estimator typically exhibits a downward bias, the size of which varies across settings and levels of the

ETI.[1] More specifically, the polynomial approach is potentially problematic because (i) it relies on visual identification of the bunching range and (ii) the nonparametric estimate is based on the observed income density excluding the bunching range, which means that optimization errors are ignored outside this bunching interval.

The recent literature has been exploring the distributional assumptions necessary to identify the ETI using the bunching method. Blomquist et al. (2021) show that the amount of bunching at a kink point cannot identify the ETI without any assumptions on the shape of the taxable income distribution. The intuitive argument is that one quantity, the number of individuals within a narrow interval around the kink point (the bunching mass), cannot identify two unknowns, the ETI and the (unrestricted) counter-factual density of taxable income at the kink.

The paper closest in spirit to ours is Bertanha et al. (2020), which also argues that nonparametric identification of both the counter-factual distribution and the ETI is impossible in general. In other words, the polynomial approach does not provide an assumption-free identification. They show that upper and lower bounds of the ETI can be identified if the steepness of the income density is bounded and suggest two semiparametric approaches to improve the bunching estimator, a Tobit estimator and a censored quantile regression estimator. Yet, none of these two estimators are designed for data that include optimization frictions which, arguably, are virtually always present. Therefore, Bertanha et al. again use polynomial smoothing to remove the optimization frictions from the data. This discussion suggests that identification of the ETI requires distributional assumptions.

This paper develops a structural model of bunching and a corresponding ML estimator. We follow earlier research in assuming that income formation is driven by a log-linear labor supply model with log-normal unobserved components.[2] The ML approach has several advantages over procedures based on polynomial smoothing of the distribution around the kink. First, and foremost, measurement errors/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 may not bunch exactly at the tax kink.

Second, the ML estimator can easily be extended to include covariates and can thus control for observable characteristics of different types of taxpayers, while still assuming common distributions of the unobserved income component and the optimization frictions, respectively. Finally, the flexibility of the approach goes beyond our particular applications, 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 nonconvex budget sets, which arise, for example, in the presence of notches in the budget constraint (see, e.g., Pudney, 1989 chapter 5 and Hausman, 1980; see also Kleven & Waseem, 2013, for a recent application), as well as to models where the ETI is heterogeneous in the population.[3]

Section 2 sets out the economic model, in which utility maximization yields a log-linear earnings function (the standard model of income formation in the ETI literature). We use this framework to present the bunching approach to the ETI. A corresponding ML estimator, which can simultaneously account for unobserved income heterogeneity and optimization errors, is presented in Section 3. We also describe the conditions under which the parameters of this model are identified as well as characterize the performance of the ML estimator compared with the conventional polynomial bunching estimator. The latter is accomplished through a Monte Carlo study. Specifically, we consider the performance of the normal–normal model, where the unobserved component of the optimal earnings as well as the optimization friction are log-normal, under different data generating processes (including the normal–normal, two-component mixture normal, and the Pareto distribution). The results show that the ML bunching estimator performs well in terms of bias and precision and that it outperforms the polynomial estimator.

In Section 4, we present an empirical application of the ML bunching estimator based on Swedish register data. In doing so, we also compare the ML estimate of the ETI with the estimate corresponding to the polynomial approach. Two samples of tax payers are analyzed: one referring to all tax units and the other to the self-employed, and the income concept is the taxable labor income. By using the ML bunching estimator, we obtain a point estimate of the ETI of about 0.007 when the data refer to all tax units, while the point estimate is roughly 10 times higher for the self-employed. In each case, both the ETI and the standard deviation of the optimization frictions are precisely estimated. The point estimates of the ETI based on the polynomial approach vary considerably depending on the choice of window when measuring the counter-factual distribution. Section 5 concludes the study. Technical details and proofs can be found in the supporting

---

[1]The results of Aronsson et al. (2022) also show that IV regression estimators are typically imprecise, and that both IV estimators and polynomial bunching estimators are outperformed by an Indirect Inference estimator new to the ETI literature.

[2]Saez (2010) and many subsequent studies rely on the same strict assumptions on individuals' preferences, as implied by the basic log-linear labor supply model, in order to identify the ETI.

[3]A model with heterogeneity in the ETI and the corresponding ML estimator are presented in a background working paper (Aronsson et al., 2021).

information appendix, where we also present an extension of the ML bunching estimator to a case with nonconvex budget sets caused by a notch.

2 | A MODEL OF BUNCHING: INTUITION AND CHALLENGES

In this section, we present the intuition behind the bunching estimator as well as some of the associated methodological challenges. We also reformulate the discussion into a coherent formal framework, which can be used in the parametric approaches taken in later sections.

The core intuition is straightforward: If the taxable income reacts to the net-of-tax rate (i.e., if the elasticity is different from zero), then the income distribution will be a function of the net-of-tax rate. This is shown in the left panel of Figure 1, which illustrates two realizations of the distribution of taxable income under two different constant net-of-tax rates, τ_1^c and τ_2^c, where $\tau_1^c > \tau_2^c$. Let z denote taxable income, and let the corresponding distributions of taxable income be represented by the density functions $\tilde{f}(z; \tau_1^c)$ and $\tilde{f}(z; \tau_2^c)$.

We assume that taxable income depends on both the net-of-tax rate and some heterogeneous individual component ω, such that $z(\tau^c; \omega)$ identifies a certain income level, and the distribution of ω determines the income distribution given the net-of-tax rate. The density of ω is given by the continuous function $f(\omega)$. If the net-of-tax rate decreases given the distribution of ω, the income distribution shifts to the left. In the left panel of the figure, and for a particular level of ω indicated by $\bar{\omega}$, the income level shifts by the distance d. Suppose that the net-of-tax rate equals τ_1^c up to the taxable income level k, which is the kink point, and τ_2^c for all income levels above k. The resulting distribution of income is shown in the right panel of Figure 1. Up to the income level k, the income distribution is given by $\tilde{f}(z; \tau_1^c)$ (the red line in the left panel), and for incomes exceeding k the income distribution equals $\tilde{f}(z; \tau_2^c)$ (the blue line in the left panel). The mass of tax payers who would otherwise have realized incomes between k and $k + d$ under the net-of-tax rate τ_1^c all realize the income level k, causing the observed distribution to spike at the kink point k. In other words, there is a tendency to bunch at the kink point, and the mass of people at k—which equals the probability to realize the income level k—is the bunching mass, B.

Saez (2010) shows that the bunching mass and the observed income densities before and after the kink point can identify the ETI, given some assumptions on individual preferences. As both B and the two densities are observable, the only assumptions that have to be made refer to the structure of individual preferences and whether individuals can optimize without frictions. We will assume frictions away for now and add them to the model later on in this section.

2.1 | Behavioral model

In order to estimate the ETI, we need behavioral assumptions. We base our estimator on the model structure used by Saez (2010) and much of the subsequent literature, which allows us to directly compare our results to those found in earlier studies. Note that the behavioral model structure is needed to identify the ETI also in so-called nonparametric estimations.
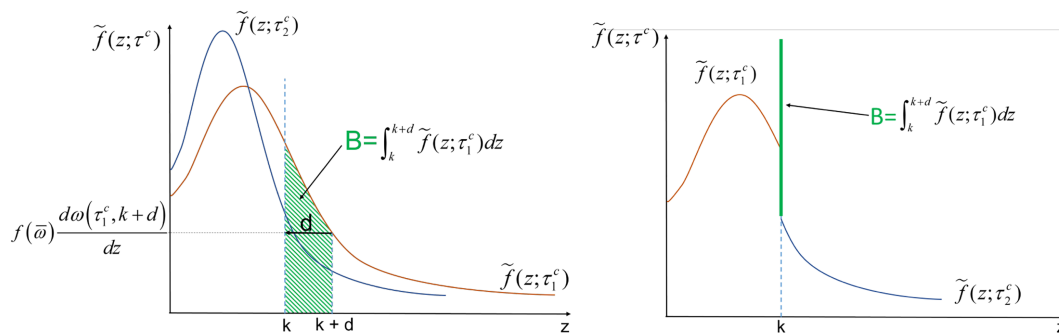
FIGURE 1 Basic intuition. _Note:_ In the left panel, $\tilde{f}(z; \tau_1^c)$ and $\tilde{f}(z; \tau_2^c)$ describe the density functions of optimal earnings under the respective marginal net-of-tax rate τ_i^c ($i = 1, 2$). k describes the income level at which the marginal tax rate changes from the lower rate τ_1 to the higher rate τ_2. d describes the difference in income levels of the marginal taxpayer under the marginal tax rate τ_1 and the marginal tax rate τ_2. The right panel depicts the observed income distribution, which equals $\tilde{f}(z; \tau_1^c)$ at income levels below k, $\tilde{f}(z, \tau_2^c)$ at income levels above k, and $\int_k^{k+d} \tilde{f}(z; \tau_1^c)dz$ at k. $f(\bar{\omega})$ identifies the marginal taxpayer whose observed income level does not exceed the income level k under τ_2. In order to correspond to the density $\tilde{f}(z; \tau_1^c)$, it has to be rescaled with $\frac{d\omega(\tau_1^c, k+d)}{dz}$ (see Appendix A, Equation A.4).

Even though we use a relatively simple model here, our approach is flexible with respect to the model framework, which is one of its advantages.

The preferences over consumption, c, and work hours, h, are represented by the quasilinear 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 elasticity of the labor supply with respect to the marginal net wage rate, $w(1 - \tau)$, and $\eta > 0$ is the disutility of work.

While earnings, $z = wh$, are assumed to be observable, both the before-tax wage rate and the disutility of work are unobservable. The optimal earnings function then takes the form:

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

where $\tau^c \equiv (1 - \tau)$ is the marginal net-of-tax rate corresponding to income z^*. The unobserved heterogeneity in wages and in the disutility of work reduces here to one single variable, $\omega = w^{(\alpha+1)}\eta$, describing an unobserved component of the individual's income. The behavioral response of earnings to an increase in the marginal net-of-tax rate is then defined as $d \ln z^*/d \ln \tau^c = \alpha$. Thus, since earnings constitute the only income source by assumption, α is interpretable as the ETI.

2.2 | Measuring the bunching mass

Given the modeling structure, it is possible to recover an estimate of the ETI from the observation of a single cross section of the earnings distribution around the kink. If the ETI is strictly positive and if the net-of-tax rate to the left of the kink is larger than the net-of-tax rate to the right of the kink, then we expect to observe some accumulation of observations at the kink. We denote the amount of bunching at k, that is, the proportion of observations with earnings equal to k, as B. The intuition is clear: the larger the bunching mass, the larger will be the ETI.

In the perfect bunching case (without optimization frictions), Saez (2010) shows that the bunching estimator of the ETI is defined implicitly as the solution in terms of α of the following equation:

$$B = g(\alpha, \tilde{f}(k_-; \tau_1^c), \tilde{f}(k_+; \tau_2^c); \tau_1^c, \tau_2^c), \tag{4}$$

where $g(.)$ is a known function of the ETI, α, and of the net-of-tax rates τ_1^c and τ_2^c. It also depends on the densities of the earnings distribution near the kink $\tilde{f}(k_-; \tau_1^c)$ and $\tilde{f}(k_+; \tau_2^c)$. We describe Saez's construction more thoroughly in the supporting information Appendix A.

2.3 | Imperfect bunching

The empirical results in Saez (2010) and in most subsequent studies acknowledge that bunching is imperfect, that is, 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. If bunching is imperfect (which virtually all studies in the literature assume), the

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.
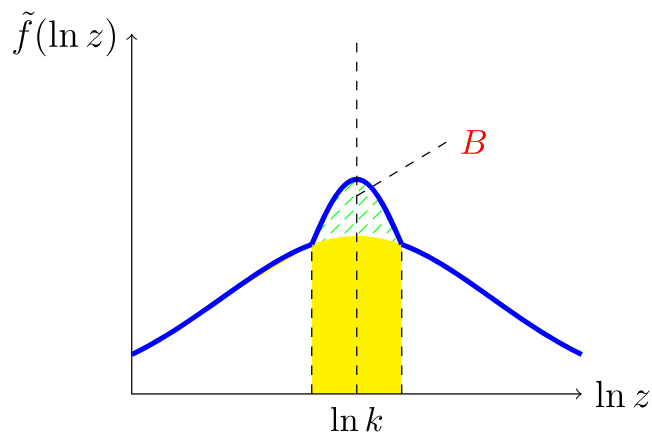
measurement of B requires both the identification of the bunching interval and an assumption about the counter-factual density around the kink.

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, that is, $\tilde{f}(k; \tau_1^c)$ 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. Several other studies proceed along these lines (e.g., Bastani & Selin, 2014, and Kleven & Waseem, 2013).

In the polynomial approach developed by Chetty et al. (2011, 2009), the earnings data are first collected into relatively narrow bins to construct a histogram of the distribution of earnings before and after the kink. Then, a polynomial is fitted through the observed income histogram excluding observations near the kink on either side. The degree of the polynomial is determined using a calibration procedure (Chetty et al. 2009, 2011, use a polynomial of order 7). Finally, the approach corrects the portion of the polynomial above the kink iteratively, such that the integral under the estimated polynomial corresponds to the number of total observations. The resulting polynomial serves as the counter-factual income distribution for identifying the bunching mass.

3 | A ML ESTIMATOR

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. However, optimization errors are typically not modeled explicitly. Instead, it is just assumed that bunching occurs in an interval around the kink. This section presents a ML estimator of the ETI under imperfect bunching, where unobserved heterogeneity and optimization errors are accounted for simultaneously. Thus, we make an explicit distinction between the observed and the optimal earnings (where the optimal earnings are generated by the underlying behavioral model).

We start 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 the single-kink case we have the following:

$$
\begin{cases}
\text{if } \alpha \ln \tau_1^c + \ln \omega < \ln k: \ \ln z^* = \alpha \ln \tau_1^c + \ln \omega, \\
\text{if } \alpha \ln \tau_2^c + \ln \omega > \ln k: \ \ln z^* = \alpha \ln \tau_2^c + \ln \omega, \\
\text{if } \alpha \ln \tau_1^c + \ln \omega > \ln k \ \text{ and } \ \alpha \ln \tau_2^c + \ln \omega < \ln k: \ \ln z^* = \ln k.
\end{cases}
\tag{5}
$$

The observed earnings z^o that the individuals experience in the end are determined by the optimal earnings z^* and a shock ϵ such that

$$
\ln z^o = \ln z^* + \ln \epsilon.
\tag{6}
$$
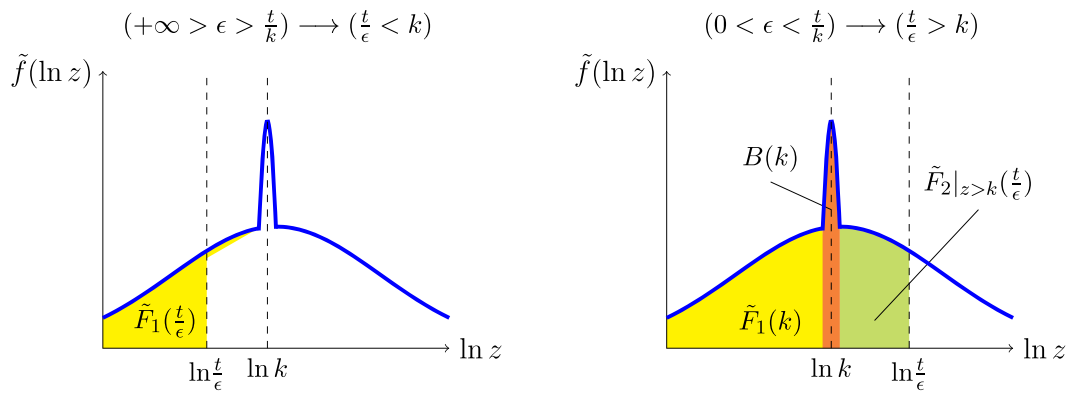
FIGURE 3 Distribution function of observed earnings. *Note:* The graphs show the densities of optimal earnings. Shaded areas depict the probability that observed earnings are below a certain level t, given the size of the shock ϵ and the distribution of optimal earnings.

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 and is log normally distributed with mean 1, which implies that the logarithm of ϵ is distributed normally with $\ln \epsilon \sim \mathcal{N}(-\frac{\sigma_\epsilon^2}{2}, \sigma_\epsilon^2)$.[4]

In order to estimate the parameters of this model using the ML method, we need to derive the density function of observed earnings. To that aim, we first derive the distribution function. We describe the distribution of observed earnings $\ln z^o$ given our assumptions concerning the distribution of optimal earnings $\ln z^*$ (which depend on the unobserved component ω, the ETI 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 optimal earnings z^* are less than $\frac{t}{\epsilon}$:[5]

$$H(t) \equiv \mathbb{P}[z^*\epsilon < t] = \mathbb{E}_\epsilon[\mathbb{P}[z^*\epsilon < t|\epsilon]] = \mathbb{E}_\epsilon\left[\mathbb{P}\left[z^* < \frac{t}{\epsilon}|\epsilon\right]\right].$$

Given the distribution of the unobserved component of optimal earnings ω, the distribution of optimal 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 optimal 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 optimal earnings given the net-of-tax rate to the right of the kink.

Let $g(\epsilon)$ denote the probability density function of the shock ϵ, $\tilde{F}_1(x) = \mathbb{P}[z^* < x]$ the distribution function of optimal earnings under the net-of-tax rate τ_1^c, and $\tilde{F}_2(x) = \mathbb{P}[z^* < x]$ the distribution function of optimal earnings under the net-of-tax rate τ_2^c. We can now decompose $\tilde{F}_2(x)$ using $\tilde{F}_1(x)$, the probability that optimal earnings are at the kink $B(k) = \mathbb{P}[z^* = k]$, and the distribution function of optimal earnings above the kink $\tilde{F}_2(x)|_{z^*>k} = \mathbb{P}[k < z^* < x]$. This is done in Equation (7). Figure 3 depicts the decomposition.

$$H(t) = \underbrace{\int_0^{\frac{t}{k}} \left(\tilde{F}_1(k) + B(k) + \tilde{F}_2|_{z^*>k}\left(\frac{t}{\epsilon}\right)\right)g(\epsilon)d\epsilon}_{\frac{t}{\epsilon} \text{ above kink}} + \underbrace{\int_{\frac{t}{k}}^{+\infty} \tilde{F}_1\left(\frac{t}{\epsilon}\right)g(\epsilon)d\epsilon}_{\frac{t}{\epsilon} \text{ below kink}}. \tag{7}$$

The density of observed earnings $h(t)$ is derived from the cumulative distribution function, $H(t)$, such that $h(t) \equiv dH(t)/dt$. Some algebra yields

$$h(t) = \frac{1}{k}B(k)\,g\left(\frac{t}{k}\right) + \int_0^{\frac{t}{k}} \tilde{f}_2\left(\frac{t}{\epsilon}\right)g(\epsilon)\frac{d\epsilon}{\epsilon} + \int_{\frac{t}{k}}^{+\infty} \tilde{f}_1\left(\frac{t}{\epsilon}\right)g(\epsilon)\frac{d\epsilon}{\epsilon}, \tag{8}$$

---

[4]The normality assumption is analytically convenient. Alternatively, the probability density function for ϵ could take the form of the density of a gamma distribution with parameter γ where the expectation is 1, that is, $g(\epsilon) = \frac{\gamma^\gamma}{\Gamma(\gamma)}\epsilon^{\gamma-1}\exp(-\gamma\epsilon)$. In the supporting information Appendix B.1, we consider the case with a log-laplace distribution for the friction.

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

FIGURE 4 Decomposing the density of observed earnings. *Note:* The dotted line shows the density function of observed earnings for an example distribution of ω and a tax system with one kink. The continuous lines describe the three components of $h(t)$ given in Equation (9).
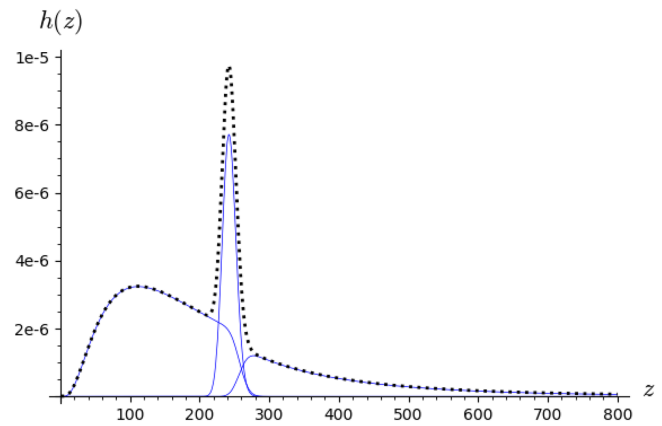
where $\tilde{f}_1$ and $\tilde{f}_2$ describe the density functions of optimal earnings given the net-of-tax rates τ_1^c and τ_2^c.

Let $\varsigma \equiv 1/\sigma_\epsilon$ and $\nu \equiv -\sigma_\epsilon/2 = -1/(2\varsigma)$. Given the normality assumptions and the specification of the optimal earnings, see Equation (3), the density of observed earnings with imperfect bunching takes the following form[6]

$$
\begin{aligned}
h(t) = &\frac{\varsigma}{t}\phi(\varsigma \ln t - \varsigma \ln k - \nu)B(k) \\
&+ \frac{s\varsigma}{St}\phi\left(\frac{1}{S}(\varsigma s \ln t - (\lambda_2\varsigma + \nu s))\right)\Phi\left[\frac{1}{S}(\varsigma^2 \ln t - S^2 \ln k + \varsigma\nu - \lambda_2 s)\right] \\
&+ \frac{s\varsigma}{St}\phi\left(\frac{1}{S}(\varsigma s \ln t - (\lambda_1\varsigma + \nu s))\right)\Phi\left[\frac{1}{S}(S^2 \ln k - \varsigma^2 \ln t + \lambda_1 s - \varsigma\nu)\right],
\end{aligned}
\tag{9}
$$

where the parameters s, S, λ_1, and λ_2 are such that:

$$
\begin{aligned}
s &\equiv \frac{1}{\sigma_\omega}, \\
S^2 &\equiv s^2 + \varsigma^2, \\
\lambda_1 &\equiv s\alpha \ln \tau_1^c + s\mu, \\
\lambda_2 &\equiv s\alpha \ln \tau_2^c + s\mu.
\end{aligned}
\tag{10}
$$

Furthermore, under our assumptions $B(k) = \Phi[s \ln k - \lambda_2] - \Phi[s \ln k - \lambda_1]$ and as a consequence, we require $\lambda_1 - \lambda_2 > 0$ to insure that some bunching exists.

Figure 4 illustrates the effect of imperfect bunching on the distribution of earnings in a simple case where the variance of the shock to optimal earnings is substantial.

3.1 | Estimation

With imperfect bunching, all earnings values have a positive density, and the likelihood can be expressed in terms of $h(t)$ only. The probability of observing a given earnings value in the range around the kink, that is , in some interval $[\underline{z}, \bar{z}]$ such that $\underline{z} < k < \bar{z}$, is given by[7]

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

---

[6]We derive the density in the supporting information Appendix B. We use the parameters s, ς, λ_1, and λ_2 instead of the underlying deep parameters σ_ω, σ_ϵ, α, and μ, when presenting the density function for the observed earnings, as this simplifies the expressions.

[7]The bunching estimator is local in the sense of relying on information in an interval around the tax kink to estimate the ETI. Note that $[\underline{z}, \bar{z}]$ is not necessarily a narrow range of observations. We show in Section 3.3 that the ML estimator is not particularly sensitive to changes in the upper and lower limits of the observed income range as long as the model is well-specified.

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:

$$\ln \mathcal{L}_{\text{IB},n} = \sum_{i=1}^{n} \ln h(z_i^o) - n \ln \mathbb{P}[\underline{z} < z^o < \bar{z}]. \tag{11}$$

The maximization of the likelihood above with respect to its parameters λ_1, λ_2, s, and ς will provide the ML estimates $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{s}$, and $\hat{\varsigma}$, from which we can calculate the estimates of the underlying structural parameters σ_ω, σ_ϵ, α, and μ.

Using the link between the statistical model and the economic structure, see the expressions (10), the estimator for the ETI takes the form:

$$\hat{\alpha}_{\text{IB},norm} = \frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\ln(\tau_1^c) - \ln\left(\tau_2^c\right)} \frac{1}{\hat{s}}. \tag{12}$$

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 thus fully control for the effects of the shock ϵ when estimating the ETI.

3.2 | Identification

As shown in Equation (6), the observed earnings, z^o, are determined by the optimal earnings, z^*, and a shock (optimization error), ϵ, such that $\ln z^o = \ln z^* + \ln \epsilon$, where the unobserved components of the optimal earnings, $\ln \omega$, and $\ln \epsilon$ are independently distributed. While maintaining the assumption that ϵ is log normal such that $\ln \epsilon \sim \mathcal{N}\left(-\frac{\sigma_\epsilon^2}{2}, \sigma_\epsilon^2\right)$, we assume here that $\ln \omega$ is characterized by the density function f_ω and distribution function F_ω. The distribution F_ω is left unspecified in this section. We also assume that the population variance of the shock/friction is bounded above, $\sigma_\epsilon < \bar{\sigma}_\epsilon$. Therefore, frictions account for a small part of the variability of observed earnings in the population.

Given the distribution of the unobserved component of the optimal earnings, F_ω, we can deduce the distribution of the optimal earnings in a given tax context. Let F^* and f^* denote the cumulative distribution function and density function, respectively, of the optimal earnings. The identification of the parameter α given the observation of the density f^* is treated in Saez (2010), Chetty et al. (2011), and more generally, in Blomquist et al. (2021), which give precise conditions on the distribution F_ω in order to deduce an estimate of α from f^*. Blomquist et al. (2021) show that the normality assumption or, more generally, a clear assumption on the smoothness of the density of the unobserved component is necessary for point identification of α. Indeed, assuming that the distribution of $\ln \omega$ is normal (as we do elsewhere in this paper), one can identify α from the knowledge of f^* (this is shown in the supporting information Appendix C).

Let us now turn to identification in the presence of a friction/optimization error, where f^o is used to denote the probability density function of the observed log-earnings. With a slight abuse of notation, because of the atom at the logarithm of the kink, we continue to use f^* to denote the density of the log of optimal earnings, while f^e denotes the probability density of the log of the friction component. The convolution operation, which generates F^o from F^e and F^*, is given by $F^o = F^* * F^e$. Finally, we assume that the characteristic functions of each distribution exist and denote them as ψ^o ψ^* and ψ^e, respectively. Thanks to the independence assumption between the optimal earnings and the friction and the definition of the observed earnings, the characteristic functions satisfy the relationship:

$$\psi^o = \psi^* \psi^e.$$

Since the main characteristic of the distribution of optimal earnings in the presence of a kink in the budget constraint is the bunching at the kink, we build the identification discussion around this property. We follow the method of proof proposed by Schwarz and Van Bellegem (2010) and show that we are able to identify σ_ϵ^2 and f^* (see Moore, 2022 who adopts a similar strategy). We therefore restrict the family of distribution of optimal earnings to the distribution which clearly bunch at the kink in the absence of frictions.

In the supporting information Appendix D.1, we analyze how the density of the optimization friction ϵ, $g(\epsilon)$, behaves as σ_ϵ changes. We show that as ϵ varies from 0 to $+\infty$, the function $g'_{\sigma_\epsilon}(\epsilon)$ is nonnegative outside of the interval $(\frac{1}{\bar{\epsilon}}, \bar{\epsilon})$ and negative inside it, where

$$\bar{\epsilon} = \exp\left(\sigma_\epsilon \sqrt{1 + \frac{\sigma_\epsilon^2}{4}}\right).$$

Moreover, outside the interval $[\underline{M}_{\sigma_\epsilon}, \bar{M}_{\sigma_\epsilon}]$, one can show that the value of the derivative $g'_{\sigma_\epsilon}(\epsilon)$ is negligible, where the bounds are such that $\underline{M}_{\sigma_\epsilon} \equiv \exp\left(-\bar{\sigma}_\epsilon\left(r + \frac{\bar{\sigma}_\epsilon}{2}\right)\right)$ and $\bar{M}_{\sigma_\epsilon} \equiv \exp\left(\bar{\sigma}_\epsilon\left(r - \frac{\bar{\sigma}_\epsilon}{2}\right)\right)$ for a large enough value of r which we can choose.

We will then impose the following restrictions on the distribution of optimal earnings:

Condition OB1. (obvious bunching of optimal earnings) For any $\sigma_\epsilon \in (0, \bar{\sigma}_\epsilon)$, for any $\eta \leq \bar{\sigma}_\epsilon\sqrt{1 + \frac{\bar{\sigma}_\epsilon^2}{4}}$, and for any t in an interval $I(\sigma_\epsilon) = [\underline{M}_{\sigma_\epsilon}, k\exp(-\eta)] \cup [k\exp(\eta), \bar{M}_{\sigma_\epsilon}]$, the distribution of optimal earnings is such that:

$$\mathbb{P}[k\exp(-\eta - u) \leq z^* < k\exp(\eta - u)] > \mathbb{P}[t\exp(-\eta) \leq z^* < t\exp(\eta)]$$

for any u in the interval $\left(-\sigma_\epsilon\sqrt{1 + \frac{\sigma_\epsilon^2}{4}}, \sigma_\epsilon\sqrt{1 + \frac{\sigma_\epsilon^2}{4}}\right)$.[8]

This condition ensures that the probability of finding the optimal earnings in any interval that contains the kink is larger than the probability of finding it in intervals of comparable size centered elsewhere but not too far away from the kink. It implies in particular that

$$P^s \equiv \max_{t \in I(\epsilon)} \mathbb{P}[t\exp(-\sigma_\epsilon) \leq z^* < t\exp(\sigma_\epsilon)]$$

is bounded above strictly by $\mathbb{P}[k\exp(-\eta) \leq z^* < k\exp(\eta)]$ for $\eta \leq \bar{\sigma}_\epsilon\sqrt{1 + \frac{\bar{\sigma}_\epsilon^2}{4}}$. This assumption is satisfied if there is bunching at k exactly.[9] Furthermore, let

$$\mathbb{P}_{min} \equiv \min_{\frac{1}{\bar{\epsilon}} < u < \bar{\epsilon}} \mathbb{P}\left[\frac{k}{u}\exp(-\eta) \leq z^* < \frac{k}{u}\exp(\eta)\right],$$

and therefore

$$\mathbb{P}_{min} > \mathbb{P}[z^* = k].$$

We now impose a second (technical) condition which ensures that the bunching is substantial enough to dominate the behavior at the tail of the friction distribution [10]:

Condition OB2. (obvious bunching of optimal earnings)

$$\mathbb{P}_{min} - P^s \geq -(1 - P^s)\frac{\int_0^{\underline{M}_{\sigma_\epsilon}} g'_{\sigma_\epsilon}(\epsilon)d\epsilon + \int_{\bar{M}_{\sigma_\epsilon}}^{+\infty} g'_{\sigma_\epsilon}(\epsilon)d\epsilon}{\int_{\frac{1}{\bar{\epsilon}}}^{\bar{\epsilon}} g'_{\sigma_\epsilon}(\epsilon)d\epsilon}.$$

Note that the right-hand side of the above inequality is positive; hence, the condition imposes a real constraint. However, the numerator of the last term is negligible whenever the bounds $\underline{M}_{\sigma_\epsilon}$ and $\bar{M}_{\sigma_\epsilon}$ are well chosen. This condition will be satisfied whenever the ETI is strictly positive.

The focus on the interval $I(\sigma_\epsilon)$ ensures that we compare the probability around the kink with nearby intervals and not intervals 'further' away where the data may be more densely distributed than around the kink. We assume therefore that the class of models of interest is such that:

---

[8]The ratio of the upper to the lower bound for the intervals around t or k is $\exp(2\eta)$ in either case. The interval around k is not centered at k but always contains it.

[9]Note moreover that in general for η such that $\eta \leq \bar{\sigma}_\epsilon\sqrt{1 + \frac{\bar{\sigma}_\epsilon^2}{4}}$

$$\mathbb{P}[z^* = k] \leq \mathbb{P}[k\exp(-\eta) \leq z^* < k\exp(\eta)].$$

[10]The support of the density of the optimization friction ranges from 0 to ∞; however, when σ_ϵ is small enough, the density is concentrated around and not too far from $\epsilon = 1$.

Condition A. Assuming $\ln \epsilon \sim \mathcal{N}(-\frac{\sigma_\epsilon^2}{2}, \sigma_\epsilon^2)$ with $\sigma_\epsilon < \bar{\sigma}_\epsilon$, and for some given η small and positive, the probability

$$\mathbb{P}\left[\exp(-\eta) \leq \frac{z^* \epsilon}{k} < \exp(\eta)\right] = \mathbb{E}_{\sigma_\epsilon}\left[\mathbb{P}\left[|\ln z^* + \ln \epsilon - \ln k| < \eta|\ln \epsilon|\right]\right]$$

is strictly decreasing with σ_ϵ over the range $[0, \bar{\sigma}_\epsilon]$.

Condition A requires that bunching at or near the kink in the distribution of observed earnings decreases as the variance of the friction process increases. It further excludes models of optimal earnings where the density of optimal earnings infinitesimally close to the kink increases fast after and/or decreases fast before the kink. If this was the case, the presence of frictions would make the identification of the density of optimal earnings based on bunching very difficult. Excluding such cases makes the identification argument direct.

The next result states that *Condition OB1* and *Condition OB2* are necessary conditions for *Condition A* (see supporting information Appendix D.2 for a proof).

Lemma 1. *Any distribution of optimal earnings which satisfies Condition OB1 and Condition OB2 satisfy Condition A.*

This result allows us to adapt the argument developed by Schwarz and Van Bellegem (2010) to the class of models of optimal earnings where the density f^* satisfies *Conditions OB1 and OB2* and therefore, assuming that the friction process is log-normal, the distribution of observed earnings satisfies *Condition A*. We first state a version of their Lemma 2.2.

Lemma 2. *Consider two distinct models for the distribution of optimal earnings F_1^* and F_2^* which satisfy Condition A, and two distinct models of frictions $\mathcal{LN}_{\sigma_1}$ and $\mathcal{LN}_{\sigma_2}$, with $0 < \sigma_1 < \sigma_2 < \bar{\sigma}$ such that the distribution of observed earnings $F_1^* * \mathcal{LN}_{\sigma_1}$ and $F_2^* * \mathcal{LN}_{\sigma_2}$ are observably equivalent, $F_1^o = F_2^o$. Then F_1^* is the convolution of F_2^* with the model of friction $\mathcal{LN}_{\sigma_3}$, where $\sigma_3 = \sqrt{\sigma_2 - \sigma_1}$.*

The proof is identical to theirs and relies on the expression of the characteristic function of the distribution of the friction: $\psi_{e,\sigma} = \exp\left(-\frac{\sigma^2}{2}(it + t^2)\right)$.

We can now state a result similar to their Theorem 2.1, which shows that the parameter of the distribution of the optimization friction is identified (which we prove in the supporing information Appendix D.3)

Proposition 1. *The distribution of optimal earnings as well as the standard deviation of the friction, σ_ϵ with $\sigma_\epsilon > 0$, are identified if Condition A on the distribution of observed earnings is satisfied.*

3.3 | Performance

We have carried out a small Monte Carlo study, where the performance of the ML bunching estimator is compared with performance of the polynomial estimator based on the approximation proposed by Chetty et al. (2011). While we estimate a model based on the assumptions that $\ln \omega$ and $\ln \epsilon$ are normally distributed, three distinct data-generating processes will be considered. Data are generated with about 14,000 observations such that the kink is located at 490, which is the position of the first kink in the Swedish tax system in 2019, measured in 1000's of SEK, and the values of the net-of-tax-rate before and after the kink are 0.6585 and 0.46585, respectively. For each data-generating process, we consider two cases where the true ETI is set at 0.05 or 0.1, respectively. The results of the Monte Carlo analysis are presented in Tables 1–3 below, and the sampling distributions are illustrated graphically in the supporing information Appendix F for the case where the true ETI is 0.05.

The first data-generating process corresponds to the normal–normal model described earlier, that is, where the unobserved component of the optimal earnings and the optimization friction are both log-normal. The mean and standard deviation of the unobserved component of the optimal earnings are set at 6 and 0.25, respectively, while the standard deviation of the optimization error is set at 0.005. We assume to begin with that the earnings are observed in an interval around the tax kink between 420 and 560. In this case, the likelihood of the normal–normal model is well specified. To apply the polynomial method of Chetty et al. (2011), data is divided into bins of unit size (141 bins). We then fit a polynomial of order 7 over the binned data and exclude observations within either 5, 10, or 15 bins around the kink. The results are presented in Table 1. We can see that the ML estimator is practically unbiased when the true ETI is 0.1, while there is a

TABLE 1   Normal–normal data, estimators sampling distribution.

	Mean	Median	Std dev	Min	Max
(A) Normal–normal data, $\alpha = 0.05$.					
Likelihood	0.0525	0.0526	0.00359	0.0404	0.0639
Chetty, 5	0.0441	0.0441	0.00357	0.0341	0.0557
Chetty, 10	0.0481	0.0480	0.00579	0.0301	0.0670
Chetty, 15	0.0460	0.0461	0.00960	0.0201	0.0739
(B) Normal–normal data, $\alpha = 0.1$.					
Likelihood	0.1040	0.1040	0.00463	0.0883	0.116
Chetty, 5	0.0869	0.0868	0.00487	0.0718	0.104
Chetty, 10	0.0977	0.0976	0.00762	0.0779	0.124
Chetty, 15	0.0956	0.0952	0.01190	0.0621	0.138

Note: The table describes descriptive statistics of the sampling distribution for a number of estimators of the ETI based on a thousand replication of the data generation described in the text. The likelihood row refers to the ML estimator assuming that the correct specification accounts for a log-normally distributed unobserved component, ω, and for a log-normally distributed friction component ϵ. Each Chetty, p row corresponds to the polynomial estimator suggested by Chetty et al. (2011) when p bins on either side of the kink are excluded. In the upper part of the table the "true" ETI is set to 0.05, while it is set to 0.1 in the bottom part of the table.

TABLE 2   Two component normal mixture–normal data, estimators sampling distribution.

	Mean	Median	Std dev	Min	Max
(A) Two component normal mixture–normal data, $\alpha = 0.05$.					
Likelihood	0.0243	0.0243	0.00223	0.0178	0.0320
Chetty, 5	0.0405	0.0403	0.00375	0.0282	0.0550
Chetty, 10	0.0401	0.0402	0.00615	0.0212	0.0606
Chetty, 15	0.0314	0.0310	0.00938	0.0006	0.0598
(B) Two component normal mixture–normal data, $\alpha = 0.1$.					
Likelihood	0.1050	0.1050	0.00384	0.0948	0.119
Chetty, 5	0.0855	0.0855	0.00428	0.0726	0.101
Chetty, 10	0.0934	0.0929	0.00672	0.0714	0.114
Chetty, 15	0.0873	0.0869	0.01020	0.0558	0.124

Note: See the note below Table 1 for further explanations.

tendency to upward bias in the other case. The sampling distribution of the polynomial estimator shows a downward bias, where the size of the bias depends on the number of bins excluded in order to calculate the counter-factual distribution. These results suggest that the ML estimator performs relatively well both in terms of bias and precision.

The second data-generating process is based on a two-component mixture normal model for the optimal earnings, where the respective mean parameters are equal to 5.95 and 6.35, with standard deviation for each type equal to 0.1. The probability to belong to the first type is set at 0.6. The optimization error is still log-normal. Figure F.2 in the supporting information Appendix F shows the distribution of earnings in the observed range for a typical replication as well as the fit of the normal–normal model. In this case, the normal–normal model will not fit the data (as it does not generate a density of the observed earnings resembling the W-shape in the figure). Table 2 illustrates the sampling distributions for the ML estimator and the polynomial estimator. Both estimators exhibit a clear downward bias when the true ETI is 0.05, and the bias of the ML estimator is larger than that of (all three versions of) the polynomial estimator. However, if the true ETI is 0.1, all estimators behave roughly as in Table 1, and the ML estimator is practically unbiased. This suggests that all estimators (and in particular the ML estimator) are more sensitive to major distributional misspecifications when the underlying ETI to be estimated is small (when the signal in the data is weak). Our conclusion is, nevertheless, that both the ML estimator and the polynomial estimator tend to perform poorly under an extreme distributional misspecification.

The third data-generating process assumes that the optimal earnings are distributed according to a Pareto distribution with an exponent equal to 3 (a credible value for the right-hand tail of the distribution of earnings in Sweden). The optimization error is log-normal with the same mean and variance as above. Table 3 illustrates the distribution over the observed range for a typical replication. In this case, the normal–normal model is misspecified (since it is not based on the Pareto distribution for optimal earnings), but the performance of both estimators is similar to the performance of the

TABLE 3 Pareto-normal data, estimators sampling distribution.

	Mean	Median	Std dev	Min	Max
(A) Pareto–normal data, $\alpha = 0.05$.					
Likelihood	0.0485	0.0484	0.00303	0.0386	0.0579
Chetty, 5	0.0434	0.0434	0.00345	0.0331	0.0550
Chetty, 10	0.0460	0.0460	0.00568	0.0301	0.0658
Chetty, 15	0.0424	0.0420	0.00882	0.0138	0.0715
(B) Pareto–normal data, $\alpha = 0.1$.					
Likelihood	0.0979	0.0980	0.00377	0.0850	0.109
Chetty, 5	0.0860	0.0859	0.00460	0.0688	0.103
Chetty, 10	0.0951	0.0949	0.00732	0.0709	0.119
Chetty, 15	0.0911	0.0905	0.01070	0.0596	0.133

Note: See the note below Table 1 for further explanations.

TABLE 4 Normal–normal data, estimators sampling distribution, different ranges.

	Mean	Median	Std dev	Min	Max
(A) Normal–normal data, $\alpha = 0.05$, Range [420,560], 14,000 observations on average.					
Likelihood	0.0525	0.0526	0.00359	0.0404	0.0639
Chetty, 5	0.0441	0.0441	0.00357	0.0341	0.0557
Chetty, 10	0.0481	0.0480	0.00579	0.0301	0.0670
Chetty, 15	0.0460	0.0461	0.00960	0.0201	0.0739
(B) Normal–normal data, $\alpha = 0.05$, range [450,530], 8,200 observations on average					
Likelihood	0.0497	0.0497	0.0034	0.0389	0.0609
Chetty, 5	0.0397	0.0394	0.0045	0.0273	0.0564
Chetty, 10	0.0415	0.0411	0.0010	0.0094	0.0777
Chetty, 15	0.0263	0.0233	0.0190	0.0002	0.1230
(C) Normal–normal data, $\alpha = 0.05$, range [475,505], 5, 460 observations on average.					
Likelihood	0.0468	0.0467	0.00352	0.03550	0.0584
Chetty, 5	0.0316	0.0312	0.00582	0.01490	0.0514
Chetty, 10	0.0189	0.0150	0.01610	0.00003	0.1170
Chetty, 15	0.0936	0.0889	0.19500	0.00134	4.9900

Note: The range of estimation varies from ±70 around the kink (sub-table A), to ±40 (sub-table B) and to ±25 (sub-table C). Each replication contributes to the entries in each of the three sub-tables: we simulate data such that the observations belong to the larger range and calculate each estimator. We then select the observations which belong to the narrower ranges and recompute the estimators. See the note below Table 1 for additional explanations.

estimators under the normal–normal data generating process. The ML estimator performs better than the polynomial method both in terms of bias and precision.

Finally, we have examined how sensitive the estimators are to variation in $[\underline{z}, \bar{z}]$, that is, the income interval in which the data are observed. To do so, we focus on the well-specified normal–normal model and the case where the true ETI equals 0.05.[11] The results are presented in Table 4. We begin with the case where $\underline{z} = 420$ and $\bar{z} = 560$ (as in the baseline in Table 1) and then gradually shrink the size of the interval by increasing $\underline{z}$ and decreasing $\bar{z}$. We can see that all estimators are affected. Specifically, there is a tendency for the mean estimate of the ETI to decrease when the data interval becomes smaller. However, the ML estimator is much less sensitive to a shrinking data interval than the other estimators. While (all versions of) the polynomial estimator exhibits a substantial bias in part c of the figure—the most narrow data range—the ML estimator still performs reasonably well.

In conclusion, our Monte Carlo analysis suggests that the ML bunching estimator based on the normal–normal model dominates the estimator based on the polynomial method whenever the distribution over the observed range can be approximated reasonably well by a truncated normal distribution. If the true data-generating process is sufficiently different from the normal–normal model, the performances of both estimators are poorer.

---

[11]We have carried out the same analysis for the case where the true ETI is set to 0.1. This change of assumption is not important: the qualitative results are similar to those presented below and are available from the authors upon request.

TABLE 5 Maximum likelihood estimates of the ETI.

	α	μ	$\log(\sigma_\omega)$	$\log(\varsigma)$
(A) All tax payers				
Range: [420,560]	0.0067	6.040	1.406	7.033
SE	(0.0001)	(0.0030)	(0.0100)	(0.0219)
Robust SE	(0.0051)	(0.0866)	(0.2795)	(0.8224)
Number of observations	1055452			
Mean log likelihood	−71.854			
Range: [450,530]	0.0063	6.146	1.959	7.071
SE	(0.0001)	(0.0013)	(0.0127)	(0.0220)
Robust SE	(0.0050)	(0.0380)	(0.3469)	(0.8122)
Number of observations	615,813			
Mean log likelihood	−66.761			
Range: [475,505]	0.0063	6.174	2.313	7.071
SE	(0.0001)	(0.0035)	(0.0672)	(0.0224)
Robust SE	(0.0052)	(0.0931)	(1.7611)	(0.8262)
Number of observations	239,682			
Mean log likelihood	−56.951			
(B) Self-employed tax payers				
Range: [420,560]	0.1021	6.200	2.026	6.777
SE	(0.0025)	(0.0018)	(0.0189)	(0.0339)
Robust SE	(0.0288)	(0.0174)	(0.1402)	(0.9680)
Number of observations	25,827			
Mean log likelihood	−69.650			
Range: [450,530]	0.0770	6.222	2.642	6.947
SE	(0.0021)	(0.0012)	(0.0199)	(0.0320)
Robust SE	(0.0231)	(0.0131)	(0.1456)	(0.8480)
Number of observations	17,705			
Mean log likelihood	−63.787			
Range: [475,505]	0.0446	6.218	3.651	7.204
SE	(0.0015)	(0.0009)	(0.0271)	(0.0327)
Robust SE	(0.0174)	(0.0097)	(0.2432)	(0.8465)
Number of observations	10013			
Mean log likelihood	−52.949			

Note: The table presents the maximum likelihood estimates for the model described in Section 3. The starting values to the optimization of likelihood are obtained by performing a sequential line search for each parameter in turn. This approach worked well in practice. We have calculated two measures of precision for the parameters: a) the 'SE' row presents the standard errors that are calculated using the observed hessian of the likelihood. This calculation is correct if the model is well-specified; b) the 'Robust SE' row presents the standard errors that are calculated using the sandwich estimator of the asymptotic variance covariance of the parameters suggested by White (1982). This accounts for the possible misspecification of the model. Gouriéroux et al. (1995) discuss the estimation of the variance-covariance of the estimated parameters that result from the maximization of the likelihood.

4 | APPLICATION BASED ON SWEDISH DATA

We apply the ML bunching estimator to Swedish register data from 2019 provided by Statistics Sweden. The income concepts are based on the tax register (IOT) and the Longitudinal Integrated Database for Health Insurance and Labour Market Studies (LISA). LISA contains the universe of the Swedish adult population, while the IOT register contains the universe of tax payers.

All incomes correspond to taxable labor income after the basic allowance that are reported in the IOT. This income concept guarantees that for every tax unit, we include all income sources that are relevant for the determination of the tax rate on labor income. Those include, for example, income from employment and related transfers (such as pensions, parental leave payments, or sickness payments), and income from self-employment that is declared as labor income. Our first sample includes all observed tax units. However, as previous studies have found that self-employed tax payers respond more to tax incentives than tax payers who are purely employed (see, e.g., Bastani & Selin, 2014), we further restrict the
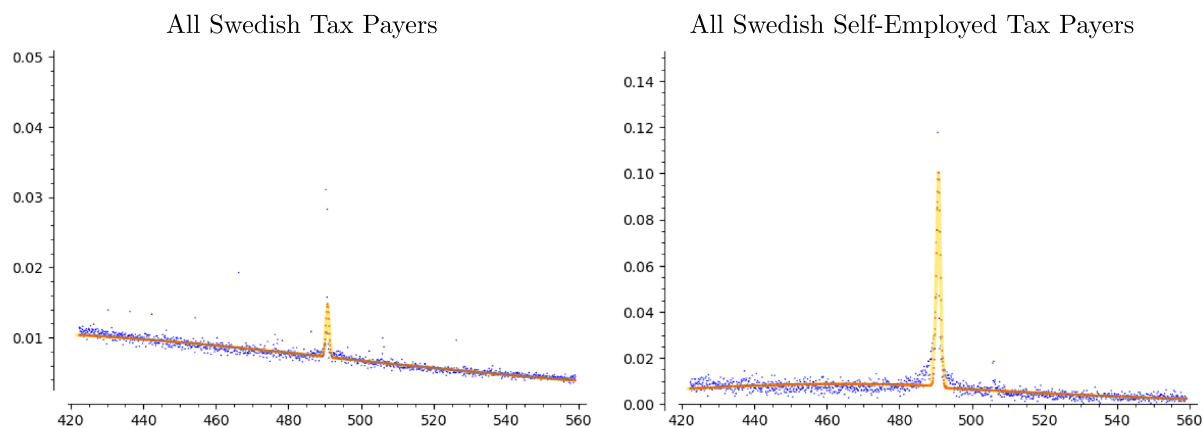
FIGURE 5 Imperfect Bunching at first Kink, 2019. _Note:_ The figure describes the distribution of earnings for all tax units (left panel, blue dots, 1055,452 observations) and all self-employed tax units (right panel, blue dots, 25,827 observations) subject to labor income taxation around the first government kink at 490.7kSEK in 2019. The predicted distribution of earnings in the interval given the best fit parameters based on the ML estimator in Equation (12) is also shown (yellow dashed line with red halo). Taxable labor income is observed in 1369 bins within a range of 68.4k SEK of the kink, at which the marginal tax rate increases by 20 percentage points. The data are grouped in income intervals of 100SEK. To compare the fit between the observed distribution and the predicted distribution, we calculate the goodness of fit χ^2 statistics. The test statistics is equal to 116,382 (all tax units) resp. 53,020 (self-employed), which rejects the null hypothesis that the two distributions are identical at all significance levels in both cases. The regression of the predicted number of observations in each bin on the theoretically expected number reports an R^2 of about 0.84 (all tax units) resp. 0.51 (self-employed).

sample to tax units who report positive income from self-employment. For this sample selection, we use self-employed incomes as reported in LISA.

The Swedish labor income tax system consists of local tax rates that vary somewhat at the municipality level and a national tax rate of 20% that only applies above the first government kink point, which is relatively high up in the income distribution (in 2019, it is at 490,700 SEK, which corresponds to the 88th percentile of all tax units). Until 2019, a second government kink point existed (at 689,300 SEK in 2019, corresponding to the 96th percentile of all tax units), where the national tax rate increased by 5 percentage points. In our application, we focus on the first government kink point, as the marginal tax rate differential is much larger and tax units display clear bunching. Estimation results based on the ML bunching approach are presented in Table 5 and Figure 5.

As expected, the self-employed are much more responsive to marginal taxation than income earners in general: the ML point estimate of the ETI (i.e., α) is more than 10 times larger for the self-employed compared to the estimate for all income earners in the baseline specification where the data range is [420,560]. All parameters are quite precisely estimated on both samples.[12]

We can also see that the ML point estimate of the ETI is not particularly sensitive to changes in the observed data range in the sample of all income earners, while it is very sensitive to such changes in the sample of self-employed. This finding is unsurprising and suggests that the underlying economic model of income formation is less suitable for the self-employed than for wage earners. The reason is that the self-employed face additional margins (such as incentives for income shifting between labor and capital) not available to the same extent for wage earners. Thus, to go further in the the analysis of income formation among the self-employed, these additional behavioral margins should be modeled explicitly and the ML bunching estimator should be modified accordingly. This is an important topic for future research.

For purposes of comparison, we present point estimates using the polynomial approach in Table 6. The point estimates based on the sample of all tax payers have a similar order of magnitude as the ML estimates in Table 5, with the qualification that the polynomial point estimates are highly sensitive to the number of bins excluded when calculating the counter-factual distribution. A similar conclusion is reached when the polynomial point estimates of the ETI based on data for the self-employed are compared with the ML estimate in Table 5. Note also that the polynomial point estimates of the ETI based on the sample of self-employed earners seem to be as sensitive to variation in the observed data range as the corresponding ML estimates.

---

[12]Bastani and Selin (2014) report an ETI close to zero for wage earner, which is reminiscent of our estimate for all earners in Table 5.

TABLE 6 Bunching polynomial estimates of the ETI.

	Estimate	B-mean	B-5%	B-25%	B-75%	B-95%	B std-dev
(A) All tax payers, range [420,560], number of observations: 1,055,452.							
Chetty, 5	0.005	0.006	0.005	0.005	0.006	0.006	0.0
Chetty, 25	0.007	0.007	0.007	0.007	0.008	0.008	0.0
Chetty, 75	0.009	0.009	0.008	0.009	0.01	0.01	0.001
(B) All tax payers, range [450,530], number of observations: 615,813.							
Chetty, 5	0.005	0.005	0.005	0.005	0.005	0.006	0.0
Chetty, 25	0.007	0.007	0.007	0.007	0.007	0.008	0.0
Chetty, 75	0.009	0.009	0.007	0.009	0.01	0.011	0.001
(C) All tax payers, range [475,505], number of observations: 239,682.							
Chetty, 5	0.005	0.005	0.005	0.005	0.005	0.005	0.0
Chetty, 25	0.006	0.006	0.005	0.005	0.006	0.006	0.0
Chetty, 75	0.014	0.015	0.003	0.009	0.02	0.028	0.008
(D) Self-employed, range [420,560], number of observations: 25,827.							
Chetty, 5	0.05	0.052	0.05	0.052	0.053	0.055	0.001
Chetty, 25	0.101	0.103	0.1	0.102	0.104	0.106	0.002
Chetty, 75	0.139	0.14	0.133	0.137	0.142	0.146	0.004
(E) Self-employed, range [450,530], number of observations: 17,705.							
Chetty, 5	0.04	0.042	0.04	0.041	0.042	0.043	0.001
Chetty, 25	0.086	0.087	0.084	0.086	0.089	0.091	0.002
Chetty, 75	0.123	0.123	0.113	0.119	0.127	0.133	0.006
(F) Self-employed, range [475,505], number of observations: 10,013.							
Chetty, 5	0.023	0.023	0.022	0.023	0.024	0.024	0.001
Chetty, 25	0.055	0.055	0.051	0.054	0.057	0.06	0.003
Chetty, 75	0.105	0.118	0.053	0.081	0.14	0.219	0.063

Note: The table describes descriptives statistics of the bootstrap distribution for the polynomial based estimators of the ETI. Each Chetty, p row corresponds to the polynomial estimator suggested by Chetty et al. (2011) when p bins on either side of the kink are excluded. The *Estimate* column reports the ETI estimates, the *B-mean* column reports the mean estimate of the bootstrap sampling distribution, the columns *B-q%* report the q percentile of the bootstrap sampling distribution, finally the column *B std-dev* reports the standard deviation of the bootstrap sampling distribution. The estimation is carried over three distinct income ranges that are symmetric around the kink, [420,560], [450,530] and [475,505]. The results of each set of estimations are presented in subtables (A) to (C) for all tax payers and in sub-tables (D) to (F) for all self-employed.

5 | SUMMARY AND DISCUSSION

In this paper, we present a structural ML alternative to the polynomial bunching estimator of the ETI. Our approach has several advantages compared to the prevailing methodology. First, the relationship between the underlying theory of income formation and the statistical problem to be solved is made clear. This means, among other things, that we can explicitly account for the presence, and estimate the size, of optimization errors. Since people tend to bunch in an interval around the tax threshold, it is key to distinguish between optimization errors (implying that individuals do not exactly reach the income levels they aim for) and random components of the optimal earnings (e.g., due to wage- and preference heterogeneity). Second, our parametric approach is flexible and can easily be extended to more comprehensive models of income formation. This is exemplified in the supporting information Appendix E, where we consider a case with non-convex budget sets. It is also straightforward to extend the analysis by adding covariates to the underlying economic model in order to control for observed heterogeneity.

We show that the parameters of the observed earnings distribution (including the variance of the friction component) are identified, if the bunching near the tax kink in the distribution of observed earnings decreases as the variance of the process of optimization frictions increases. This result does not assume any specific distribution for the unobserved component of the optimal earnings, albeit a log-normal distribution for the friction component. Our results also suggest that the ML bunching estimator based on the normal–normal data-generating process outperforms the estimator based on the polynomial method whenever the distribution over the observed range can be approximated by a truncated normal distribution. Finally, we present an application of the ML bunching estimator using Swedish data, allowing us to compare

the ML estimate of the ETI with that of the polynomial approach. The ML point estimate of the ETI is around 0.007 for the full sample and between 0.05-0.1 in a subsample of self-employed individuals. As expected, the estimated standard deviation of the optimization error is small relative to the estimated standard deviation of the unobserved component in the optimal earnings. For the polynomial approach, the point estimate of the ETI is very sensitive to the exact procedure of measuring the counter-factual earnings distribution.

Future research may take several directions, and we briefly discuss two of them here. One would be to use the methodology proposed here for estimating 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 savings behavior. Note that this issue is relevant regardless of whether labor income and capital income are taxed jointly or separately. Part of the challenge of estimating more comprehensive models is also to examine the consequences of other distributional assumptions. In addition, the methodological framework developed here is suitable for analyzing the effects of complex tax reforms on income formation and welfare. In our view, this is one of the major advantages of a structural approach to income formation.

OPEN RESEARCH BADGES

This article has been awarded Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. Data is available at https://doi.org/10.15456/jae.2023278.1038307623.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

Aronsson, T., Jenderny, K., & Lanot, G. (2021). Maximum likelihood bunching estimators of the ETI. Working Paper, Umeå Economic Studies no 987.

Aronsson, T., Jenderny, K., & Lanot, G. (2022). The quality of the estimators of the ETI. *Journal of Public Economics*, *212*, 104679. https://www.sciencedirect.com/science/article/pii/S0047272722000810

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

Bertanha, M., McCallum, A. H., & Seegert, N. (2020). Better Bunching, Nicer Notching. SSRN WP 3144539.

Blomquist, S., Newey, W., Kumar, A., & Liang, C.-Y. (2021). On bunching and identification of the taxable income elasticity. *Journal of Political Economy*, *129*(8), 2320–2343.

Chetty, R., Friedman, J., Olsen, T., & Pistaferri, L. (2009). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: evidence from danish tax records. (*15617*), Working Paper Series: National Bureau of Economic Research.

Chetty, R., Friedman, J., Olsen, T., & Pistaferri, L. (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *Quarterly Journal of Economics*, *126*(2), 749–804.

Feldstein, M. (1995). The effect of marginal tax rates on taxable income: A panel study of the 1986 tax reform act. *Journal of Political Economy*, *103*(3), 551–72.

Gouriéroux, C., Monfort, A., & Vuong, Q. (1995). *Statistics and econometric models*, Statistics and Econometric Models 2 volume set: Cambridge University Press. https://books.google.se/books?id%3DgqI-pAP2JZ8C

Gruber, J., & Saez, E. (2002). The elasticity of taxable income: evidence and implications. *Journal of Public Economics*, *84*(1), 1–32.

Hausman, J. A. (1980). The effect of wages, taxes, and fixed costs on women's labor force participation. *Journal of Public Economics*, *14*(2), 161–194. http://www.sciencedirect.com/science/article/pii/0047272780900390

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

Moore, D. T. (2022). Evaluating tax reforms without elasticities: What bunching can identify: University of Michigan. https://dylantmoore.github.io/website/JMP _DylanTMoore.pdf

Pudney, S. (1989). *Modelling individual choice: The econometrics of corners, kinks, and holes.* Oxford and New York: Blackwell.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, *2*(3), 180–212.

Schwarz, M., & Van Bellegem, S. (2010). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters*, *80*(3), 236–241. https://www.sciencedirect.com/science/article/pii/S0167715209003988

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

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*(1), 1–25. http://www.jstor.org/stable/1912526

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of the article.

How to cite this article: Aronsson T., Jenderny K., & Lanot G. (2024). A maximum likelihood bunching estimator of the elasticity of taxable income. *Journal of Applied Econometrics*, 1–17. https://doi.org/10.1002/jae.3015