

DISCUSSION PAPER SERIES

IZA DP No. 13141

**The Ability Gradient in Bunching**

Spencer Bastani  
Daniel Waldenström

APRIL 2020

## DISCUSSION PAPER SERIES

IZA DP No. 13141

# The Ability Gradient in Bunching

**Spencer Bastani**

*Linnaeus University and CESifo*

**Daniel Waldenström**

*IFN, CEPR, CESifo, IZA and WIL*

APRIL 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9  
53113 Bonn, Germany

Phone: +49-228-3894-0  
Email: [publications@iza.org](mailto:publications@iza.org)

[www.iza.org](http://www.iza.org)

## ABSTRACT

---

# The Ability Gradient in Bunching\*

We analyze the relationship between cognitive ability and bunching in the context of a large and salient kink point of the Swedish income tax schedule. Using population-wide register data from the Swedish military enlistment and administrative tax records, we find that high-ability individuals bunch more than low-ability individuals. This ability gradient is stronger for the self-employed, but is also present among wage earners. We also use high-school GPA and math grades to analyze gender differences, finding a stronger ability gradient among men.

**JEL Classification:** H21, H24, J22, J24

**Keywords:** bunching, ability, skills, complexity, optimal taxation

**Corresponding author:**

Daniel Waldenström  
Research Institute of Industrial Economics (IFN)  
Grevgatan 34  
10215 Stockholm  
Sweden  
E-mail: [daniel.waldenstrom@ifn.se](mailto:daniel.waldenstrom@ifn.se)

---

\* We are grateful to Felix Bierbrauer, Emanuel Hansen, Håkan Selin and seminar participants at the Free University Berlin and IFN Stockholm for helpful comments. Financial support from the Swedish Research Council is gratefully acknowledged.

# 1 Introduction

According to the modern optimal tax literature, following Mirrlees (1971), the goal of the tax system is to redistribute between individuals with different abilities to generate income. In actual economies, such redistribution is often achieved by employing piecewise linear tax structures where marginal tax rates increase discontinuously at certain kink points. However, whether such tax structures are efficient in redistributing between high- and low-skill individuals or if high-skill individuals are able to avoid sharply increasing marginal tax rates has not been thoroughly investigated.

In this paper, we analyze how bunching at a large and salient kink point of the Swedish central government tax schedule is related to an individual's ability. To this aim, we use population-wide administrative tax records and a unique data set on individual ability from the military enlistment in Sweden.

Our analysis builds on a large and growing literature in public finance that estimates behavioral responses to tax changes by quantifying the extent to which taxpayers bunch at convex kinks of the income tax schedule. The basic idea of this estimation approach, which has become known as “bunching estimation”, is to compare the actual empirical distribution of taxable income with an estimated counter-factual distribution locally around a kink point to trace out the effects of increased marginal tax rates on individual behavior. The bunching approach has gained enormous popularity in the literature, a major reason being that it is a genuinely visual technique that provides transparent evidence of how people react to tax incentives.

The relationship between ability and bunching is important to study for at least three reasons. First, if the understanding of nonlinear tax incentives is systematically related to ability, this is important to uncover since it means that low-skill agents are not only penalized for having a low earnings capacity, but also for not making the best choices in relation to the tax system. Hence, complex tax incentives, motivated by a desire to redistribute from high-skill to low-skill individuals, might not work as intended if they are perceived differently by low-ability and high-ability people, or if high-ability individuals are able to circumvent higher marginal tax rates altogether.<sup>1</sup> Second, the optimal tax literature has typically assumed that individuals differ only in their earnings capacity, and not in terms of their labor-leisure preferences. If preferences are systematically related to ability, central results regarding the structure of optimal taxes need to be qualified.<sup>2</sup> Third, if it is the case that high-skill individuals respond more strongly to tax incentives, the reduction of effort by high-ability people might be more costly to society than a reduc-

---

<sup>1</sup>This is similar to the issue of heterogeneous take-up in the design of welfare programs, see for example, Kleven and Kopczuk (2011).

<sup>2</sup>For example, the seminal result of Atkinson and Stiglitz (1976) on the optimality of uniform commodity taxation in the presence of an optimal nonlinear labor income tax assumes that preferences are homogeneous and unrelated to ability.

tion in effort by low-ability people if the effort of high-skill individuals makes low-skill individuals more productive through complementarities in production.<sup>3</sup>

The key advantage of using military enlistment data is that they provide register-based, population-wide, measures of ability in young adulthood, before individuals have entered the labor market or enrolled in higher education. The focus of our analysis is thus on how non-acquired skills are related to taxpayer behavior. Ability is measured at age 18, and we measure taxable income several decades later.

Our main finding is that individuals with higher measured cognitive ability are substantially more likely to bunch at the kink. This means that there is an ability gradient in bunching. The previous literature has shown that taxable income elasticities tend to be larger for individuals with higher income (see e.g., Saez et al. 2012).<sup>4</sup> Our results indicate that responses are higher for individuals with higher skill among people with the *same* taxable income.

We complement our bunching analysis by running individual-level regressions estimating the marginal effect of cognitive ability on the probability that an individual has a taxable income *exactly* at the kink point. The regression analysis confirms the bunching results. We find a positive and statistically significant marginal effect of ability on the likelihood of sharp bunching at the kink point. This is a somewhat remarkable finding, considering the fact that we link ability measured at age 18 with incomes recorded more than 30 years later.

The ability gradient in bunching is particularly strong for the self-employed. This hints to the importance of income shifting, which is particularly relevant in the context of Sweden's dual income tax system, as documented earlier for the entire Swedish taxpayer population by Bastani and Selin (2014). We also analyze gender differences in the relationship between ability and bunching. In this analysis we use high-school GPA and math grades as proxies for cognitive ability since the military enlistment data includes very few women. Using this auxiliary data source, we find a clear ability gradient in bunching for men but almost no gradient for women.

Our study relates to several strands of the literature on behavioral responses to income taxation. First, it relates to the bunching literature that begun with the seminal contribution of Saez (2010), and was recently surveyed by Kleven (2016). In terms of the empirical setting, the most closely related paper is Bastani and Selin (2014) who studied bunching in Sweden during the period 1999-2005. They found virtually non-existent

---

<sup>3</sup>See e.g., Stiglitz (1982) who studies an optimal tax problem where high-skill labor supply makes low-skill labor supply more productive, and highlights cases where the labor supply of high-ability people should be subsidized at the margin in an optimal tax structure.

<sup>4</sup>This finding refers to behavioral responses along the *intensive* margin. Along the *extensive* margin, behavioral responses are typically found to be higher among low-income individuals, see e.g., Bastani et al. (2020b) who analyze heterogeneous extensive margin responses to a transfer program reform. In this paper, we analyze individuals with relatively high income and a strong attachment to the labor market, where the extensive margin is less relevant.

bunching among wage earners, which was remarkable given the large size of the kink point. In our paper, we study the same kink point during a more recent sample period. However, we are not concerned with the absolute magnitude of bunching, or the size of elasticities, but rather differences in bunching between ability groups.<sup>5</sup>

Our paper also contributes to the literature on how individuals understand and respond to the complexity of tax rules (see Chetty et al. 2009, Abeler and Jäger 2015, Taubinsky and Rees-Jones 2018) and papers highlighting the "regressive" nature of complexity, such as Aghion et al. (2018) who find, in the context of self-employed individuals in France, that more educated individuals adopt better tax-filing strategies.<sup>6</sup> Finally, as ability can be important for the possibilities to overcome optimization frictions, our paper is also related to papers that have analyzed the role of optimization frictions for observed bunching behavior (see, for example, Chetty 2012, Kleven and Waseem 2013, SØgaard 2019, Kosonen and Matikka 2019, and Gelber et al. 2020).

The rest of the paper is organized as follows. In section 2 we outline a theoretical framework that can be used to interpret our results and briefly describe our estimation strategy. Section 3 describes the data sources that we use and the institutional setting for our analysis. Section 4 describes our baseline bunching results and section 5 presents analyses of potentially important mechanisms and some extensions. Section 6 concludes.

## 2 Analytical framework

### 2.1 A bunching model with differing abilities

We consider a simple extension of the standard bunching model of Saez (2010) and Kleven (2016). In contrast to their model, we assume that individuals not only differ in terms of their ability  $\theta$ , but also in terms of their preference for working  $\xi$ . An individual's choice of taxable income  $z$  is guided by the following optimization problem:

$$\max_z \left\{ z - T(z) - \frac{z_p(\theta, \xi)}{1 + \frac{1}{e(\theta)}} \left( \frac{z}{z_p(\theta, \xi)} \right)^{1 + \frac{1}{e(\theta)}} \right\}. \quad (1)$$

The function  $z_p$  depends on both  $\theta$  and  $\xi$  and we assume that they are continuously distributed according to a smooth density function  $f(\theta, \xi)$ . The parameter  $e$  is a prefer-

---

<sup>5</sup>Another related Swedish study of bunching behavior is Seim (2017) who analyzed the wealth tax threshold in Sweden during the early 2000s. While not being a specific focus of his study, Seim conducted a heterogeneity analysis of wealth tax responses dividing taxpayers into below- and above-median cognitive ability using the same Swedish military enlistment data as in our study. The results did not indicate any significant differences across these two halves of the population, but the estimated excess mass at the wealth tax threshold was higher for the more able group.

<sup>6</sup>See also Bhargava and Manoli (2015), Hoopes et al. (2015), Feldman et al. (2016), and Bastani et al. (2020a). Bernheim and Taubinsky (2018) provide a recent overview of the emerging field of behavioral public economics.

ence parameter that depends on skill  $\theta$ , but not on  $\xi$ . Along a linear segment of the tax schedule  $T$ , with marginal tax rate  $\tau$ , the solution to (1) takes the familiar form:

$$z = z_p(\theta, \xi)(1 - \tau)^{e(\theta)}, \quad (2)$$

where it can be seen that  $z^p$  has the convenient interpretation as the taxable income of a  $(\theta, \xi)$ -individual in the absence of taxation ( $\tau = 0$ ). In this more general setting, there are differences in skill at each income level because income is determined by a combination of skill and preference for working.

With two dimensions of heterogeneity, there is not a one-to-one mapping between ability and the slope of individuals' indifference curves in the consumption-income space. However, the slope of individuals' indifference curves can still be used to characterize the set of individuals who bunch. Specifically, the set of bunchers at an income level  $\hat{z}$  are characterized by the region in  $(\theta, \xi)$ -space such that

$$1 - \tau - \Delta\tau \leq \left( \frac{\hat{z}}{z_p(\theta, \xi)} \right)^{\frac{1}{e(\theta)}} \leq 1 - \tau.$$

Each income level corresponds to a combination of  $\theta$  and  $\xi$ . Therefore, there will be heterogeneous elasticities at each income level, since the elasticity is a function of  $\theta$ . Formally, the joint distribution of  $\theta$  and  $\xi$ ,  $f(\theta, \xi)$ , determines a joint distribution of  $z$  and  $e$ ,  $\tilde{h}_0(z, e)$ . Using the results in Kleven and Waseem (2013), or more recently, Gelber et al. (2020), online appendix A.3, we can then write observed bunching in the presence of heterogeneous elasticities as follows:

$$\begin{aligned} B &= \int_e \int_{\hat{z}}^{\hat{z} + \Delta\hat{z}_e} \tilde{h}_0(z, e) dz de \\ &= h_0(\hat{z}) \int_e \int_{\hat{z}}^{\hat{z} + \Delta\hat{z}_e} \frac{\tilde{h}_0(z, e)}{h_0(\hat{z})} dz de \approx h_0(\hat{z}) E[\Delta\hat{z}_e]. \end{aligned}$$

Here we have adopted the simplifying assumption that  $\tilde{h}_0$  is constant in  $z$  on the bunching interval  $(\hat{z}, \hat{z} + \Delta\hat{z})$  for all  $e$ . This implies that the mass of individuals who move to the kink is just the average width of the bunching interval ( $E[\Delta\hat{z}_e]$ ) times the height of the counter-factual distribution, assumed to be equal to  $h_0(\hat{z})$ .

An alternative is to express the counter-factual income distribution in terms of  $z$  and  $\theta$ , such that:

$$B(E[\Delta\hat{z}_\theta]) = \int_{\theta \in \Theta} \int_{\hat{z}}^{\hat{z} + \Delta\hat{z}_\theta} \tilde{h}_0(z, \theta) dz d\theta \approx h_0(\hat{z}) E[\Delta\hat{z}_\theta].$$

In particular, we can estimate bunching for subgroups  $S$  of  $\Theta$ :

$$B_S(E_S[\Delta\hat{z}_\theta]) = \int_{\theta \in S} \int_{\hat{z}}^{\hat{z} + \Delta\hat{z}_\theta} \tilde{h}_0(z, \theta) dz d\theta \approx h_0(\hat{z}) E_S[\Delta\hat{z}_\theta]. \quad (3)$$

We define the *excess mass* at income level  $\hat{z}$  in skill group  $S$  as follows:

$$b_S = \frac{B_S(E_S[\Delta\hat{z}_\theta])}{h_0(\hat{z})}. \quad (4)$$

We can also define the compensated taxable labor income elasticity locally at  $\hat{z}$  for individuals belonging to skill group  $S$  as

$$e_S = \frac{E_S[\Delta\hat{z}_\theta]}{\hat{z}} \bigg/ \frac{\Delta(1-\tau)}{(1-\tau_1)}. \quad (5)$$

Inserting (3) into (5) and rearranging gives

$$e_S = \frac{B_S(E_S[\Delta\hat{z}_\theta])}{\hat{z} \times h_0(\hat{z}) \times \frac{\Delta(1-\tau)}{(1-\tau_1)}}. \quad (6)$$

In our empirical analysis, we mainly focus on skill groups based on the deciles of the distribution of cognitive ability.

## 2.2 Estimation approach

Bunching estimation amounts to comparing the density of the empirical distribution of taxable income with the density of an estimated counterfactual distribution locally around a kink point. The key methodological challenge is to construct the counterfactual distribution, that is, the distribution of taxable income that would prevail in the absence of a kink. In this paper, we follow Chetty et al. (2011) and fit a polynomial to the observed income distribution, omitting an income band around the kink.

We express taxable income in terms of the distance to the kink point  $\hat{z}$ . Data are collapsed into bins of width 1000 SEK (roughly 100 EUR) and each bin  $j$  is represented by an income level  $Z_j$ , defined as the mean absolute income distance between the observations falling within income bin  $j$  and the kink point. We then specify a "doghnut-shaped" region around the kink consisting of the disconnected set  $[-R, \hat{z} - \delta] \cup [\hat{z} + \delta, R]$  which contains the observations that will be used to estimate the counterfactual distribution. Here,  $[-R, R]$  refers to the "wide" bunching window and  $[-\delta, \delta]$  refers to the "small" bunching window. The idea is that the small bunching window should capture exactly those individuals who bunch at the kink. These individuals are then excluded when estimating the counterfactual distribution. Since we estimate bunching for various subgroups of the population and for different years, we do not choose  $\delta$  based on visual inspection (which is commonly done in the literature), but instead fix a baseline  $\delta$  in our analysis

and then report extensive robustness checks with respect to this parameter.<sup>7</sup> Consistent with Bastani and Selin (2014), our baseline analysis focuses on the wide bunching window  $[-50k, 50k]$  and the small bunching window  $[-5k, 5k]$ , but sensitivity analysis (see appendix figure A3) shows that varying these windows does not alter the main results.

The counterfactual distribution is estimated using the following regression model:

$$C_j = \sum_{i=0}^q \beta^i Z_j^i + \sum_{s=-\delta}^{\delta} \gamma_s \mathbf{I}[Z_j = s] + \eta_j \quad (7)$$

where  $C_j$  is the number of individuals in income bin  $j$ ,  $q$  is the degree of the polynomial in  $Z_j$ ,  $\beta_i$  is the regression coefficient on the  $i$ :th order polynomial term, and  $\gamma_s$  are dummy variables for observations within the small bunching window, finally,  $\eta_j$  accounts for the error of the polynomial fit.

Denote by  $\hat{C}_j$  the predicted values from regression (7). Bunching is estimated as the number of taxpayers at the kink (denoted by  $\hat{B}$ ) relative to the average height of the counterfactual distribution in the band  $[-\delta, \delta]$ . Formally, we have:

$$\hat{b} = \frac{\hat{B}}{\sum_{j=-\delta}^{\delta} \frac{\hat{C}_j}{2\delta+1}} \quad \text{where} \quad \hat{B} = \sum_{j=-\delta}^{\delta} (C_j - \hat{C}_j). \quad (8)$$

The quantity  $\hat{b}$  in (8) is the empirical *excess mass*, namely, the empirical counterpart of (4). We compute standard errors using bootstrap on binned data by sampling from the empirical distribution function associated with the observed income distribution, computing  $\hat{b}$  repeatedly.<sup>8</sup>

Finally, we also run individual-level regressions to complement our bunching analysis. In addition to linking cognitive ability and taxpayer behavior at the individual level, the regressions allow us to study bunching using a different approach to constructing counterfactual outcomes (a linear multivariate regression instead of a polynomial in income, as in the bunching method). The regression approach also enables us to run placebo regressions (see section 5).

### 3 Data and institutional setting

We use data from several population-wide administrative registers in Sweden. From the population register, we retrieve the full population living in Sweden born in the years 1951-1975. From the military enlistment register, we retrieve cognitive ability scores for all men born in the same years. This means that most test scores are observed between 1969 and 1993. About two-thirds of all males have ability scores and the remaining men

<sup>7</sup>Diamond and Persson (2017), page 17 use a similar reasoning but provide an automated approach.

<sup>8</sup>We use the Stata program `bunchcount`, estimating the counterfactual distribution using a seven-degree polynomial and bootstrap standard errors with 500 replications.

have missing scores for different reasons, mainly because they did not have a Swedish citizenship, were chronically ill, incarcerated or some other extraordinary reason.

From the income tax register, we obtain data on individual taxable labor and capital income for wage earners and for self-employed individuals owning either a closely held corporation or a sole proprietorship.<sup>9</sup> The analysis focuses on the income years 2012-2016. This is the latest period for which we have income tax records when all the men in our sample are of working age (the common pension age in Sweden is 65). Furthermore, the studied period is largely devoid of tax reforms, and follows a series of smaller income tax reforms launched in 2007-2011. It is also a period where the Swedish economy was stable, having stabilized following the great recession in 2008-2009. We provide a broad picture of the income distribution, and the location of the kink point, for each year between 2012 and 2016 in figure A1 in the Appendix. To complete our data-set, we add information from the education register about educational attainment (number of years of education and field of highest degree at the 3-digit level) and high-school GPA, observed for everyone born 1955 and later, and final math grades, observed for everyone born 1966 and later. Descriptive statistics and sample attrition are presented in the appendix (tables A1 and A2).

The measurement of cognitive ability in the military enlistment took place at around age 18. There were four different ability tests: (i) inductive ability (reasoning), (ii) verbal comprehension, (iii) spatial ability (metal folding), and, (iv) technical comprehension. In order to ensure comparability across the sub-tests and also over time, the enlistment authorities transformed the test scores on each of these tests into a nine-degree normal distribution, a so-called stanine scale, and finally generated an overall cognitive ability stanine score based on the four individual test stanines. In our main analysis, we use an unweighted average of the stanine scores across the sub-tests. This is coherent with the overall test score constructed by the military enlistment, but gives us a slightly more detailed measure of cognitive ability (in  $9 \times 4 = 36$  levels) since we avoid rounding off numbers, which is helpful when we divide the population into ability decile groups.

A key advantage of the military enlistment data on ability is that they measure skills in young adulthood, before enrollment into college or occupational choices. A large number of scholars have used these ability scores in different applications and found them to be coherent over time and robustly correlated with a range of important economic outcomes later in life.<sup>10</sup> In relation to hourly wages (that are often used to proxy ability in empirical applications), the enlistment data provides us with a measure that is

---

<sup>9</sup>Information about ownership of non-listed corporations comes from a specifically matched firm-individual ownership database, FRIDA (we use the register variables `bkufoab` or `bfoab`). Sole proprietorships are associated with individuals reporting income from business activities (variable `nakte`) or limited liability partnerships (`nakthb`).

<sup>10</sup>See, for example, Lindqvist and Vestman (2011), Edin et al. (2018) who analyze the relationship between cognitive ability and labor market outcomes.

more closely related to the skill-measure envisioned by the optimal income tax literature (Mirrlees 1971).

The Swedish military enlistment data cover almost exclusively men. In order to analyze gender differences, we make a supplementary analysis using high-school GPA and final high-school math grades as proxy for cognitive ability. These school grades are measured at around the same time in life as the military enlistment records, but are imperfect substitutes for our main ability measure since they reflect individual education effort and are sensitive to the institutional details of the school system.<sup>11</sup> Apart from the ability measure, we draw information about women from the same administrative registers described above. Notice that taxable income has been assessed individually in Sweden since 1971 and labor force participation at age 35-65 (the age span in our main analysis of outcomes 2012-2016) was 91 percent for men and 87 percent for women.<sup>12</sup>

An important aspect of our contribution is that we conduct our analysis of skill-group differences in the context of the first kink point of the Swedish central government tax schedule, previously analyzed by Bastani and Selin (2014). In some respects, this is an ideal laboratory to examine differential responses in bunching behavior. The kink point is one of the largest kinks that has been studied in the bunching literature (an increase in the marginal tax rate of 20 percentage points in most years) which implies that optimization frictions and salience concerns should be less of an issue as compared to other bunching settings. Moreover, it is located in the upper middle part of the income distribution where many taxpayers are located who have a strong attachment to the labor market, weakening concerns about the identification of the counter-factual distribution.

## 4 Baseline results

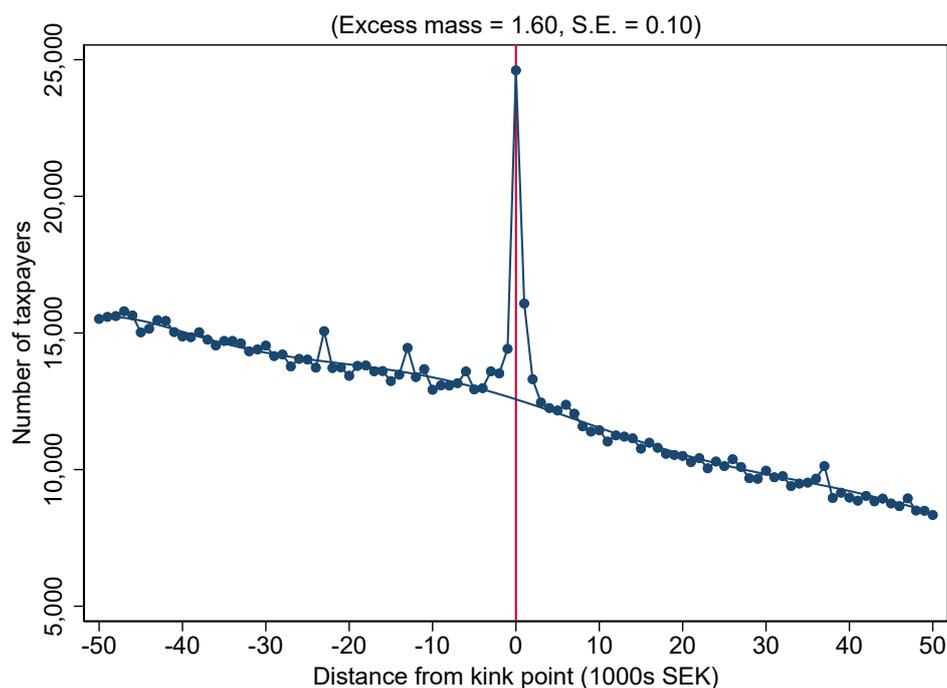
Figure 1 shows bunching results for our main sample of men with taxable income measured 2012-2016. Income is expressed in bins of thousands of Swedish kronor (SEK), which is equivalent to hundreds of euros or US dollars, and is measured relative to the location of the kink point. The figure depicts statistically significant bunching at the kink point, and the estimated excess mass is 1.60 with a bootstrapped standard error of 0.10. This finding confirms the fiscal importance of the large break in the Swedish government income tax schedule and is also in line with the findings of bunching in Sweden during the 2000s by Bastani and Selin (2014).

---

<sup>11</sup>School effort can be important both along the intensive margin (working harder to acquire better grades) and the extensive margin (whether to acquire a high-school degree or not).

<sup>12</sup>Note that the sample population for men in the analysis of gender differences deviates somewhat from the population in our main analysis since we do not need to condition the sample on the availability of data from the military enlistment.

Figure 1: Bunching at the kink, 2012-2016

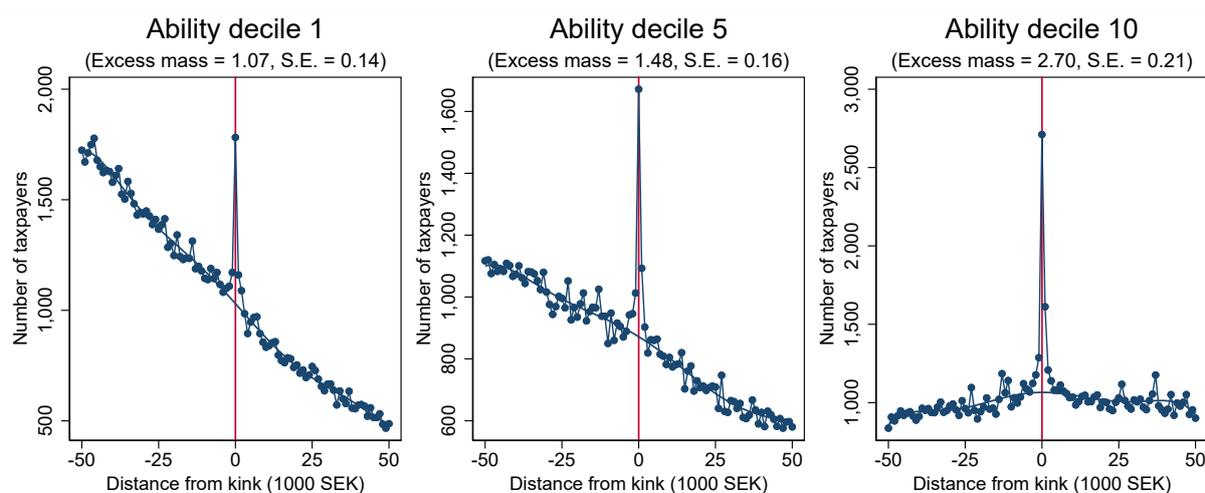


*Note:* The figure shows bunching among all men (born 1951-1975) at the largest kink of the Swedish income tax schedule (payment of central government income tax) for taxable labor incomes during 2012-2016 (pooled annual data).

Next, we turn to the main question of the paper, namely, whether bunching behavior differs systematically for people with different levels of cognitive ability. We exploit the fine-grained register data on individual ability to analyze this issue, and in this section, we do this by splitting the population into deciles of cognitive ability, ranked from the lowest (decile 1) to the highest (decile 10). Within each decile, we estimate bunching around the same statutory kink point. This means that we compare the decile-specific observed mass of income earners around the kink point with the estimated decile-specific counterfactual density around the kink point.

Figure 2 shows bunching estimates for three of the ten ability deciles (the appendix figure A2 shows all ten deciles). Ability decile 1 has an excess mass of 1.07 (standard error 0.13), and there is thus statistically significant bunching within this group. Ability decile 5 also displays bunching, but with an excess mass of 1.48 (0.15) and the top ability decile, decile 10, has an excess mass of 2.70 (0.22). All these ability groups are thus associated with statistically significant bunching at the kink point, and the results add up to the population-wide excess bunching with an estimated excess mass of 1.60 shown in figure 1. However, the magnitudes of the excess masses are not the same in these ability groups, and the next step is to examine whether there is a systematic pattern in these differences across the ability distribution.

Figure 2: Bunching across cognitive ability deciles

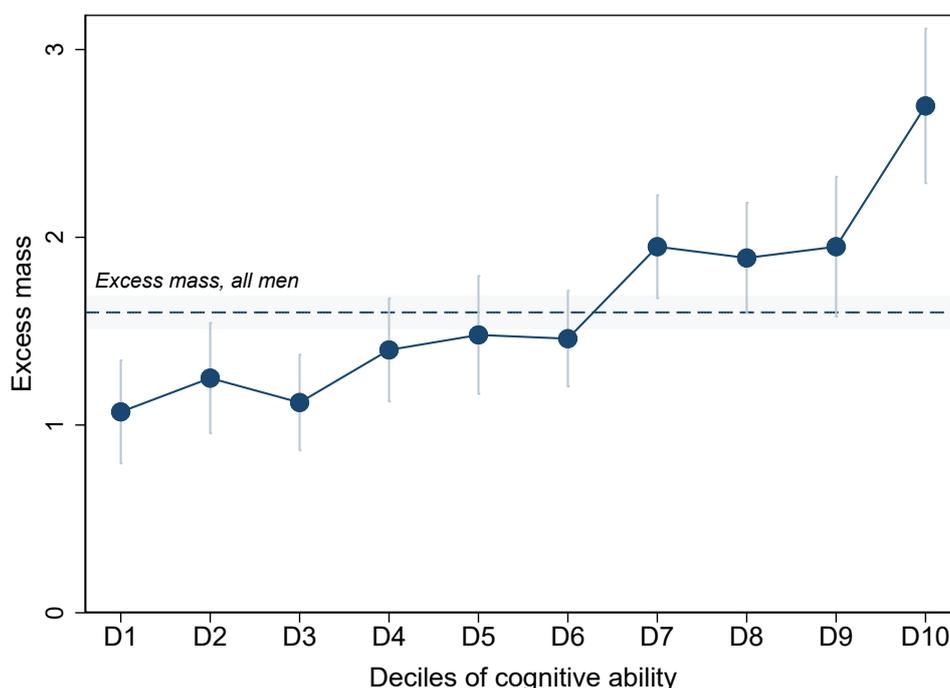


Note: The graph shows bunching at the kink for the period 2012-2016 (pooled data) for all adult men born 1951-1975 divided into deciles of cognitive ability.

Figure 3 depicts bunching estimates for all the ten ability deciles. It also includes a dashed line showing the estimated bunching in the full male population. The figure shows that there is a clear, almost monotonic, increase in bunching in the level of individuals' cognitive ability. Ability decile 10 has an excess mass at the kink point that is almost three times as large as the excess mass in ability decile 10 and twice as high as the excess mass in the full male population. These differences are statistically significant. Looking at the other deciles, we note that deciles seven and higher have a higher bunching than the in the population at large, whereas all deciles from six and below have less bunching.

These results make clear that there is an *ability gradient* in bunching at the kink point. To our knowledge, this relationship between cognitive ability and bunching has not been shown before in the literature. In the appendix (section A.2), we present results that show that the ability gradient is highly robust to perturbations in the estimation framework, in particular using different sizes for the small and wide windows around the kink point.

Figure 3: Ability gradient in bunching



*Note:* The graph displays the excess mass and 95% confidence intervals ( $\pm 1.96$  times standard error, bootstrapped with 500 replications) at the kink point estimated separately for each decile in the male cognitive ability distribution (for all men in our main sample population) and for labor incomes earned during 2012-2016 (pooled data). The underlying bunching estimation for these ten excess mass estimates is presented in appendix figure A2.

## 5 Extensions and potential mechanisms

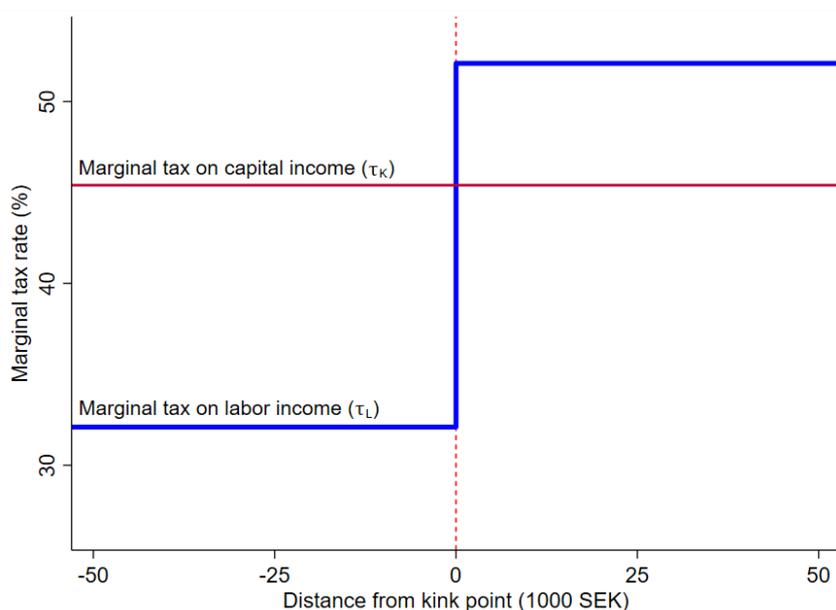
The previous section showed that bunching is higher among high-ability groups than among low-ability groups. In this section, we examine some of the potential mechanisms through which this relationship may be operating. First, we divide the population into wage earners and self-employed. The importance of this distinction in the bunching context was emphasized by Chetty et al. (2011). Second, we run individual-level regressions where we regress a dummy variable for having an income exactly at the kink point on individuals' ability, cohort and income year fixed effects and, in some specifications, controls for whether the individual is self-employed, has positive divided income and a number of dummies reflecting level and type of educational attainment.

### 5.1 Self-employed vs. wage earners

The division between self-employed and wage earners has been studied extensively in the bunching literature, motivated by the fact that self-employment occupations allow for a greater influence over income flows from capital and labor than is the case for wage earners. In this section, we estimate ability gradients separately for self-employed and wage earners. The separation between self-employed and wage earners groups is

particularly motivated in the context of the Swedish dual tax system where the taxation of labor income is separated from the taxation of capital income. While such a tax system has several merits (for example, by allowing the marginal tax rate on capital income to be different than that applying to labor income), it provides incentives for the shifting of income from the labor income to the capital income tax base in cases where the tax rate applying to capital income is lower than the marginal tax rate on labor income. As the incentive for income shifting depends on the difference between the marginal tax rates applying to labor and capital income, income shifting incentives arise discontinuously at the kink point, as illustrated by figure 4. Thus, the observed bunching at the kink most likely reflects a combination of traditional labor supply responses and income shifting.

Figure 4: Marginal tax rates on labor and capital income around the kink point.



*Note:* Statutory marginal tax rates in 2016. Labor income marginal tax rate below kink point equals the average municipal income tax rate (32.1%) and to the right it equals this average municipal income tax rate plus the central government tax rate on labor income, 20%. No adjustment is made for the tax content in social security contributions. The capital income marginal tax rate equals the combined effect of the corporate income tax (22.6%) and the tax rate on dividend income from listed firms (30%).

Figure 5 shows bunching around the kink point for self-employed and wage earners, and in line with previous studies, we find more bunching activity among the self-employed than among wage earners. However, it is notable that we do encounter significant bunching also among wage earners, contrasting the previous findings of Bastani and Selin (2014) who found no visible bunching for wage earners in the earlier time period, 1999-2005. This suggests there could be time trends in bunching and, more generally, dynamic taxpayer responses, such as learning of the the tax code, or inter-generational differences.

Figure 5: Bunching: Wage earners vs. self-employed.

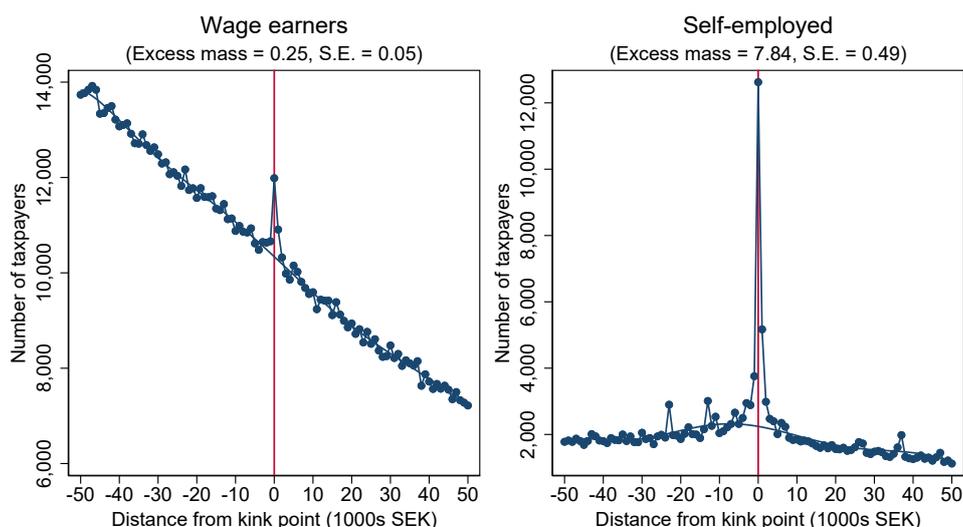


Figure 6 examines differences in bunching between different ability deciles of wage earners and self-employed. Like in our baseline analysis above, we begin by presenting results for the 1st, 5th and 10th ability deciles. Among wage earners, there is no bunching in the bottom decile, bunching appears larger in the middle decile, and is statistically significant, yet small, in the top ability decile. For the self-employed, bunching is many times larger across the entire ability distribution, but there is still a similar pattern in the sense that the estimated excess mass is the highest in the top ability decile.

Figure 6: Bunching across ability deciles: Wage earners vs self-employed

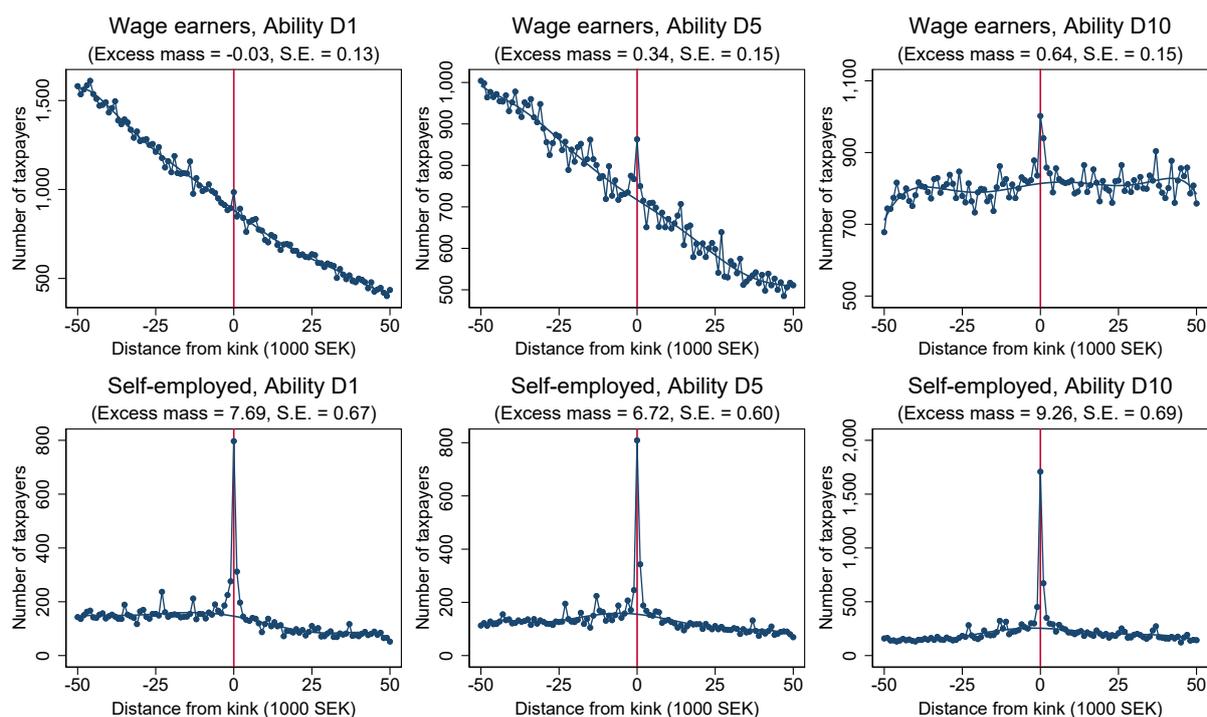
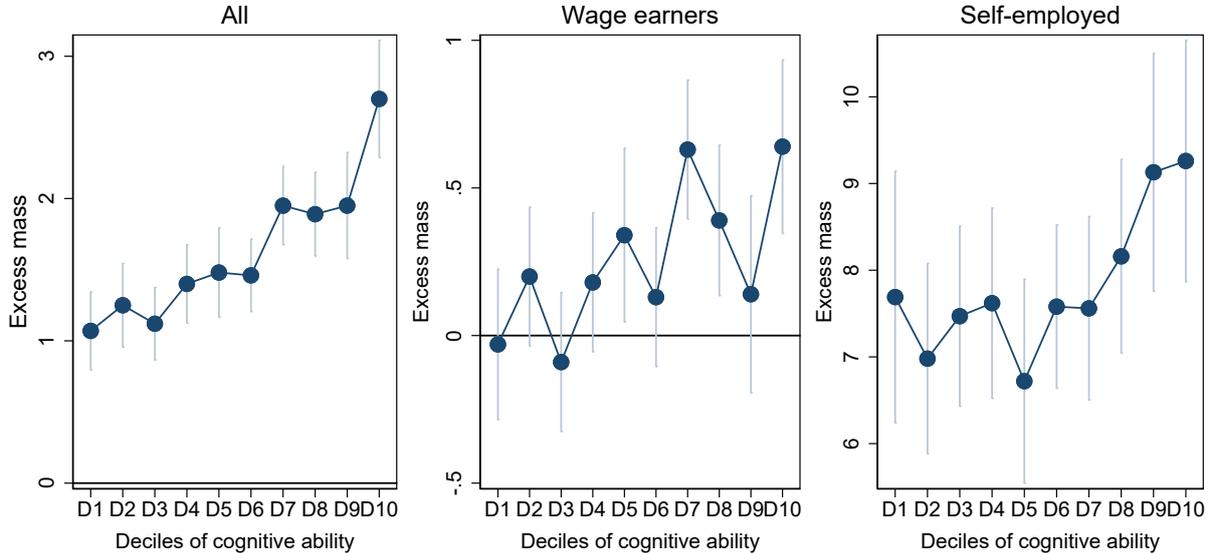


Figure 7 presents excess mass estimates for all ten deciles, which allows to explore if there are systematic patterns. The figure suggests that both groups exhibit positive ability gradients. The smaller sample sizes when analyzing decile partitions within the two occupational subcategories imply that estimates are less precise than when using the full sample. However, differences are statistically significant at least when comparing the top and bottom ability deciles. A potential explanations for the observed pattern in the case of wage earners could be that highly skilled individuals substitute time in the regular market with time devoted to financial investment, in order to secure a higher return, consistent with the models of Gahvari and Micheletto (2016) and Gerritsen et al. (2019). That is, while income shifting could be a major source of bunching in general and also related to the ability gradient in bunching, the results suggest that bunching also could arise through other channels.

Figure 7: Ability gradient in bunching: Wage earners vs self-employed



## 5.2 Individual-level regressions

The preceding analyses were conducted on binned data and were based on standard bunching estimation. An alternative way to estimate ability gradients in bunching is to use individual-level data and estimate the marginal effect of cognitive ability on the probability of a person locating *exactly* at the kink point.

We run three sets of individual-level bunching regressions. The first specification uses our *cognitive ability z-score*,  $Cog_i$ , to capture the ability effect:

$$I\{y_{iat} = \hat{y}\} = \eta_a + \lambda_t + \beta \cdot Cog_i + \varepsilon_{iat}. \quad (9)$$

where  $I\{y_{iat} = \hat{y}\}$  indicator variable for locating *exactly* at the kink point  $\hat{y}$ ,  $\eta_a$  represents cohort fixed effects, and  $\lambda_t$  are year fixed effects. Incomes are the same as in the bunch-

ing analysis, measured in the period 2012-2016, and standard errors are clustered at the individual level. Since cognitive ability is measured at age 18, we study incomes realized around 30 years later (at ages 40-65) and the estimated ability coefficient can therefore be given a causal interpretation. We include controls for potentially relevant mechanisms: self-employment status (indicator variable  $Self_{it}$ ), having dividend income (indicator  $Div_{it}$ ) and educational attainment (dummies for education length and field of education,  $Educ_i$ ).<sup>13</sup> We view these additional factors as potential mediators that indicate how much of the reduced form-effect of ability on bunching that operates through occupational choice, capital income and skills acquired through the education system.

Table 1 shows the results from this first specification, and they show that ability has a statistically significantly positive effect on the probability to earn a taxable income at the kink point. This confirms the ability gradient result from our bunching analysis. The regression coefficient is not large: one standard deviation increase in cognitive ability leads to about a three hundredth of a percent increase in the likelihood of earning exactly at the kink. However, when interpreting the size of the coefficients, one should keep in mind that the number of people locating exactly at the kink point each year is roughly two thousand individuals out of a total of around 1.2 million annual observations. The statistically significant and positive ability effect remains, but is reduced, after including controls for self-employment status and having positive dividend income.

---

<sup>13</sup>The control for self-employment is a dummy equal to one if an individual either the main income from a sole proprietorship or owns a closely held corporation.

Table 1: The effect of cognitive ability on the probability to locate exactly at the kink

	(1)	(2)	(3)	(4)	(5)
	Probability of earning at the kink $I\{y_{iat} = \hat{y}\}$				
<i>Cog</i>	0.00028*** (0.00002)	0.00021*** (0.00002)	0.00018*** (0.00002)	0.00006*** (0.00001)	0.00007*** (0.00002)
<i>Self</i>		0.00482*** (0.00017)		0.00339*** (0.00014)	0.00332*** (0.00013)
<i>Div</i>			0.00132*** (0.00004)	0.00058*** (0.00003)	0.00047*** (0.00003)
<i>Self</i> × <i>Cog</i>				0.00058*** (0.00015)	0.00058*** (0.00015)
<i>Div</i> × <i>Cog</i>				0.00008* (0.00003)	0.00006 (0.00003)
<i>Self</i> × <i>Div</i>				0.00234*** (0.00022)	0.00222*** (0.00022)
<i>Self</i> × <i>Div</i> × <i>Cog</i>				0.00003 (0.00023)	0.00006 (0.00023)
Cohort FE	Yes	Yes	Yes	Yes	Yes
Income year FE	Yes	Yes	Yes	Yes	Yes
Educ controls	No	No	No	No	Yes
Obs	6,004,200	6,004,200	6,004,200	6,004,200	5,706,505
R <sup>2</sup>	0.0001	0.0022	0.0004	0.0025	0.0027

Note: Pooled regressions for the period 2012-2016. *Cog* denotes standardized cognitive ability, and the coefficient shows the effect of increasing *Cog* by one standard deviation on the probability of having an annual income exactly at the kink point. *Self* is a dummy variable equal to one if individual is self-employed a particular year and *Div* is a dummy variable equal to one if the individual has a positive dividend income. Standard errors are clustered on individuals. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001.

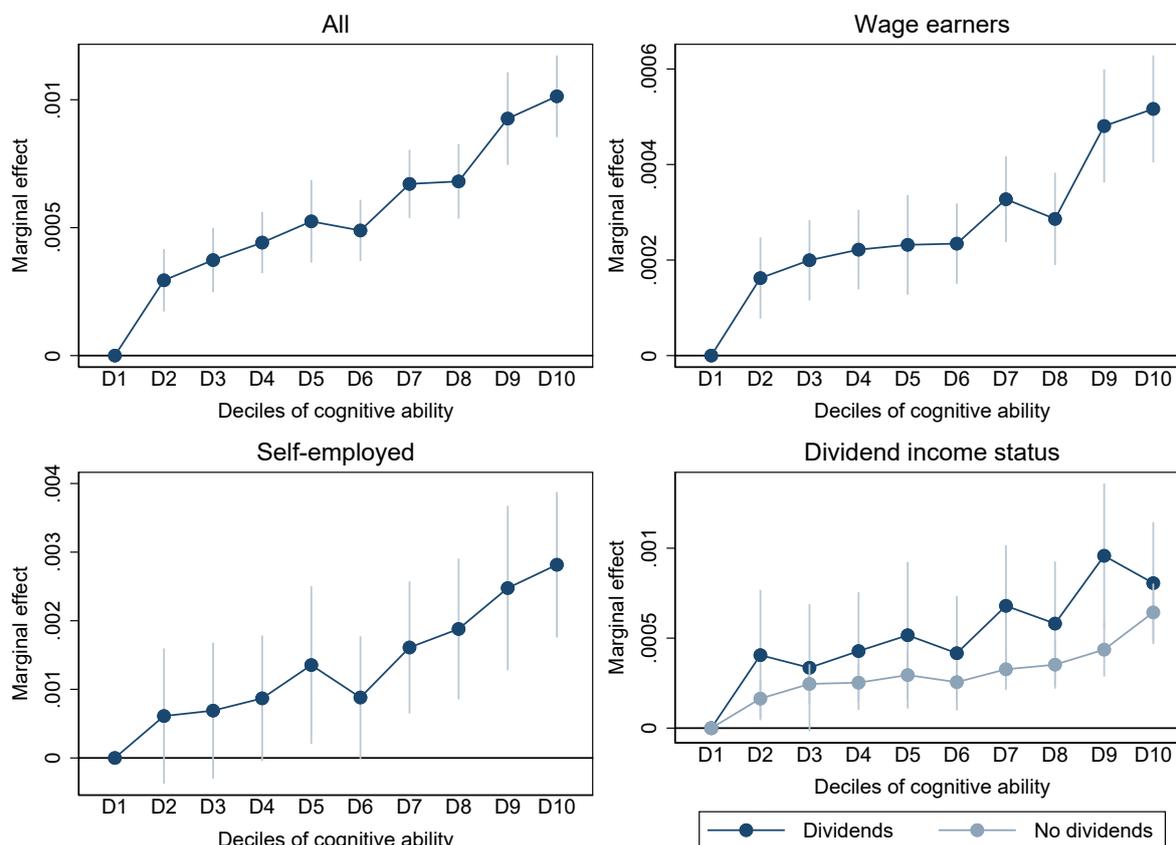
The second specification uses *ability decile dummies*,  $CogDec_i^d, d = 1, 2, \dots, 10$  that take on the value 1 if individual  $i$  belongs to decile  $d$  of the distribution of cognitive ability. This allows for a flexible form for the relationship between ability and sharp bunching. Formally, we estimate the following equation:

$$I\{y_{iat} = \hat{y}\} = \eta_a + \lambda_t + \sum_d \beta^d \cdot CogDec_i^d + \varepsilon_{iat}. \quad (10)$$

Figure 8 depicts the estimated marginal effects of each decile dummy relative to ability decile 1 for five different samples: All men, wage earners, self-employed, individuals with no dividend income, and individuals with positive dividend income (regression tables are found in appendix table A3). The results are very consistent with the results for the ability gradient found in our bunching estimation. Marginal effects are almost monotonically increasing in ability in all specifications. In the full male population, individuals in deciles 9 and 10 are twice as likely to bunch as individuals in the median ability deciles 5 and 6 are. Self-employed individuals are in general more likely to bunch

than wage earners are, but both groups exhibit positive ability gradients. We see a similar pattern in the bottom-right panel, where we divide the sample based on whether individuals have dividend income or not. These results suggest that income-shifting cannot completely explain the ability gradient.

Figure 8: Regression-based ability gradient in bunching



Note: The figure shows estimated marginal effects from regressions (10) where the only difference between the four panels is the samples which are used (stated in panel headings).

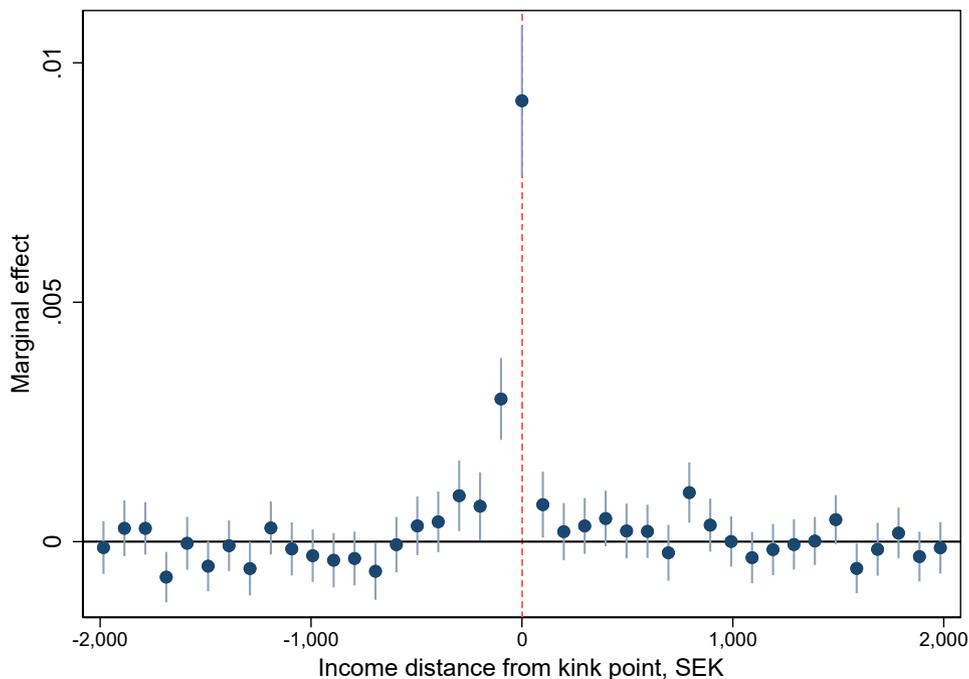
In the third specification of this section, we run *placebo regressions* by estimating rolling regressions with indicators  $I\{y_i = y\}$  for different income levels  $y$  around the kink point as the dependent variable:

$$I\{y_{iat} = y\} = \eta_a + \lambda_t + \beta \cdot Cog_i + \varepsilon_{iat}, \quad y \in \{\hat{y} - 2000, \hat{y} - 1900, \dots, \hat{y} + 2000\} \quad (11)$$

Notice that this exercise limits the sample to individuals with income levels  $y$  in the interval  $[\hat{y} - 5000, \hat{y} + 5000]$ , which implicitly acts as controlling for income in a broad sense. Figure 9 presents marginal effects of cognitive ability on the probability of locating at different income levels around the kink point. The results show that ability has a much larger effect on the probability of locating *exactly* at the income level of the kink point relative to the effect of ability on the probability of locating at adjacent income

levels. This placebo test thus reinforces our main results. In the appendix (see appendix figure A4), we extend the placebo analysis by including controls for self-employment status and an indicator for having positive dividend income, resulting in very similar patterns.

Figure 9: Placebo regressions: Ability effect on locating at incomes around the kink



Note: The figure presents estimated marginal effects of  $Cog$  on the probability to earn an income of  $y \in \{\hat{y} - 2000, \hat{y} - 1900, \dots, \hat{y} + 2000\}$  where  $\hat{y}$  is the kink point income level (see equation (11))

### 5.3 Gender differences using school grades

Is the ability gradient in bunching exclusive for the male population or is it also relevant among women? A limitation of the military enlistment data is that it contains data for very few women. In this section, we use high-school grade-point averages (GPA) and high-school math grades as proxies for ability to enable an analysis of gender differences. As discussed in section 3, there are several shortcomings associated with using grades as a proxy for ability, such as the potential confounding effect of education effort. Despite these shortcomings, we estimate bunching in taxable income during 2012-2016 for men and women divided up into ten GPA deciles and five math grade groups.<sup>14</sup>

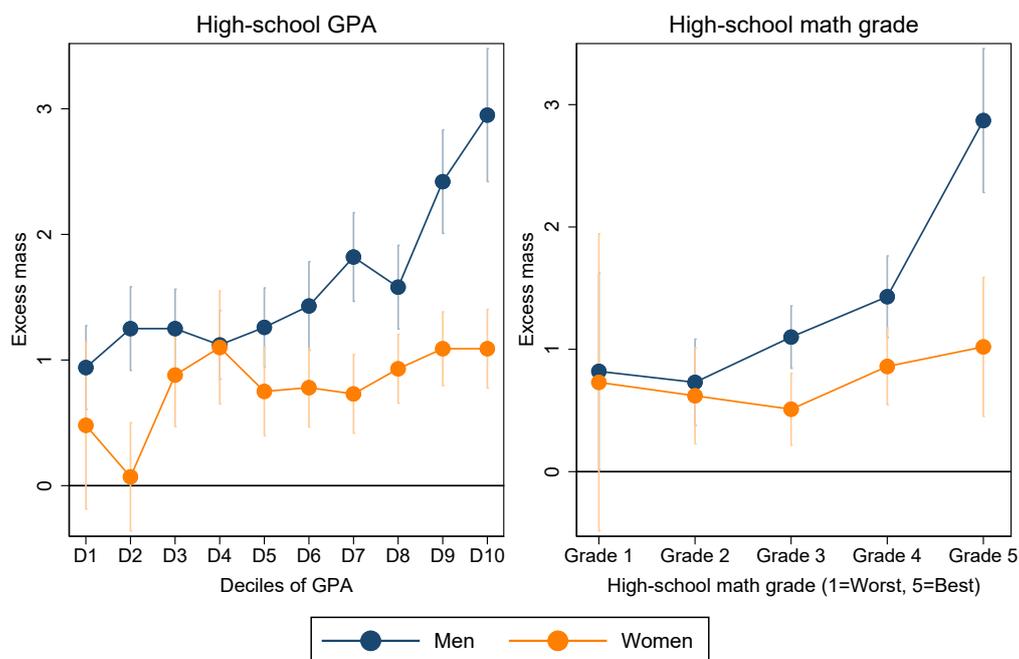
Figure 10 presents the estimated bunching excess mass for the ten GPA deciles and five math grade groups.<sup>15</sup> The main finding is that while there is an ability gradient

<sup>14</sup>The Swedish GPA in our sample comprises an almost continuous score between 1 and 5 (the average of 10-15 different subject grades between 1 and 5). Math grades also range between 1 and 5, but each grade represents a different share of the students since Sweden practiced relative grading schemes (the shares of students were allocated as: 7 percent received grade 5, 24 percent grade 4, 38 percent grade 3, 24 percent received 2 and 7 percent received grade 1).

<sup>15</sup>In general, men bunch more than women in our sample. The excess mass at the kink is 1.49 for men

among men, that has a similar shaped to the gradient based on the military enlistment scores, there is no ability gradient among women. In other words, the difference in bunching between high-skill men and low-skill men is much larger than the difference in bunching between high-skill women and low-skill women. The results look roughly the same independently of whether we use GPA or math grades.

Figure 10: Gender and ability gradient in bunching: High-school GPA and math grades



Note: Bunching during income years 2012-2016 in the sample of men and women born 1955-1975.

School grades are determined by many different factors and it is therefore not obvious how one should explain the seemingly stronger ability gradient among men. Bunching was found in the previous analysis to be larger for the self-employed relative to wage earners. Hence, one explanation could be that highly skilled men select into self-employment to a greater extent than low-skill men, whereas selection into self-employment is more uniform across the ability distribution for women. Figure 11 divides the population into self-employed and wage-earning men and women and repeats the analysis. We find a gender difference in the ability gradient both among wage earners and the self-employed, but it is clearer among the self-employed. An interesting finding is that the gender difference in bunching is strongest among the self-employed with the highest math grades.

---

and 0.90 for women, as compared to 1.27 for the total sample under consideration in this section. See table A2 for more information about the difference between the sample studied in this section, and the sample studied in the main analysis.

Figure 11: Gender and ability gradient in bunching: Self-employed and wage earners



Note: Bunching during income years 2012-2016 in the sample of men and women born 1955-1975.

## 6 Conclusions

We have studied the relationship between cognitive ability and bunching at kink points of the income tax schedule using population-wide military enlistment registers and administrative tax data. Our approach relies on relating ability, measured during young adulthood, to taxable income observed several decades later.

The analysis suggests that behavioral responses to taxation depend on the cognitive ability of individuals. Bunching is found broadly in the population, but individuals in the top decile of the ability distribution are twice as likely to bunch as compared to the average individual and three times more likely than individuals in the bottom ability decile. This ability gradient in bunching is almost monotonic and also robust to changes in the estimation strategy. Exploring potential mechanisms, we confirm previous findings that self-employed individuals are more likely to bunch, likely because of tax-driven incentives to shift income from labor to capital, but we also find bunching among wage earners, and there is a clear ability gradient for both self-employed and wage earners.

We have also run individual-level panel regressions, estimating the marginal effect of cognitive ability on the probability of locating exactly at the kink point, controlling for background characteristics, that confirmed the findings of the bunching estimation. In a final analysis, we used school-grade data as proxies for cognitive ability and studied gender differences, finding a weaker ability gradient among women than among men.

An interpretation of these findings is that the awareness of complex tax incentives could have an ability gradient. If high skill individuals understand the incentives inherent in the tax system better than low-skill individuals, this introduces regressivity into the tax system, as low-skill individuals are penalized not only for having a low earnings capacity, but also for not making the best choices in relation to the tax system. Another aspect is that, if high-skilled individuals are more sensitive to taxation, as the bunching evidence suggests, there are efficiency gains to be had if these individuals could be made subject to more lenient taxation or be better targeted in other ways. If high-skill individuals reduce their labor effort relatively more than low-skill individuals in response to tax changes, the efficiency costs of tax changes could be underestimated if a reduction in high-skilled effort makes low-skill workers less productive. The focus on ability here is crucial, as the increased effort of high-*income* individuals could instead be associated with rent-seeking activities, as suggested by Lockwood et al. (2017).

As a final set of remarks, we would like to highlight some limitations of our analysis. First, while we document a clear relationship between ability and taxpayer behavior using a large administrative register data set, we cannot fully pinpoint the underlying mechanisms. Our occupational decomposition in the bunching analysis, and the inclusion of controls in the individual-level regressions, provided some clues. However, to provide a precise identification of mechanisms, a carefully designed experiment would be required. Second, high-skill individuals have higher incomes and therefore could have more to gain from bunching, which could account for their higher propensity to bunch. This aspect could be relevant if there is a fixed cost associated with locating at the kink (such as the cost of setting up a firm that enables income shifting). These are interesting research issues that we hope will attract further attention in the future.

## References

- Abeler, J. and Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, 7(3):1–28.
- Adam, S., Browne, J., Phillips, D., and Roantree, B. (2019). Frictions and taxpayer responses: evidence from bunching at personal tax thresholds. *IFS Working Paper W17/14*.
- Aghion, P., Akcigit, U., Lequien, M., and Stantcheva, S. (2018). Tax simplicity and heterogeneous learning. *Banque de France Working Paper #665*.
- Atkinson, A. and Stiglitz, J. (1976). The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, 6(1):55 – 75.
- Bastani, S., Giebe, T., and Miao, C. (2020a). Ethnicity and tax filing behavior. *Journal of Urban Economics*, 116:103215.
- Bastani, S., Moberg, Y., and Selin, H. (2020b). The anatomy of the extensive margin labor supply response. *Scandinavian Journal of Economics (forthcoming)*.
- Bastani, S. and Selin, H. (2014). Bunching and non-bunching at kink points of the Swedish tax schedule. *Journal of Public Economics*, 109:36–49.
- Bernheim, B. D. and Taubinsky, D. (2018). Behavioral public economics. *NBER Working Paper No. 24828*.
- Bhargava, S. and Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review*, 105(11):3489–3529.
- Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, 80(3):969–1018.
- Chetty, R., Friedman, J. N., Olsen, T., and 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.
- Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.
- Diamond, R. and Persson, P. (2017). The long-term consequences of teacher discretion in grading of high-stakes tests. *Working Paper, Stanford University*.
- Edin, P.-A., Fredriksson, P., Nybom, M., and Öckert, B. (2018). The rising return to non-cognitive skill. *IFAU Working Paper 2018:18*.
- Feldman, N. E., Katuščák, P., and Kawano, L. (2016). Taxpayer confusion: Evidence from the child tax credit. *American Economic Review*, 106(3):807–35.
- Gahvari, F. and Micheletto, L. (2016). Capital income taxation and the Atkinson–Stiglitz theorem. *Economics Letters*, 147:86 – 89.

- Gelber, A. M., Jones, D., and Sacks, D. W. (2020). Estimating adjustment frictions using non-linear budget sets: Method and evidence from the earnings test. *American Economic Journal: Applied Economics* (Forthcoming).
- Gerritsen, A., Jacobs, B., Rusu, A., and Spiritus, K. (2019). Optimal taxation of capital income when capital returns are heterogeneous. *Working Paper, Erasmus University Rotterdam*.
- Hargaden, E. P. (2018). Cycles and frictions in taxpayer behavior. *Working Paper, University of Tennessee, Knoxville*.
- Harju, J. and Matikka, T. (2016). The elasticity of taxable income and income-shifting: what is “real” and what is not? *International Tax and Public Finance*, 23(4):640–669.
- Hoopes, J. L., Reck, D. H., and Slemrod, J. (2015). Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy*, 7(3):177–208.
- Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8:435–464.
- Kleven, H. J. and Kopczuk, W. (2011). Transfer program complexity and the take-up of social benefits. *American Economic Journal: Economic Policy*, 3(1):54–90.
- Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *Quarterly Journal of Economics*, 128(2):669–723.
- Kosonen, T. and Matikka, T. (2019). Discrete earnings responses: empirical evidence and implications. *Working Paper, VATT Institute for Economic Research*.
- Kreiner, C. T., Leth-Petersen, S., and Skov, P. E. (2014). Year-end tax planning of top management: Evidence from high-frequency payroll data. *American Economic Review*, 104(5):154–58.
- Kreiner, C. T., Leth-Petersen, S., and Skov, P. E. (2016). Tax reforms and intertemporal shifting of wage income: Evidence from danish monthly payroll records. *American Economic Journal: Economic Policy*, 8(3):233–57.
- le Maire, D. and Schjerning, B. (2013). Tax bunching, income shifting and self-employment. *Journal of Public Economics*, 107:1 – 18.
- Lindqvist, E. and Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *American Economic Journal: Applied Economics*, 3(1):101–28.
- Lockwood, B. B., Nathanson, C. G., and Weyl, E. G. (2017). Taxation and the allocation of talent. *Journal of Political Economy*, 125(5):1635–1682.
- Mirrlees, J. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38:175–208.
- Mortenson, J. A. and Whitten, A. (2020). Bunching to Maximize Tax Credits: Evidence from Kinks in the U.S. Tax Schedule. *American Economic Journal: Economic Policy* (Forthcoming).

- Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3):180–212.
- Saez, E., Slemrod, J., and Giertz, S. H. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, 50(1):3–50.
- Seim, D. (2017). Behavioral responses to wealth taxes: Evidence from Sweden. *American Economic Journal: Economic Policy*, 9(4):395–421.
- Søgaard, J. E. (2019). Labor supply and optimization frictions: Evidence from the danish student labor market. *Journal of Public Economics*, 173:125 – 138.
- Stiglitz, J. E. (1982). Self-selection and Pareto efficient taxation. *Journal of Public Economics*, 17(2):213 – 240.
- Taubinsky, D. and Rees-Jones, A. (2018). Attention variation and welfare: Theory and evidence from a tax salience experiment. *Review of Economic Studies*, 85(4):2462–2496.

# A Appendix

## A.1 Descriptive statistics

Table A1: Descriptive and distributional statistics

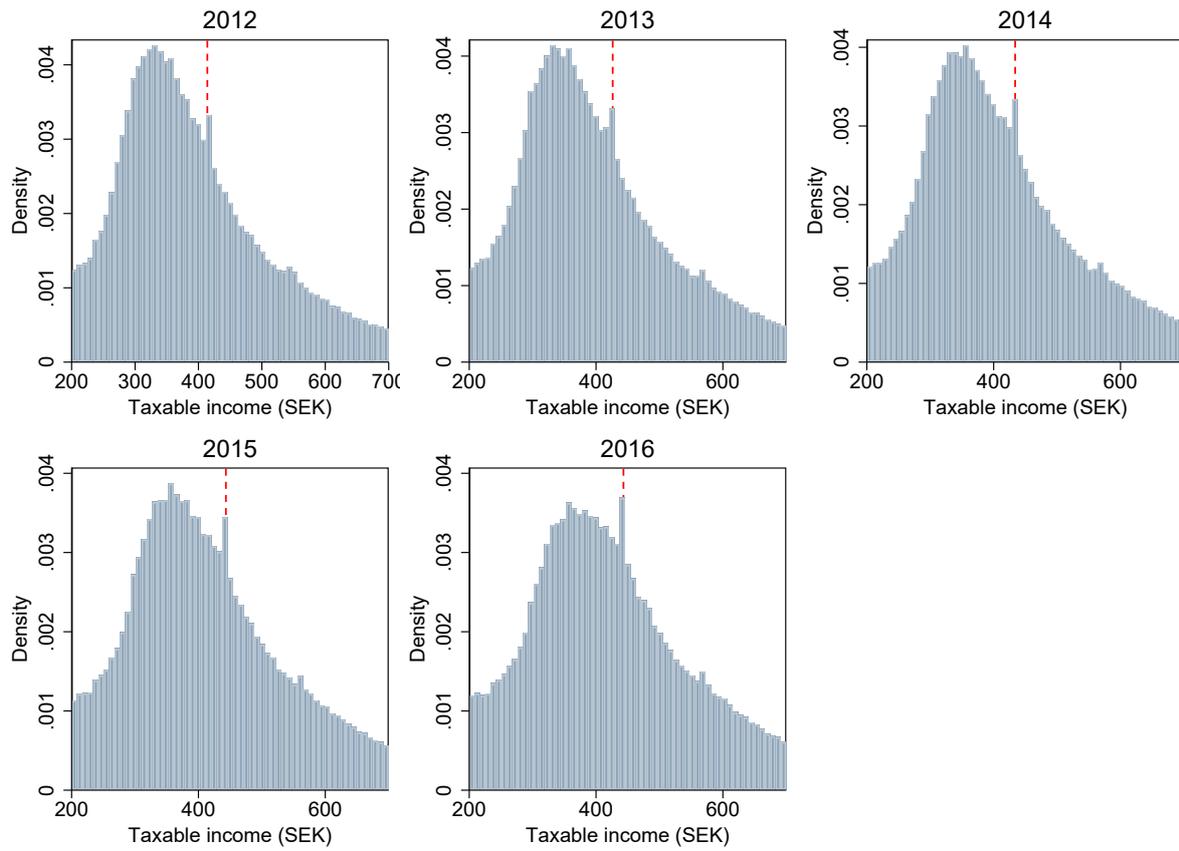
	Mean	S.D.	Min	P25	P50	P90	P99	Max
Cog. ability (z-score)	0.0226	1	-2.64	-0.699	0.11	1.4	2.05	2.54
Birth year	1963	7.25	1951	1957	1964	1973	1975	1975
Labor income (1000 EUR)	200	170	0	134	185	330	686	55,006
Dividend income (1000 EUR)	13.4	272	0	0	0.0052	6.5	240	167,533
GPA (z-score)	0	0.993	-4.85	-0.661	-0.0129	1.32	2.36	2.87
Math grade (z-score)	0	1	-3.12	-0.214	-0.214	1.72	1.72	1.72
Wage earner (%)	87.3							
Self-employed (%)	12.7							
Sole proprietor (%)	7.3							
Corporate owner (%)	7.2							

*Note:* Cognitive ability, high-school GPA and math grade are in z-scores (standard normal, mean = 0 and standard deviation = 1). Incomes are in thousands of euros and wage-earner vs self-employment status are averaged for the 2012-2016 period. Sole proprietors and corporate owners together make up all of the self-employed.

Table A2: Attrition in the sample population (number of individuals)

Category	Observations
<i>Main sample: Men born 1951-1975</i>	
(1) Men born 1951-1975 with cognitive ability score in military enlistment	1,283,254
(2) In (1) born 1955- (cohorts with GPA data coverage)	1,075,500
(3) In (2) and observed high-school GPA	786,032
(4) In (3) born 1966- (cohorts with math data coverage)	534,770
(5) In (4) and observed high-school math grade	292,262
<i>Supplementary sample: Men and women born 1955-1975</i>	
(6) Men and women born 1955-1975 with high-school GPA	1,677,654
(7) Men and women born 1966-1975 with high-school math grade	632,954

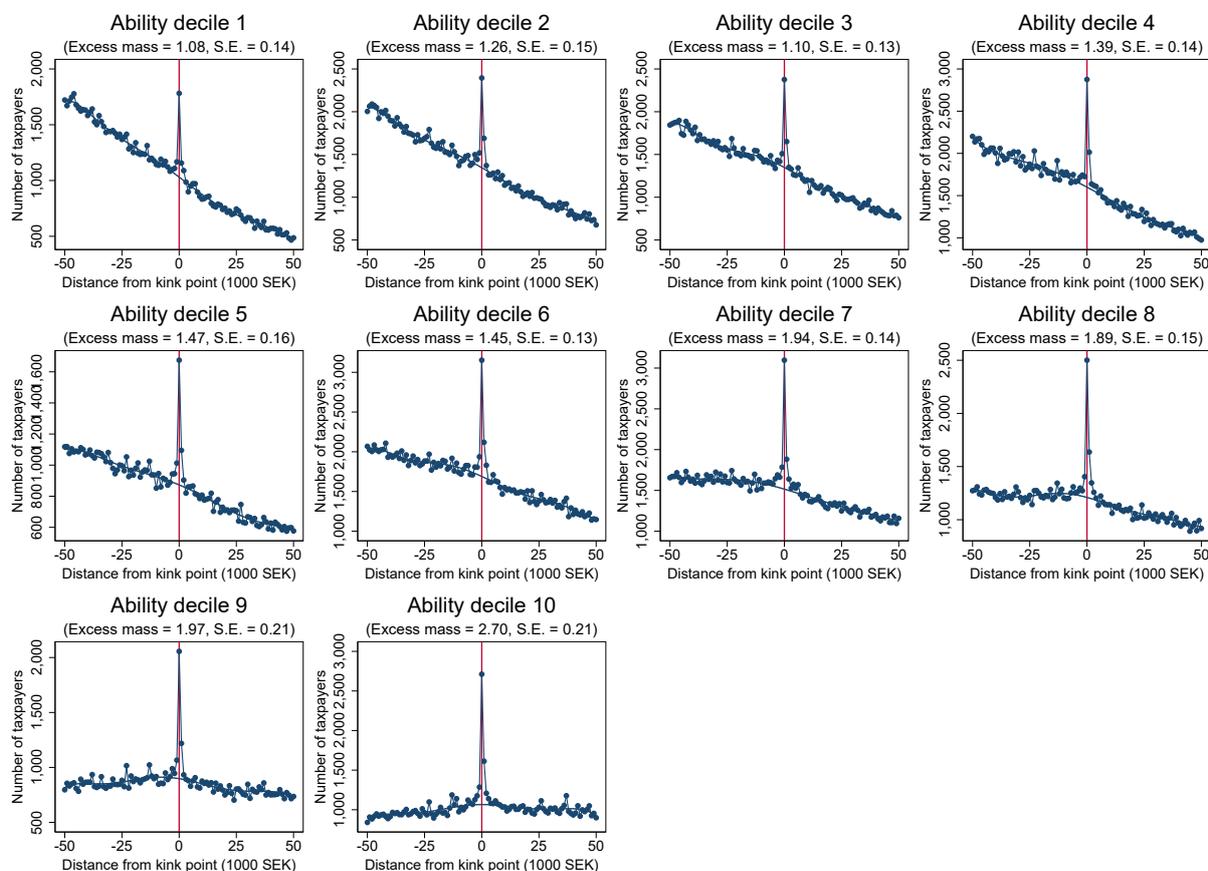
Figure A1: Distribution of taxable income around the kink point, Swedish men.



*Note:* Annual taxable labor income distributions around the kink point. Male population, born 1951-1975.

## A.2 Additional bunching results

Figure A2: Bunching by cognitive ability deciles



*Note:* Graph shows bunching during years 2012-2016 at the marginal tax kink point (when government income started being paid) for all adult men in different deciles of cognitive ability.

Figure A3: Sensitivity analysis with respect to the wide and small bunching windows (cognitive ability)

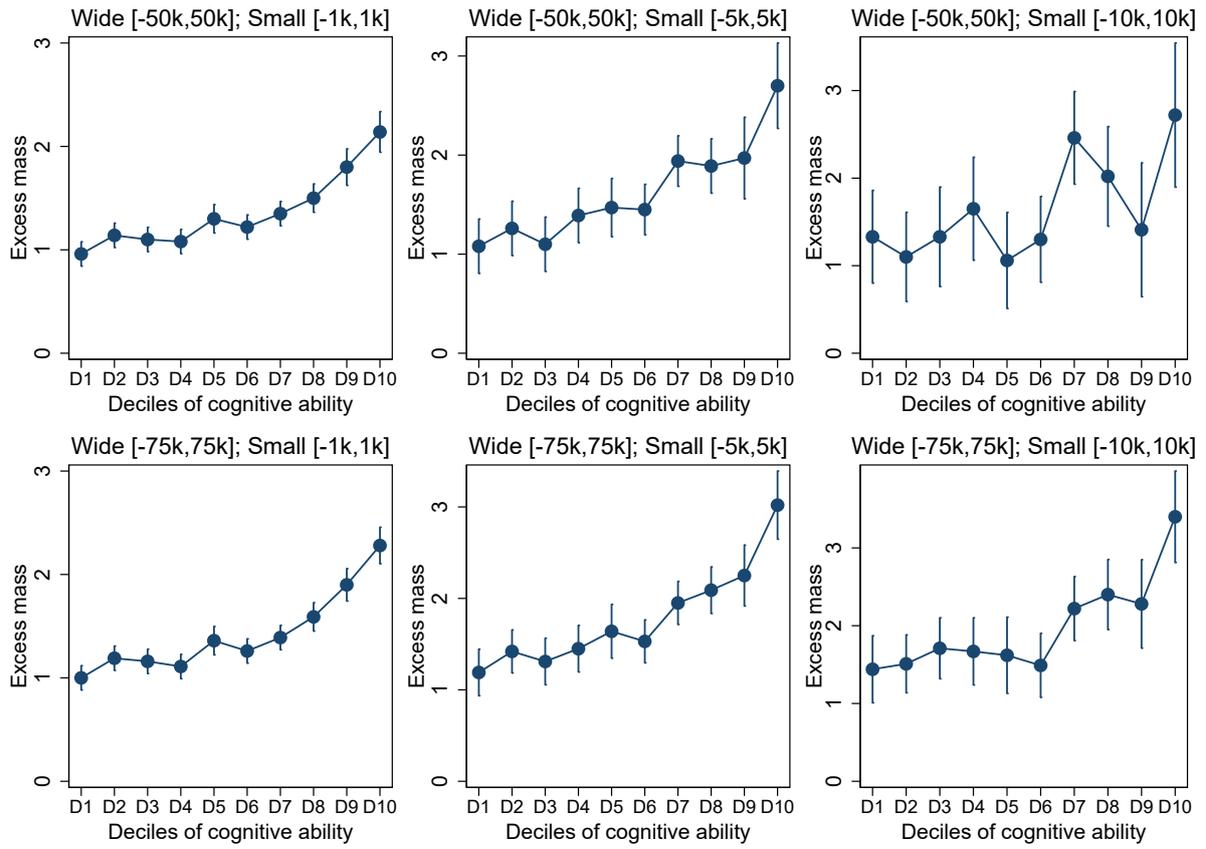
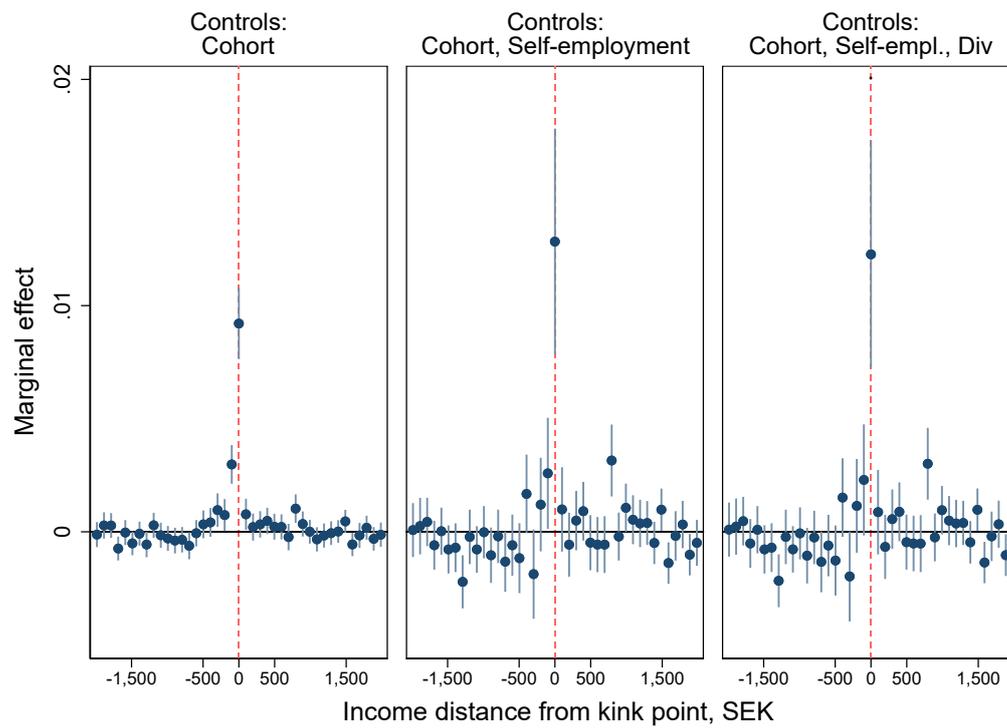


Table A3: Attrition in the sample population (number of individuals)

	(1) All	(2) Wage earners	(3) Self- employed	(4) No dividends	(5) Positive dividends
Ability decile 2	0.000294*** (0.000062)	0.000162*** (0.000043)	0.000611 (0.000503)	0.000164** (0.000054)	0.000406* (0.000185)
Ability decile 3	0.000373*** (0.000064)	0.000200*** (0.000043)	0.000690 (0.000504)	0.000246*** (0.000056)	0.000336 (0.000180)
Ability decile 4	0.000442*** (0.000061)	0.000222*** (0.000042)	0.000869 (0.000466)	0.000253*** (0.000053)	0.000428* (0.000167)
Ability decile 5	0.000525*** (0.000082)	0.000232*** (0.000053)	0.001355* (0.000585)	0.000295*** (0.000070)	0.000516* (0.000207)
Ability decile 6	0.000489*** (0.000061)	0.000234*** (0.000043)	0.000882 (0.000455)	0.000255*** (0.000054)	0.000416* (0.000162)
Ability decile 7	0.000671*** (0.000068)	0.000327*** (0.000046)	0.001610** (0.000490)	0.000327*** (0.000058)	0.000680*** (0.000172)
Ability decile 8	0.000681*** (0.000074)	0.000286*** (0.000049)	0.001880*** (0.000523)	0.000353*** (0.000068)	0.000581*** (0.000176)
Ability decile 9	0.000927*** (0.000092)	0.000481*** (0.000060)	0.002477*** (0.000612)	0.000436*** (0.000076)	0.000958*** (0.000205)
Ability decile 10	0.001013*** (0.000081)	0.000516*** (0.000057)	0.002816*** (0.000541)	0.000643*** (0.000082)	0.000806*** (0.000173)
Observations	6,004,200	5,299,431	704,769	3,871,042	2,133,158
R-squared	0.000	0.000	0.000	0.000	0.000
Cohort FE	Yes	Yes	Yes	Yes	Yes
Income year FE	Yes	Yes	Yes	Yes	Yes

Note: Robust standard errors in parentheses. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001.

Figure A4: Placebo regressions with controls: Ability effects on locating at incomes around the kink



*Note:* Estimated marginal effect of  $Cog$  on the probability to earn an income of  $y \in \{\hat{y}-2000, \hat{y}-1900, \dots, \hat{y}+2000\}$  where  $\hat{y}$  is the kink point income level. The first panel is identical to figure 9.