CESifo

AREA CONFERENCES 2020

Public Economics

Online, 8–9 December 2020

The Ability Gradient in Tax Responsiveness

Spencer Bastani and Daniel Waldenström

CESifo

The Ability Gradient in Tax Responsiveness*

Spencer Bastani† Daniel Waldenström‡

October 16, 2020

Abstract

We analyze the relationship between ability and tax responsiveness using administrative tax and military enlistment registers from Sweden. Studying income levels locally around a large and salient kink in the tax code, we find that high-ability individuals react more strongly to tax incentives than low-ability individuals, pointing to the importance of considering taxpayer skill heterogeneity conditional on income. These results are found both among wage earners and the self-employed. Exploring possible channels, we document larger hours adjustments among high-ability individuals, and higher capital income among high-ability self-employed. Using school grades, we also analyze gender differences and spousal effects.

Keywords: Skills, optimal taxation, labor supply, income shifting, bunching
JEL codes: H21, H24, J22, J24

*We are grateful to Felix Bierbrauer, Emanuel Hansen, Håkan Selin and seminar participants at the Free University Berlin, Research Institute for Industrial Economics (IFN) in Stockholm, The Swedish Institute for Social Research (SOFI), Stockholm University, for helpful comments. Financial support from the Swedish Research Council is gratefully acknowledged.

†Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala and Linnaeus University; Uppsala Center for Fiscal Studies (UCFS), Uppsala Center for Labor Studies (UCLS), CESIfo, Germany. E-mail: spencer.bastani@ifau.uu.se.

‡Research Institute for Industrial Economics (IFN), Stockholm; CEPR, CESifo, IZA, WIL. E-mail: daniel.waldenstrom@ifn.se

1 Introduction

One of the most fundamental areas of economic life are the interactions people have with the tax system. A vast empirical literature studies how people react to tax policy, with a heightened recent interest in how people understand, reason and learn about the tax code.[1] The existing literature mainly highlights how acquired skills, such as education and information transmission among peers, matter for taxpayer behavior. However, less is known about the role of innate traits and non-acquired skills. The purpose of this paper is to fill this gap by analyzing how different measures of ability constructed in young adulthood are related to how strongly people react to tax incentives. We do this by studying the way in which individuals with different abilities react differently to a large and salient discontinuity in the marginal tax rate inherent in the Swedish tax schedule.

Our analysis relies on a combination of bunching estimation and regression-based methods, using population-wide administrative tax record that are linked with a unique data set on individual ability from the Swedish military enlistment. 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 react to the large marginal tax change in the tax schedule. This means that there is an ability gradient in tax responsiveness. The previous literature has shown that tax responsiveness tends to be larger for individuals with higher income. However, our results indicate that responses are higher for individuals with higher skill among people with the *same* taxable income. This challenges the common assumption in the optimal tax literature that individuals with different earnings capacities react in the same way to tax changes, and also points to the importance of considering taxpayer heterogeneity conditional on labor income.[2] Furthermore, our findings highlight a central conflict in optimal tax design. From the optimal tax literature, we know that the government would like to set high taxes on high-ability individuals and on those with low elasticities. Our results show that high-ability individuals may be the ones who respond the most to progressive income taxation.

While the main purpose of our paper is to analyze the relationship between abil-

[1] See, for example, Saez et al. (2012), Bernheim and Taubinsky (2018), and Stantcheva (2020).

[2] 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 relies on the assumption that preferences (and thereby tax elasticities) are homogeneous and unrelated to ability. In the presence of multidimensional heterogeneity, optimal marginal tax rates depend on the average behavioral response among taxpayers who earn the same income, as conjectured by Saez (2001), and formally shown by Jacquet and Lehmann (2020).

ity and tax responsiveness, the taxpayer skill heterogeneity conditional on income that we document, can be interpreted as providing direct empirical evidence on the so-called mimicking issues emphasized in the optimal tax literature, following Mirrlees (1971).[3] The goal of the tax planner is to redistribute between individuals with different abilities to generate income, but since since earnings abilities are not observable for tax purposes, the government must achieve redistribution indirectly through distortionary income taxation. The design of the income tax is formulated as a second-best problem, constrained by the incentive constraints that ensure that high-skill individuals are not tempted to replicate the earned income of low-skill individuals. While these constraints have been the subject of hundreds, if not thousands, of papers in the literature, their empirical relevance has not been assessed.[4]

Our main estimation method builds on the 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, referred to as bunching estimation.[5] This estimation method cannot typically uncover the full responses of taxpayers to income taxation, since bunching estimates tend to be downwards biased due to optimization frictions (such as restricted labor supply choices). However, it is ideally suited for providing transparent and compelling visual evidence on the existence of behavioral responses and for estimating differences in tax responsiveness between subgroups of the population.

We complement the 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 in the marginal tax rate schedule. We then compare this estimate with the marginal effects of ability on the probability of locating at adjacent income levels, basically running a large set of rolling regressions. This analysis confirms the bunching results. We estimate a positive and statistically significant marginal effect of ability on the likelihood of sharp bunching at the kink point, with no such effect on adjacent income levels, pointing to the special importance of taxpayer responses that place individuals exactly at the kink point.

In a series of extensions, we explore possible channels through which the effect that we document might operate. We find that the ability gradient in tax responsiveness is particularly strong for the self-employed. We also find a spike in the distribution of

[3]These mimicking issues are most transparent in the discrete adaptations of the Mirrlees model, introduced by Stiglitz (1982) and Stern (1982) as well as Guesnerie and Seade (1982) who considered an arbitrary finite number of types.

[4]Our analysis is facilitated by an empirical setting with a kinked tax schedule where mimicking occurs in equilibrium as well as data on ability closely related to the skill measure originally envisioned by Mirrlees (1971). Ebert (1992) provides a discussion of bunching in the context of optimal income taxation.

[5]The basic idea of 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.

capital income exactly at the kink point for the self-employed, but no such spike for wage earners. This hints to the importance of income shifting, which is relevant in the context of Sweden's dual income tax system where the marginal tax rate on capital income is lower than the marginal tax rate on labor income for most mid-to-high income earners.[6] Using an auxiliary data set on hours of work, we also find a downwards spike in hours of work exactly at the kink point, suggesting labor supply responses. In other words, the tax responses we observe are most likely driven by a combination of real responses and income shifting.

The military enlistment data also provide us with information on non-cognitive traits and measures of physical ability. We find that the ability gradient in tax responsiveness using non-cognitive traits is, in comparison to the gradient obtained using cognitive ability, much weaker in the bottom of the ability distribution, but follows a similar pattern for higher levels of ability. In terms of our physical ability measures, we find no gradient, or possibly even a negative gradient for work capacity (aerobic fitness), but moderately positive gradients for grip strength and taxpayer height.

A limitation of the military enlistment data is that they are only available for men. We therefore provide a supplementary analysis of gender differences in the relationship between ability and tax responsiveness using high-school GPA and math grades as proxies for cognitive ability. Our results show a clear ability gradient in tax responsiveness for men but almost no gradient for women. These results hold independently of whether we consider wage earners or self-employed, or whether we use GPA or math grades. We also analyze married couples and use the data on school grades to investigate how responses depend on household ability. These results indicate that households where both spouses have a high ability tend to respond more.

Our study relates to the literature on behavioral responses to income taxation and the literature on the optimal design of tax systems. Methodologically, it builds on the bunching literature that began with Saez (2010), and was recently surveyed by Kleven (2016). In terms of the empirical setting, a related paper is Bastani and Selin (2014) who studied bunching in Sweden during an earlier period 1999–2005, but who did not study the relationship between ability and tax responsiveness.[7]

We also contribute to the literature on how individuals understand and respond to tax rules (see Chetty et al. 2009, Abeler and Jäger 2015, Taubinsky and Rees-Jones

[6] Income shifting can be either in the form of tax avoidance, or in terms of a real re-allocation of labor effort towards activities that earn a higher rate of return on savings. The latter channel is examined by Christiansen and Tuomala (2007) who argues that it is an argument to tax capital income.

[7] 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.

2018) and papers highlighting the "regressive" nature of tax 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.[8] 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 frictions for observed bunching behavior (see Chetty 2012, Kleven and Waseem 2013, Søgaard 2019, Kosonen and Matikka 2019, Mortenson and Whitten 2020, Gelber et al. 2020). There is also a related literature on the role of income shifting for taxpayer behavior in the context of dual income tax systems (see, for example, Pirttilä and Selin 2011, le Maire and Schjerning 2013, and Harju and Matikka 2016).[9]

The remainder of the paper is organized as follows. In section 2, we outline a theoretical and empirical framework that can be used to interpret our results. It also briefly describes our estimation strategy. Section 3 presents our data sources and the institutional setting. Section 4 describes our baseline bunching results and in section 5, we present analyses of potentially important mechanisms and some extensions as well as our regression analysis. Section 6 concludes.

2 Analytical framework

2.1 A bunching model with two dimensions of heterogeneity

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 θ, but also in terms of some other characteristic ξ that affects an individual's earnings capacity.[10] The purpose of this extension is to have an empirically relevant model where ability is not the only determinant of an individual's income, since we observe heterogeneity in ability conditional on income not only at kink points, but also along interior segments of the piece-wise linear tax schedule. An individual's choice of taxable income z is guided by the following optimization problem:

$$\max_{z} \Big\{ z - T(z) - v(z, \theta, \xi) \Big\}, \tag{1}$$

[8] See also Bhargava and Manoli (2015), Hoopes et al. (2015), Feldman et al. (2016), and Bastani et al. (2020a).

[9] See also Tazhitdinova (2020) for evidence of income shifting and business entry responses to taxes in the UK.

[10] For example, it could represent an individual's preference for leisure. It could also represent measurement error in the skill θ. The parameter ξ is assumed to be a scalar for tractability. The framework can be extended to allow ξ to be a vector without affecting the qualitative results.

where $z - T(z)$ is the consumption level and v represents the disutility of earning z for a (θ,ξ)-type agent, which we assume takes the specific form:

$$v(z,\theta,\xi) = \frac{z_p(\theta,\xi)}{1+\frac{1}{e(\theta)}}\left(\frac{z}{z_p(\theta,\xi)}\right)^{1+\frac{1}{e(\theta)}}. \tag{2}$$

The function z_p in this expression depends on both θ and ξ, which are continuously distributed according to a smooth joint density function $f(\theta,\xi)$. The parameter e is a preference parameter that depends on skill θ, but not on ξ. Along a linear segment of the tax schedule T, with marginal tax rate τ, the solution to (1) takes the familiar form:

$$z(\theta,\xi) = z_p(\theta,\xi)(1-\tau)^{e(\theta)}, \tag{3}$$

where it can be seen that z^p has the convenient interpretation as the taxable income of a (θ,ξ)-individual in the absence of taxation ($\tau = 0$) and $e(\theta)$ is the elasticity of z w.r.t. $(1-\tau)$ for an individual with skill θ. In this more general setting, there are differences in skill at each income level because income is not only determined by skill, but also by other factors. This implies that there will be a distribution of elasticities at each income level, which is an important property in our context.

Consider a baseline tax system that exposes individuals to the constant marginal tax rate τ. Based on equation (3), the joint distribution of θ and ξ, $f(\theta,\xi)$, determines a baseline joint distribution of earnings z and skills θ that we denote $\tilde{h}_0(z,\theta)$. The baseline marginal earnings distribution is obtained through integration as $h_0(z) = \int_\theta \tilde{h}_0(z,\theta)$.

With multiple 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. Consider the introduction of a convex kink at the income level $\hat{z}$ that lowers the net-of-tax rate at all income levels $z \geq \hat{z}$ from $1-\tau$ to $1-\tau-\Delta\tau$, where $\Delta\tau > 0$. The set of bunchers at $\hat{z}$ is characterized by the region in (θ,ξ)-space such that

$$1-\tau-\Delta\tau \leq \left(\frac{\hat{z}}{z_p(\theta,\xi)}\right)^{\frac{1}{e(\theta)}} \leq 1-\tau. \tag{4}$$

We refer to the set of (θ,ξ)-values that imply that the left inequality of (4) is satisfied as an equality as the *marginal bunchers*. They are marginal in the sense that if they would have marginally flatter indifference curves, they would not bunch, and instead locate in the interior of the second segment of the kinked tax schedule. For each skill θ, there is a unique marginal buncher that can be identified by finding the ξ that solves $1-\tau-\Delta\tau = \left(\frac{\hat{z}}{z_p(\theta,\xi)}\right)^{\frac{1}{e(\theta)}}$. We denote this value by $\hat{\xi}_\theta + \Delta\xi_\theta$ and the associated potential income by $z_p(\theta, \hat{\xi}_\theta + \Delta\hat{\xi}_\theta)$. Here, $\hat{\xi}_\theta$ is the value of ξ that places an individual with skill θ exactly at $\hat{z}$ in the absence of a kink, and $[\hat{\xi}_\theta, \hat{\xi}_\theta + \Delta\hat{\xi}_\theta]$ is the range of ξ-values that bunch. The taxable income of the marginal buncher, before and after the reform, is, respectively:

$$\hat{z} + \Delta\hat{z}_\theta = z_p(\theta, \hat{\xi}_\theta + \Delta\hat{\xi}_\theta) \times (1-\tau)^{e(\theta)} \tag{5}$$

$$\hat{z} = z_p(\theta, \hat{\xi}_\theta + \Delta\hat{\xi}_\theta) \times (1-\tau-\Delta\tau)^{e(\theta)}. \tag{6}$$

Hence, the income response of of the marginal buncher with skill θ, denoted by $\Delta\hat{z}_\theta$, is given by:

$$\Delta\hat{z}_\theta = z_p(\theta, \hat{\xi}_\theta + \Delta\hat{\xi}_\theta) \times \left[(1-\tau)^{e(\theta)} - (1-\tau-\Delta\tau)^{e(\theta)}\right]. \tag{7}$$

Notice that $\Delta\hat{z}_\theta$ is just the standard interior income response to a change in the net-of-tax rate from $1-\tau$ to $1-\tau-\Delta\tau$ for an individual with characteristics θ and $\hat{\xi}_\theta + \Delta\hat{\xi}_\theta$, following formula (3).

An expression for the elasticity can be derived by recognizing that:

$$\frac{\hat{z} + \Delta\hat{z}_\theta}{\hat{z}} = \left(\frac{1-\tau}{1-\tau-\Delta\tau}\right)^{e(\theta)}, \tag{8}$$

which implies

$$e(\theta) = -\frac{\log(1+\Delta\hat{z}_\theta/\hat{z})}{\log(1-\Delta\tau/(1-\tau))}. \tag{9}$$

We can relate the quantity $\Delta\hat{z}_\theta$ to the amount of bunching at the kink among individuals with skill θ, by integrating over the pre-reform income distribution:

$$B_\theta = \int_{\hat{z}}^{\hat{z}+\Delta\hat{z}_\theta} \tilde{h}_0(z,\theta)dz \approx \tilde{h}_0(\hat{z},\theta)\Delta\hat{z}_\theta, \tag{10}$$

where we have assumed that $\tilde{h}(z,\theta)$ is constant in z in the bunching segment $[\hat{z}, \hat{z}+\Delta\hat{z}_\theta]$. We define the *excess mass* at $\hat{z}$ for individuals with skill θ as

$$b_\theta = \frac{B_\theta}{h_0(\hat{z},\theta)} \approx \Delta\hat{z}_\theta. \tag{11}$$

Thus, an empirical estimate of $e(\theta)$ can be obtained by replacing $\Delta\hat{z}_\theta$ with the empirically observable quantity b_θ into formula (9).

We can also compute average elasticities that do not condition the population on a particular value of θ. Let Θ denote the support of θ. Using the results in Kleven and Waseem (2013), or more recently, Gelber et al. (2020), (online appendix A.3), we can write observed bunching in terms of an integral over the pre-reform income distribution:

$$B = \int_{\theta\in\Theta}\int_{\hat{z}}^{\hat{z}+\Delta\hat{z}_\theta} \tilde{h}_0(z,\theta)dzd\theta \tag{12}$$

$$= h_0(\hat{z})\int_{\theta}\int_{\hat{z}}^{\hat{z}+\Delta\hat{z}_\theta} \frac{\tilde{h}_0(z,\theta)}{h_0(\hat{z})}dzd\theta \approx h_0(\hat{z})E[\Delta\hat{z}_\theta]. \tag{13}$$

Here we have adopted, the simplifying assumption that $\tilde{h}_0$ is independent of θ, namely $\tilde{h}_0(z,\theta) = h_0(\hat{z})$, for $z \in [\hat{z}, \hat{z}+\Delta\hat{z}_\theta]$.[11] This implies that the mass of individuals who move to the kink is just the average income response across all skill levels, $E[\Delta\hat{z}_\theta]$, times the height of the counter-factual distribution, assumed to be equal to $h_0(\hat{z})$.

In our empirical analysis, we estimate bunching for different subgroups (skill groups) S of Θ. Thus, it is relevant to derive an expression for the average elasticity within a subgroup. We define bunching in subgroup S as follows:

$$B_S = \int_{\theta\in S}\int_{\hat{z}}^{\hat{z}+\Delta\hat{z}_\theta} \tilde{h}_0(z,\theta)dzd\theta \approx h_{0,S}(\hat{z})E_{\theta\in S}[\Delta\hat{z}_\theta], \tag{14}$$

under the assumption that $\tilde{h}_0(z,\theta) = h_{0,S}(\hat{z})$ for $\theta \in S$ and $z \in [\hat{z}, \hat{z}+\Delta\hat{z}_\theta]$.[12] This allows us to define the excess mass at income level $\hat{z}$ in skill group S as the ratio between bunching and the height of the counterfactual distribution in skill group S:

$$b_S = \frac{B_S}{h_{0,S}(\hat{z})} \approx E_{\theta\in S}[\Delta\hat{z}_\theta]. \tag{15}$$

Thus, an empirical estimate of the average elasticity in skill group S, denoted e_S, can be obtained by replacing $\Delta\hat{z}_\theta$ with b_S in formula (9).[13]

Finally, we would like to make three remarks. First, the optimization problem (1) implies a relationship between the two dimensions of heterogeneity (θ,ξ), and the distribution of income z, for a given nonlinear tax system $T(z)$. Along a linear segment of the tax schedule $T(z)$, equation (3) provides us with a simple expression for this relationship. At the kink $\hat{z}$, there will be bunching of taxpayers reflecting both the set of (ξ,θ)-values that imply an optimal income level of exactly $\hat{z}$ when the tax schedule is linear, as well as the set of (ξ,θ)-values associated with those who bunch at $\hat{z}$ when the kink is introduced (who would have otherwise located to the right of the kink under a purely linear tax system).

Second, one suitable interpretation of ξ is that it captures preferences for leisure. For example, among taxpayers with the same income, there will likely be some high-ability

[11] This computation is similar to the one used by Kleven and Waseem (2013) to compute average bunching responses in the presence of heterogeneous elasticities. A weaker assumption will be exploited below.

[12] Notice that this assumption is strictly weaker than the one used to derive equation (13) since it only needs to hold at the sub-group level.

[13] Notice that formula (15) is applied at the sub-group level, so there is no requirement that the counterfactual distribution must be the same in different skill groups.

individuals with a high preference for leisure who work few hours, and some low ability workers with a low preference for leisure who work many hours.[14] The assumptions underlying the analysis imply that having a high preference for leisure lowers an individual's taxable income, independently of what the tax rate is, but does not directly affect the size of the behavioral response to a tax change.

Third, notice that $e(\theta)$ is not a deep structural parameter but instead a reduced-form parameter that we interpret as capturing the various mechanisms through which ability might influence behavioral responses to taxes. For example, it could capture that low-ability workers have very inflexible work schedules that make it difficult for them to adjust their incomes, whereas high-skill workers work in occupations that offer more flexibility. It could also capture that high-skill workers are more aware of tax incentives, and therefore respond more to tax changes.

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. Whether this estimation strategy is compelling depends on the data, and in our case, we find that the polynomial approximation fits the Swedish income distribution quite nicely.[15]

We express annual 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 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, which are then excluded when estimating the counter-factual distribution. Since we estimate bunching for various subgroups of the population and for different years, we do not choose δ based on visual inspection (which

[14]This will be the case in our model under the plausible assumption that z_p in (3) is increasing in θ but decreasing in ξ, and provided $e'(\theta)$ is sufficiently small, as it guarantees a positive correlation between θ and ξ at any given income level along a linear segment of the tax schedule.

[15]We have also studied data in earlier years, before the kink point was introduced, verifying that there were no anomalies in the income distribution in these income regions during that time. More sophisticated discussions of identification and inference in the bunching context are provided by Blomquist et al. (2017) and Bertanha et al. (2019).

is commonly done in the literature), but instead fix a baseline δ in our analysis and then report extensive robustness checks with respect to this parameter.[16] 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 A5) shows that varying these windows does not alter the main results.

The counterfactual distribution is estimated using the following regression model:

$$C_j = \sum_{i=0}^{q} \beta^i Z_j^i + \sum_{s=-\delta}^{\delta} \gamma_s \mathbf{I}[Z_j = s] + \eta_j, \tag{16}$$

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

Denote by $\widehat{C}_j$ the predicted values from regression (16). 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:

$$\hat{b} = \frac{\hat{B}}{\sum_{j=-\delta}^{\delta} \frac{\widehat{C}_j}{2\delta+1}} \quad \text{where} \quad \hat{B} = \sum_{j=-\delta}^{\delta} (C_j - \hat{C}_j). \tag{17}$$

The quantity $\hat{b}$ in (17) is the empirical *excess mass*, namely, the empirical counterpart of (15). 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.[17]

In section 5, we complement our bunching analysis with regressions using individual-level data. More specifically, we adopt a linear probability model and estimate the marginal effect of cognitive ability on the probability of locating exactly at the kink point, and compare this estimate with the effect on adjacent (placebo) income levels. In that analysis, we restrict the sample to individuals in a smaller neighborhood around the kink, allowing us to analyze the effect of cognitive ability on the extent of *sharp* bunching at the kink point.

3 Data and institutional setting

We use individual data from several population-wide administrative registers in Sweden. From the population register, we retrieve the full population living in Sweden born in the

[16] Diamond and Persson (2017), page 17 use a similar reasoning but provide an automated approach.

[17] We use the Stata program `bunchcount`, estimating the counterfactual distribution using a seven-degree polynomial and boostrap standard errors with 500 replications. We also use a correction to make sure that the counter-factual distribution integrates to one.

years 1951-1975. For these years, we retrieve measures of cognitive and physical ability, as well as non-cognitive traits, from the military enlistment register. 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 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 an incorporated or unincorporated firm.[18] 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). 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. From the education register, we add records on 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 cognitive 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.

The assessment of non-cognitive traits was made in personal interviews by psychologists who followed the same systematic evaluation procedure for all cohorts in the analysis (Lindqvist and Vestman 2011). This procedure resulted in scores of an individual's social maturity, psychological energy, intensity and emotional stability. We use the total score for all of these traits, measured along a stanine scale. Physical ability was evaluated in several tests during the enlistment, and we use three representative outcomes: height (in centimeter), work capacity in a bicycling test (measured in watts), and hand grip strength (maximum pressure, measured in Newton, exerted by the strongest hand

[18]Information on ownership of non-listed corporations comes from a specifically matched firm-individual ownership database, FRIDA (we use the register variables `bkufoab` or `bfoab`). Unincorporated firm ownership is based on whether an individual has any income from a sole proprietorship (variable `nakte`) or is connected to a limited liability partnership (variable `nakthb`).

squeezing a dynamometer). We generate decile dummies from the distributions of each of the three physical abilities.

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.[19] 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 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.[20] 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.[21]

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.

[19] See, for example, Lindqvist and Vestman (2011), Edin et al. (2018) who analyze the relationship between cognitive ability and labor market outcomes.

[20] 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).

[21] 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.

4 Main results

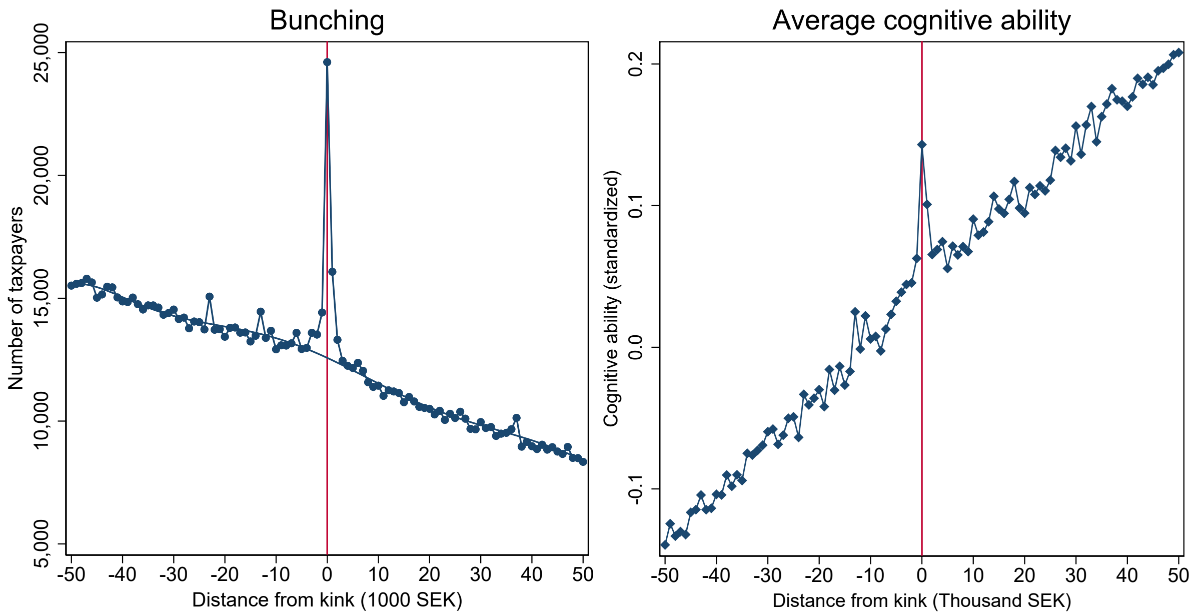
We begin by showing two key stylized facts in figure 1: the local shape of the income distribution and the average standardized cognitive ability around the tax kink point. The data underlying the figure is the total sample of men with taxable income during 2012-2016, expressed in bins of thousands of Swedish kronor (SEK) (equivalent to hundreds of euros or US dollars), and measured relative to the location of the kink point. The left panel of the figure depicts statistically significant bunching at the kink point with an estimated excess mass of 1.60 and a bootstrapped standard error of 0.10.[22] The purpose of our paper is to analyze the relationship between ability and bunching. The right panel of figure 1, provides a first indication of our results by showing a clear spike in average ability exactly at the kink point.

The fact that we find an over-representation of high-ability individuals at the kink can be interpreted as empirical evidence of high-skill agents mimicking low-skill agents in order to reduce their tax burden. This represents the fundamental constraint on the design of the income tax as emphasized in the optimal tax literature, but has not previously been empirically tested in this direct way.

Next, we turn to investigate 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 at the same statutory kink point by comparing the decile-specific observed mass of income earners around the kink point with the estimated decile-specific counterfactual density around the kink point.

[22]That bunching is found for the total population at the first central government kink point is in line with the findings of bunching in Sweden during the 2000s by Bastani and Selin (2014).

Figure 1: Bunching and average cognitive ability at the kink, 2012-2016

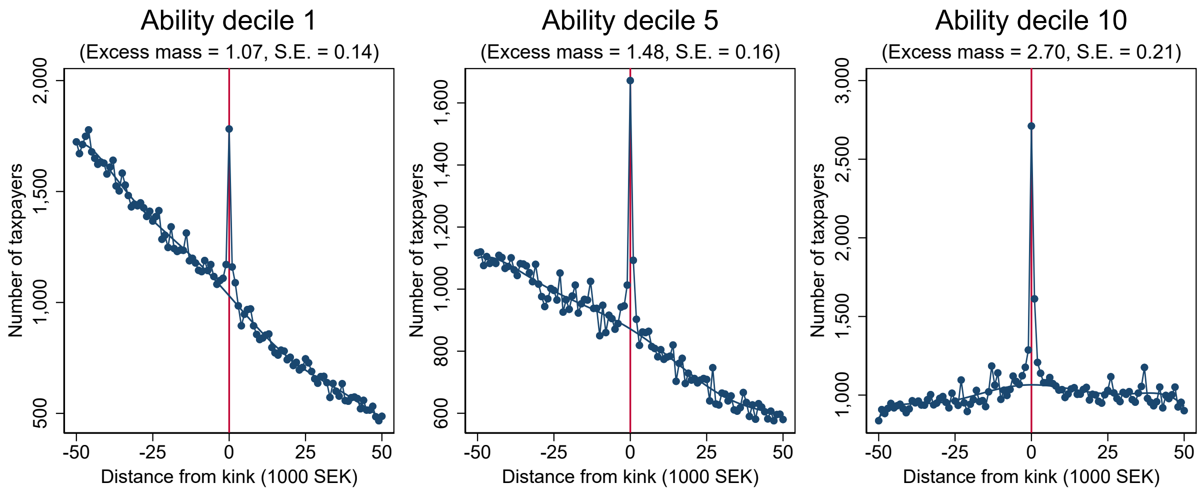


Note: The figure shows bunching among all men (born 1951-1975) at the largest kink of the Swedish income tax schedule (payment of central government income tax) for taxable labor income during 2012-2016 (pooled annual data). The smooth line in the left panel is an estimate of the counter-factual distribution, i.e., an estimate of how the taxable income distribution would look like in the absence of a kink.

Figure 2 shows bunching estimates for three of the ten ability deciles (appendix figure A4 shows all ten deciles).[23] 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 three 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.

[23]Appendix figure A2 compares the overall income distributions for the ten decile groups.

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.

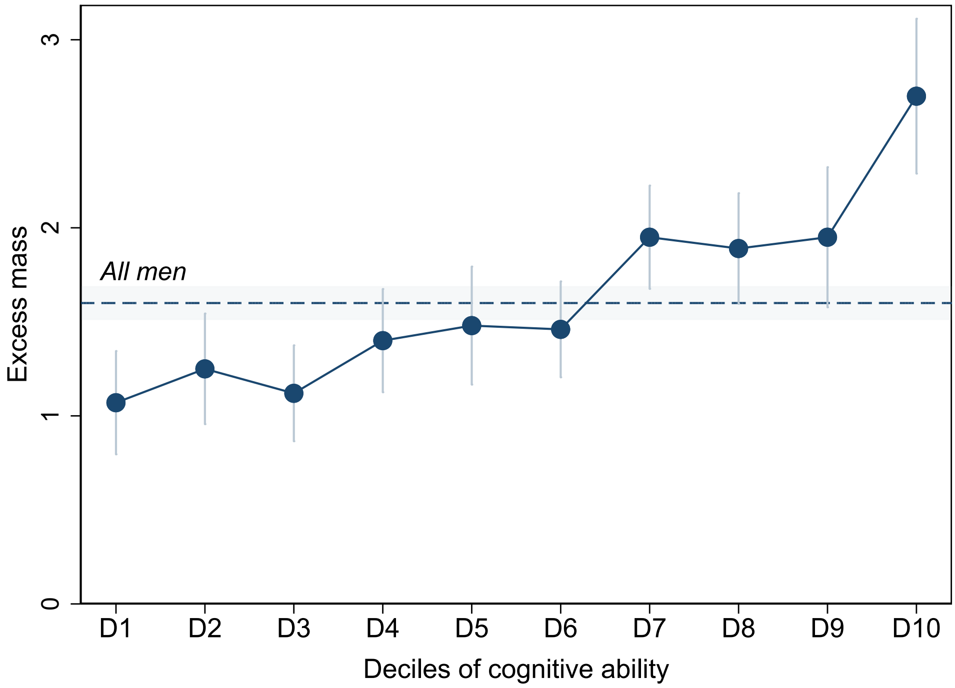
Figure 3 depicts bunching estimates for all the ten ability deciles. It also includes a dashed line showing the estimated bunching in the full male population. The figure shows that there is a clear, almost monotonic, increase in bunching in the level of individuals' cognitive ability. Ability decile 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 tax responsiveness at the kink point. To our knowledge, this relationship has not been shown before in the literature. Previous papers have shown that taxable income elasticities tend to be larger for individuals with higher income (see, for example, Saez et al. 2012).[24] Our results indicate that responses are higher for individuals with higher skill among people with the *same* taxable income.

In appendix figure A4, we present results that show that the ability gradient is robust to perturbations in the estimation framework, in particular using different sizes for the small and wide windows around the kink point.

[24]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, for example, Bastani et al. (2020b). 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.

Figure 3: Ability gradient in tax responsiveness



Note: Excess mass and 95% confidence intervals (+/- 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 dashed line is the estimated (average) bunching in the full male population. The underlying bunching estimation for the ten excess mass estimates is presented in appendix figure A4.

5 Extensions

The previous section established that bunching is larger among high-ability groups than among low-ability groups. In this section, we shed further light on this relationship through a series of extensions. First, we divide the population into wage earners and self-employed.[25] Second, we use auxiliary data on hours worked to analyze the relationship between bunching and labor supply. Third, we investigate the role of income shifting and capital income, by documenting the distribution of capital income locally around the kink point. Fourth, we analyze gender differences and the role of household ability using supplementary data on school grades. Fifth, we analyze the role of non-cognitive traits and physical ability for bunching outcomes. Finally, we run a robustness analysis of the baseline bunching results of section 4, where we analyze the effects of ability on bunching using individual (non-aggregated) data. In this analysis, the outcome variable is a dummy indicating whether an individual has an income exactly at the kink point

[25]Providing a separate bunching analysis for wage earners and self-employed is common in the literature due to the self-employed having greater flexibility to adjust their taxable income (Chetty et al. 2011). Notice that the use of professional tax preparers is relatively uncommon in Sweden. Moreover, for most people, there are limited possibilities to use deductions to locate at the kink (see Paetzold 2018 for the role of deductions in the bunching context).

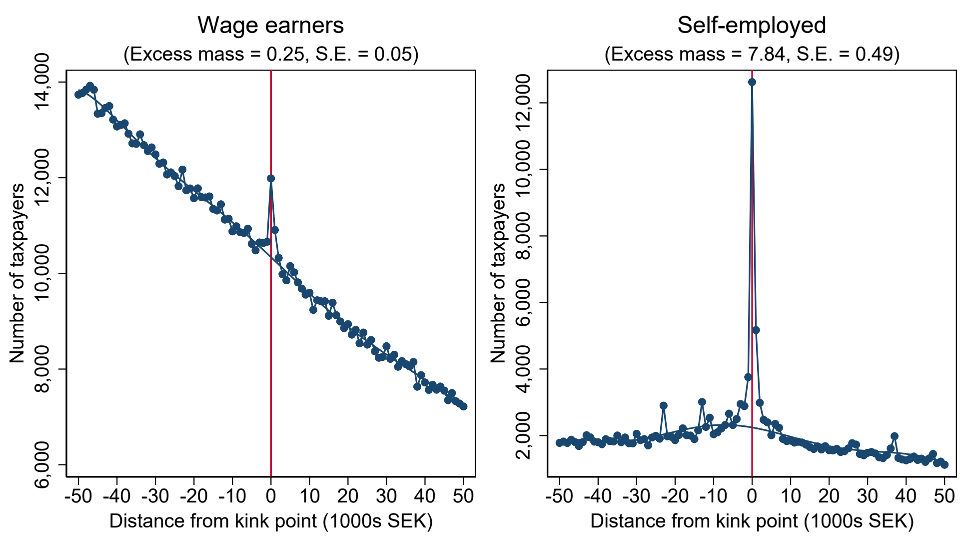
and we restrict the sample to individuals who locate in a close neighborhood around the kink. We also include cohort and income-year fixed effects.

5.1 Wage earners and self-employed

The division between wage earners and self-employed has been studied extensively in the bunching literature, motivated by the fact that the self-employed have greater control over their taxable income.

Figure 4 shows bunching around the kink point for wage earners and self-employed individuals. In line with previous studies, we find more bunching among the self-employed. However, it is noticeable that we encounter significant bunching also among wage earners, contrasting the previous findings of Bastani and Selin (2014) who found no visible bunching for wage earners.[26]

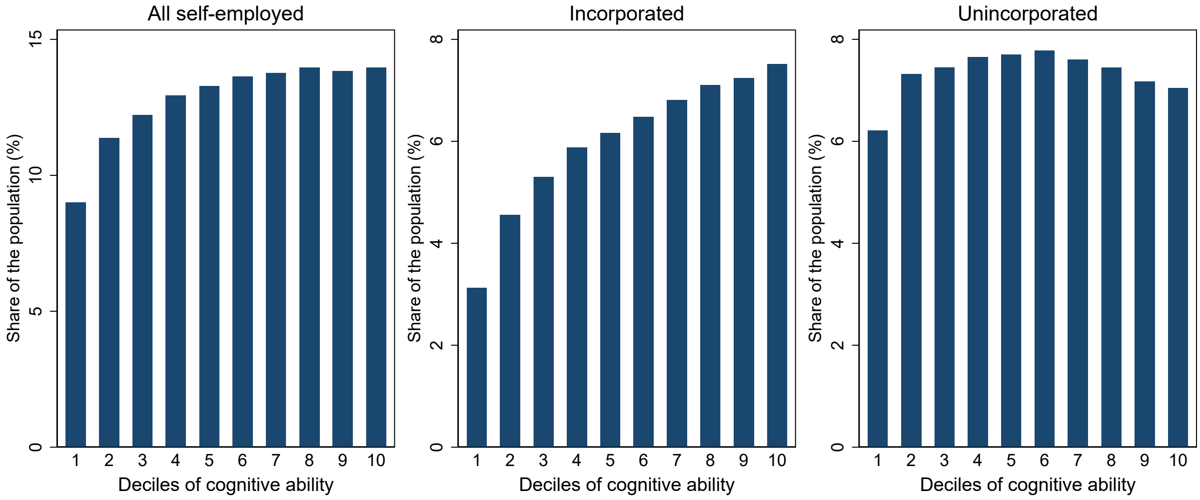
Figure 4: Bunching: Wage earners vs. self-employed.



Is there a selection of high ability-people into self-employment? Figure 5 shows that those with high ability are more likely to be self-employed, which is likely to be one explanation for the higher responsiveness of high ability individuals documented in figure 3, and the higher average responsiveness for the self-employed documented in the right panel of figure 4. We also see that over-representation of high-ability individuals among the self-employed is driven by incorporated business owners.

[26]Notice that we study the time period 2012-2016 whereas Bastani and Selin (2014) studied the time period 1999-2005. Our sample is also different. We focus on males with data from the military enlistment. A careful investigation of whether bunching estimates in Sweden have changed over time is left for future research.

Figure 5: Share of self-employed across the ability distribution



Note: Incorporated self-employed are owners of closely held corporations, whereas unincorporated self-employed run other types of businesses, such as sole proprietorships.

We proceed to estimate ability gradients separately for self-employed and wage earners. Figure 6 examines differences in bunching between different ability deciles of wage earners and self-employed. Like in our baseline analysis in figure 2, we begin by presenting results for the 1st, 5th and 10th ability deciles. The results show that among wage earners, there is no bunching in the bottom decile, whereas bunching appears to be larger in the middle decile, and is statistically significant, yet small, in the top ability decile. For the self-employed, overall 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

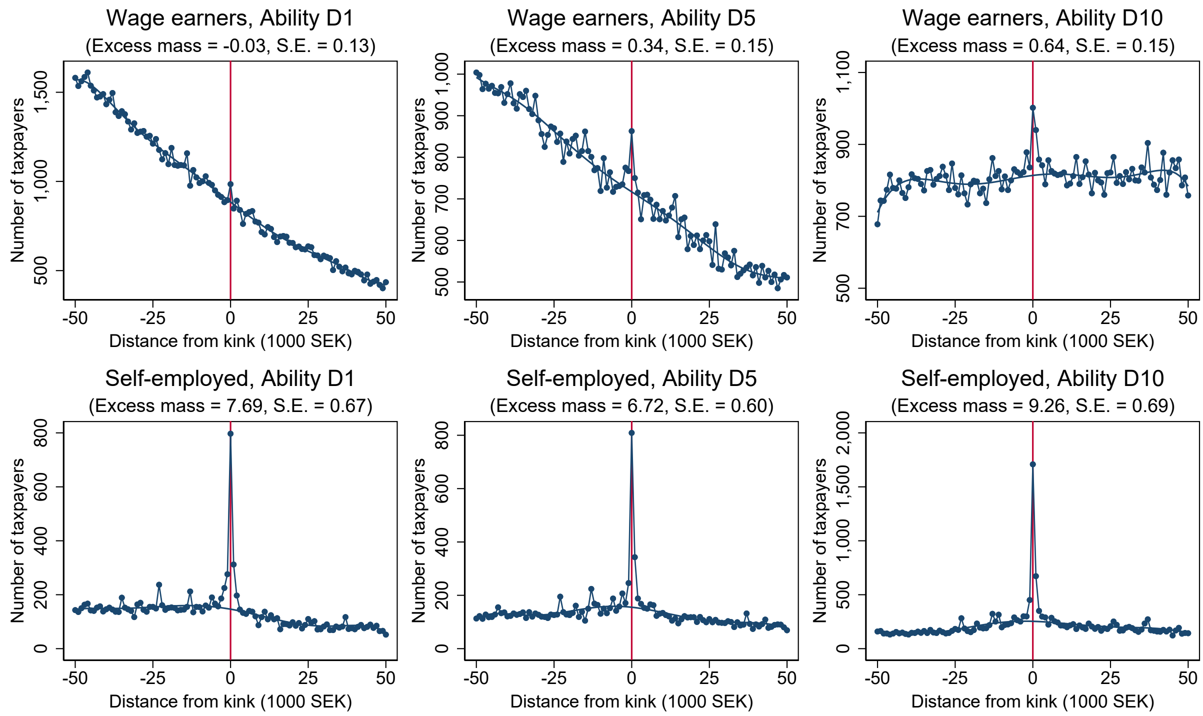


Figure 7 turns to examine the ability gradient across all ten deciles, separately for wage earners and the self-employed, in a figure similar to figure 3 (which is reproduced in the first panel for ease of comparison). Figure 7 suggests that that there is a positive ability gradient for both wage earners and the self-employed. However, the smaller sample sizes when analyzing decile partitions within the two occupational subcategories imply that the estimates are less precise than when using the full sample. Nonetheless, the differences are statistically significant, at least when comparing the top and bottom ability deciles.

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

One might also wonder whether the statistically significant bunching we find for wage earners, and the ability gradient that we find for this group, could be driven by some connection between these workers and self-employment activities. To further sharpen our definition of wage earners, we use the panel dimension of our data, and check whether the wage earners we use in our bunching estimation have a previous history of self-employment. Figure 8 re-computes the ability gradient for wage earners, separating the analysis between those who were previously self-employed and those who have no self-employment history. As can be seen from the right panel of figure 8, even though the general level of bunching appears to be somewhat higher among those who have been previously self-employed, the ability gradient is clearest among those who have no self-employment history. This gives us some assurance that the ability gradient that we find for wage earners is unrelated to self-employment activities.

Figure 8: Role of previous self-employment for the ability gradient among wage earners

Note: Additional graphs showing the underyling estimation are shown in appendix figure A6.

5.2 Contracted hours of work

In this subsection, we match our main dataset with a register-based dataset containing information on contracted hours. The purpose is to investigate to which extent the excess bunching that we observe for high-ability individuals is associated with part-time work, representing a possibly real labor supply response to the sharp discontinuity in the marginal tax rate.

The data set covers all workers in the public sector and around 50 percent of all workers in the private sector.[27] The contracted hours of work is a variable expressed as a percentage of full-time, where full-time work (100%) corresponds to working 40 hours during a typical work-week.[28] For public workers, contracted hours is equal to actual hours of work, whereas for private workers the connection between contracted and actual hours is somewhat weaker.[29]

[27]Around 33% of all Swedish workers work in the public sector. The data is administered by the Swedish National Mediation Office. All employees in private firms with more than 500 employees are covered, whereas private firms with less than 500 employees are represented by a stratified, representative, sample of around 8500 firms each year. The measurement date is September each year for the private sector and government employees, and November for workers employed by the local government.

[28]In 2018, around 20% of workers in the private sector had some kind of part-time employment.

[29]The main reason for the discrepancy is that over-time work is not observed in the private sector.

Figure 9: Average contracted hours around the kink point (as a percentage of full-time)

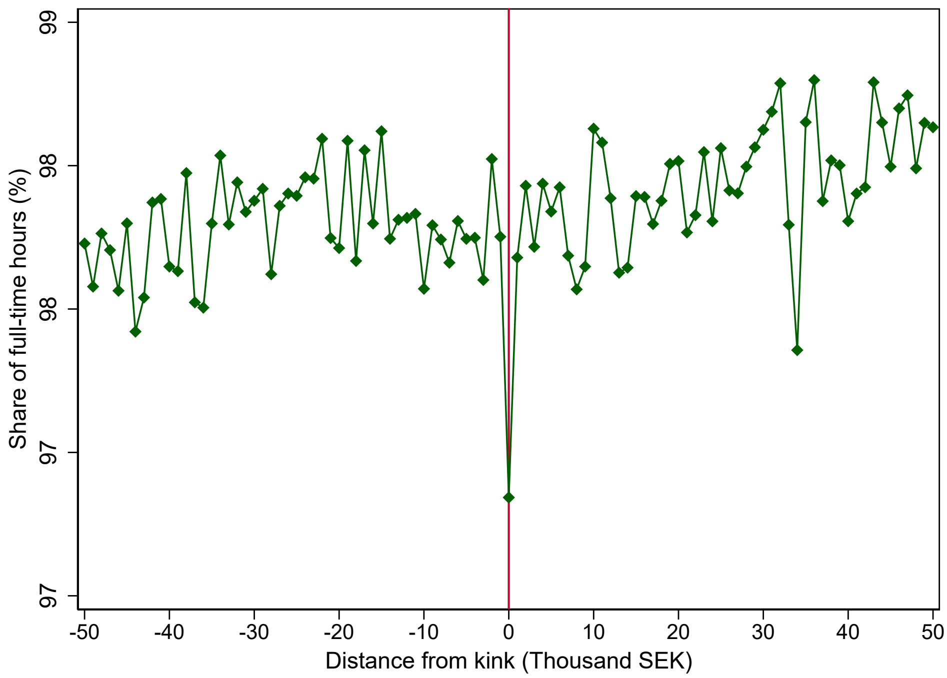
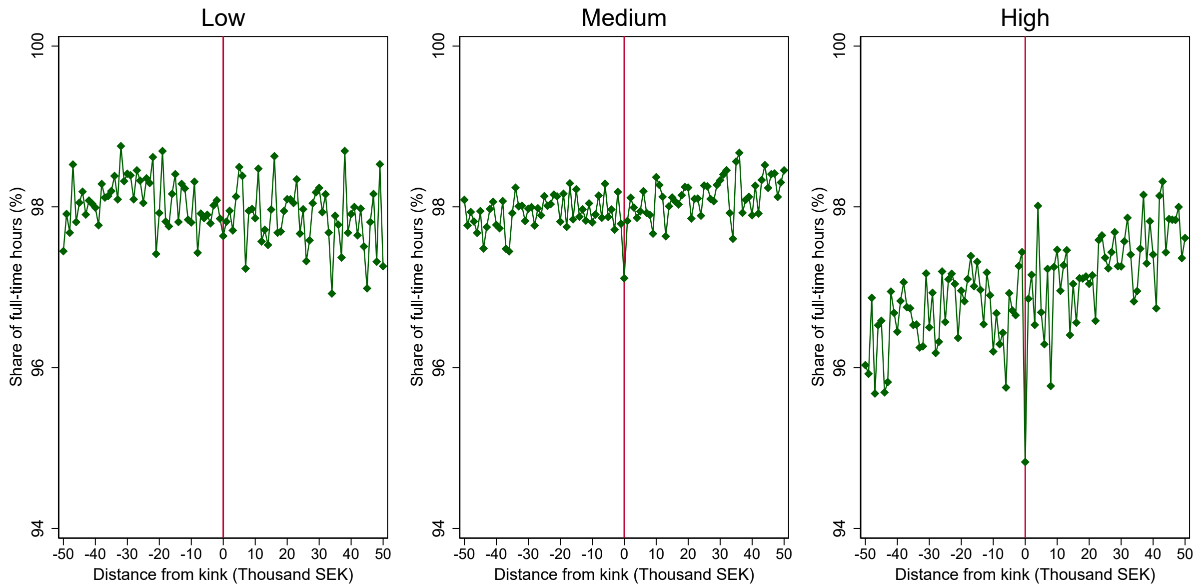


Figure 9 shows the average contracted hours (as a share of full-time) for workers with different income levels around the kink point. As can be seen from the figure, most workers work full-time, which is expected given that the kink point is located relatively high up in the income distribution. However, we can also see that there is a sharp downwards spike exactly at the kink point, showing that those who bunch have lower contracted hours. There also appears to be a general tendency for the mass to be lower in a wider interval around the kink.

Is there a link between ability and labor supply at the kink point? Figure 10 repeats the descriptive analysis in figure 9, dividing the population into three ability groups.

Figure 10: Distribution of contracted hours around the kink for three ability groups



Note: Low = ability deciles 1-2, Medium = ability deciles 3-8, High = ability deciles 9-10.

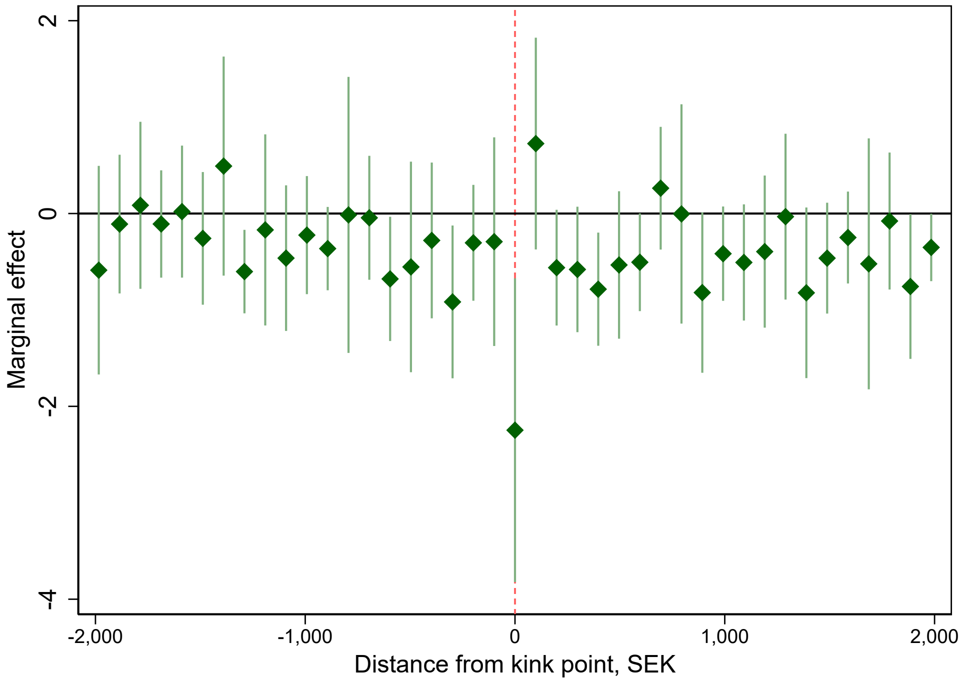
As can be seen from the figure, the lower hours appear to be predominantly concentrated among individuals belonging to the group with high cognitive ability. However, given the smaller sample size of the hours data, and the noisy nature of the data in figure 10, we now turn to a regression approach to quantify the relationship between ability and hours adjustments at the kink point. For this purpose, we examine the conditional correlation between contracted hours and cognitive ability at different income levels by running the following regression:

$$hours_{iat} = \eta_a + \lambda_t + \rho \cdot Cog_i + \varepsilon_{iat}, \tag{18}$$

where $hours_{iat}$ is contracted hours of individual i of age a in income year t, and Cog_i is standardized cognitive ability. We estimate this regression repeatedly for different subsamples of the population with different income levels y, and denote the corresponding estimates by $\hat{\rho}^y$. The results are shown in figure 11 and demonstrate that the combination of having a high ability and low hours is especially prevalent at the kink. This indicates that some of the ability gradient in tax responsiveness could be driven by real labor supply responses.[30]

[30]The data used in this section does not allow a meaningful separation between wage earners and the self-employed.

Figure 11: Marginal effect of ability on hours at different income levels



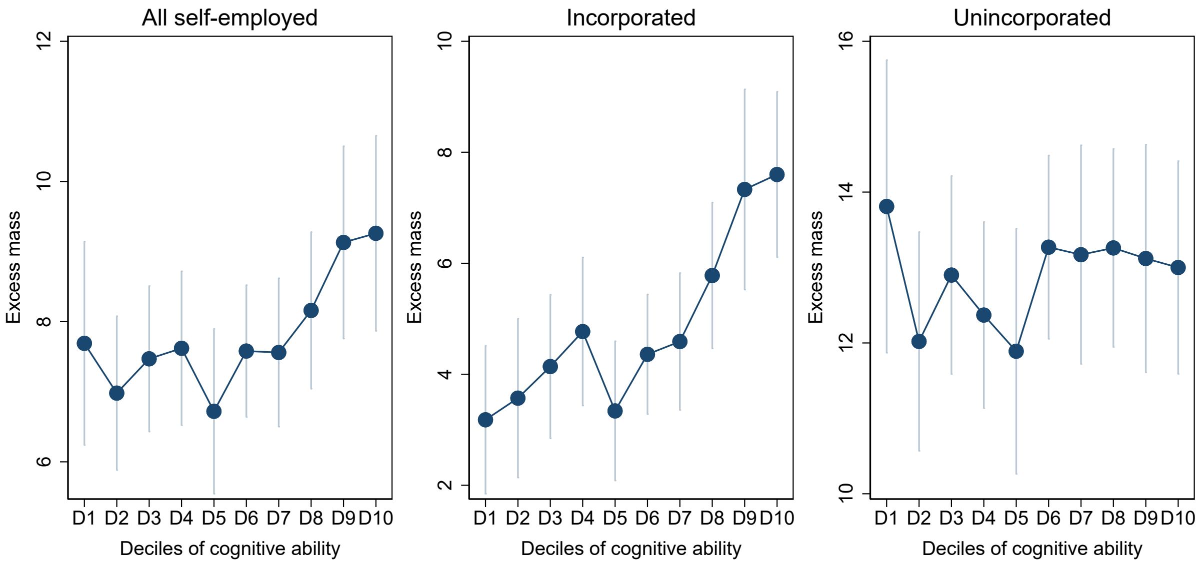
5.3 The role of income shifting and capital income

In Sweden, the taxation of capital income is separated from the taxation of labor income within the context of the Swedish dual income tax system, with a typically lower marginal tax rate applying to capital income. This provides incentives for individuals to reduce their labor income and increase their capital income to reduce their total tax burden. An important observation is that the incentive for income shifting changes discontinuously at the kink point. Thus, the observed bunching at the kink is likely to reflect a combination of traditional labor supply responses in response to the progressiveness of the labor income tax code, as well as efforts to reduce labor income and increase capital income in response to the sharp discontinuity in the tax differential between the labor and capital income tax bases.

There are two main channels this type of tax-driven behavior can occur. One possibility to take advantage of the differential between the marginal tax on labor and capital income is to 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). This opportunity is available both to wage earners and the self-employed. Another possibility, available mainly to self-employed individuals who run an incorporated business, is to re-classify, what is essentially labor income, as capital income, in order to reduce the total tax burden. While self-employed individuals, in comparison to wage earners, have greater control over their labor income flows, self-employed individuals who run their own corporations, have greater control over both their labor and capital income flows.[31]

Figure 12 shows that the ability gradient is generally much steeper among corporate owners relative to unincorporated business owners. Thus, income shifting among high ability individuals who own a corporation is likely to be one important driving factor behind the ability gradient among the self-employed. Among unincorporated business owners, the overall magnitude of bunching is larger, but we find no ability gradient for this group.

Figure 12: Ability gradient, incorporated vs. unincorporated business owners

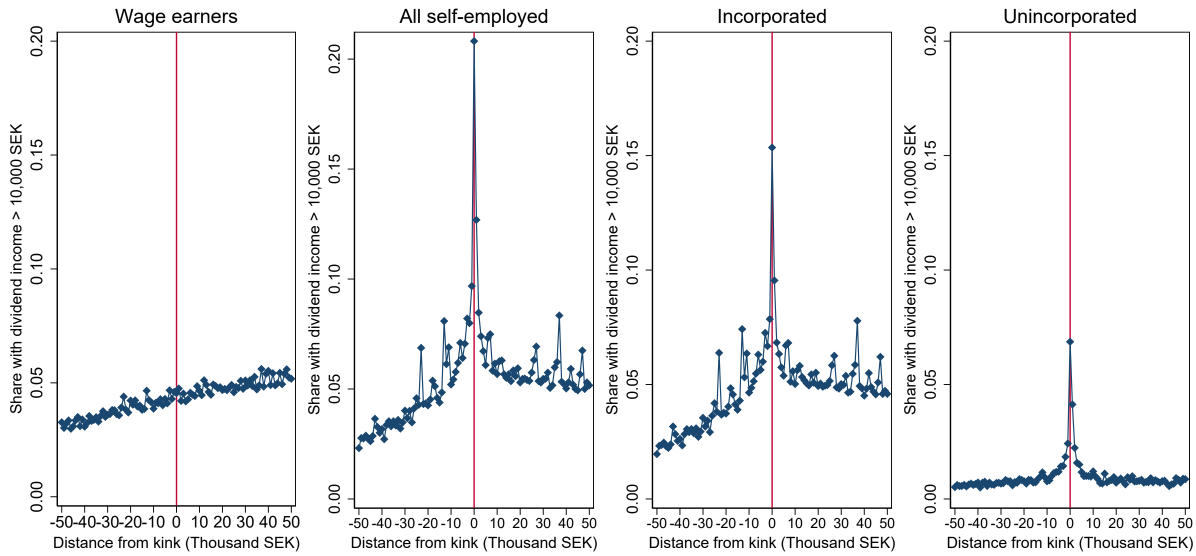


Note: For underlying bunching graphs, see appendix figure A7.

If income shifting between the labor and capital income tax bases is an important channel through which individuals bunch at the kink, we expect individuals with a labor income at the kink to have excess dividends. Figure 13 shows that this is indeed the case. The spike in dividends at the kink is largest for corporate owners, but is also present among individuals who run an unincorporated business. However, for wage earners, there is no clear spike at all.

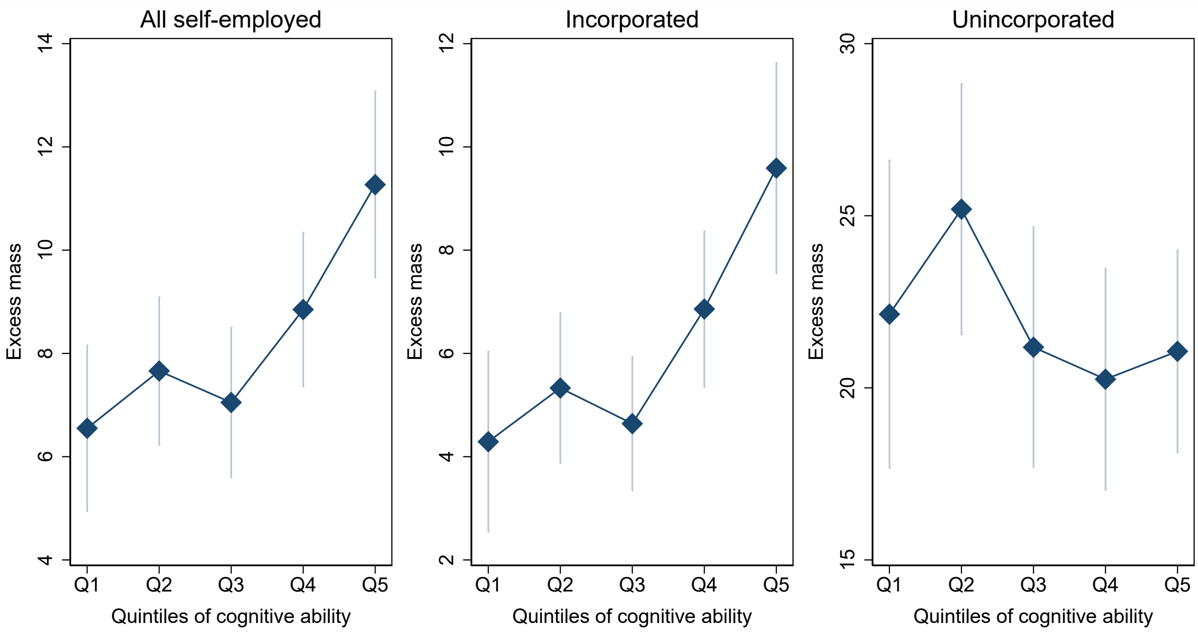
[31] An illustration of the sharp discontinuity in the tax differential between the labor and capital income tax base at the kink point is provided in appendix figure A3.

Figure 13: Propensity for high dividends around the kink point



Next, we investigate if individuals with high ability, who we know are more likely to bunch at the kink in the labor income tax, also are more likely to have high dividend income. Motivated by figure 13, we focus here on the self-employed. Figure 14 applies the bunching estimator to ability subgroups of the three groups of self-employed individuals shown in figure 13. The results in figure 14 show that it is indeed the case that high ability self-employed individuals are not only more likely to bunch at the kink, but also much more likely to have high dividends. Moreover, the results in figure 14 also show that this result is primarily driven by incorporated business owners.

Figure 14: Ability gradient in the propensity for high dividends among bunchers



Note: A visualization of the underlying bunching estimation is provided in appendix figure A8.

5.4 Gender differences and the role of household ability

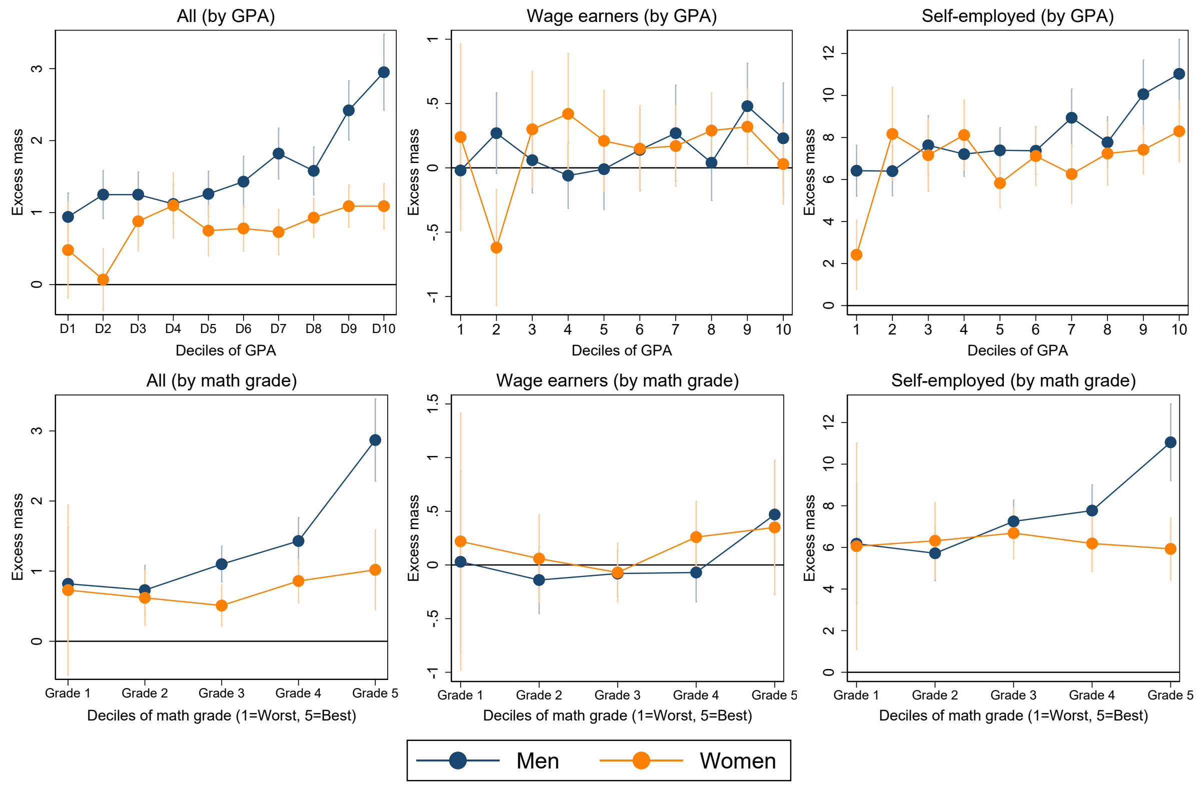
Is the ability gradient in tax responsiveness exclusive for the male population, or is it also present among women? Does bunching depend on the joint ability of spouses in married couples? Since the military enlistment data includes very few women, we analyze in this section high-school grade-point averages (GPA) and high-school math grades as proxies for ability.[32]

We begin by analyzing gender differences in the ability gradient in tax responsiveness. We focus on bunching in taxable income during 2012-2016 and divide the male and female population into ten GPA deciles and five math grade groups. We also present separate results for wage earners and self-employed individuals, in line with the distinction introduced in section 5.1. Figure 15 shows the results.[33] The main finding is that while there is an ability gradient among men, that has a similar shape 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. Moreover, the absence of an ability gradient for women holds true both among wage earners and the self-employed.

[32]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. The Swedish GPA in our sample is represented by an almost continuous score between 1 and 5 (the average of 10-15 different subject grades that take on discrete values 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 grades were allocated as follows: 7 percent received grade 5, 24 percent grade 4, 38 percent grade 3, 24 percent received 2 and 7 percent received grade 1).

[33]In general, men bunch more than women in our sample. The excess mass at the kink is 1.49 for men 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 15: Ability gradients in bunching for men and women



Note: Bunching during income years 2012-2016 in the sample of men and women born 1955-1975. For underlying bunching graphs, see appendix figure A9.

Even though the tax unit in Sweden is the individual, there is a large literature emphasizing that decisions about labor supply and taxable income are determined at the household level rather than at the individual level. Therefore, one could argue that what matters is not only individual ability, but also *household ability*. To shed light on this issue, we modify our sample by restricting our analysis to married couples and investigate bunching outcomes based on different combinations of ability of the two spouses. To this aim, we categorize each spouse into one of three ability groups: bottom (ability deciles 1-2), middle (ability deciles 3-8), and top (ability deciles 9-10).[34] Figure 16 shows bunching as a function of household ability for husbands and wives using either GPA (top panel) or math grade (bottom panel) as our ability measure. We would like to point out already from the outset, that most of the estimates are not very precisely estimated, hence the results should be interpreted with some caution.

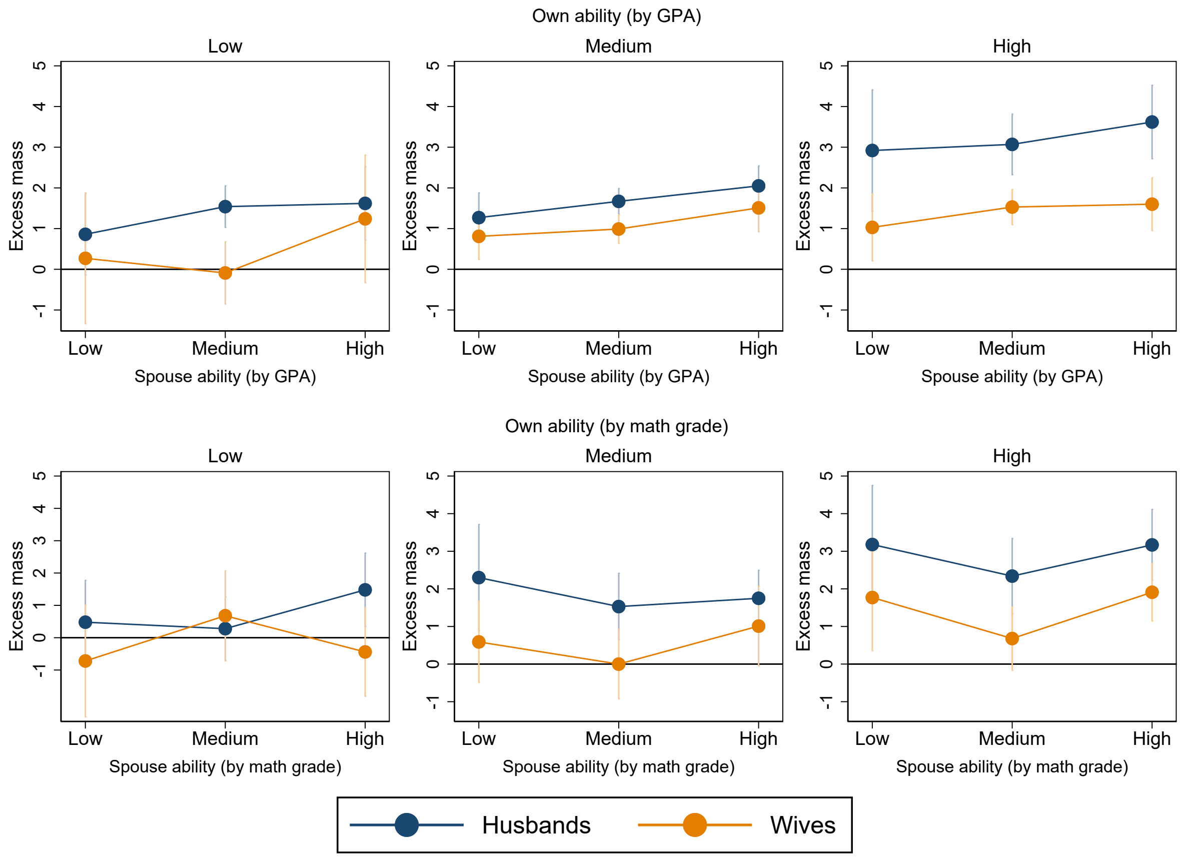
Let us start discussing the top panel of figure 16 which uses GPA. The first graph restricts attention to husbands and wives of low ability and investigates how much their bunching depends on the ability of their spouse. The second graph restricts attention to bunching among husbands and wives of medium ability, and the third panel restricts

[34]The reason we use a coarser classification of ability here is due to sample size considerations when analyzing subgroups of the population based on the joint ability of both spouses.

attention to bunching among high ability husbands and wives. As a first observation, we may note that as we move across the three graphs, we see that the general level of bunching becomes higher for husbands. This reflects our baseline results that bunching is an increasing function of the ability of men, and the results here indicate that this holds true more or less independently of the ability of the wife. The undoubtedly more interesting pattern, which appears in all three graphs, is that there appears to exist some complementarity between the skills of the two spouses: having a higher ability spouse contributes positively to individual bunching independently of whether the own ability of bunchers are low, medium or high, and independently of whether we consider husbands or wives.

The bottom panel of figure 16 that uses math grades paints, however, a more complex picture. For instance, the first graph shows that among low ability wives, bunching is highest among those with husbands of medium ability. The third graph shows that bunching among individuals with high ability is highest among those with spouses who are of either low or high ability. In sum, the results indicate that there are complementarities between the abilities of household members in the production of bunching outcomes.

Figure 16: Ability gradients in bunching as a function of household ability



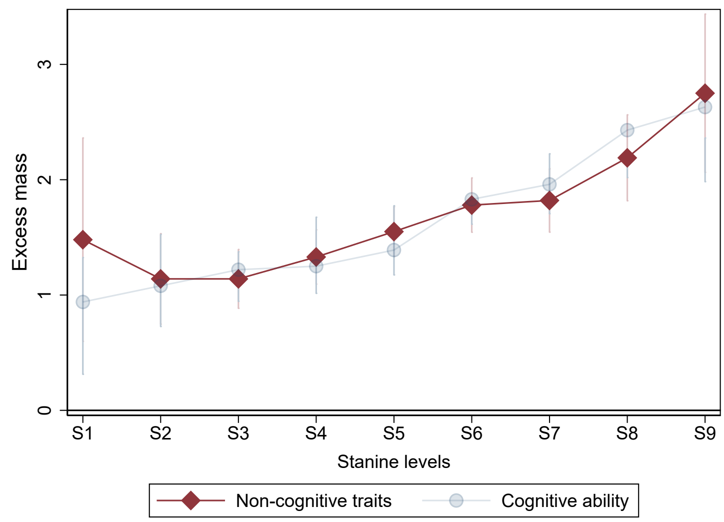
Note: For underlying bunching graphs, see appendix figures A10-A13.

5.5 The role of non-cognitive traits and physical ability

One might argue that ability, as typically interpreted in the optimal income tax literature, not only relates to cognitive ability, but also other aspects of ability that might be relevant for individuals' capacity to earn income, such as non-cognitive traits and physical ability.[35]

Figure 17 compares the ability gradient in tax responsiveness for cognitive ability with the ability gradient for non-cognitive traits. We see that the ability gradient for non-cognitive traits has a very similar shape and level as the ability gradient for cognitive ability, with the exception of the lowest decile. This is perhaps not so surprising, given that the cognitive and non-cognitive ability measures are highly correlated, as documented by the previous literature.[36]

Figure 17: The ability gradient in non-cognitive ability



Note: Bunching during income years 2012-2016 in the sample of men born 1955-1975. The underlying bunching estimation is visualized in appendix figure A14.

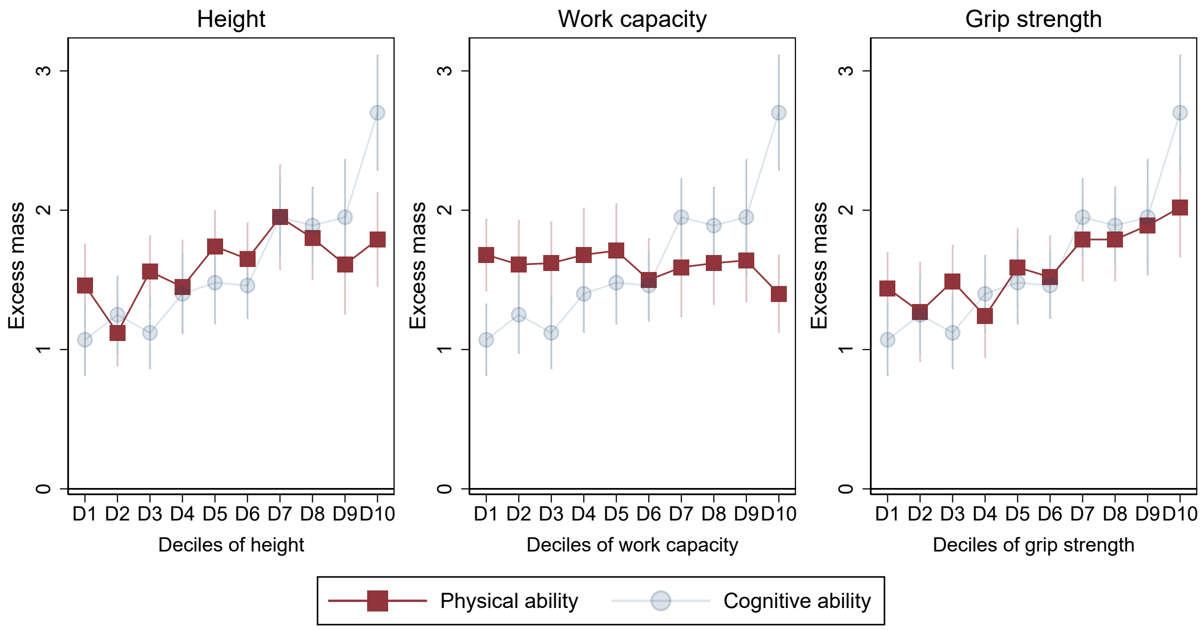
Figure 18 compares the ability gradient using cognitive ability, with the ability gradient obtained by dividing the population based on three different measures of physical ability: height, work capacity and grip strength. The results show no ability gradient for work capacity, and mild ability gradients for height and grip strength.[37] It appears thus that cognitive ability is much more strongly associated with bunching as compared to measures of physical ability.

[35]For example, taxpayer height has received attention in the optimal tax literature as a robust correlate with earnings ability, see Mankiw and Weinzierl (2010).

[36]See, for example, Lindqvist and Vestman (2011).

[37]The effect of height on earnings in the Swedish context has previously been investigated by Lundborg et al. (2014).

Figure 18: Ability gradient and physical ability



Note: Bunching during income years 2012-2016 in the sample of men born 1951-1975. The underlying bunching estimation is visualized in appendix figure A15.

5.6 Individual-level regressions

The bunching analysis relies on aggregated (binned) data and a polynomial (or non-parametric) curve fitting technique to approximate counter-factual outcomes. In this section, we adopt a set of alternative strategies and estimate ability gradients using individual-level data in a regression framework.

5.6.1 Regression-based ability gradients and placebo gradients

We generate *ability decile dummies*, $CogDec_i^d, d = 1,2,\ldots,10$ that take on the value 1 if individual i belongs to decile d in the distribution of cognitive ability. We then estimate the effect of cognitive ability on the probability of locating in different income intervals $\mathscr{Y}$, by estimating the following equation:

$$\mathrm{I}\{y_{iat} \in \mathscr{Y}\} = \eta_a + \lambda_t + \sum_d \beta^d \cdot CogDec_i^d + \varepsilon_{iat}, \tag{19}$$

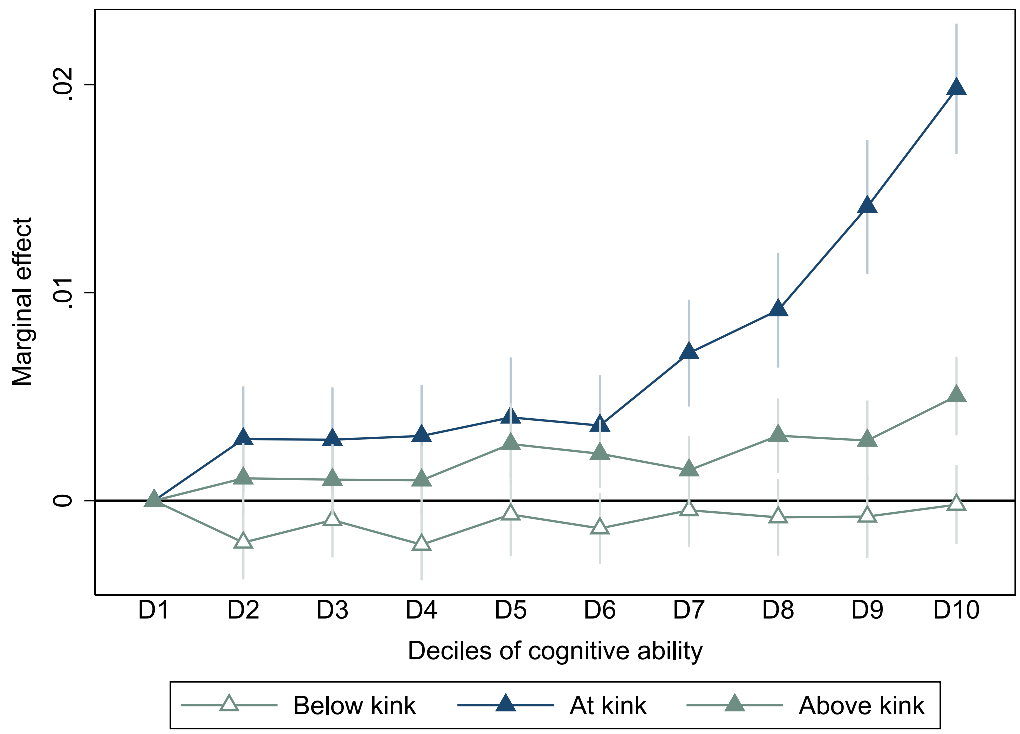
where $\mathrm{I}\{y_{iat} \in \mathscr{Y}\}$ is an indicator variable taking on the value of 1 if an individual's income y_{iat} falls within the income interval $\mathscr{Y}$, η_a represent cohort fixed effects, λ_t are year fixed effects, and ε_{iat} are error terms.[38] Using a notation similar to that used in section 2.2, we let $\mathscr{Y}$ be equal either to a small window centered at $\hat{z}$, $[-\delta,\delta]$, an

[38]The income measure is the same as in the bunching analysis, covering the period 2012-2016, and standard errors are clustered at the individual level. Notice that, just like in the bunching analysis, the regressions focus on a local part of the income distribution.

interval to the left of $\hat{z}$, $[-3\delta,-\delta]$, or an interval to the right of $\hat{z}$, $[\delta,+3\delta]$. The idea will be to compare the ability gradient obtained from the estimates β^d, $d = 1,2,\dots,10$, when the outcome variable is a dummy equal to one for observations falling within the small income interval centered at the kink, with the *placebo-gradients* β^d obtained when the outcome variable is a dummy equal to one for observations located in either of the two neighboring intervals. In this analysis, we restrict the overall estimation sample to observations in the interval $[-25k,25k]$. In our baseline results, we choose $\delta = 0.5k = 500SEK$.

The results shown in figure 19 are very consistent with the ability gradient found in the bunching estimation above (figure 3). The marginal effects are very close to being monotonically increasing in ability for individuals at the kink point, with no such gradient for individuals in the two placebo-intervals (a regression table is found in appendix A, table A3). Since cognitive ability is measured at age 18, and we study incomes realized around 30 years later (at ages 40-65), the estimated ability coefficients can be given a causal interpretation.

Figure 19: Regression-based ability gradients in bunching and placebo-gradients



Note: Estimated coefficients from regressions in (19) for the three income intervals described in the text.

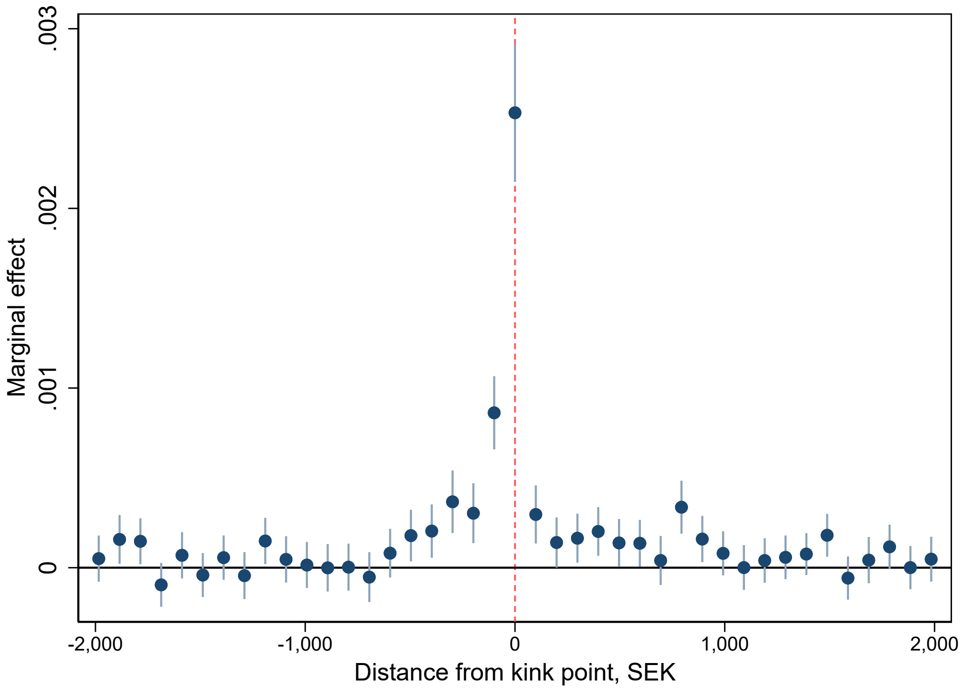
5.6.2 Marginal effect of cognitive ability on the probability of locating exactly at different income intervals around the kink

We next turn to characterize sharp bunching behavior *exactly* at the kink point. In this analysis, we maintain as our estimation sample the interval $[-\phi,\phi] = [-25k,25k]$, and analyze the effect of cognitive ability on the probability of locating exactly at different income levels within this window in steps of 100 SEK.[39] Due to sample-size considerations, we focus on a linear specification with a continuous measure of ability rather than the flexible dummy-variable specification used earlier. Formally, we estimate:

$$\mathrm{I}\{y_{iat} = y\} = \eta_a + \lambda_t + \beta \cdot Cog_i + \varepsilon_{iat}, \quad y \in \{\hat{y} - \phi, \hat{y} - \phi + 100, \ldots, \hat{y} + \phi - 100, \hat{y} + \phi\}, \qquad (20)$$

where Cog_i is a continuous measure of cognitive ability (the z-score) and $\hat{y}$ represents the kink point income level. Notice that these are *rolling* regressions, with one separate regression for each value of $y \in \{\hat{y} - \phi, \hat{y} - \phi + 100, \ldots, \hat{y} + \phi - 100, \hat{y} + \phi\}$. The results are shown in Figure 20.

Figure 20: Placebo regressions in a small neighborhood around the kink point



Note: The figure presents estimated marginal effects of Cog on the probability to earn an income of $y \in \{\hat{y}-\phi, \hat{y}-\phi+100, \ldots, \hat{y}+\phi-100, \hat{y}+\phi\}$ where $\hat{y}$ is the kink point income level and 2ϕ is width of the estimation window around the kink in equation (20). For the purpose of graphical exposition, the figure focuses on the narrow interval $[-2k, 2k]$.

The figure shows that ability has a much larger effect on the probability of locating *exactly* at the income level of the kink point relative to the probability of locating at adjacent income levels. This placebo test thus reinforces our main results and also points to the special importance of the income level exactly at the kink point.

[39]The results in this section are robust to the selection of the overall estimation sample $[-25k, 25k]$, and virtually identical results are obtained when selecting narrower intervals.

6 Concluding remarks

We have analyzed the relationship between ability and tax responsiveness by studying the amount of bunching at a large and salient kink point in the Swedish income tax schedule. Our main result is that individuals in the top decile of the ability distribution react twice as strongly to the discontinuity as compared to the average individual, and three times more strongly than individuals in the bottom ability decile. These are somewhat remarkable findings, considering the fact that we link ability measured at age 18 with incomes recorded more than 30 years later. The ability gradient in tax responsiveness is almost monotonic and robust to changes in the estimation strategy.

In a series of extensions, we have shed further light on the link between ability and tax responsiveness, documenting downwards adjustments in hours of work, and excess densities in the distribution of capital income. We have also analyzed gender-differences in the ability gradient using data on high-school grades, finding a stronger ability gradient among men than among women. Furthermore, we show that tax responsiveness might be related to household ability and matching patterns in the marriage market. Finally, even though our main analysis has focused on the role of cognitive ability, we have found a similar ability gradient for non-cognitive traits. However, when examining the role of physical ability, we find only a moderate ability gradient for taxpayer height and grip strength, and no ability gradient for work capacity.

While the primary purpose of our paper has been to analyze the relationship between non-acquired skills and taxpayer behavior, our findings also shed light on the second-best nature of the problem of designing the income tax. The fact that we find substantial skill heterogeneity conditional on income at the kink, can be interpreted as providing direct empirical evidence of high-skill individuals mimicking low-skill individuals in order to escape progressive income taxation. To our knowledge, this is the first time these mechanisms have been tested empirically.

The relationship between ability and tax responsiveness highlight a conflict in optimal tax design. Governments want to tax high-ability individuals more and high-elasticity individuals less. However, our results indicate that high-ability individuals might also be high-elasticity individuals, which questions the efficiency of using labor income taxation to achieve skill-based redistribution. Instead, there seems to be a need to complement taxes on labor income with other taxes and policies in order to achieve ability-based redistribution. We leave the interesting question of what other taxes and policy tools ideally should be employed to achieve this aim as a topic for future research.

References

Abeler, J. and Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy*, 7(3):1–28.

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). Chapter 5 - Behavioral Public Economics. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics - Foundations and Applications 1*, volume 1, pages 381 – 516. North-Holland.

Bertanha, M., McCallum, A. H., and Seegert, N. (2019). Better bunching, nicer notching. *Unpublished manuscript.*

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.

Blomquist, S., Newey, W., Kumar, A., and Liang, C.-Y. (2017). On bunching and identification of the taxable income elasticity. Working Paper 24136, National Bureau of Economic Research.

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*.

Ebert, U. (1992). A reexamination of the optimal nonlinear income tax. *Journal of Public Economics*, 49(1):47 – 73.

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*, 12(1):1–31.

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*.

Guesnerie, R. and Seade, J. (1982). Nonlinear pricing in a finite economy. *Journal of Public Economics*, 17:157–179.

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*, 23:640–669.

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.

Jacquet, L. and Lehmann, E. (2020). Optimal Income Taxation with Composition Effects. *Journal of the European Economic Association*. jvaa022.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8:435–464.

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*.

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.

Lundborg, J., Nystedt, P., and Rooth, D.-O. (2014). Height and earnings: The role of cognitive and non-cognitive skills. *Journal of Human Resources*, 49(1):141–166.

Mankiw, N. G. and Weinzierl, M. (2010). The optimal taxation of height: A case study of utilitarian income redistribution. *American Economic Journal: Economic Policy*, 2(1):155–76.

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 us tax schedule. *American Economic Journal: Economic Policy*, 12(3):402–32.

Paetzold, J. (2018). How do taxpayers respond to a large kink? Evidence on earnings and deduction behavior from Austria. *International Tax and Public Finance*, pages 1–31.

Pirttilä, J. and Selin, H. (2011). Income shifting within a dual income tax system: Evidence from the Finnish tax reform of 1993. *The Scandinavian Journal of Economics*, 113(1):120–144.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *Review of Economic Studies*, 68(1):205–229.

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.

Stantcheva, S. (2020). Understanding economic policies: What do people know and learn? *Harvard University Working Paper, 2020.*

Stern, N. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, 17(2):181 – 211.

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.

Tazhitdinova, A. (2020). Are changes of organizational form costly? Income shifting and business entry responses to taxes. *Journal of Public Economics*, 186:104187.

A Appendix

A.1 Descriptive statistics and additional figures

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)

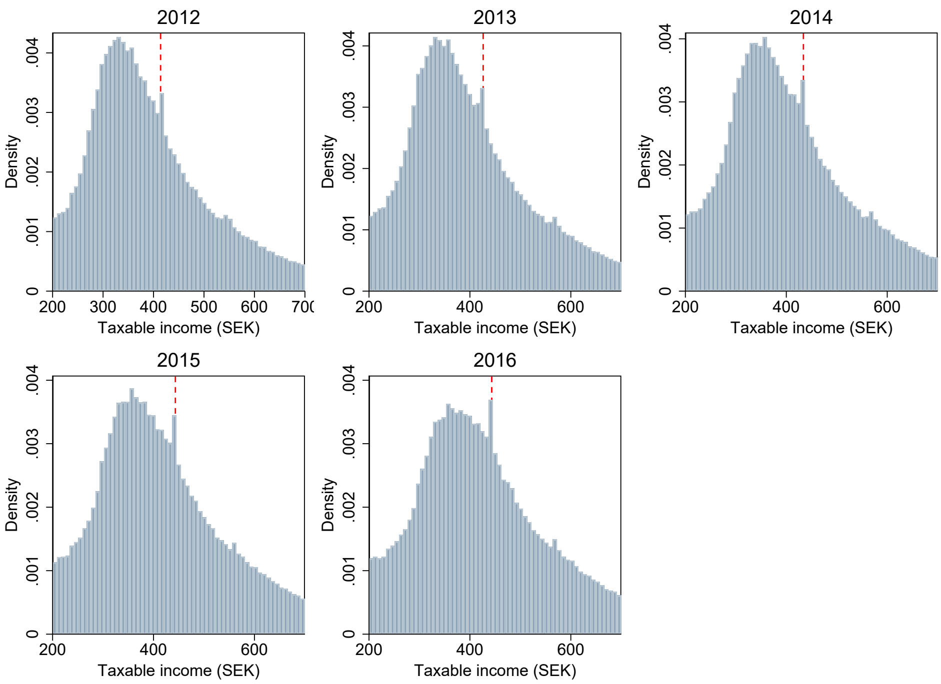
Category	Observations
Main sample: Men born 1951-1975	
(1) Men born 1951-1975 with cognitive ability score in military enlistment	1,283,254
(2) In (1) born 1955- (cohorts with GPA data coverage)	1,075,500
(3) In (2) and observed high-school GPA	786,032
(4) In (3) born 1966- (cohorts with math data coverage)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

Table A3: Regression-based ability gradient in tax responsiveness, estimation results

	Below kink	At kink	Above kink
D2	-0.002*	0.003*	0.001
	(0.001)	(0.001)	(0.001)
D3	-0.001	0.003*	0.001
	(0.001)	(0.001)	(0.001)
D4	-0.002*	0.003*	0.001
	(0.001)	(0.001)	(0.001)
D5	-0.001	0.004**	0.003**
	(0.001)	(0.001)	(0.001)
D6	-0.001	0.004**	0.002**
	(0.001)	(0.001)	(0.001)
D7	-0.000	0.007***	0.001
	(0.001)	(0.001)	(0.001)
D8	-0.001	0.009***	0.003**
	(0.001)	(0.001)	(0.001)
D9	-0.001	0.014***	0.003**
	(0.001)	(0.002)	(0.001)
D10	-0.000	0.020***	0.005***
	(0.001)	(0.002)	(0.001)
Observations	622,842	622,842	622,842

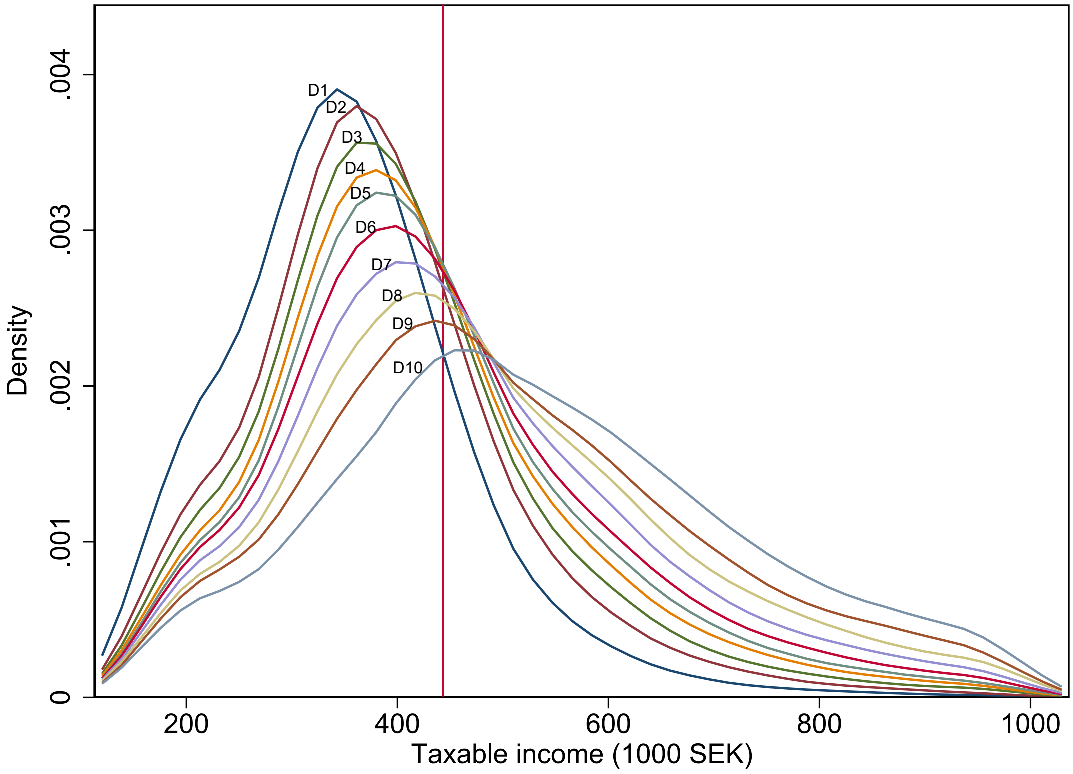
Note: Regression results for equation (19). *, **, *** denote statistical significance at the 5%-, 1%-, and 0.1%-level. Standard errors are clustered by individual and year.

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



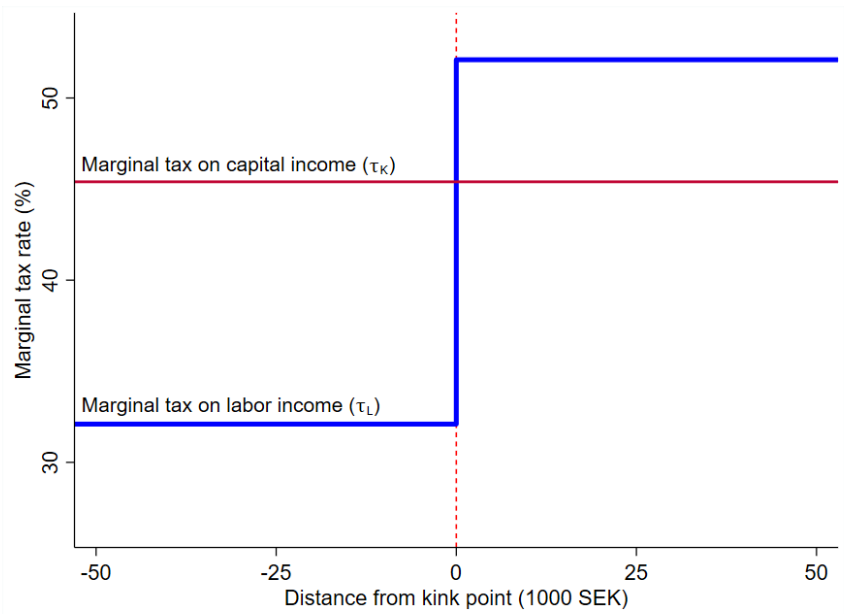
Note: Annual taxable labor income distributions around the kink point. Male population, born 1951-1975.

Figure A2: Kernel density estimation of taxable income distributions by ability decile.



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

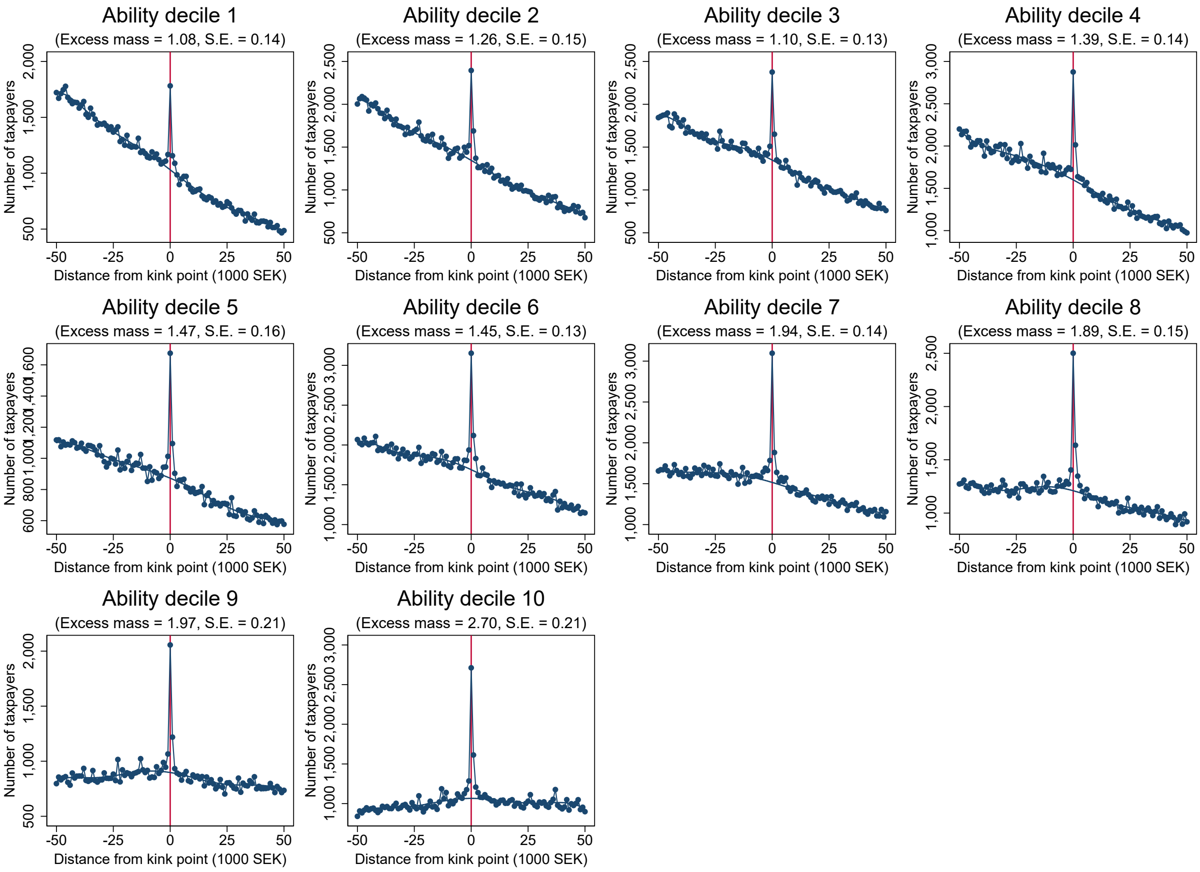
Figure A3: Increase in incentives for income shifting at the kink point.



Note: Illustration using the tax system 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%).

A.2 Additional bunching results

Figure A4: 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 A5: Sensitivity analysis with respect to the wide and small bunching windows (cognitive ability)

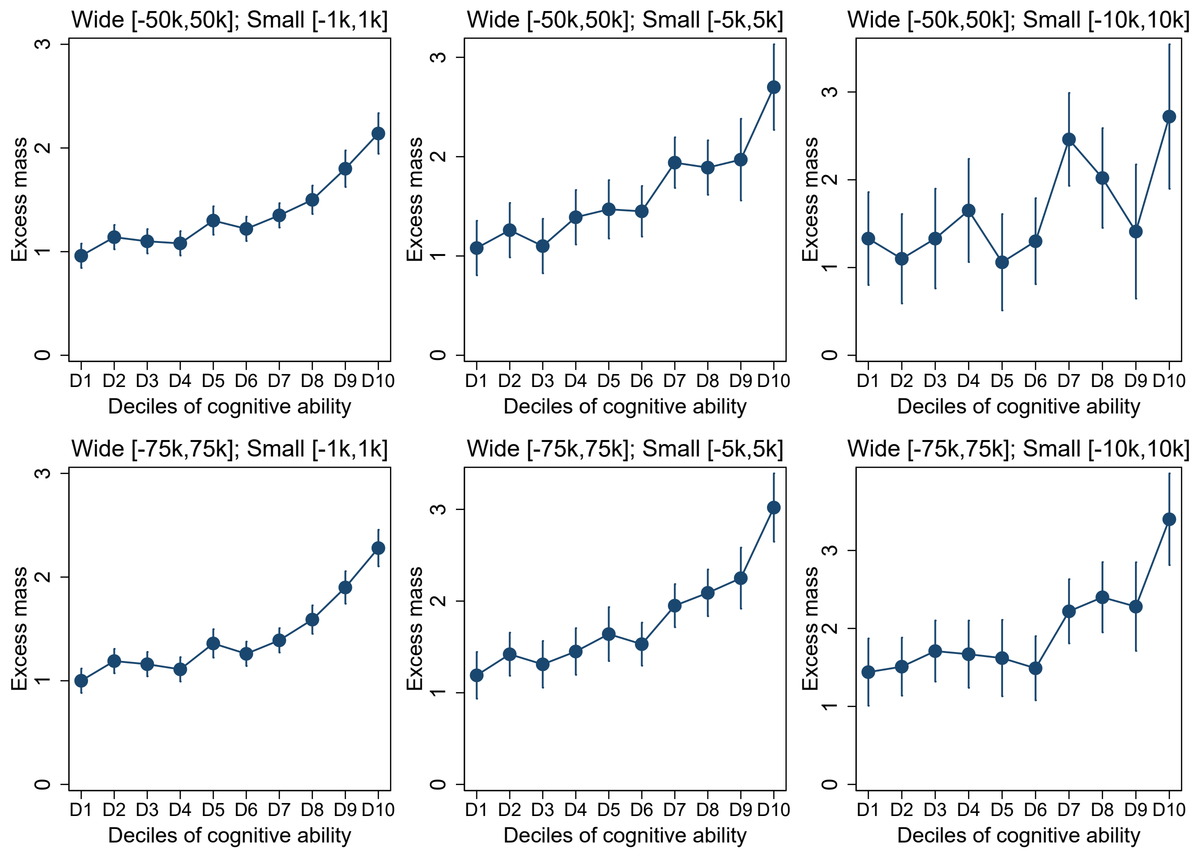


Figure A6: Bunching of wage earners with or without self-employment history

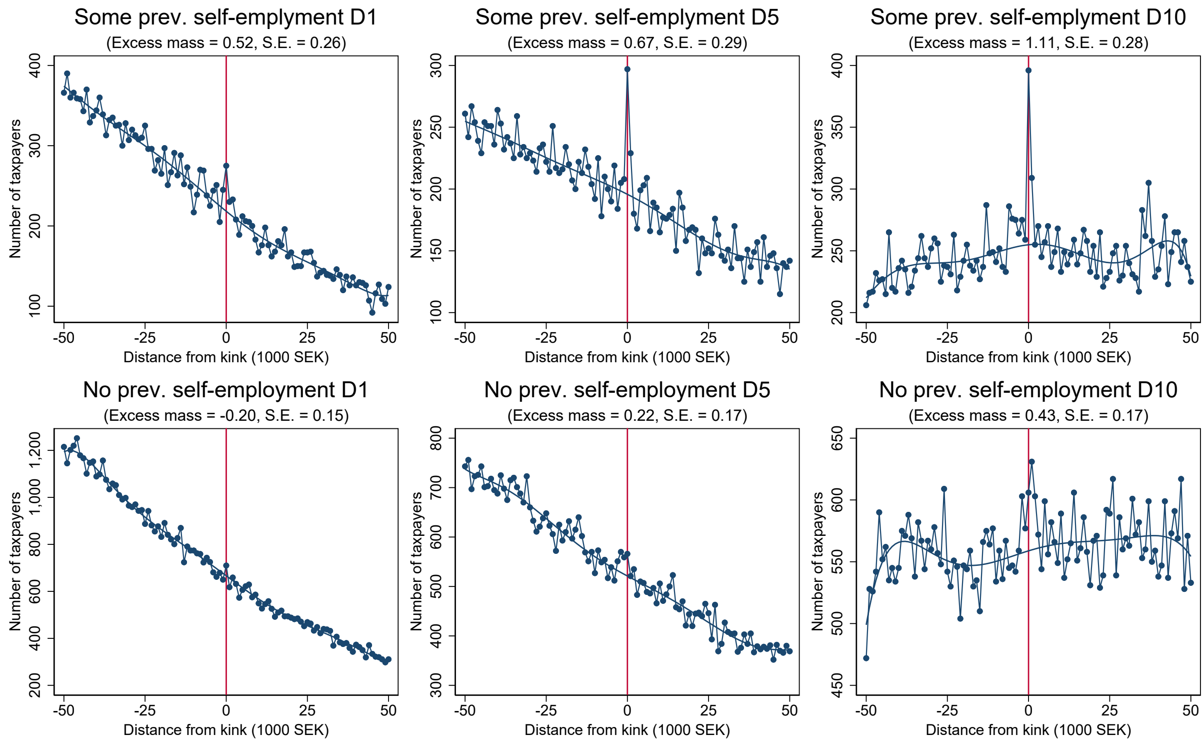


Figure A7: Bunching among the self-employed

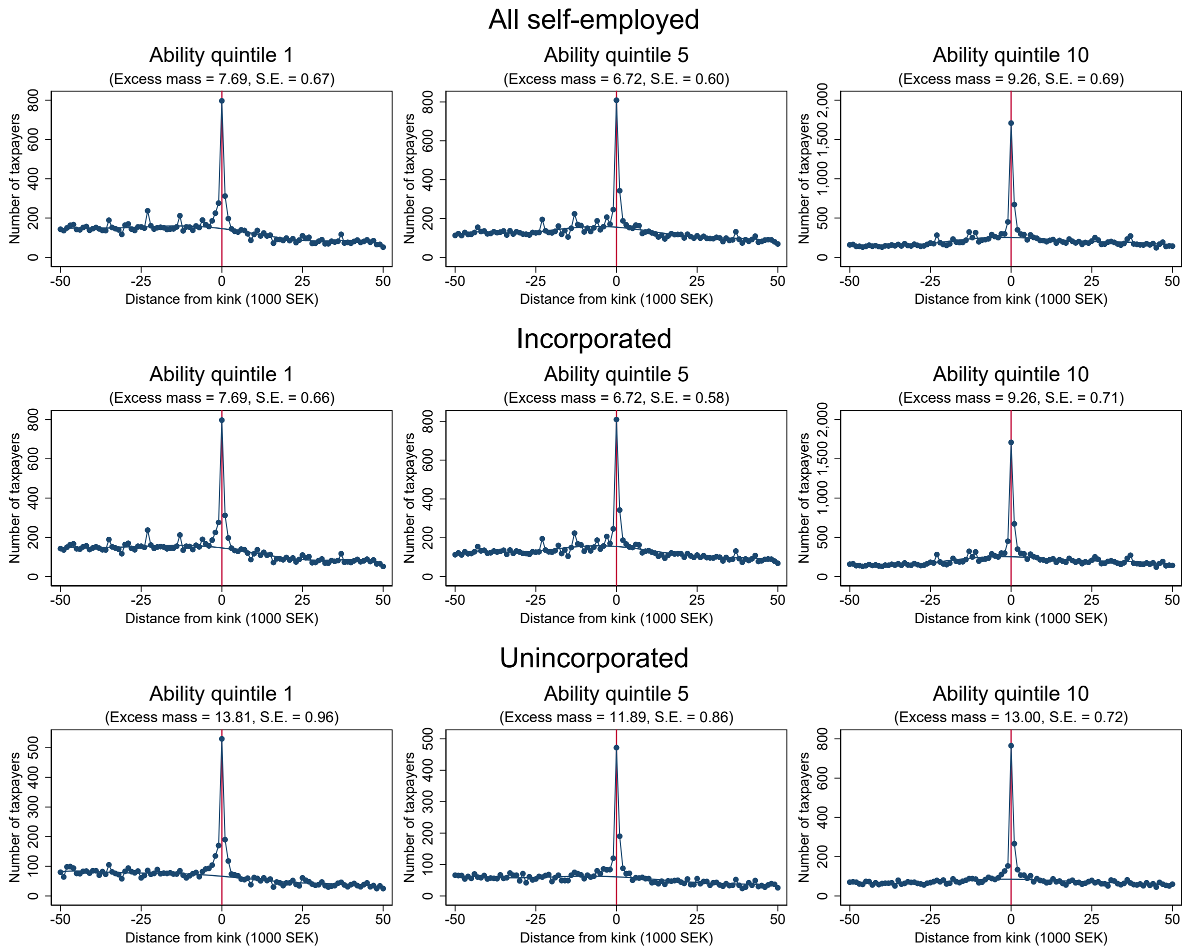
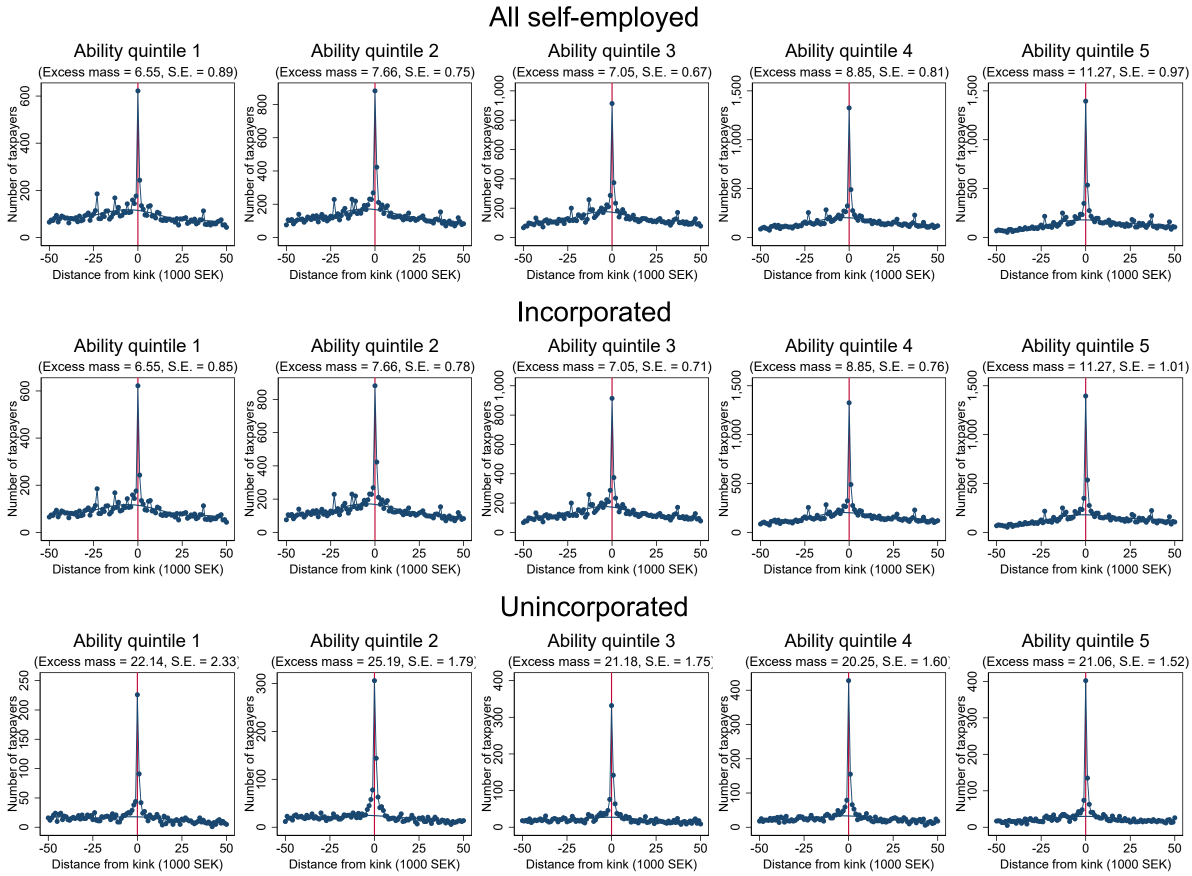


Figure A8: Bunching estimation applied to dividends among the self-employed



Note: Bunching estimation is applied to the frequency distribution of the propensity of having high dividends (> 10000 SEK) around the kink point for different ability groups.

Figure A9: Bunching of men and women: high-school GPA and math grades

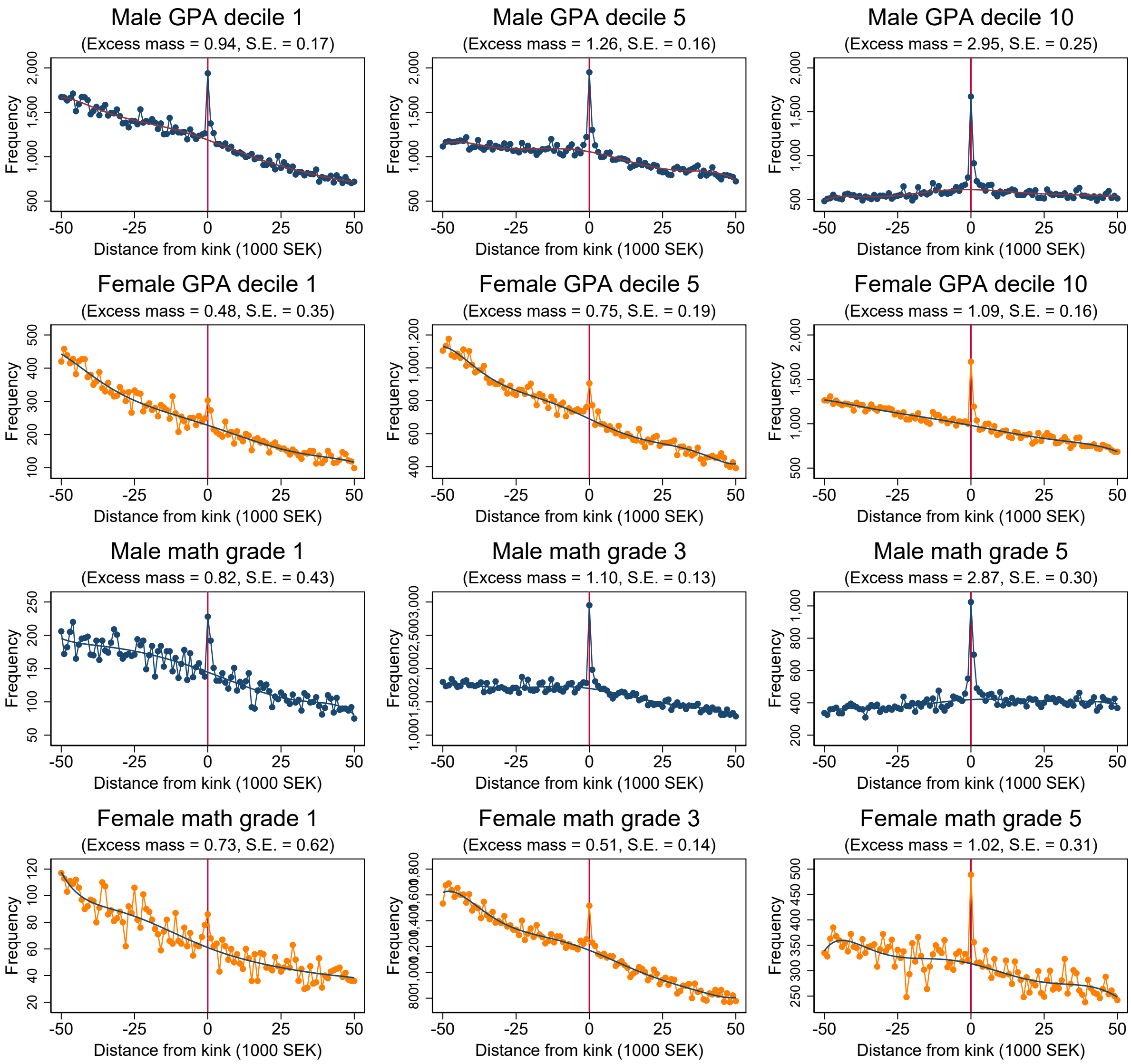


Figure A10: Male bunching as a function of household ability: GPA

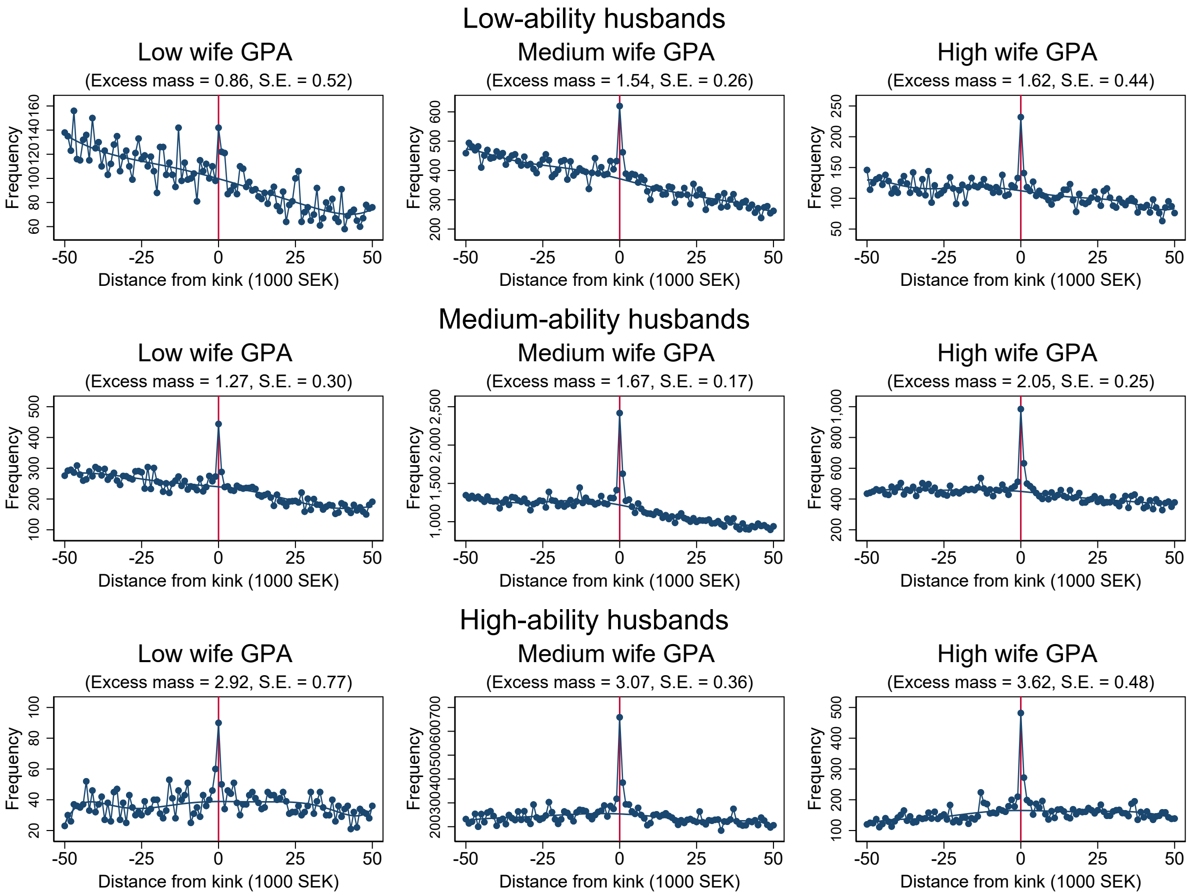


Figure A11: Female bunching as a function of household ability: GPA

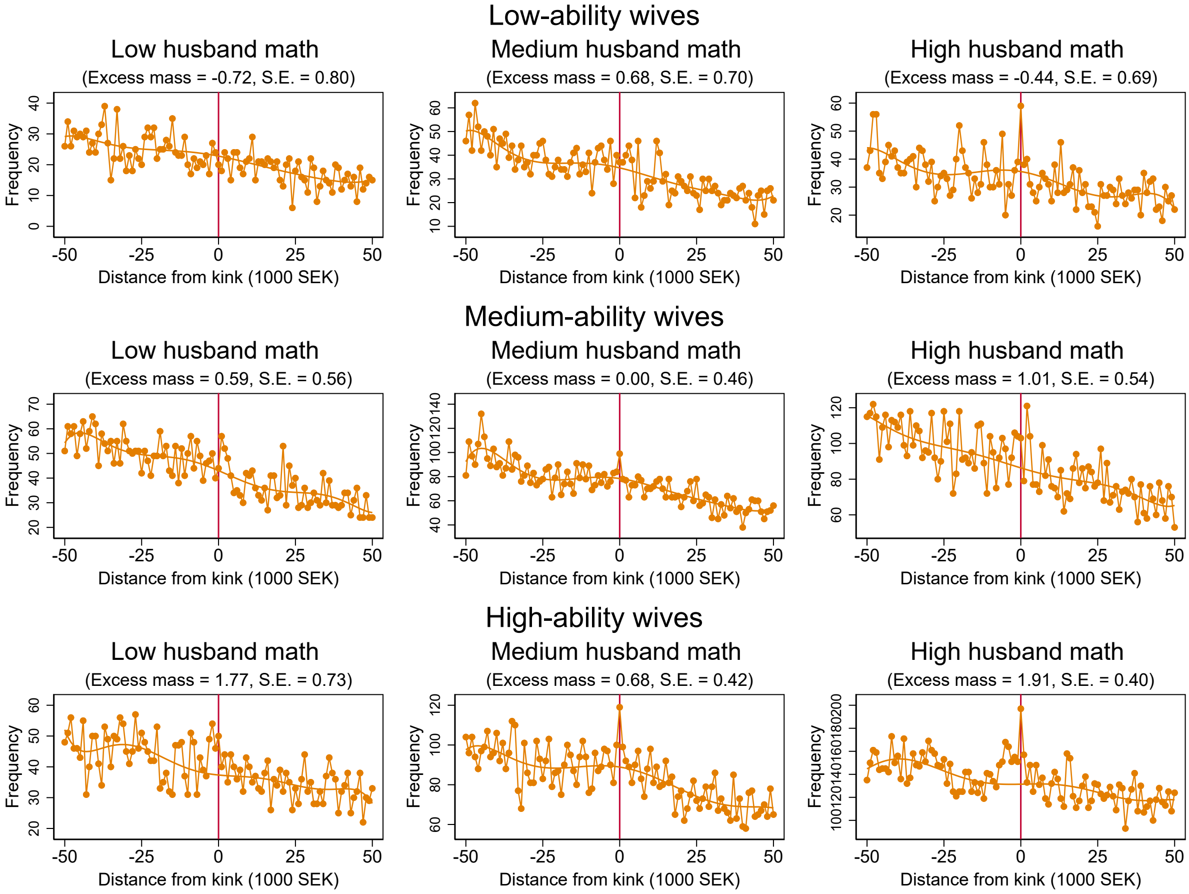


Figure A12: Male bunching as a function of house ability: math grades

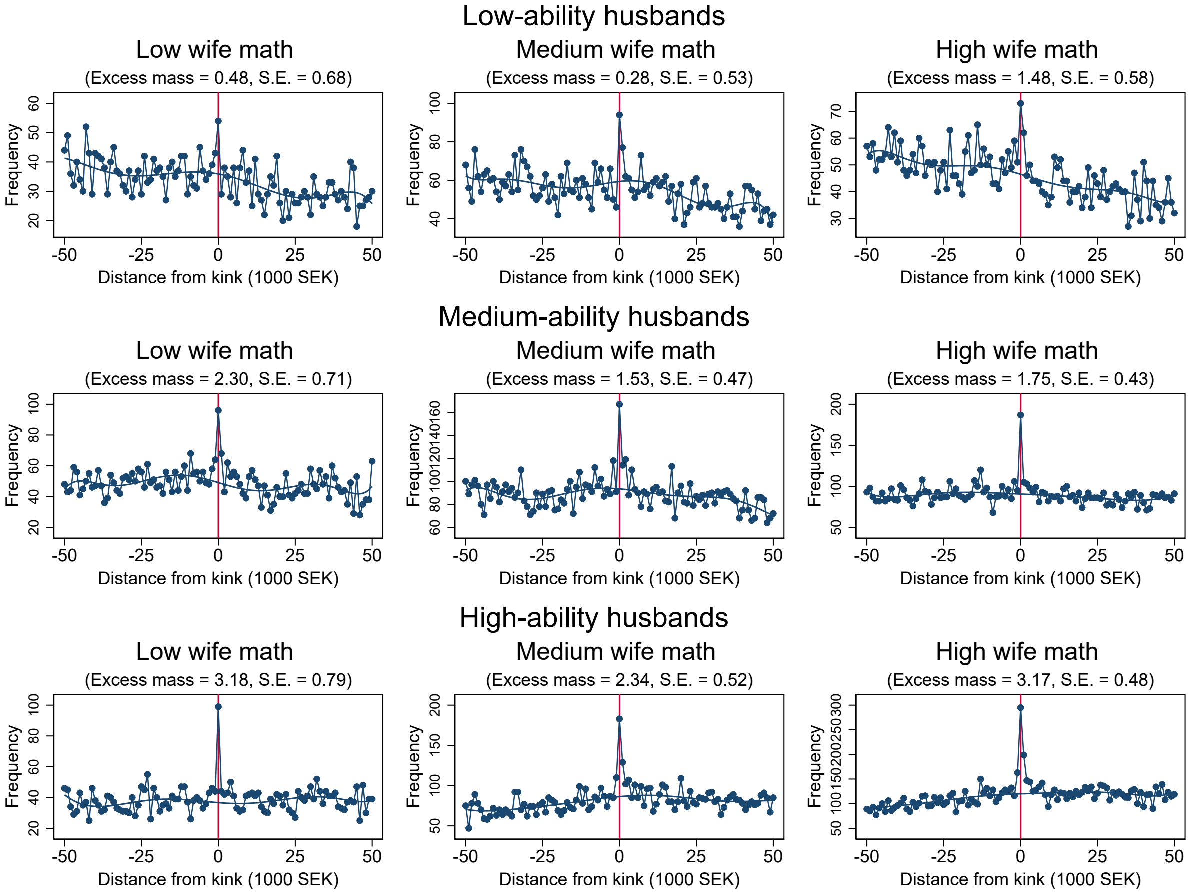


Figure A13: Female bunching as a function of household ability: math grade

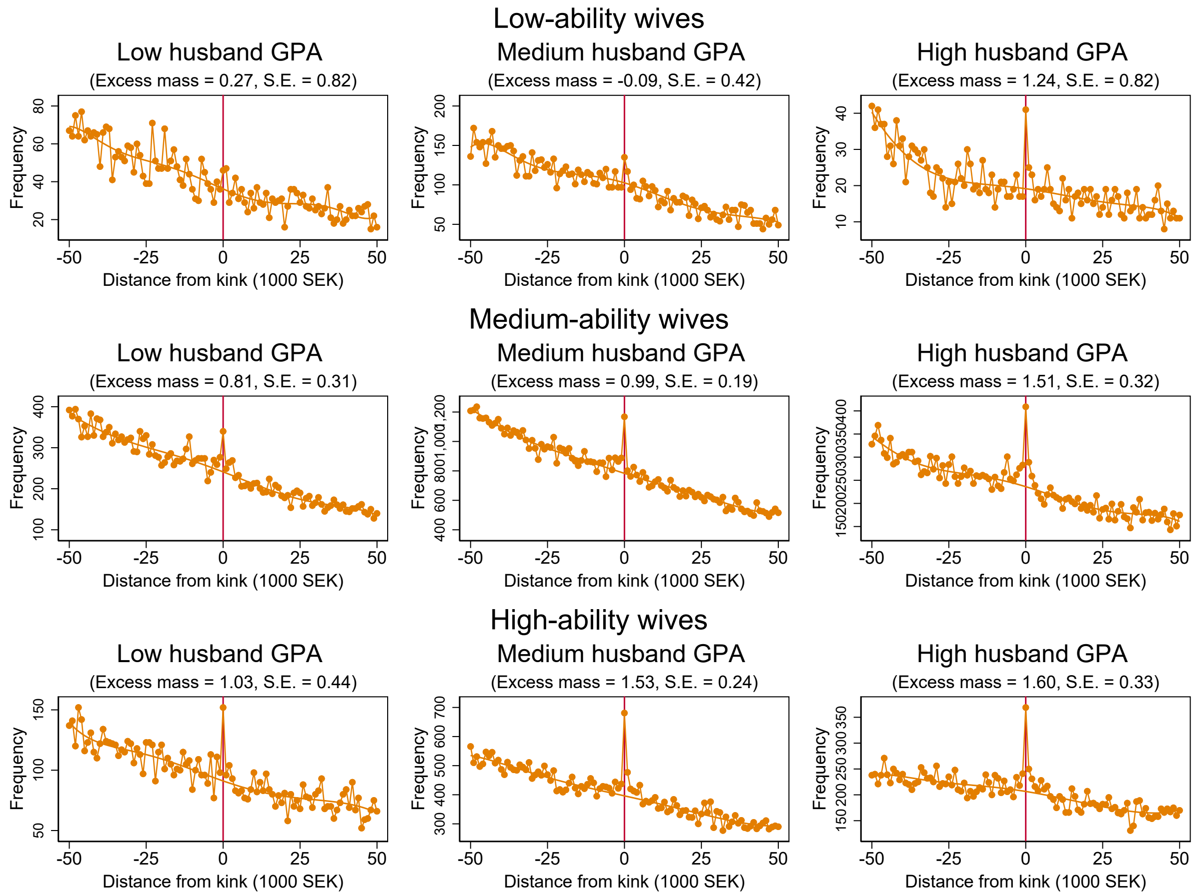


Figure A14: Bunching in cognitive and non-cognitive traits: Stanine groups

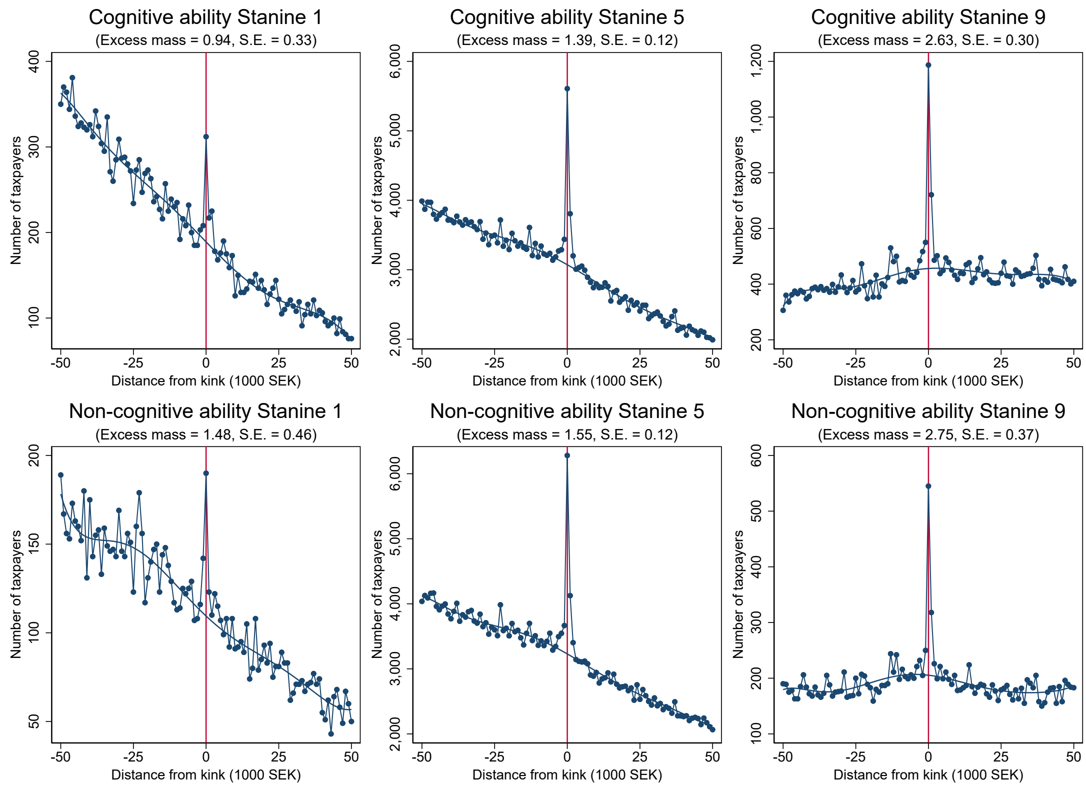


Figure A15: Bunching in grip strength and physical work capacity