The Ability Gradient in Tax Responsiveness*

Spencer Bastani† Daniel Waldenström‡

March 21, 2021

Abstract

We study the role of cognitive ability for individuals' tax responsiveness using linked administrative tax and military enlistment registers. Our main finding is that individuals in the top decile of the ability distribution react twice as strong to a large and salient kink point in the Swedish tax code as compared to the average individual, and three times as strong as individuals in the bottom ability decile. This ability gradient is mainly driven by income shifting among high-ability owners of incorporated businesses, but we also find evidence of labor-supply responses among high-ability wage earners.

Keywords: cognitive ability, bunching, income shifting, labor supply, optimal taxation

JEL codes: H21, H24, J22, J24

*We are grateful to Felix Bierbrauer, Emanuel Hansen, Håkan Selin, Joel Slemrod and seminar participants at the University of Michigan, The Free University of Berlin, The Research Institute of Industrial Economics (IFN), The Swedish Institute for Social Research (SOFI), Stockholm University, and the CESifo Area Conference on Public Economics 2020 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 Research Institute for Industrial Economics (IFN), Stockholm; 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 taxpayer responses are related to cognitive ability measured in young adulthood. We do this by studying the way in which individuals with different abilities react differently to a large and salient discontinuity in the Swedish marginal tax rate schedule.

Our analysis relies on a combination of bunching estimation and a regression-based approach, using population-wide administrative tax records that are linked to 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, between age 30 and 65.

The paper's main finding is that individuals with high measured cognitive ability are substantially more likely to react to the large marginal tax change in the tax schedule. As this pattern gradually gets stronger as one moves up in the ability distribution, it 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. Our results indicate that responses are higher for individuals with higher ability among people with the *same* taxable income. This challenges the common assumption in the theoretical and empirical tax literature that individuals with different earnings capacities react in the same way to tax changes, and points to the importance of considering taxpayer heterogeneity, conditional on labor income, when designing the tax system or estimating behavioral responses to taxation.[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

[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).

the most to progressive income taxation.

The ability gradient in tax responsiveness is sharpest and quantitatively most important for the self-employed. Examining the distribution of capital income along the labor income distribution shows that self-employed individuals who bunch at the kink are much more likely than wage earners to have high dividends. This pattern suggests that the ability gradient for the self-employed is driven by income shifting.[3] Since income shifting mainly is available to owners of closely help corporations, we decompose the self-employed population according to the incorporation status of their firm. This decomposition reveals that the ability gradient for the self-employed is driven by incorporated business owners. High-ability owners of closely held corporations take advantage of the lower tax rate on capital income within the context of the Swedish dual income tax system by reclassifying what is essentially labor income, as capital income, thereby circumventing the intentions of the progressive labor income tax. These results thus provide evidence of an ability gradient in tax avoidance.

The ability gradient among wage-earners is not as sharp as for the self-employed, and quantitatively less important. However, using a special register-based data set on contracted hours of work, we can measure the existence of an ability-gradient also in real responses to the discontinuity in the tax schedule.[4] This analysis shows evidence of sharp downwards labor supply adjustments at the kink in the tax schedule, but only for middle-high ability individuals. In other words, the tax responses we observe for the total population are therefore driven by a combination income shifting and (to a smaller extent) real labor supply responses.

We provide two robustness checks to our main analysis. First, we calculate ability gradients using other measures of ability that are contained in the military enlistment data, non-cognitive ability and physical status. In comparison to the gradient obtained using cognitive ability, the ability gradient using non-cognitive traits is weaker in the bottom of the ability distribution, but follows a similar pattern for higher levels of ability. Interestingly, it is increasing for wage earners but flat for the self-employed. The physical ability measures almost do not display any ability gradient at all, especially among the self-employed, where the ability gradient for physical ability sometimes even is negative. These findings show that it is cognitive ability in particular that drive our results. Second, we use high-school grades as proxies for cognitive ability. This analysis

[3]Income shifting refers to the legal tax optimization in order to reduce tax burden. In the United States, which has a comprehensive income tax system, shifting between the corporate and personal tax bases has previously been documented by Gordon and Slemrod (2000). In Sweden, the taxation of labor and capital income is separated within the context of the Swedish "dual" income tax system, with a lower tax rate on capital income. This gives incentives for taxpayers to start up closely held corporations and transfer income from the labor income to the capital income tax base.

[4]Notice that the income shifting that we document for the self-employed could in principle also be a real response to the tax differential between labor and capital income if it induces individuals to make a real re-allocation of labor effort towards activities that earn a higher rate of return on savings.

not only serves as an additional robustness check, but also allows us to analyze gender differences since the military enlistment data essentially only covers men. The results show a sharper ability gradient in tax responsiveness for men, a result which holds independently of whether we use grade-point-averages (GPA) or math grades. We also examine married couples and how tax responses depend on household ability. These results indicate that individuals belonging to households where both spouses have a high ability tend to have a higher tax responsiveness.

The main estimation method in the paper is bunching estimation, which 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.[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, and this is how we apply the bunching estimator in our paper.

As a complement to the bunching analysis we run individual-level regressions where we regress the probability that an individual has a taxable income at the kink point on a set of cognitive ability decile dummies. In this way, we estimate a regression-based ability gradient in tax responsiveness. The main benefit of these regressions is that they are less data-intensive and allow the straightforward inclusion of covariates. They also serve as a robustness check for the bunching results.

Our study contributes to the literature on behavioral responses to income taxation and the literature on the optimal design of tax systems. In terms of the empirical setting, our paper relates to 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.[6] We also contribute to the literature on how individuals understand and respond to tax rules (see Chetty et al. 2009, Abeler and Jäger 2015, Bhargava and Manoli 2015, Taubinsky and Rees-Jones 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.[7] 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

[5]The bunching literature began with Saez (2010), and has recently been surveyed by Kleven (2016).

[6]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.

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

observed taxpayer responses (see, for example, Chetty 2012, Kleven and Waseem 2013, Søgaard 2019, Kosonen and Matikka 2019, Mortenson and Whitten 2020, Gelber et al. 2020, Kostøl and Myhre 2020). There is also a related literature on the role of income shifting in the context of dual income tax systems (see, for example, Pirttilä and Selin 2011, le Maire and Schjerning 2013, Alstadsæter and Jacob 2017, Harju and Matikka 2016, as well as Tazhitdinova (2020) who analyze income shifting and business entry responses to taxes in the UK).

Finally, while the main purpose of our paper is to analyze the relationship between ability and tax responsiveness, the taxpayer ability heterogeneity conditional on income that we document also provides empirical evidence on the so-called mimicking issues emphasized in the optimal tax literature.[8] While these constraints have been the subject of hundreds (if not thousands) of papers in the literature, they have not been analyzed empirically. We can shed light on this by virtue of our 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).

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 main estimation strategy. Section 3 presents our data sources and the institutional setting. Section 4 describes our baseline bunching results. In section 5, we analyze the role of occupations, income shifting and hours of work. Section 6 presents a regression-based framework to analyze ability gradients. Section 7 presents robustness checks, estimating ability gradients using non-cognitive traits, physical ability, and high-school grades. Finally, section 8 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.[9] 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

[8] These mimicking issues are most transparent in the discrete adaptations of the Mirrlees (1971) model introduced by Stiglitz (1982) and Stern (1982) as well as Guesnerie and Seade (1982) who considered an arbitrary finite number of types.

[9] 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.

of taxable income z is guided by the following optimization problem:

$$\max_z \left\{ z - T(z) - v(z,\theta,\xi) \right\}, \tag{1}$$

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 $z \in [\hat{z}, \hat{z}+\Delta\hat{z}_\theta]$.[10] 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 $z \in [\hat{z}, \hat{z}+\Delta\hat{z}_\theta]$.[11] 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:

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

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).[12]

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

[10] 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.

[11] 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.

[12] 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.

example, among taxpayers with the same income, there will likely be some high-ability 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.[13] 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.[14]

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

[13] 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.

[14] 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. (2019) and Bertanha et al. (2019).

population, and for different years, we do not choose δ based on visual inspection (which 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.[15] 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 C2) 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.[16] In section 5, we complement our bunching analysis with regressions using individual-level data where 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.

3 Institutional setting and data

In Sweden, taxable income has been assessed individually since 1971. In addition, the taxation of capital income is separated from the taxation of labor income through the Swedish "dual" income tax system, where a lower proportional tax rate applies to capital income. The 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.

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

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

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. 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.[17]

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 years 1951-1975. For these cohorts, we retrieve measures of cognitive and physical ability, as well as non-cognitive traits. The measurement of cognitive ability in the military enlistment took place at around age 18. 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, or were incarcerated.

From the income tax register, we obtain data on individual taxable labor and capital income for wage earners and for the self-employed. 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). To obtain more detailed information on self-employed individuals, we match the income register to another population-wide register containing individual-level tax data linked to firm-specific information (FRIDA). We distinguish between two groups of self-employed individuals: (i) incorporated business owners (sometimes referred to as owners of closely held corporations), and, (ii) unincorporated business owners. The first group is defined as owner-managers of incorporated businesses who receive labor income from their firm (sometimes referred to as *active* owners of closely held corporations). The second group is defined as those who receive income either from a sole proprietorship or a partnership.[18] Notice that sole proprietorships and partnerships are taxed at the individual level, whereas closely held corporations are separate taxable entities. We define wage earners as those who are neither incorporated nor unincorporated business owners. Notice that wage earners, according to our definition, might own shares in closely held corporations (as well as publicly listed firms), as long as they do not receive labor income from these firms (they can, however, receive dividend income from these firms).

[17]The exact location of the kink point varied between 414,000 and 433,900 SEK (roughly 41.4–43.4 thousand EUR) during this period. There was hardly any inflation during this period.

[18]We define owners of closely held corporations through the variable `bkufoab`. Unincorporated business owners are defined using the income variables `nakte` and `nakthb`.

The assessment of cognitive ability was made through 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 used by the military enlistment and offers a measure of cognitive ability with higher numerical variability ($9 \times 4 = 36$ distinct numerical levels), 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), fitness (work capacity in a bicycling test, measured in watts), and strength (grip strength exerted by the strongest hand squeezing a dynamometer, measured in Newton). 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).

A limitation of the military enlistment data is that they essentially only cover men. Therefore, our paper also exploits high-school grade-point-averages (GPA) and final high-school math grades. These grades are obtained from the education register and are available for everyone born in 1966 and later. We use them as a proxy for cognitive ability, partially to serve as a robustness check to the main analysis, and partially as a way to shed light on the ability gradient for women.[20] The school grades are measured at around the same time in life as the military enlistment records and are positively correlated with

[19] See, for example, Lindqvist and Vestman (2011) and Edin et al. (2021) who analyze the relationship between cognitive ability and labor market outcomes.

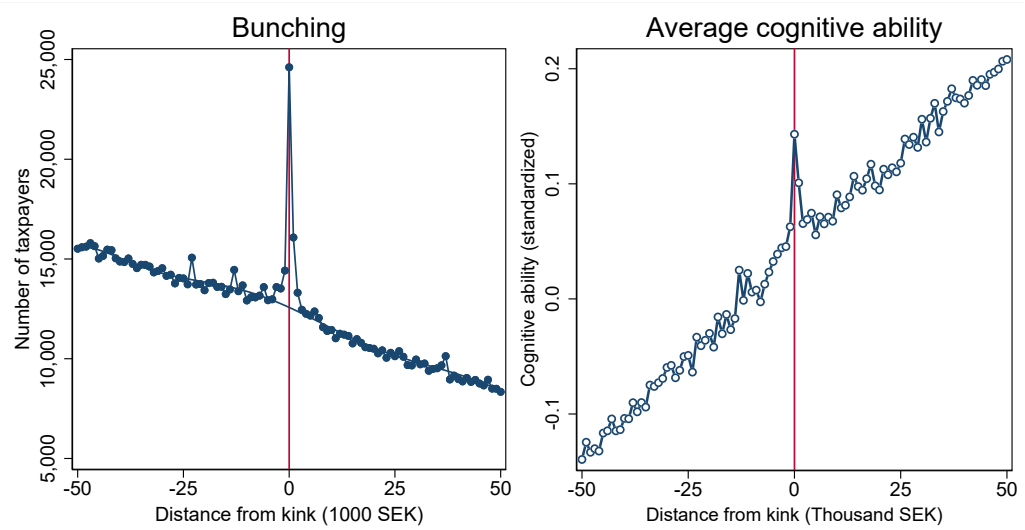
[20] The sample population for men in this analysis differs slightly from the population in our main analysis since we do not need to condition the sample on the availability of data from the military enlistment.

cognitive ability (a significant correlation coefficient of around 0.45 for both grade measures, a correlation matrix is provided in appendix figure A2). The main drawback is that they reflect individual education effort and are sensitive to the institutional details of the school system.[21] In similarity to cognitive ability, we divide men and women into GPA deciles using the fact that GPA is an almost continuous score between 1.00 (lowest grade) and 5.00 (highest grade). In addition to the data sources described above, we use occupational data and third-party reported contracted hours of work. These data sources are described in greater detail in sections 5.1 and 5.3, respectively. Descriptive statistics and sample attrition are presented in the appendix (tables A1 and A3).[22]

4 Main results

Figure 1 captures a fascinating stylized pattern. For 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, we observe statistically significant bunching in the income distribution and a spike in the average standardized cognitive ability at the kink point.[23]

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



Note: The left graph 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 graph 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.

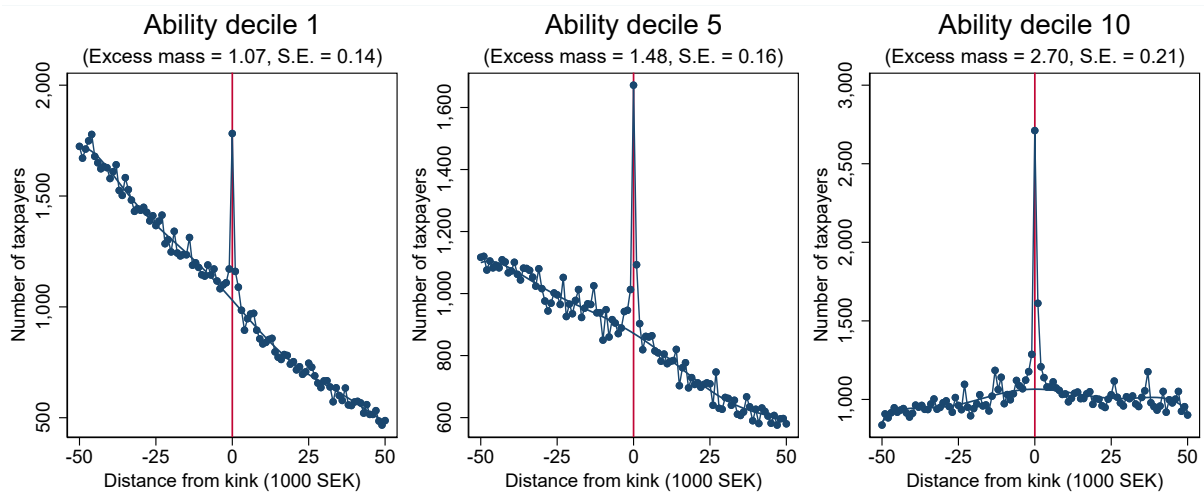
[21]School effort can be important both along the intensive margin (working harder to acquire better grades) and the extensive margin (whether to acquire a degree or not).

[22]Further details about our data sources can be found in: Krigsarkivet, Enlistment register (1996), Statistics Sweden, Income and tax register (2016), Statistics Sweden, Higher education register (2016), Statistics Sweden, Occupation register (2016), Statistics Sweden, Monthly wage register (2016).

[23]The estimated excess mass is 1.60 with a bootstrapped standard error of 0.10. 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 1999–2005 by Bastani and Selin (2014).

Turning to a more systematic examination of how tax responses differ across people with different levels of cognitive ability, we exploit the fine-grained register data and split 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 by comparing the decile-specific observed mass of income earners around the kink with the estimated decile-specific counterfactual density around the kink. Figure 2 shows bunching estimates for three of the ten ability deciles (appendix figure C1 shows all ten deciles).[24] 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.

Figure 2: Bunching across cognitive ability deciles



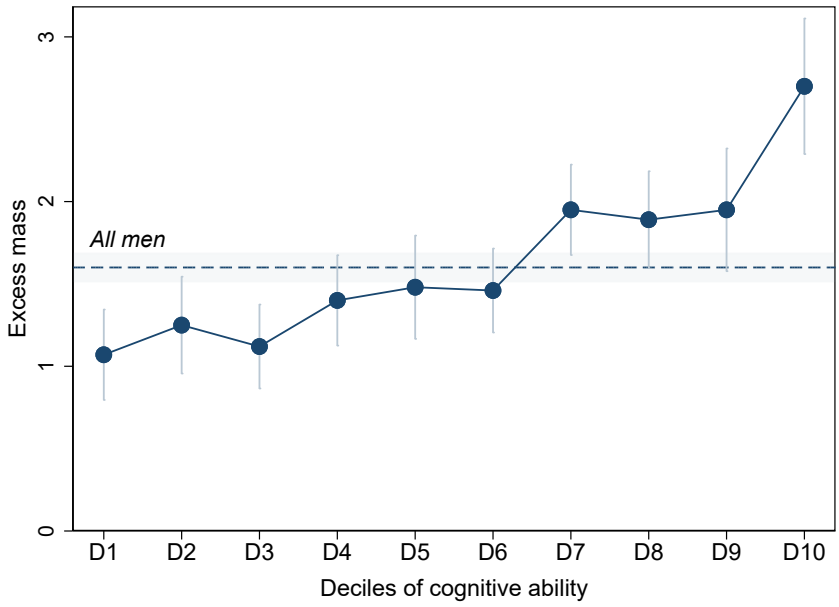
Note: The graph shows bunching at the kink for the period 2012-2016 (pooled data) for all adult men born 1951-1975 divided into deciles of cognitive ability.

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

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

bunching than the in the population at large, whereas all deciles from six and below have less bunching.

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 C1.

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).[25] Our results indicate that responses are higher for individuals with higher skill among people with the *same* taxable income.

In appendix figure C2, 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. Appendix figure C3 shows corresponding ability gradients for the four different sub-tests of cognitive ability.

[25]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. (2021). 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.

5 What drives the ability gradient in bunching?

The previous section established that bunching is larger among high-ability groups than among low-ability groups. In this section, we investigate what drives this ability gradient in tax responsiveness. For this purpose, we begin by analyzing the role of occupations, starting by dividing the population into wage earners and the self-employed, and then continuing with finer occupational categories. This analysis aims to shed light on the relationship between the ability gradient and occupational sorting as well as the potential role of occupation-specific optimization frictions for the observed taxpayer responses.[26] Second, we investigate the role of income shifting from the labor income to the capital income tax base within the context of the Swedish "dual" income tax system. In this analysis, we decompose the self-employed into owners of incorporated and unincorporated businesses, and relate the labor and capital income distributions. Third, we use auxiliary data on hours worked to analyze the role of labor supply in explaining the observed taxpayer responses. Fourth, and finally, we evaluate the relative importance of the different mechanisms through a "horse-race" regression analysis.

5.1 The role of occupations

The distinction between wage earners and self-employed is common in the bunching literature given that the self-employed have greater control over their taxable income. Figure 4 shows bunching around the kink point estimated separately for wage earners and self-employed individuals. In line with previous studies, we find more bunching among the self-employed.[27]

[26]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).

[27]It is noticeable that we also encounter significant bunching among wage earners, contrasting the previous findings of Bastani and Selin (2014). Here it should be noticed that we study the time period 2012–2016, whereas Bastani and Selin (2014) studied the time period 1999–2005. Our sample is also different as we focus on males with data from the military enlistment. In appendix section C.7, we show that bunching for our male sample has increased over time, suggesting that tax responsiveness might have increased in Sweden over the past two decades.

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

Figure 5 proceeds to estimate ability gradients separately for self-employed and wage earners. Like in our baseline analysis (figure 2), we present 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 the pattern is similar in the sense that the estimated excess mass is the highest in the top ability decile.

Figure 5: Bunching across ability deciles: Wage earners vs self-employed

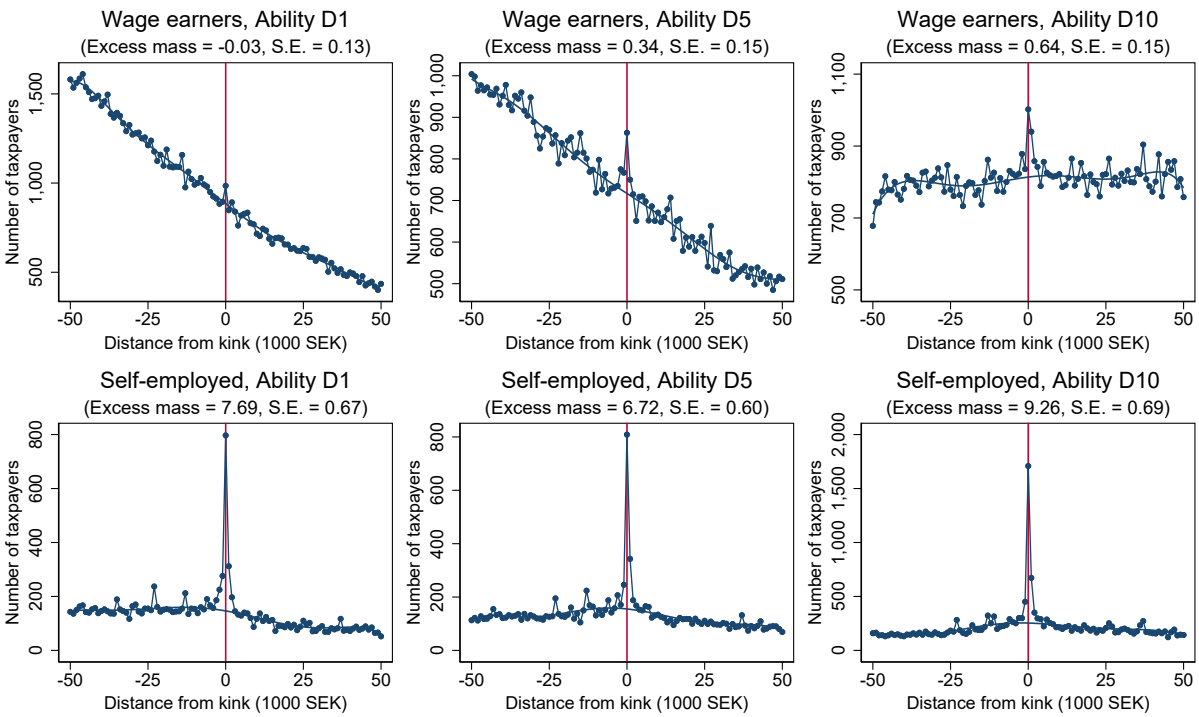


Figure 6 examines the ability gradient across all ten deciles, separately for wage earners and the self-employed, in a figure similar to figure 3 (which is reproduced in the first panel for ease of comparison) and shows that there is a positive ability gradient both among wage earners and the self-employed. 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.

One might 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 residual connection to self-employment activities. In appendix C.3 we show that there is an ability gradient even for wage earners with no history of self-employment activities.

Figure 6: Ability gradient in tax responsiveness: Wage earners vs self-employed

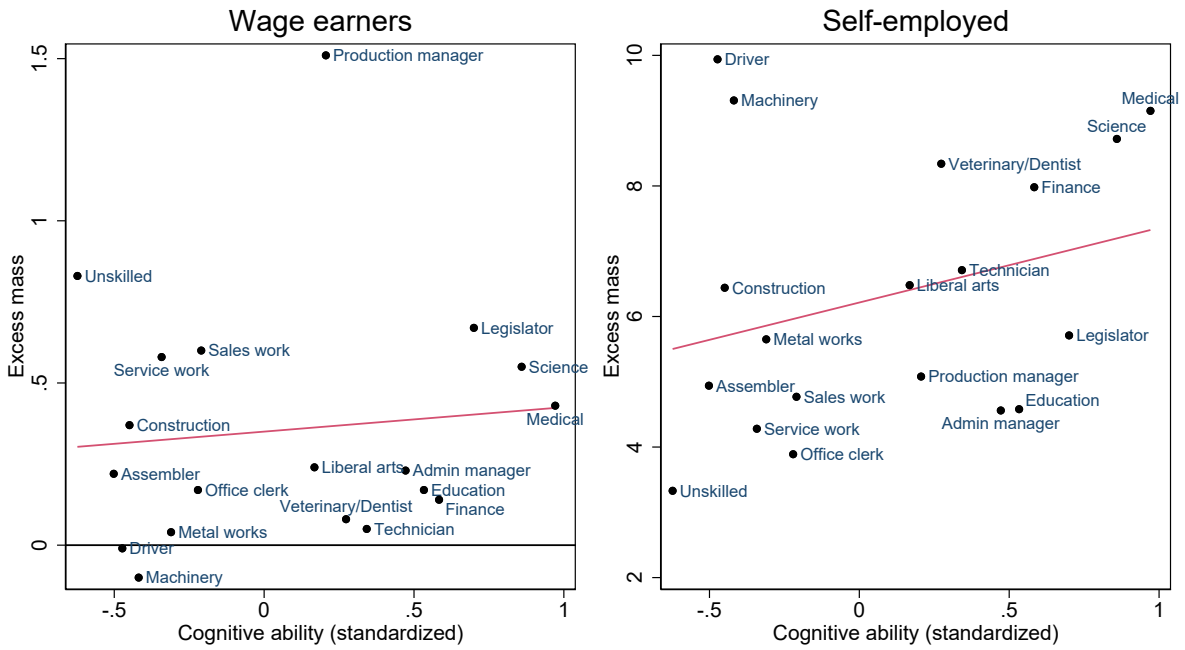
The fact that the self-employed bunch much more than wage earners (figure 4) suggests that one explanation for the ability gradient in tax responsiveness in the total population (figure 3) is that high-skill individuals select into self-employment which provides them with technologies to bunch at the kink (such as flexible working arrangements). We will later show a figure that supports this interpretation, showing that there is a positive selection of high-ability individuals into self-employment (see figure 8). However, as shown in figure 6, we find an ability gradient also *within* the groups of wage earners and the self-employed.

Figure 7 provides a further decomposition of wage earners and the self-employed into different occupation categories following the International Standard Classification of Occupations (ISCO) available in Swedish occupational data.[28] The figure reveals sub-

[28]Workers in the agricultural sector are excluded from this specific occupational level analysis.

stantial heterogeneity across occupations, both in terms of average ability, and in terms of the propensity to bunch at the kink. The relationship appears to be positive, suggesting an ability gradient in bunching at the occupational level. A limitation of the figure is that due to sample size considerations, we use ISCO classifications at a 2-digit level, which implies quite broad occupational categories. Notice that several occupations are represented in both panels, which naturally reflects that many occupations can be carried out either in the form of wage employment or in the form of self-employment. As can be seen, there are several interesting patterns, such as the high average ability and bunching among financial, medical and scientific occupations.

Figure 7: Bunching by occupation, ordered by average ability



Note: The figure includes occupations using the 2-digit level ISCO/08 classification for which there are at least 1,000 taxpayers within the baseline $[-50k, 50k]$ window around the kink point. Average ability is computed using all taxpayers in the whole population.

5.2 The role of income shifting

Sweden has a "dual" income tax system where the proportional tax rate on capital income historically has been lower than the marginal tax rate on high labor income. This provides incentives for individuals to reduce their labor income and increase their capital income in order to reduce their total tax burden, referred to as income shifting. Importantly, the incentive for income shifting changes discontinuously at the kink point (where there is a jump in the marginal capital income tax rate but not in the labor income tax rate). Thus, the observed bunching at the kink reflects a combination of traditional labor supply responses in response to the progressiveness of the labor income tax code, and efforts to reduce labor income and increase capital income in response to the sharp

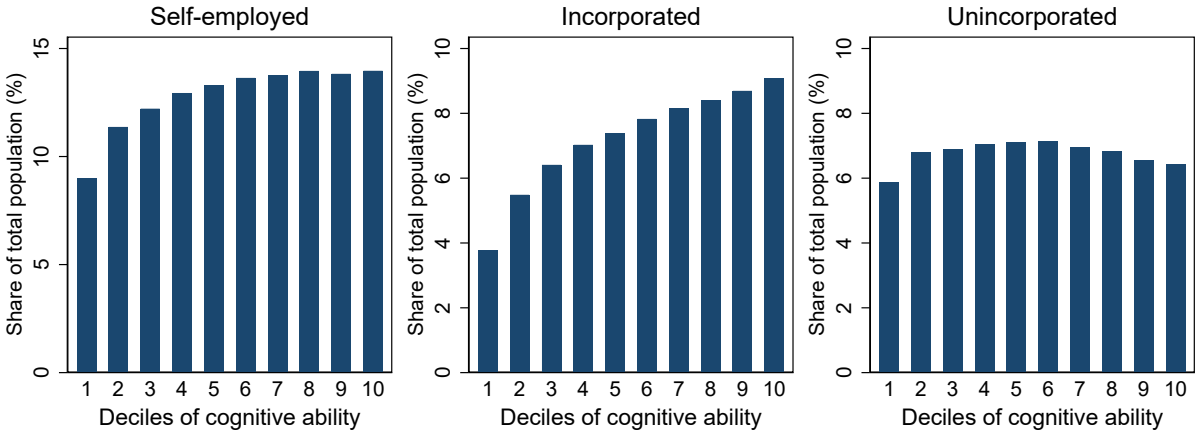
discontinuity in the tax differential between the two tax bases. We illustrate the sharp discontinuity in the marginal tax differential between labor income and capital income arising at the kink in appendix figure A3.[29]

Income shifting occurs through two channels, broadly speaking. The first channel is to substitute time in the regular labor 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). The other channel is to re-classify labor income as capital income in order to reduce the total tax burden, typically only available to self-employed individuals.

A particular aspect of the Swedish setting is that the tax system allows owners of closely held corporations to receive dividend income from their own firms taxed at the reduced rate of 20%, which is substantially lower than the standard Swedish capital income tax rate of 30%.[30] This low-tax dividend scheme makes it important to distinguish between *incorporated* and *unincorporated* business owners.

We begin by documenting the share of self-employed in the total population for different deciles of cognitive ability in figure 8. This figure serves to illustrate the overall self-employment rate in our sample, as well as the skill composition of the two sub-groups of self-employed. We see that there is an over-representation of high-ability individuals among the self-employed, driven by incorporated business owners.

Figure 8: Share of different types of self-employed by ability decile



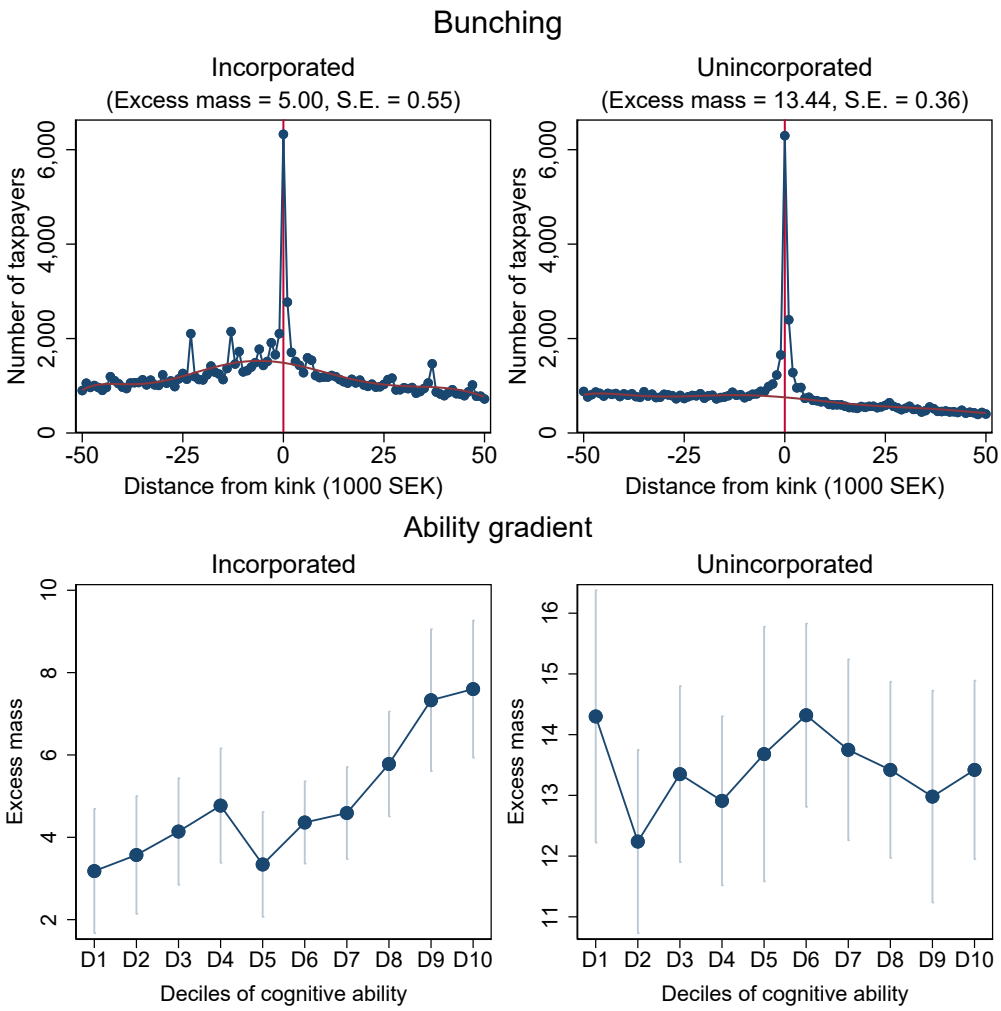
Note: Incorporated refers to self-employed active owners of closely held corporations and unincorporated refers to self-employed individuals who do not own an incorporated business but who receive taxable income from a sole proprietorship or a partnership.

[29]Income shifting in Sweden is regulated by the so-called 3:12–rules, which become much more lenient following a reform in 2006. This reform has been studied by Alstadsæter and Jacob (2016) who found increases in dividends in connection with the reform, mainly consisting of re-classified labor income.

[30]There are also possibilities for owners of unincorporated business owners to transfer income from the labor income to the capital income tax base (through what is referred to as "positive interest allocation"). However, the amount that is allowed to be transferred between the tax bases is typically smaller than for incorporated business owners, and the relevant capital income tax rate is 30% and not 20%.

Figure 9 proceeds to decompose the ability gradient analysis of the self-employed (presented above in figure 6) into incorporated and unincorporated business owners. The top panel reveals sharp bunching in the labor income tax for both incorporated and unincorporated business owners. The bottom panel shows a sharp ability gradient in this bunching for corporate owners, but no such gradient whatsoever among unincorporated business owners. This result is especially noteworthy given that incorporated business owners is a group of the population with high average ability, as documented by figure 8.

Figure 9: Incorporated vs. unincorporated business owners



Note: Incorporated refers to self-employed owners of closely held corporations and unincorporated refers to self-employed who do not own an incorporated business but receive income from a sole proprietorship or a partnership. See appendix figure C5 for bunching estimation graphs for a selection of deciles of cognitive ability.

Is the *bunching* in the top panel of figure 9 driven by income shifting? To investigate this, we examine the joint distribution of labor income and capital income by examining the propensity to have high dividends for individuals locating in a neighborhood of the kink point in the labor income tax. The top panel of figure 10 shows that those who bunch at the labor income tax are much more likely to have high dividends in compari-

son to individuals locating at adjacent income levels.[31] There is a large and clear spike in the propensity to have high dividends for both incorporated and unincorporated business owners. Thus, income shifting is an important driving factor behind the observed bunching. There is also a clear spike in dividends for wage earners. This can be evidence of wage earners responding to the tax kink by reducing their labor supply and increasing their investment effort resulting in higher capital income.[32] Is the *ability gradient* in the bottom panel of figure 9 driven by income shifting? The bottom panel of figure 10 provides affirmative evidence by showing a clear ability gradient in the excess propensity to have high dividends among owners of closely held corporations who bunch at the labor income tax kink, but not among wage earners and unincorporated business owners.[33] Is the ability gradient for incorporated business owners in the bottom panel of figure 9 driven by *low-tax* income shifting? In figure 10 we did not distinguish between the source or tax status of the observed dividend income. In appendix figure B3, we show that the ability gradient in dividends for owners of closely held corporations shown in the bottom panel of 10 is entirely driven by low-tax dividends.[34]

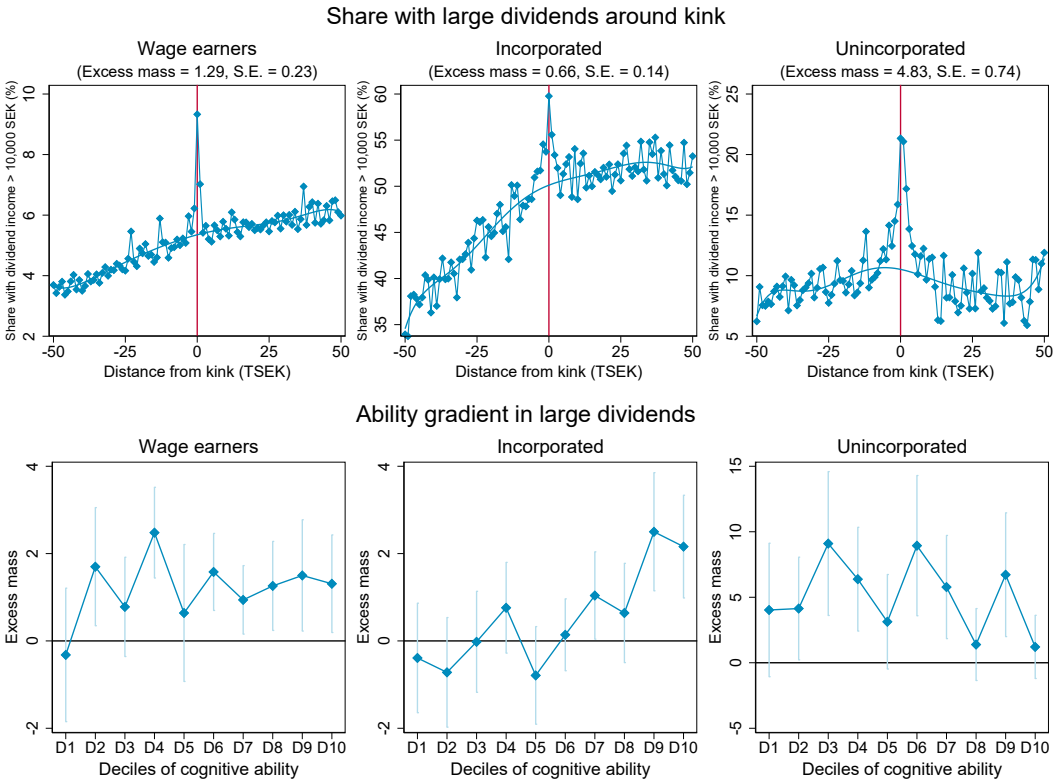
[31]We choose the threshold 10,000 SEK to focus on those who shift meaningful amounts and to avoid the noise that would be introduced by including individuals who receive small amounts of dividend income from e.g., listed corporations, which mainly impacts wage earners. We include small dividends in appendix figure B2.

[32]Notice that according to our definition of wage earners, they can be *passive* owners of closely held corporations, defined as those who own an incorporated business but who do not receive any labor income from their firm. Recall also our definition of incorporated business owners as individuals who receive labor income from the business they own. This definition allows incorporated business owners to receive labor income from other firms.

[33]In appendix figure B1 we analyze the ability gradient in excess dividends defined in terms of *total* dividend income. This analysis reveals a similar pattern as that in figure 10.

[34]We also present an analysis of total low-tax dividends in appendix figure B4.

Figure 10: Ability gradient in large dividends



Note: The top panel shows the excess propensity to have high (> 10,000 SEK) dividends for individuals who locate around the kink in the labor income tax schedule. The bottom panel shows the ability gradient in this excess propensity. The dividend variable is the sum of all types of dividend income (from listed and non-listed corporations). A corresponding figure where we include small dividends is provided in appendix figure B2. A visualization of the underlying bunching estimation is provided in appendix figure C9.

To sum up, we conclude that income shifting among high-ability corporate owners is a very likely explanation for the ability gradient in tax responsiveness among the self-employed that we document in this paper. We consider this to be an important result, not only since it suggests an ability gradient in tax avoidance, but also because it suggests that high-ability people are able to circumvent the sharply increasing marginal tax rates in the Swedish tax code by participating in income shifting through closely held corporations. Thus, achieving the desired degree of progressiveness (in terms of ability-based redistribution) of the overall tax system critically requires the appropriate use of capital income taxation. However, whether the marginal tax rate on capital income should be the same as on labor income (such as in comprehensive income tax systems), or be lower than the labor income tax rate, is an open question.[35] Given the frequently assumed higher mobility of capital income, a lower capital income tax rate is likely to be optimal. However, as we show in our paper, this implies ability-based income shifting.[36]

[35]In Bastani and Waldenström (2020), we provide an extensive overview of the literature on optimal capital taxation, shedding light on this issue.

[36]Selin and Simula (2020) analyze income shifting in an optimal tax framework and argue that the gov-

5.3 The role of labor supply

How do wage earners land on a taxable labor income exactly at the kink point? We match our data on annual taxable income with a smaller register-based data-set containing information on contracted hours. The purpose is to investigate whether the excess bunching we observe for high-ability wage earners is associated with part-time work, representing a real labor supply response to the 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.[37] The contracted hours of work is a variable expressed as a percentage of full-time, where full-time work (100 percent) corresponds to working 40 hours during a typical work-week.[38] 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, mainly due to the fact that over-time work is not observed in the private sector.

The top panel of figure 11 focuses on the total sample and shows average contracted hours (as a share of full-time) for workers with different income levels around the kink point. As can be seen, 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.[39] There also appears to be a slight general tendency for the mass to be lower in a wider interval around the kink. This strongly suggest that individuals move to the kink from the right of the kink in the taxable income distribution.

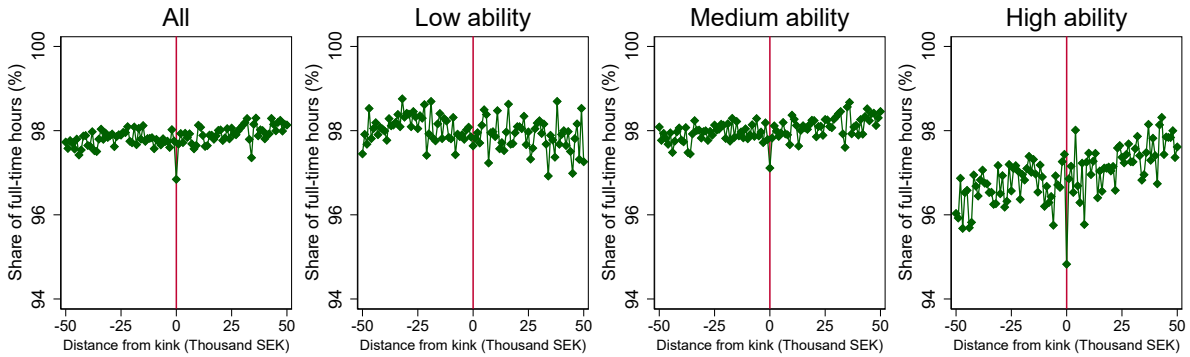
ernment should not necessarily combat income shifting when people who shift income have more elastic labor supply. Our empirical results suggest that this argument should take into account that high-elasticity individuals also are high-ability individuals.

[37] Around one-third 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. Only workers with an agreed upon wage and regular working hours are included, which implies that the vast majority of individuals in the sample are wage-earners.

[38] In 2018, around 20 percent of workers in the private sector were part-time employed.

[39] We have also checked whether the contracted wages, evaluated at full-time work, place individuals at the kink point, but we find no evidence of this.

Figure 11: Lower labor supply for high-ability people who bunch at the kink



Note: The graphs show average shares of full-time employment for different levels of annual taxable income around the kink point. Low = ability deciles 1-2, Medium = ability deciles 3-8, High = ability deciles 9-10.

From the bottom panel of figure 11 it can also be seen that the lower hours appear to be predominantly concentrated among individuals belonging to the group with high cognitive ability, providing strong evidence that labor supply adjustments play a role in explaining the ability gradient in tax responsiveness for wage earners. However, given the noisy nature of the data in figure 11 driven by the fact that the hours data only cover a subset of the population, we will use a regression approach to quantify the relationship between ability and hours adjustments at the kink point. This analysis is provided in the next section after we have introduced our regression framework.

6 Regression analysis

The bunching estimation uses aggregated (binned) data of individual income earners and a polynomial curve fitting technique to approximate counter-factual outcomes. In this section, we instead estimate ability gradients using individual-level regressions. In this analysis, the outcome variable is a dummy indicating whether an individual has an income at the kink point and we restrict the sample to individuals who locate in a neighborhood around the kink (and therefore have comparable income levels).

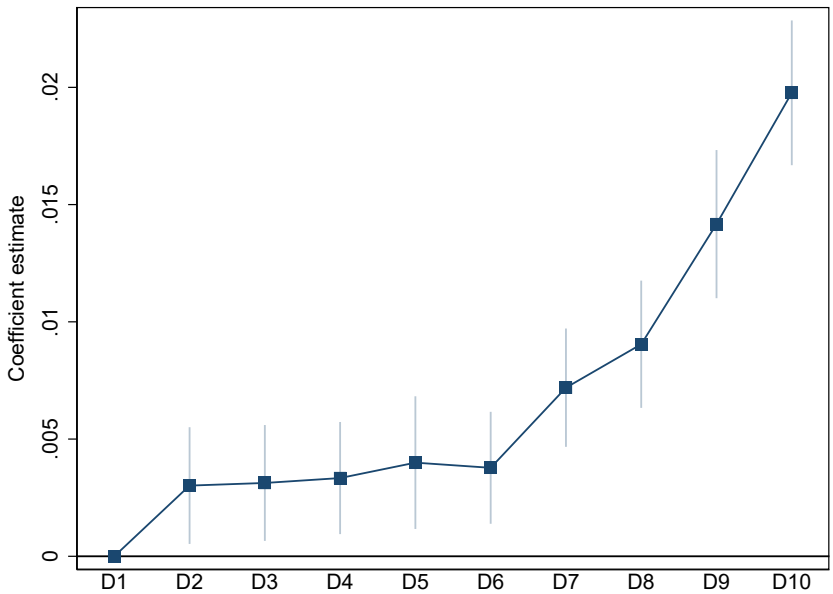
6.1 Regression-based ability gradient

Let $CogDec_i^d, d = 1,2,\ldots,10$ be *ability decile dummies* that take on the value 1 if individual i belongs to decile d in the distribution of cognitive ability. We estimate the effect of cognitive ability on the probability of locating in a small income interval $\mathscr{K}$ centered at the kink, by estimating the following equation:

$$\mathrm{I}\{y_{iat} \in \mathscr{K}\} = \eta_a + \lambda_t + \sum_d \beta^d \cdot CogDec_i^d + \delta' X_{iat} + \varepsilon_{iat}, \tag{18}$$

where $\mathrm{I}\{y_{iat} \in \mathscr{K}\}$ is an indicator variable taking on the value of 1 if an individual's income y_{iat} falls within the income interval $\mathscr{K}$, η_a represent cohort fixed effects, λ_t are year fixed effects, X_{iat} is a vector of potential covariates, and ε_{iat} are error terms.[40] Using a notation similar to that used in section 2.2, we let $\mathscr{K}$ be equal to a small window centered at $\hat{z}$, given by $[-\delta, \delta]$, while we restrict the overall estimation sample to the neighborhood of observations given by the interval $[-\phi, \phi] = [-25k, 25k]$. In our baseline analysis, we choose $\delta = 0.5k = 500SEK$. We interpret the coefficients β^d, $d = 1, 2, \ldots, 10$, as representing a regression-based ability gradient in tax responsiveness. The estimation results are shown in figure 12 and the regression-based gradient is consistent with the ability gradient found in the bunching estimation above (figure 3), with marginal effects very close to being monotonically increasing in ability.[41] Since cognitive ability is measured at age 18, and we study incomes realized around 20-50 years later (at ages 37-65), the estimated ability coefficients can be given a causal interpretation.

Figure 12: Regression-based ability gradient in bunching



Note: Estimated coefficients from the regression in (18).

In appendix D.2, we provide an additional analysis where we reduce the bin size to 100 SEK, zoom in on the kink, and estimate the marginal effect of cognitive ability on the probability of locating at the kink. We 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 main bunching results, with

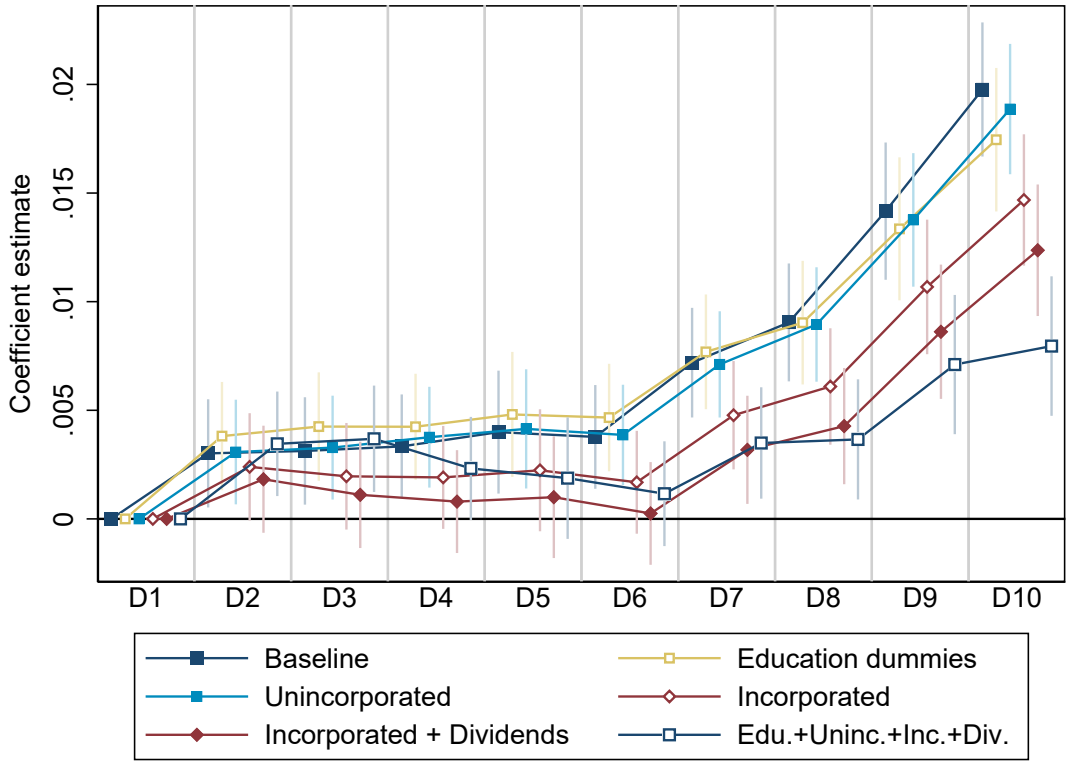
[40]The income measure is the same as in the bunching analysis, covering the period 2012-2016, and standard errors are clustered at the individual level. Notice that, just like in the bunching analysis, the regressions focus on a local part of the income distribution.

[41]In Appendix D.1 we perform a robustness test, repeating this analysis for two adjacent income levels, to the left, and to the right of the kink, thereby obtaining "placebo" gradients. We find no ability gradient for the two placebo-intervals.

a positive and statistically significant ability effect on the likelihood of sharp bunching at the kink point but no such effect on adjacent income levels. This points to the specific importance of taxpayer responses that place individuals exactly at the kink point.

Next, we employ a so-called horse-race regression where we include covariates in (12), one-by-one, and examine how the estimates $\hat{\beta}^d$ are affected.

Figure 13: Regression-based ability gradient in bunching: The role of controls



Note: Estimated coefficients from the regression in (18) with the sequential inclusion of controls. "Education dummies" controls for education length (in number of years) and field (3-digit ISCED 97 levels), "Unincorporated" and "Incorporated" are dummies for respective form of self-employment and "Dividends" denotes the receipt of positive dividends. Adding controls slightly re-shapes the sample. A figure with constant sample is provided in appendix D2.

The results in figure 13 show that the sharply increasing slope of the ability gradient in figure 12 from D6 and onward is robust to the inclusion of controls, and experiences the greatest drop when including controls for being an owner of a closely held corporation and having high dividends, in line with the income shifting story of section 5.2. In contrast, controlling for incorporated business ownership, or flexible education controls, has only a minor impact on the ability gradient.

6.2 Regression-based ability gradient in labor supply responses

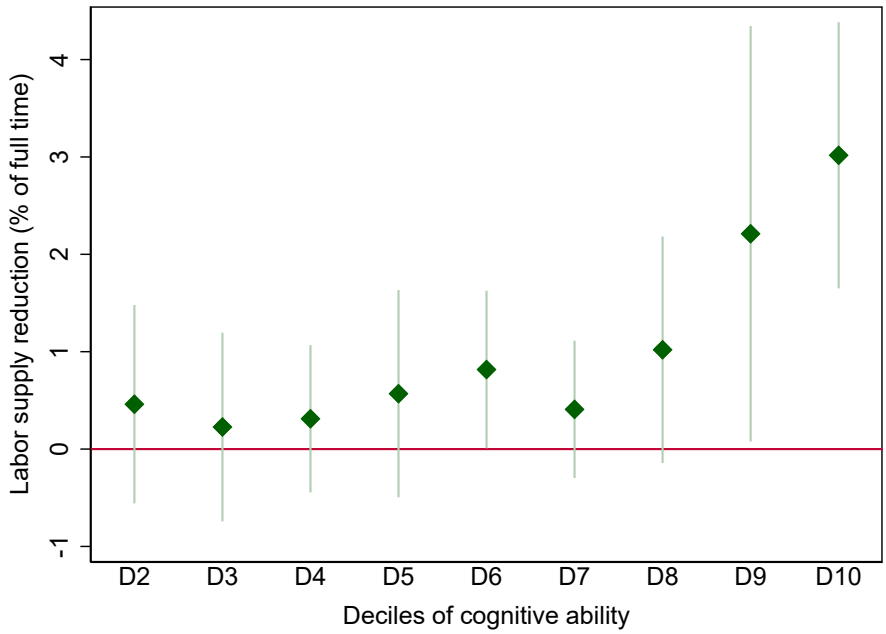
We now continue our analysis in section 5.3, turning to a regression analysis of the role of labor supply in explaining the ability gradient. For this purpose, we estimate the

following equation:

$$100 - hours_{iat} = \sum_d \rho^d \cdot \mathrm{I}\{y_{iat} \in \mathscr{K}\} \cdot CogDec_i^d + \eta_a + \lambda_t + \varepsilon_{iat}, \tag{19}$$

where we use a notation similar to that used in equation (18) of section 6.1. Here, $hours$ is the share of full-time hours in percentage terms (0-100) as in figure 11, and $\mathscr{K}$ is a small window centered at the kink $\hat{z}$, given by $[-0.5k, 0.5k]$. Notice that the outcome variable is 100 *minus* hours of work (measured as percent of full-time). The results in figure 14 show that there are statistically significant downwards labor adjustments at the kink pertaining to deciles 9 and 10 relative to the baseline (decile 1).

Figure 14: Ability gradient in labor supply responses ($100-hours$)



Note: The figure shows regression coefficients of ρ^d in equation 19.

In appendix section D.3, we zoom in on the kink, and provide an analysis of the conditional correlation between contracted hours and ability at different income levels in a narrow interval around the kink. The analysis shows a clear spike in this correlation at the kink, but no such spike at adjacent income levels, proving additional evidence on the role of labor supply adjustments that place high-ability individuals at the kink.

7 Robustness using other ability measures

In this section, we extend the analysis to test the sensitivity of our main results. First, we estimate ability gradients using non-cognitive traits and physical ability. Second, we estimate ability gradients for men and women using high-school school grades.

7.1 Ability gradients in non-cognitive traits and physical ability

The military enlistment register contains test scores for other ability dimensions: cognitive traits (assessed through an interview with a psychologist) and three physical abilities (height, fitness and strength). We estimate ability gradients using these other ability measures and present the results in figure 15.[42]

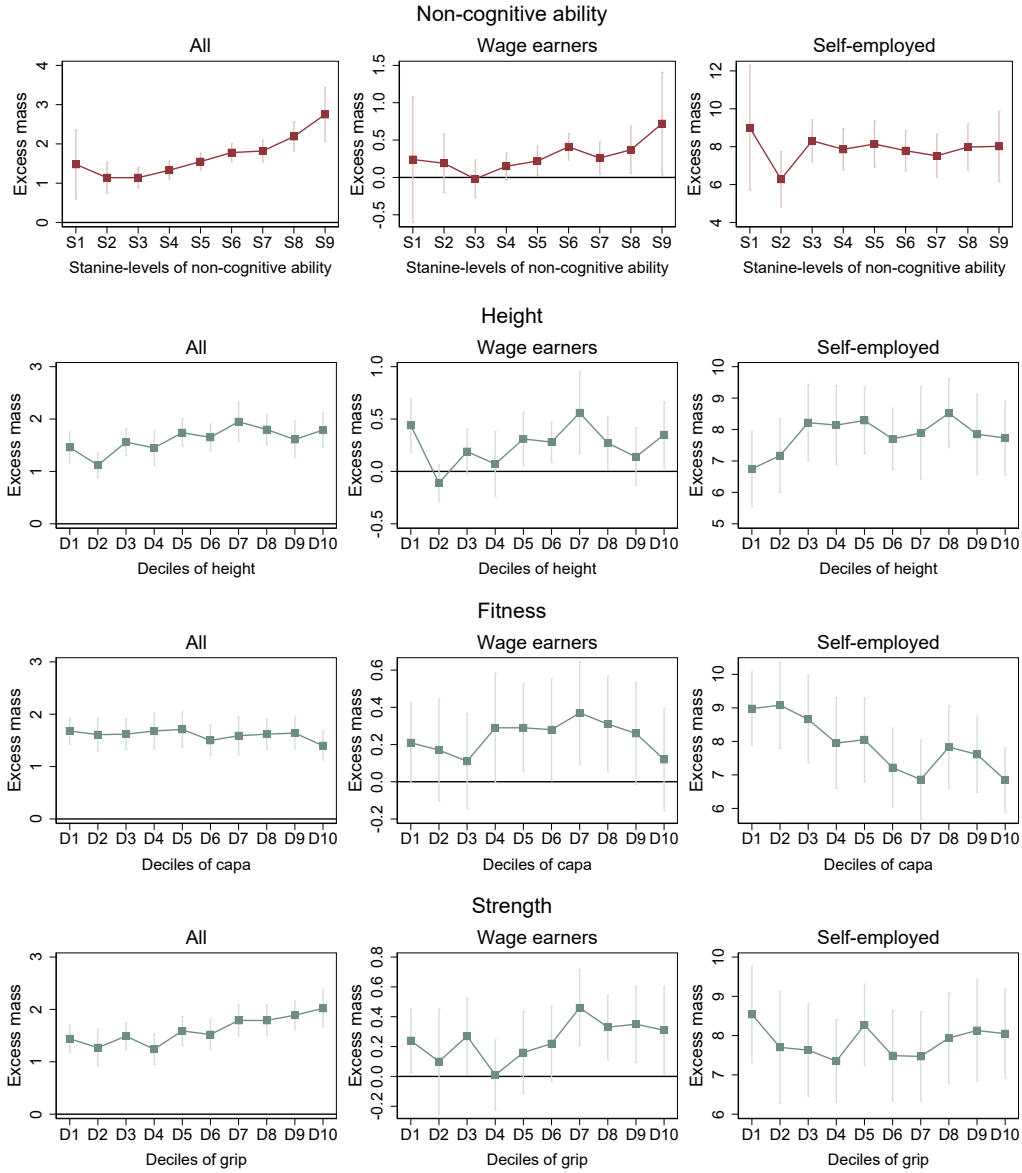
Let us begin with non-cognitive traits. The first thing to notice is that the ability gradient for this ability measure has a similar shape and level as the ability gradient for cognitive ability. This is perhaps not that surprising given that the cognitive and non-cognitive ability measures are highly correlated (see appendix figure A2). However, when decomposing the analysis into wage earners and self-employed, we can make an interesting observation. There is a positive gradient in non-cognitive ability for wage earners, but not for the self-employed. Thus, comparing these results with the results in figure 6, shows that cognitive and non-cognitive ability appear to have different effects on tax responsiveness for the self-employed. Since our analysis has revealed income shifting as the main driver of the ability gradient among the self-employed, this suggests that it is cognitive and not non-cognitive ability that is primarily associated with such responses.[43]

Turning to height, fitness and strength, figure 15 reveals that there is no clear ability gradient for any of the physical ability measures. The ability gradients are particularly flat (and sometimes even negative) for the self-employed. Thus, even though these measures are positively correlated with taxable income, they are not in any clear way correlated with tax responsiveness.

[42]A further division of the self-employed into incorporated and unincorporated business owners is presented in appendix figure C21.

[43]The fact that non-cognitive ability is more strongly associated with bunching among wage earners than among the self-employed could reflect that wage earners with high non-cognitive ability are better than wage earners with low non-cognitive ability in negotiating with their employer income levels that place them at the kink (perhaps in addition receiving some fringe benefits that we not observe in the data). In contrast, self-employed do not have to engage in social interactions to adjust their income.

Figure 15: Ability gradients using non-cognitive traits and physical abilities



Note: Analysis on incomes earned 2012-2016 by men born 1955-1975. Stanine levels show the population divided into fixed distribution shares with the following exact percentiles: S1 (P1-4, "the bottom 4%"), S2 (P5-11), S3 (P12-23), S4 (P24-40), S5 (P41-60, "the middle 20%"), S6 (P61-77), S7 (P78-89), S8 (P90-96), S9 (P97-100, "the top 4%"). See appendix figures C17–C20 for the underlying bunching estimation.

7.2 Ability gradients for men and women using high-school GPA

Is the ability gradient in tax responsiveness exclusive for the male population, or is it also present among women? How does bunching depend on the joint ability of spouses in married couples? Since the military enlistment data includes very few women, we here make an attempt to analyze gender differences in the ability gradient using high-school grade-point averages (GPA) as proxies for cognitive ability. The analysis in section can

also be viewed as an extra robustness check to the main analysis in the paper.[44]

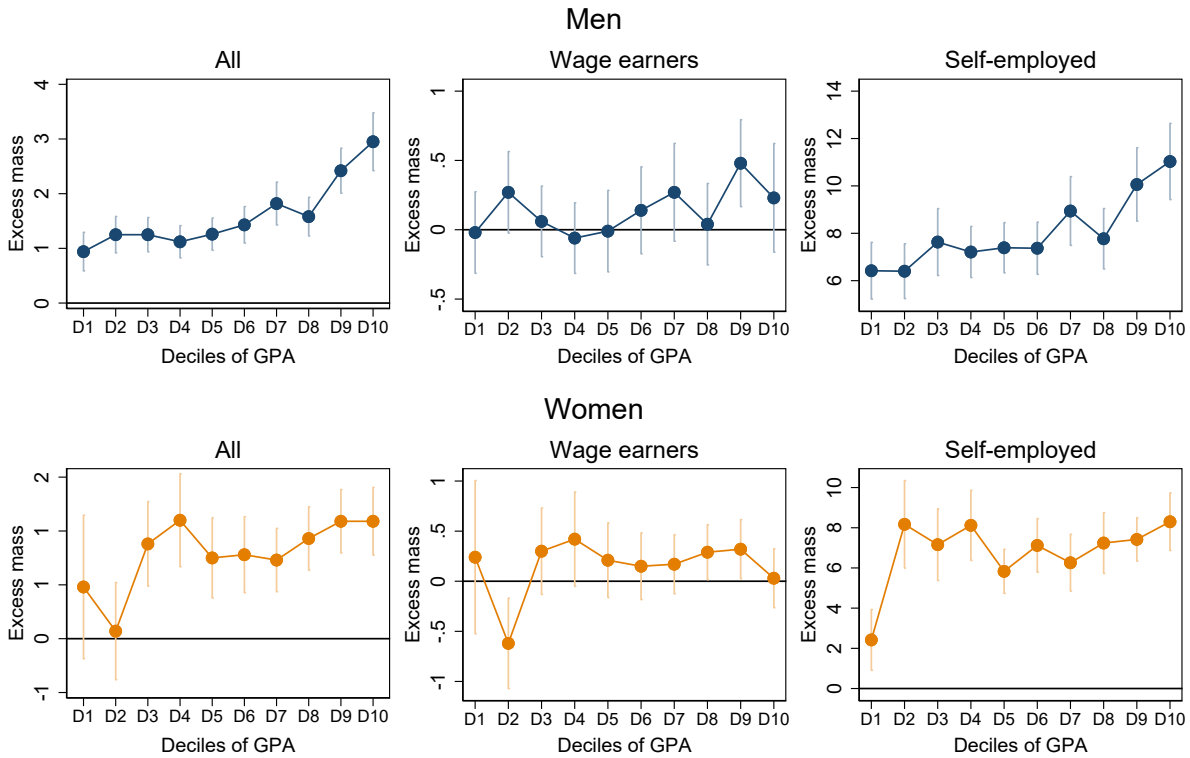
We begin by analyzing gender differences in the ability gradient in tax responsiveness focusing on bunching in taxable income during the same period as before, 2012–2016, and divide the male and female population into ten GPA deciles. We also present separate results for wage earners and the self-employed, in line with the distinction introduced in section 5.1, and between incorporated and unincorporated business owners, in line with the distinction introduced in section 5.2. Figure 16 shows the results.[45] The main finding is that the ability gradient is stronger among men than among women. As can be seen from the figure, this result is mainly driven by the self-employed.[46] One potential explanation is that fewer women select into self-employment where we know from our previous analysis that the ability gradient is stronger. However, as shown in the bottom-left panel of figure 16, the ability gradient for men appears to be stronger even when considering men and women who are owners of closely held corporations.

[44]GPA and cognitive ability are positively related with a correlation of around 0.45. In appendix section C.5, we use math grades as an alternative proxy for cognitive ability. One should keep in mind that school grades are imperfect proxies for ability as they also reflect school effort and aspects of the school system (see the discussion in section 3).

[45]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 A3 for more information about the difference between the sample studied in this section, and the sample studied in the main analysis.

[46]Among wage earners, there is hardly any difference in the ability gradient between men and women. In appendix figures C16 and D5 we provide an analysis along the lines of sections 5.3 and 6.2 that show that there is an ability gradient in labor supply reductions at the kink for both men and women.

Figure 16: Gender differences in the ability gradient using GPA

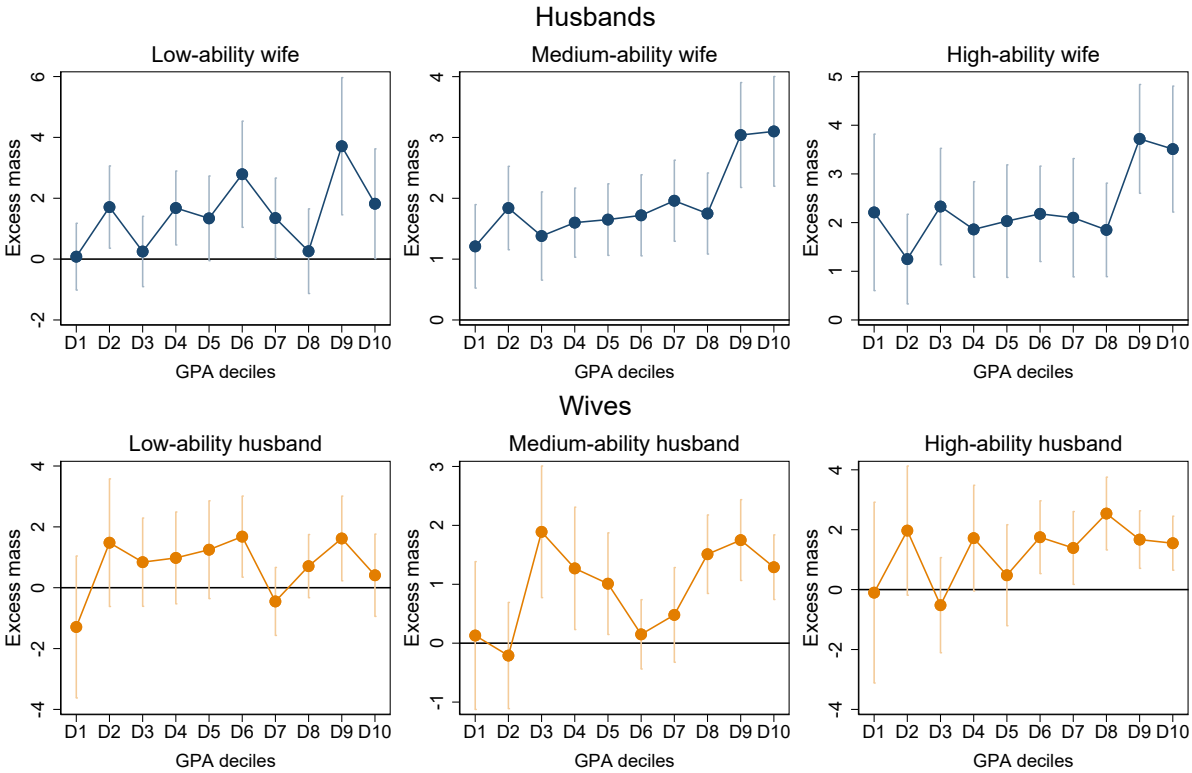


Note: Responses 2012-2016 in sample of men and women born 1955-1975. For underlying estimation, see appendix figures C10 and C11. For a decomposition of the self-employed into incorporated and unincorporated business owners, see appendix figure C13. For similar graphs using math grades, see appendix figure C14.

Even though the unit of taxation 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 for husbands and wives separately depending on whether they have a spouse that belong to one of three ability groups: bottom (ability deciles 1-2), middle (ability deciles 3-8), and top (ability deciles 9-10).[47] The results are shown in figure 17 and do not suggest that the ability of the spouse has any major influence on own bunching among husbands and wives. There is some evidence of an ability gradient for women with high-ability husbands and there appears to be some complementarity between the skills of spouses in producing bunching, but the results are imprecisely estimated.

[47]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. Also due to sample size considerations, we do not perform the corresponding analysis using math grades.

Figure 17: The role of household ability for the ability gradient using GPA



Note: "Low-ability" spouses belong to GPA deciles 1-2, "Medium-ability" spouses are in GPA deciles 3-8, and "High-ability" spouses are in GPA deciles 9-10. For underlying bunching graphs, see appendix figure C12.

8 Concluding remarks

We have shown that high-ability individuals react stronger to tax incentives than low-ability individuals by studying bunching at a large and salient kink point in the Swedish income tax schedule. Our results show that individuals in the top decile of the ability distribution react twice as strong to the discontinuity as compared to the average individual, and three times as strong as individuals in the bottom ability decile. These are remarkable findings, considering the fact that we link ability measured at age 18 with income records between 20 and 50 years later. This ability gradient in tax responsiveness is monotonic and highly robust to changes in the estimation strategy. It is present both among wage earners and the self-employed, but is stronger within the latter group, driven by owners of closely held corporations pursuing income-shifting opportunities.

The income shifting story is supported by two observations. First, high-ability incorporated business owners are much more likely to locate at the kink in the labor income tax schedule, with no such pattern for high-ability unincorporated business owners. Second, high-ability incorporated business owners who locate at the kink are much more likely to have high capital income, with no such pattern for high-ability unincorporated business owners. In other words, highly able self-employed individuals appear to take

advantage of the lower tax rate on capital income within the context of the Swedish dual income tax system by reclassifying labor income as capital income, thereby circumventing the intentions of the progressive labor income tax code.

The ability gradient among wage earners is not as sharp as for the self-employed, and quantitatively less important. However, using a special register-based data set on contracted hours of work, we have found evidence of labor-supply adjustments for high ability individuals who respond to to the sharp discontinuity in the income tax code. This provides evidence that the ability gradient that we observe for the total population, is driven by a combination of labor supply and income shifting.

The above results highlight the importance of non-acquired skills for taxpayer behavior, the prevalence of ability-based tax avoidance, and the need to consider ability heterogeneity conditional on income when designing the tax system or estimating behavioral responses to taxation. The results also shed light on the second-best nature of the problem of designing the income tax. Both the income shifting we observe among the self-employed and the labor-supply adjustments observed for wage earners, provide empirical evidence of high-skill individuals mimicking low-skill individuals in order to escape progressive income taxation. These mechanisms have attracted much attention in the optimal tax literature, but have not been tested empirically in a direct way using a kinked budget constraint and data on ability.

The relationship between ability and tax responsiveness highlights a conflict in 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 progressive labor income taxation to achieve skill-based redistribution. This result is especially important if reduced labor effort by high-ability people is more costly to society than a reduction in effort by low-ability people, which is the case when high-skill effort makes low-skill effort more productive through complementarities in production (Stiglitz 1982).

From a tax policy perspective, our findings highlight the crucial need to complement taxes on labor income with other taxes and policies in order to achieve the desired distributional objectives. In the context of the Swedish dual income tax system, a higher tax rate on capital income would be one way to enhance the extent of ability-based redistribution (Christiansen and Tuomala 2008). Another possibility would be to improve enforcement (Slemrod 1994). We leave the interesting question of how policy optimally should respond to the ability-gradient in tax responsiveness that we have uncovered 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*.

Alstadsæter, A. and Jacob, M. (2017). Who participates in tax avoidance? Evidence from Swedish microdata. *Applied Economics*, 49(28):2779–2796.

Alstadsæter, A. and Jacob, M. (2016). Dividend taxes and income shifting. *The Scandinavian Journal of Economics*, 118(4):693–717.

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. (2020). Ethnicity and tax filing behavior. *Journal of Urban Economics*, 116:103215.

Bastani, S., Moberg, Y., and Selin, H. (2021). The anatomy of the extensive margin labor supply response. *Scandinavian Journal of Economics*, 123(1):33–59.

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.

Bastani, S. and Waldenström, D. (2020). How should capital be taxed? *Journal of Economic Surveys*, 34(4):812–846.

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. (2019). 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.

Christiansen, V. and Tuomala, M. (2008). On taxing capital income with income shifting. *International Tax and Public Finance*, 15:527–545.

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. (2021). The rising return to non-cognitive skill. *American Economic Journal: Applied Economics (forthcoming)*.

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*.

Gordon, R. H. and Slemrod, J. (2000). Are "real" responses to taxes simply income shifting between corporate and personal tax bases? In Slemrod, J., editor, *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*. Harvard University Press.

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*.

Kostøl, A. R. and Myhre, A. S. (2020). Labor supply responses to learning the tax and benefit schedule. *IZA DP No. 13900*.

Krigsarkivet, Enlistment register (1968-1996). Military archives, enlistment register, https://www.riksarkivet.se/kontakter-pa-krigsarkivet (Accessed March 21, 2021).

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.

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.

Selin, H. and Simula, L. (2020). Income shifting as income creation? *Journal of Public Economics*, 182:104081.

Søgaard, J. E. (2019). Labor supply and optimization frictions: Evidence from the Danish student labor market. *Journal of Public Economics*, 173:125 – 138.

Slemrod, J. (1994). Fixing the leak in Okun's bucket optimal tax progressivity when avoidance can be controlled. *Journal of Public Economics*, 55(1):41–51.

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

Statistics Sweden, Higher education register (1975-2016). https://www.scb.se/vara-tjanster/bestalla-mikrodata/vilka-mikrodata-finns/individregister/registret-over-befolkningens-utbildning/ (Accessed March 21, 2021).

Statistics Sweden, Income and tax register (2012-2016). https://www.scb.se/vara-tjanster/bestalla-mikrodata/vilka-mikrodata-finns/individregister/inkomst--och-taxeringsregistret-iot/ (Accessed March 21, 2021).

Statistics Sweden, Monthly wage register (2012-2016). https://www.scb.se/vara-tjanster/bestalla-mikrodata/vilka-mikrodata-finns/individregister/lonestrukturstatistik/ (Accessed March 21, 2021).

Statistics Sweden, Occupation register (2012-2016). https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-yrkesklassificering-ssyk/ (Accessed March 21, 2021).

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 For online publication: Descriptive statistics

Table A1: Descriptive and distributional statistics

	Mean	S.D.	Min	P25	P50	P90	P99	Max
Cog. ability (z-score)	0	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)	41.7	33.8	0	28.8	37.6	66.4	139	11,001
Dividend income (1000 EUR)	2.82	56.5	0	0	.002	1.48	50.2	33,507
Labor income (z-score)	0	1	-1.23	-.382	-.123	.729	2.89	324
Dividend income (z-score)	0	1	-.0499	-.0499	-.0499	-.0238	.838	593
GPA (z-score)	0	0.993	-4.85	-0.661	-0.0129	1.32	2.36	2.87
Math grade (z-score)	0	1	-3.08	-0.69	-0.212	1.70	1.70	1.70
Non-cog traits (z-score)	0	1	-2.25	-0.625	-0.0539	1.09	2.23	2.23
Height (cm)	179	6.53	138	175	179	187	195	225
Height (z-score)	0	1	-6.29	-.621	-.00863	1.22	2.44	7.03
Fitness (z-score)	0	1	-2.38	-.201	.165	.978	1.61	7.52
Strength (z-score)	0	1	-5.75	-.59	-.0277	1.19	2.41	3.62
Wage earner (%)	87.3							
Self-employed (%)	12.7							
Incorporated (%)	7.1							
Unincorporated (%)	5.6							

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/self-employment status are averaged over the 2012-2016 period.

Table A2: Correlation matrix

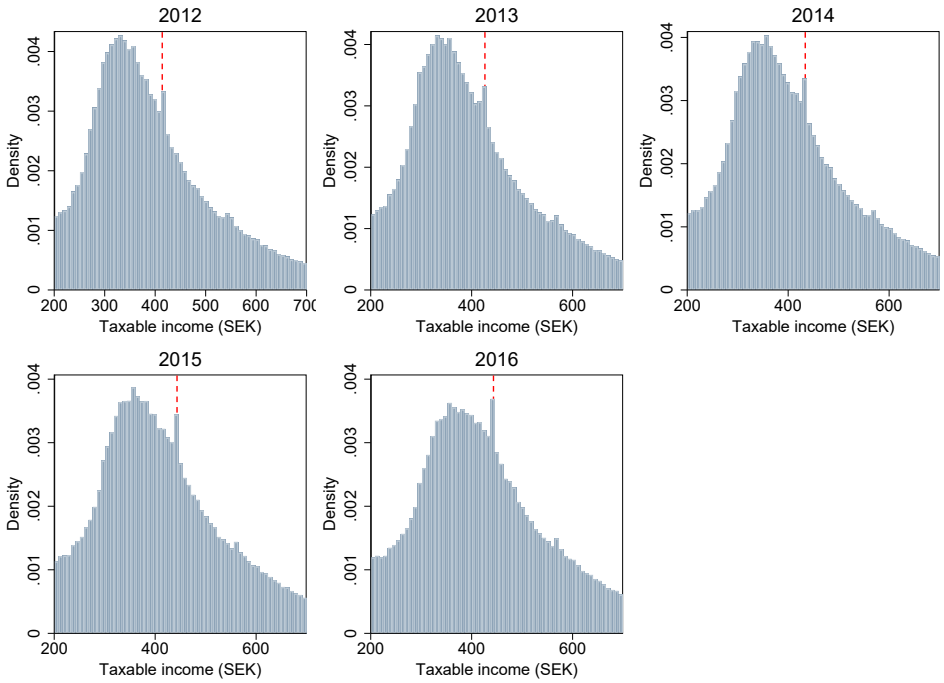
	Labor inc.	Cog	GPA	Math	Non-cog	Height	Fitness	Strength
Labor income	1							
Cog	0.233	1						
GPA	0.203	0.443	1					
Math	0.151	0.445	0.769	1				
Non-cog	0.229	0.382	0.256	0.139	1			
Height	0.085	0.155	0.076	0.061	0.124	1		
Fitness	0.129	0.194	0.177	0.104	0.407	0.175	1	
Strength	0.044	0.051	0.035	-0.021	0.215	0.240	0.253	1

Note: Number of observations is 1.2 million for all entries except the correlations with "Math" which are based on 745,000 observations. Labor income is averaged over the 2012-2016 period. All correlations are statistically significant.

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

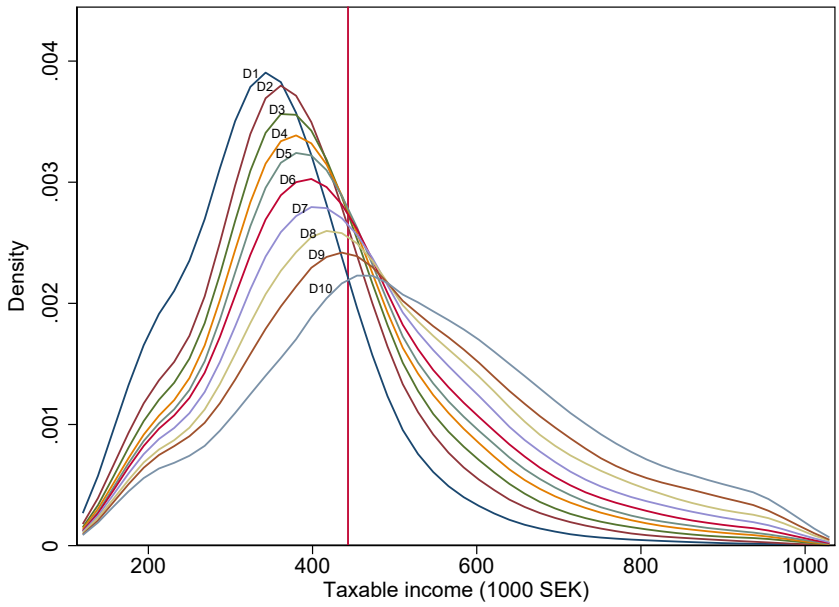
Category	Observations
Main sample: Men born 1951-1975	
(1) Men born 1951-1975 with cognitive ability score in military enlistment	1,283,254
(2) In (1) born 1955- (cohorts with GPA data coverage)	1,075,500
(3) In (2) and observed high-school GPA	786,032
(4) In (3) born 1966- (cohorts with math data coverage)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

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



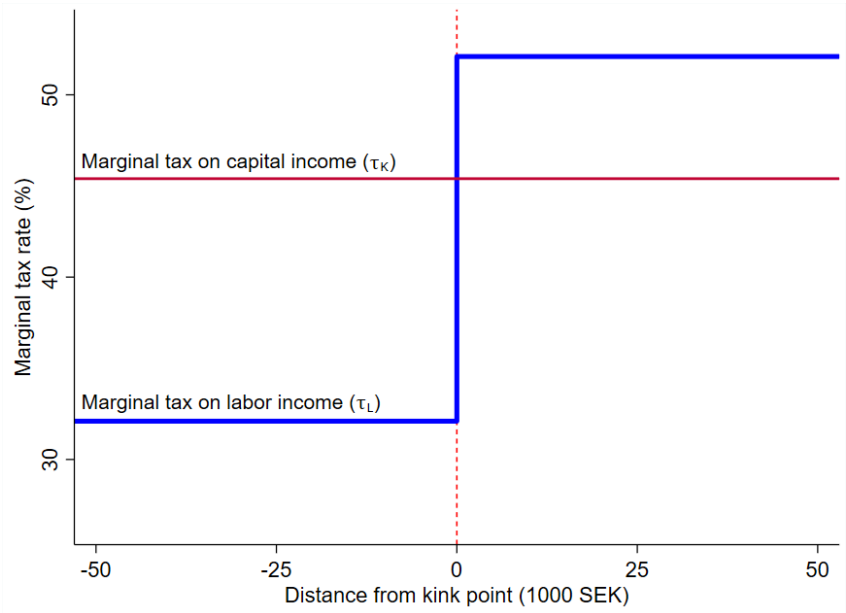
Note: Annual taxable labor income distributions around the kink point. Male population, born 1951-1975. The smaller spike to the right of the kink is associated with round-number bunching (exactly 50,000 SEK per month).

Figure A2: Kernel density estimation of taxable income distribution 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

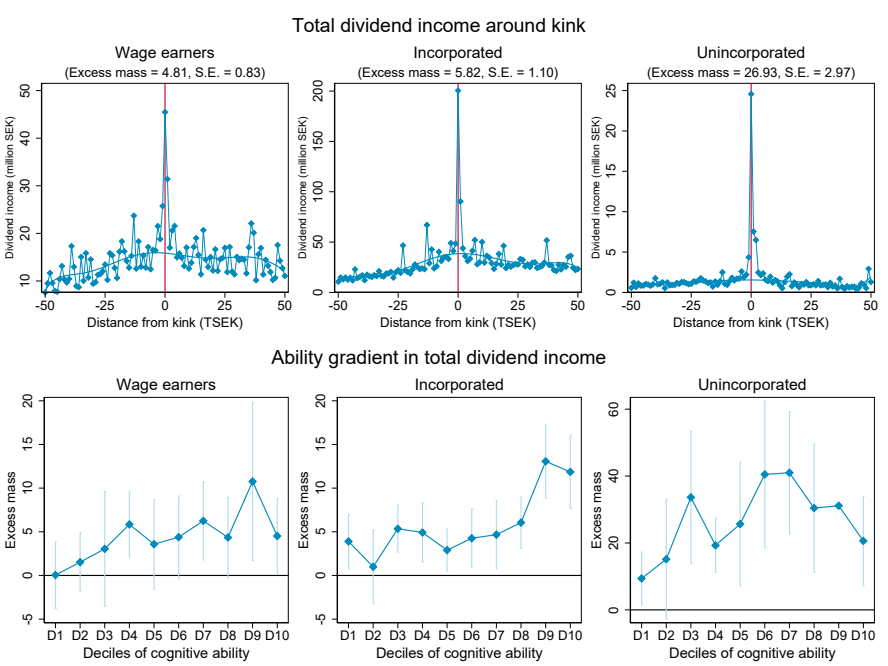


Note: Illustration using the tax system in 2016. Labor income marginal tax rate below kink point equals the average municipal income tax rate (32.1%) and to the right it equals this average municipal income tax rate plus the central government tax rate on labor income, 20%. No adjustment is made for 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%). The exact location of the kink point varied between 414,000 and 433,900 SEK (roughly 41.4–43.4 thousand EUR) during this period.

B For online publication: Low-tax dividends among bunchers

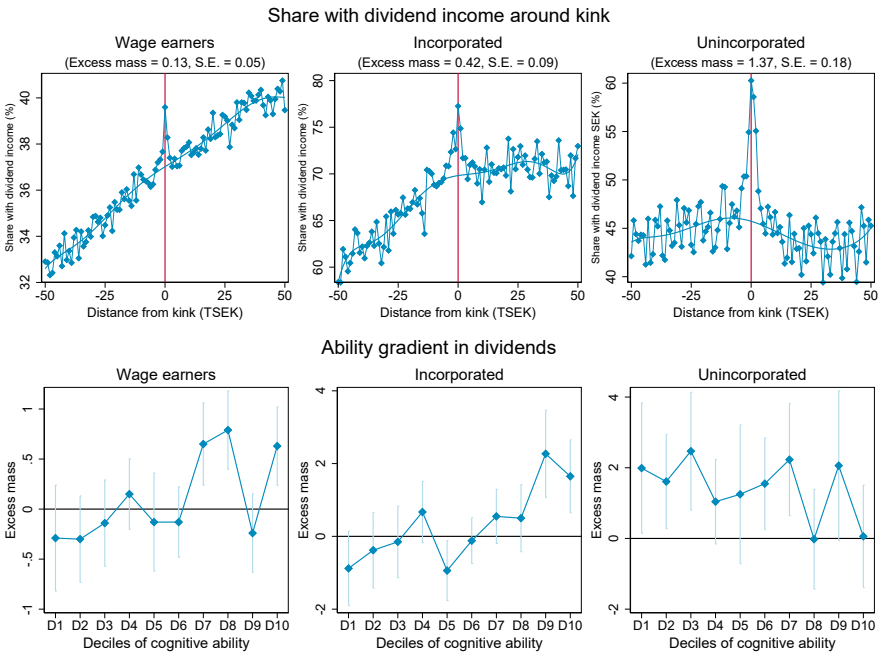
We first present two additional graphs that complement figure 10. Figure B1 produces a similar graph where the outcome variable is total dividends. Figure B2 produces a similar graph including also small dividends ($< 10,000$ SEK).

Figure B1: Ability gradient in total dividend income



Note: The top panel shows bunching in total dividend income among those who bunch at the kink in the labor income tax. The bottom three graphs show the ability gradient in the bunching in total dividends. The dividend variable is the sum of all types of dividend income (from listed and non-listed corporations).

Figure B2: Ability gradient in dividends (including small dividends)

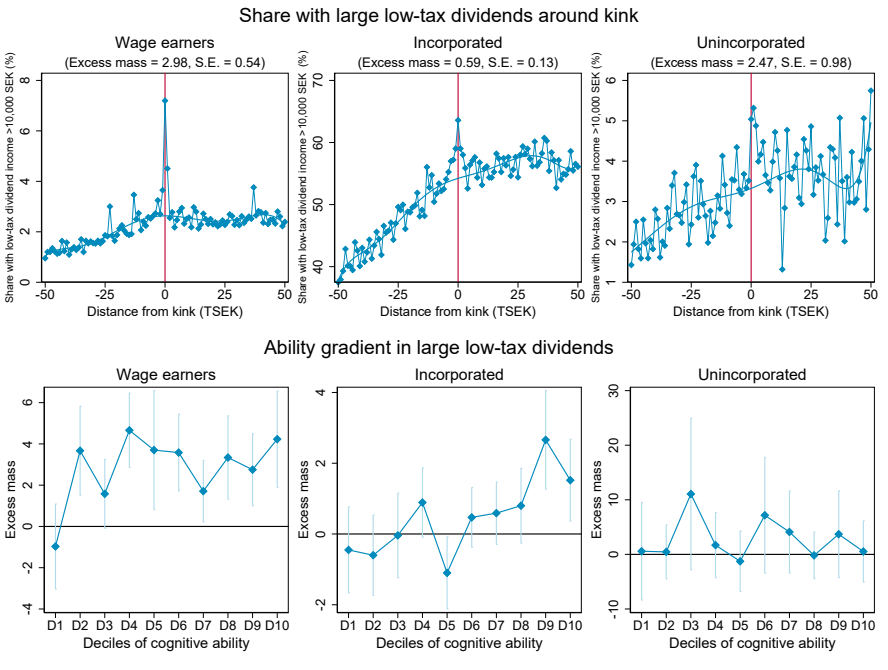


Note: The top panel shows the excess propensity to have dividend income for individuals who locate around the kink in the labor income tax schedule. The bottom three graphs show the ability gradient in this excess propensity. The dividend variable is the sum of all types of dividend income (from listed and non-listed corporations).

Next, we present an analysis similar to the one in figure 10, but focusing only on low-tax dividends. The standard marginal tax rate on capital income in Sweden is 30%. However, as discussed in section 5.2, owners of closely held corporations (and their close relatives) have opportunities to receive dividends taxed at a reduced tax rate of 20%. Notice that not only *active* owners of closely held corporations, but also *passive* owners of closely held corporations (who do not receive any labor income from the firm, and hence according to our definition can be either wage earner or unincorporated business owners) can receive dividends from closely held corporations. Figure B3 shows that the ability gradient in dividends for incorporated business owners documented in figure 10 is entirely driven by low-tax dividends. A similar picture emerges when focusing on total low-tax dividends, as shown in figure B4.[48]

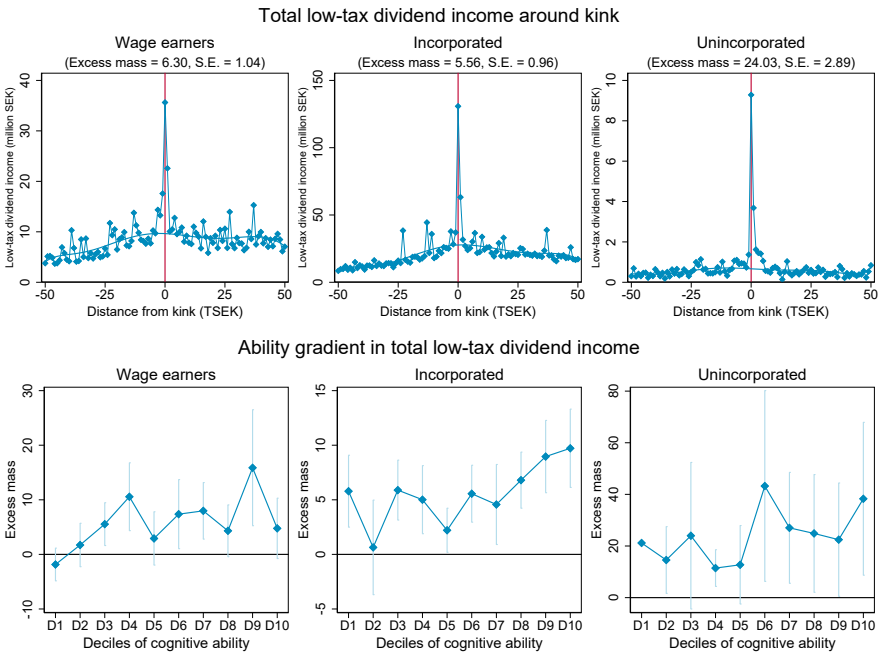
[48]The outcome variable in figures B3 and B4 is defined on the basis of the sum of box 1.15 and box 2.16 in the K10 form used by owners of closely held corporations when filing their income taxes. In Sweden, capital income is third-party reported by firms, banks and financial institutions. Dividends from closely held corporations are not only reported by the firm (through the KU31 form) but also by the individual taxpayer, through the K10 form. The tax authority requires the K10 form to determine eligibility for the lower tax rate.

Figure B3: Ability gradient, large (low-tax) dividends, owners of closely held corporations



Note: The top panel shows the excess propensity to have large (> 10,000 SEK) low-tax dividends among those who bunch at the kink in the labor income tax. The bottom panel shows the ability gradient in this excess propensity. The dividend variable is based on boxes 1.15 and 2.16 in the K10 tax-return for owners of closely held corporations (both passive and active).

Figure B4: Ability gradient, total (low-tax) dividend income, owners of closely held corporations



Note: The top panel shows excess propensity in total low-tax dividend income among those who bunch at the kink in the labor income tax. The bottom panels show the ability gradient in this excess propensity. The dividend variable is based on boxes 1.15 and 2.16 in the K10 tax-return for owners of closely held corporations (both passive and active).

The amount of dividends that can be received at the low tax rate by owners of closely held corporations is restricted by the *dividend allowance*. Below, we describe how this dividend allowance is calculated.

There are two options available to owners of closely held corporations who wish to receive low-tax dividends from their business: the *simplification rule* and the *general rule*. The simplification rule enables every firm to pay out a total fixed amount of lightly taxed dividend income (taxed at 20% instead of 30%) to its owners each year. In 2020, this fixed amount was equal to 2.75 × the income price base amount for 2019 = 177,100 SEK. This amount is shared by all owners of the closely held corporation, and a person who is listed as an owner of multiple companies, can only use the rule to receive dividends from one of these firms.

The general rule is targeted towards owners of larger closely held corporations and provides opportunities for owners to receive large amounts of leniently taxed dividend income provided the acquisition cost of the shares or the labor costs in the firm are sufficiently large.[49] The general rule implies that the dividend allowance is equal to the weighted sum of the acquisition cost of the shares and the total salaries paid out by the

[49]The purpose of the rule is to lower the effective tax rate on entrepreneurs while taking measures to prevent income shifting through firms that are established with the sole purpose of tax avoidance (since such firms would neither have many employees nor have a high valuation).

company the previous year plus any unused dividend allowances from the previous year (in 2020, the calulation is based on 9% of the acquision cost of the shares plus 50% of total salaries). Importantly, the salary component of the dividend allowance can only be used provided the tax-filing owner had a sufficiently high salary from the firm the previous year, known as the *salary requirement*.[50]

The salary requirement states that an owner must have, in the year previous to the current tax declaration, a labor income from the firm equal to the minimum of two values. The first value is 9.6 income price base amounts (618,240 SEK in 2019 when filing taxes for 2020) and the second value is 6 income price base amounts (386,400 SEK in 2019 when filing taxes for 2020) + 5% of the total gross salaries of the firm. Since the first value exceeds the location of the first central government kink point, the main rule is only useful as part of an income shifting strategy in the context of the first central government kink point if the total salaries of the firm fall below a threshold w^{Σ} given by the solution to the equation $6\text{IPBA} + 0.05 \times w^{\Sigma} = \hat{z}$ where $\hat{z}$ is the location of the kink point.[51] Using values from the 2020 tax declaration, we have $\hat{z} = 523,200$ and obtain $w^{\Sigma} = 2,458,000$, which corresponds roughly to having 4.7 employees paid at the kink point income level. An owner of a firm with $w^{\Sigma} = 2,458,000$ will have a salary-based component of the dividend allowance equal to $0.5 \times 2,458,000 \approx 1{,}2$ million SEK provided the owner receiving the dividends has a salary of at least $\hat{z} = 523,200$.[52] Thus, for owners of closely held corporations that employs a sizable number of people, and where the owner takes a sufficiently large salary from the firm, the dividend allowance is much larger as compared to when using the simplification rule.

[50] In addition, the ownership share must be at least 4%.

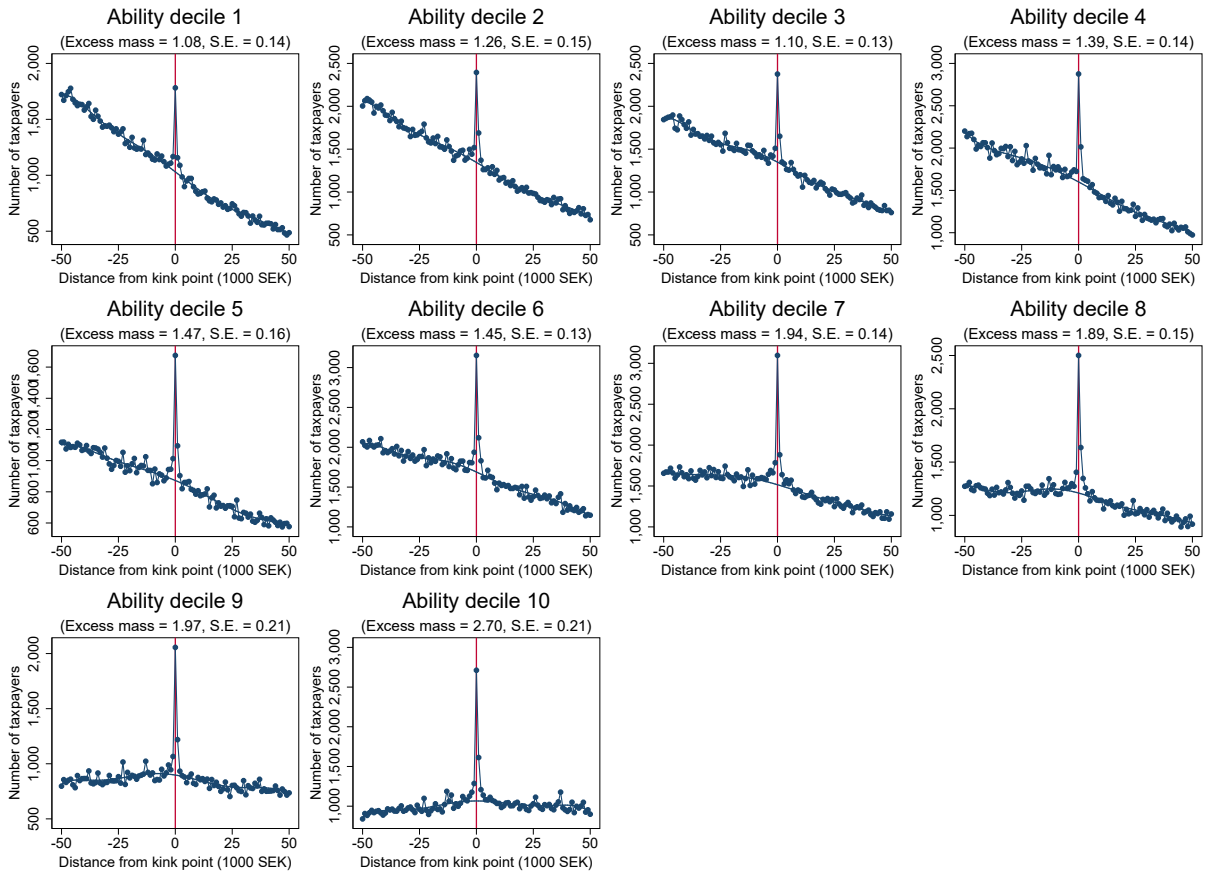
[51] For owners of medium-sized firms with labor costs that are likely to have exceeded $\frac{9.6\text{IPBA}-6\text{IPBA}}{0.05} \approx$ 4.64 million SEK in 2019 must earn a labor income of 9.6 income price base amounts (618,240 SEK). This is however to the right of the kink, providing no incentive to bunch at the first central government kink point.

[52] Failure to meet this requirement forfeits the whole dividend allowance.

C For online publication: Additional bunching results

C.1 Main bunching results and sensitivity checks

Figure C1: Bunching by cognitive ability decile



Note: The estimation underlying figure 3. The figure shows bunching during the years 2012-2016 for all adult men in different deciles of cognitive ability.

Figure C2: Sensitivity analysis, wide and small bunching window (cognitive ability)

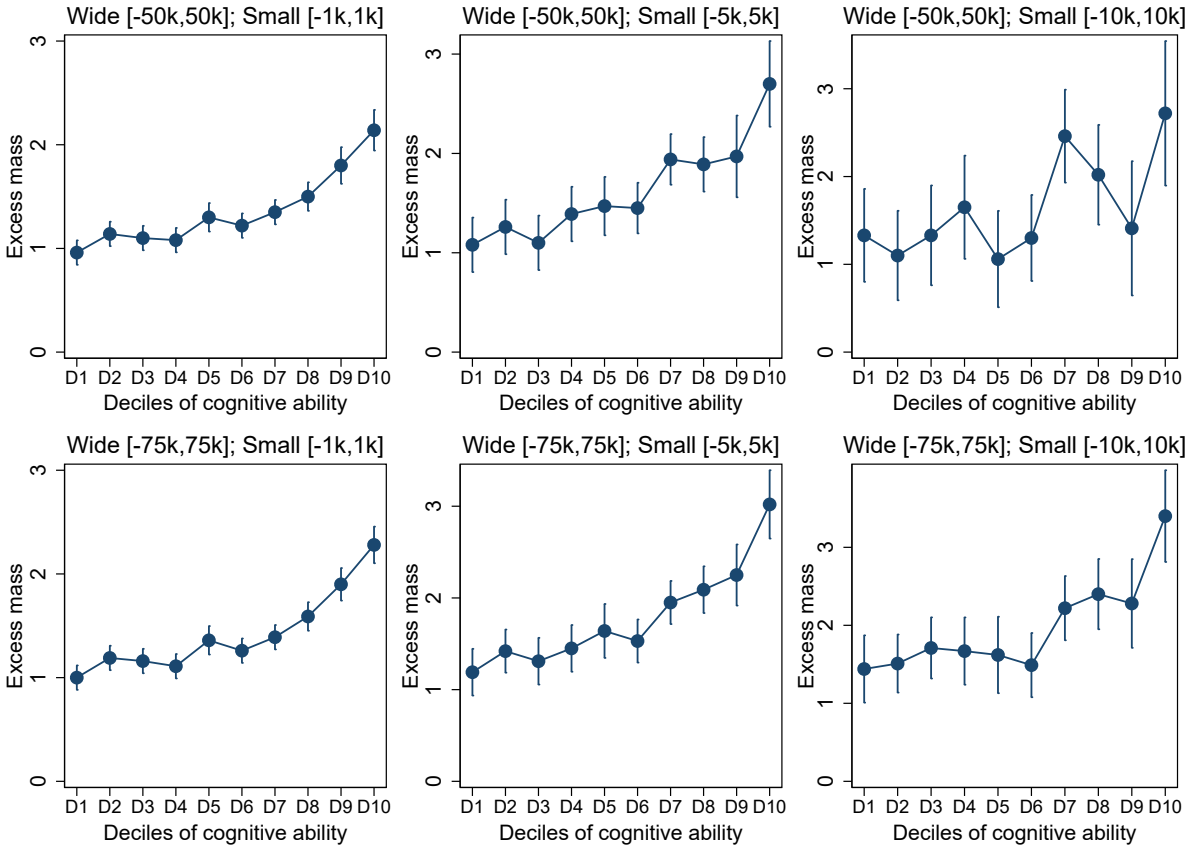
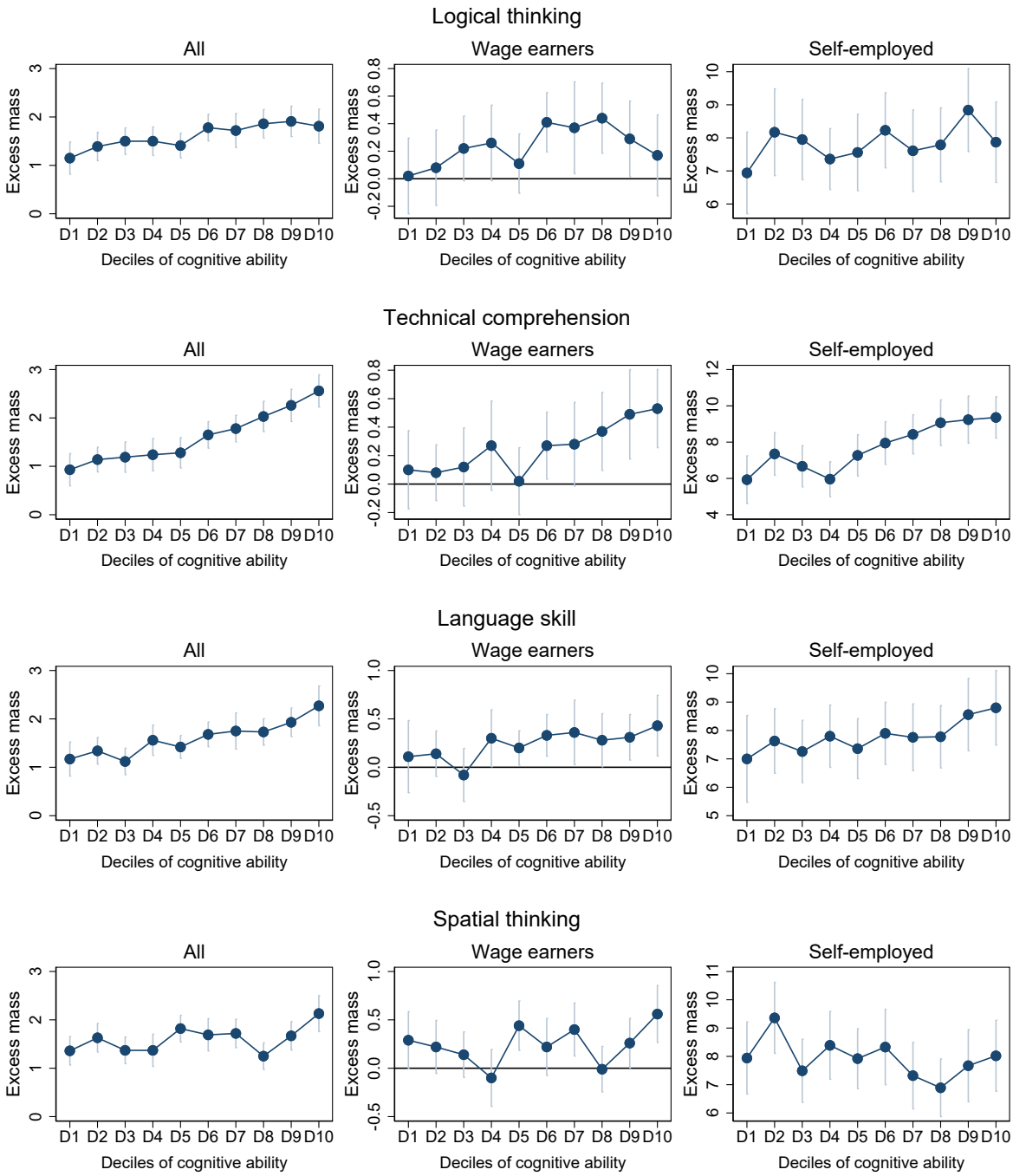
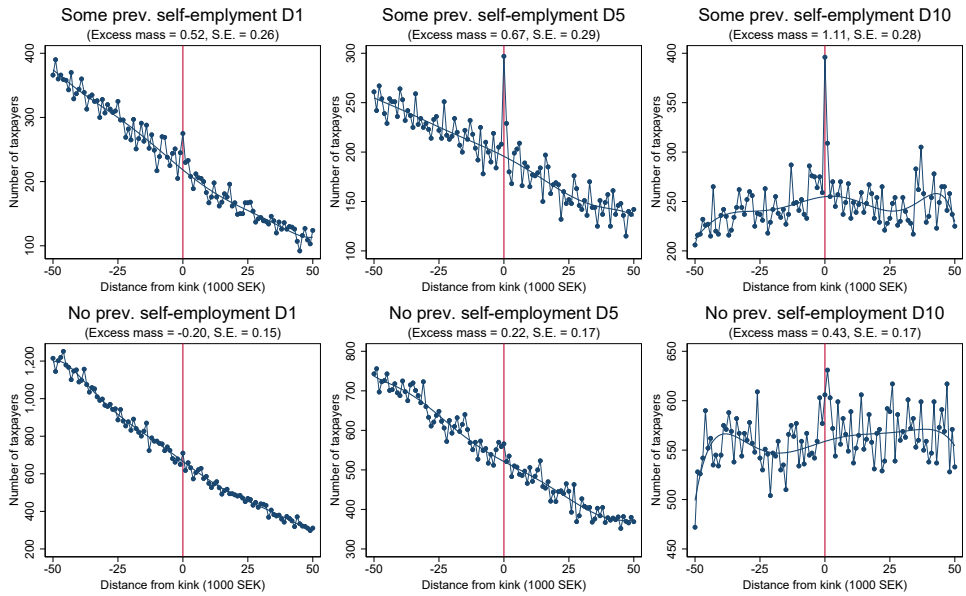


Figure C3: Ability gradients using the four sub-tests of cognitive ability



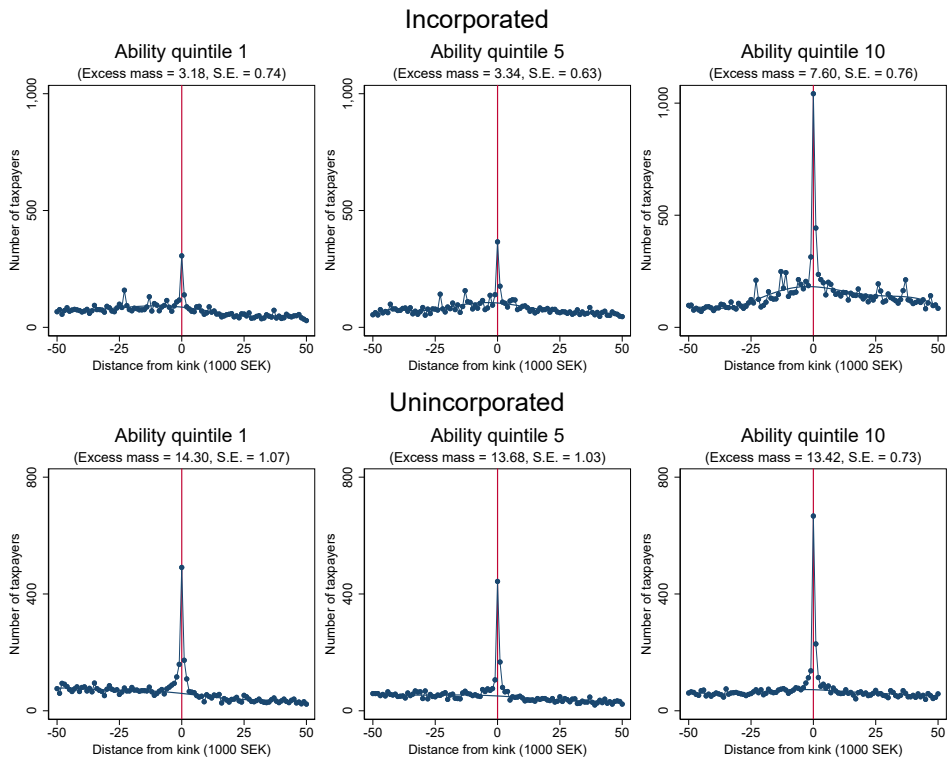
C.2 Additional bunching and occupation

Figure C4: Bunching of wage earners with or without self-employment history



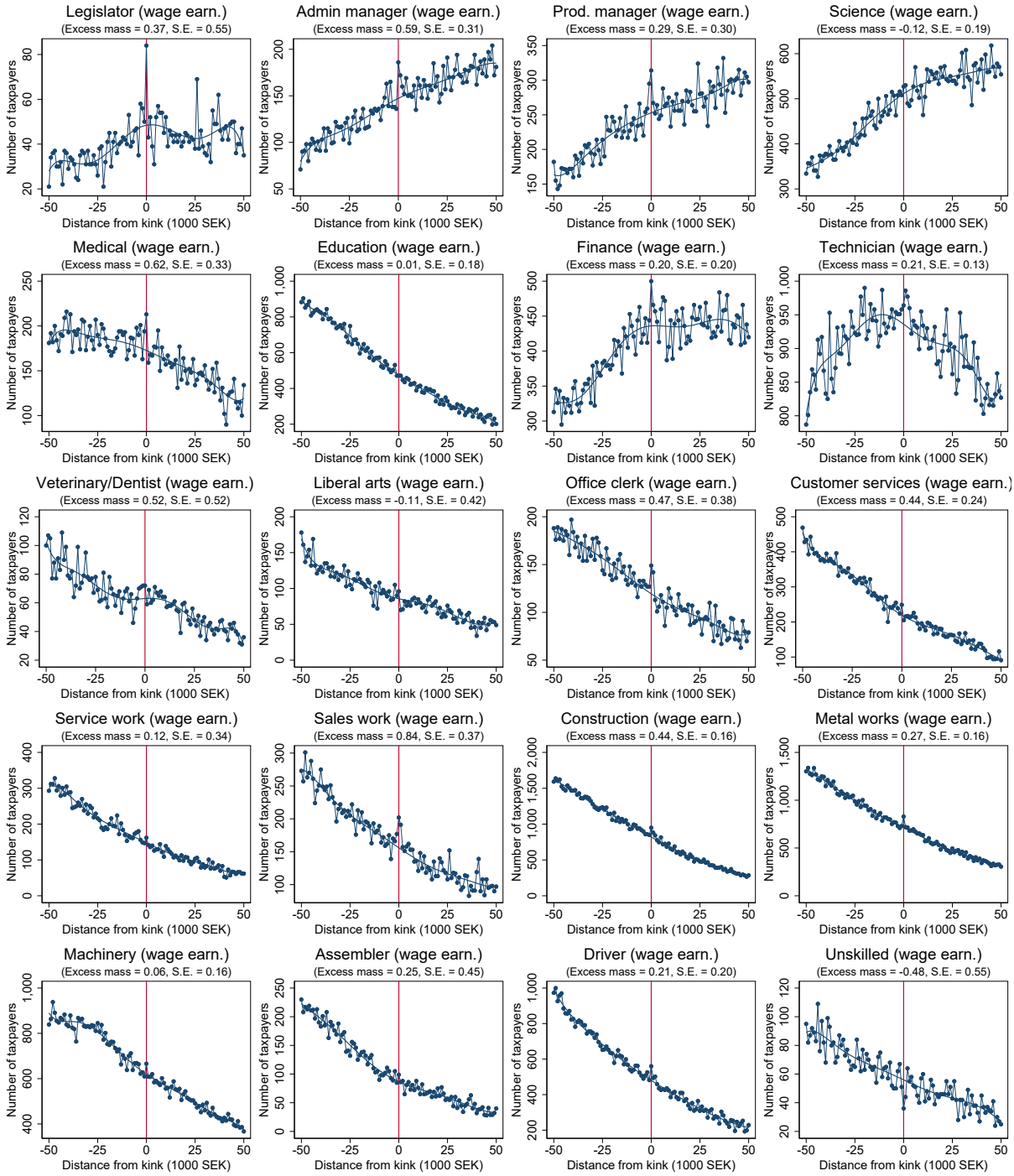
Note: A selection of the estimation underlying the ability gradient in figure C8.

Figure C5: Bunching among the self-employed



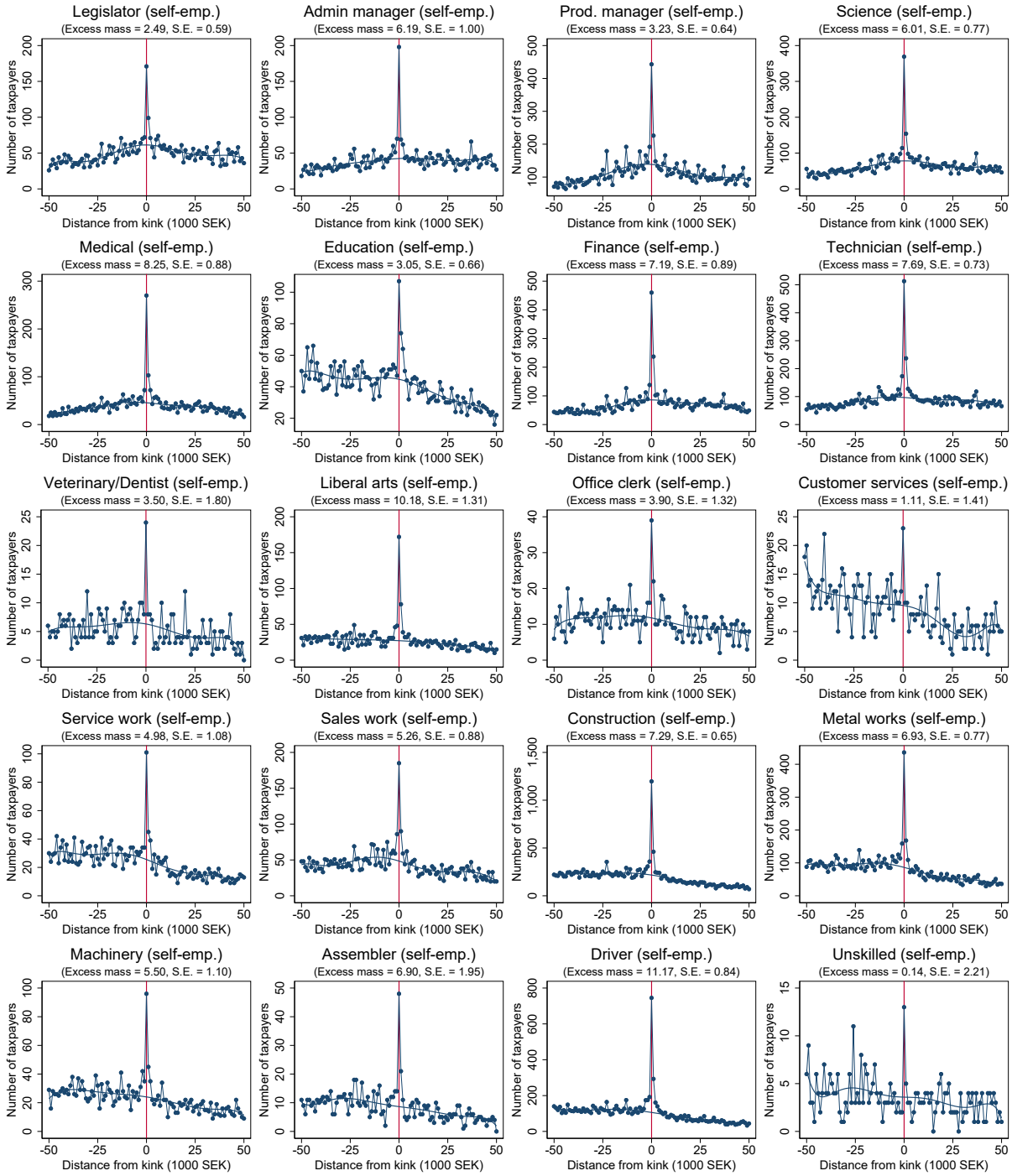
Note: A selection of the estimation underlying the ability gradient in figure 9.

Figure C6: Bunching within occupations: Wage earners



Note: Bunching estimation underlying figure 7.

Figure C7: Bunching within occupations: Self-employed



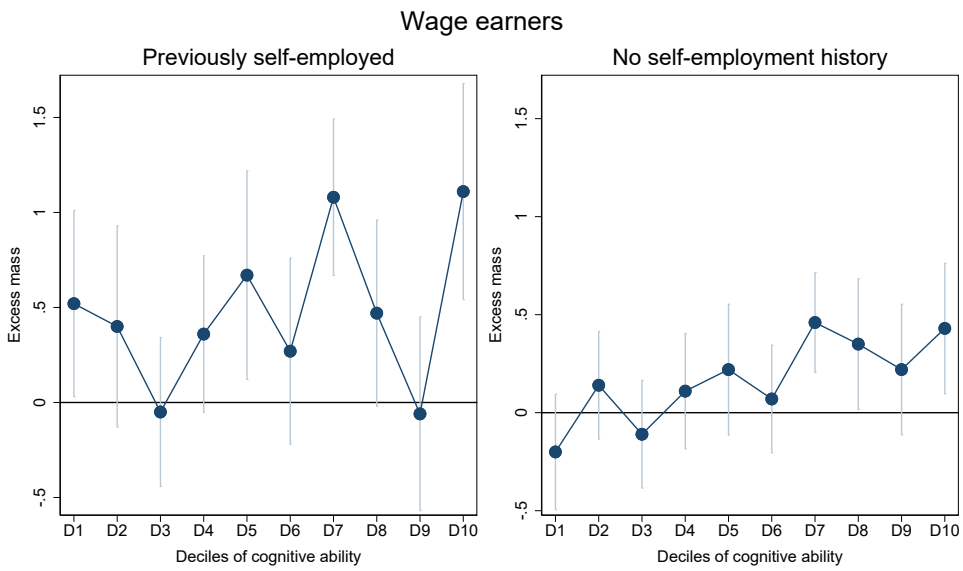
Note: Bunching estimation underlying figure 7.

C.3 The role of self-employment history among wage-earners

One might 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 residual connection to 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 C8

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 C8, 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.

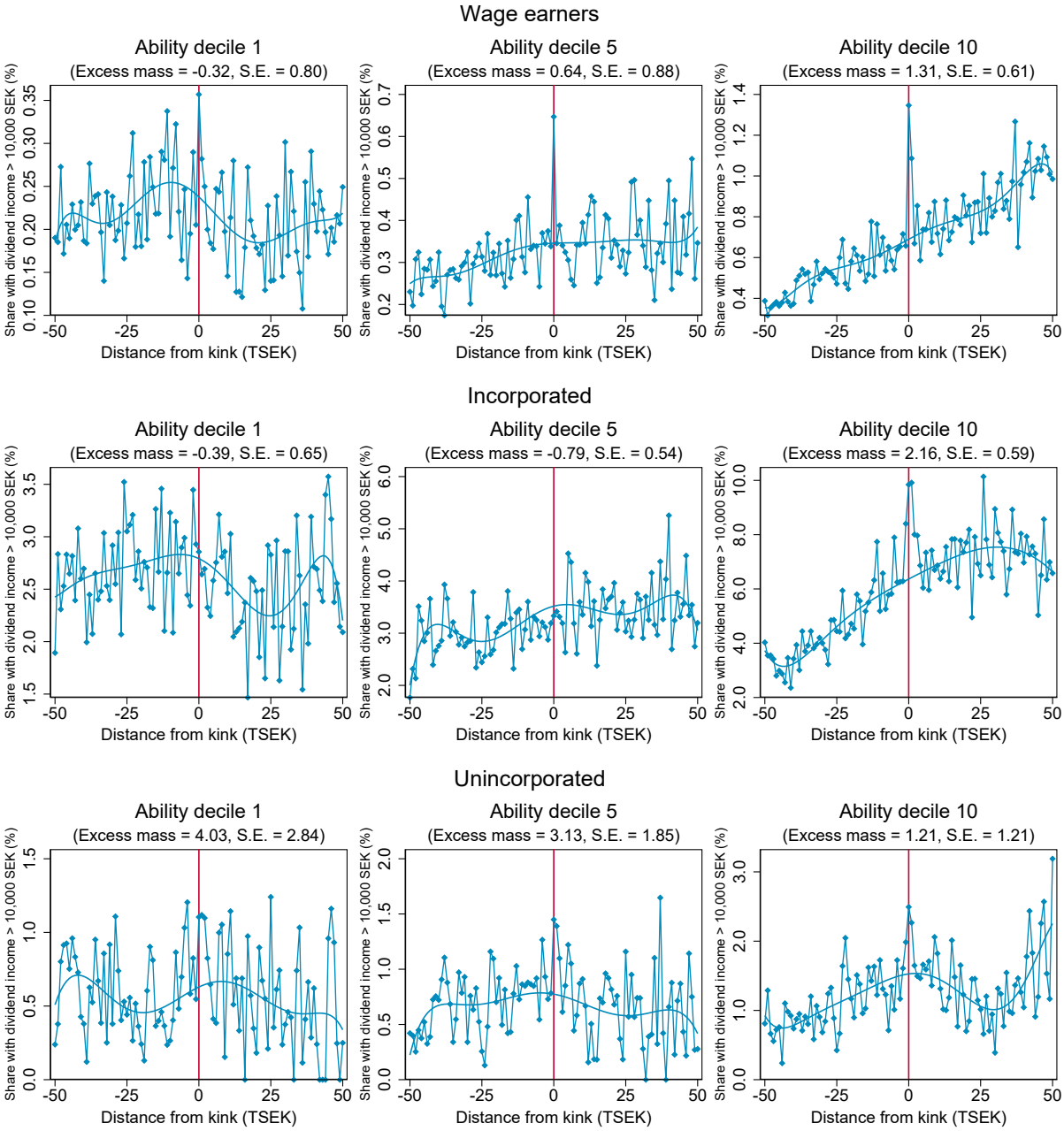
Figure C8: Role of previous self-employment for the ability gradient among wage earners



Note: Additional graphs showing the underyling estimation are shown in appendix figure C4.

C.4 Additional bunching and income shifting

Figure C9: Dividend bunching across ability deciles



Note: A selection of the estimation underlying the ability gradient in figure 10.

C.5 Additional bunching and gender

Figure C10: Bunching and gender using GPA: All, wage earners, self-employed

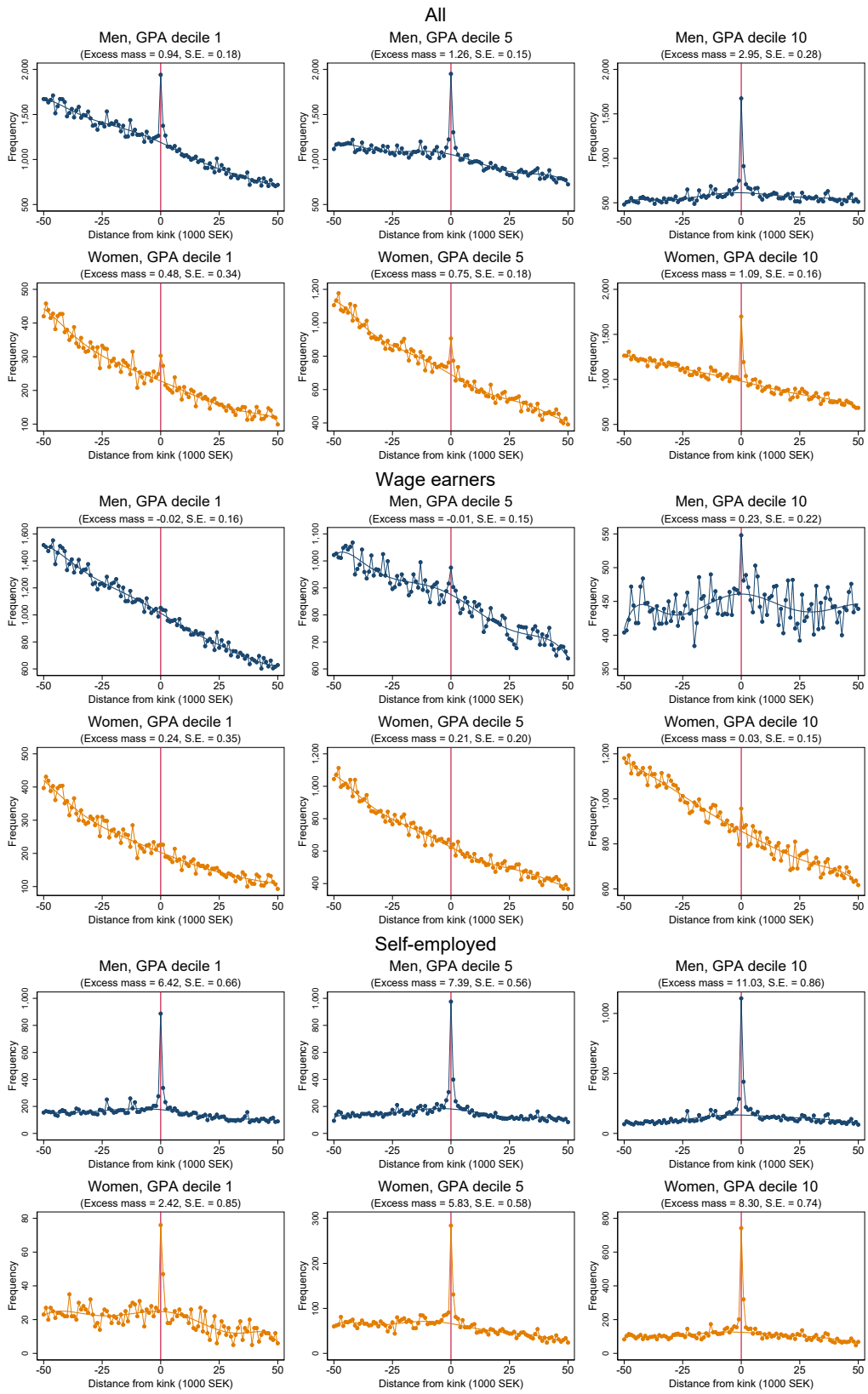


Figure C11: Bunching and gender using GPA: Incorporated and unincorporated

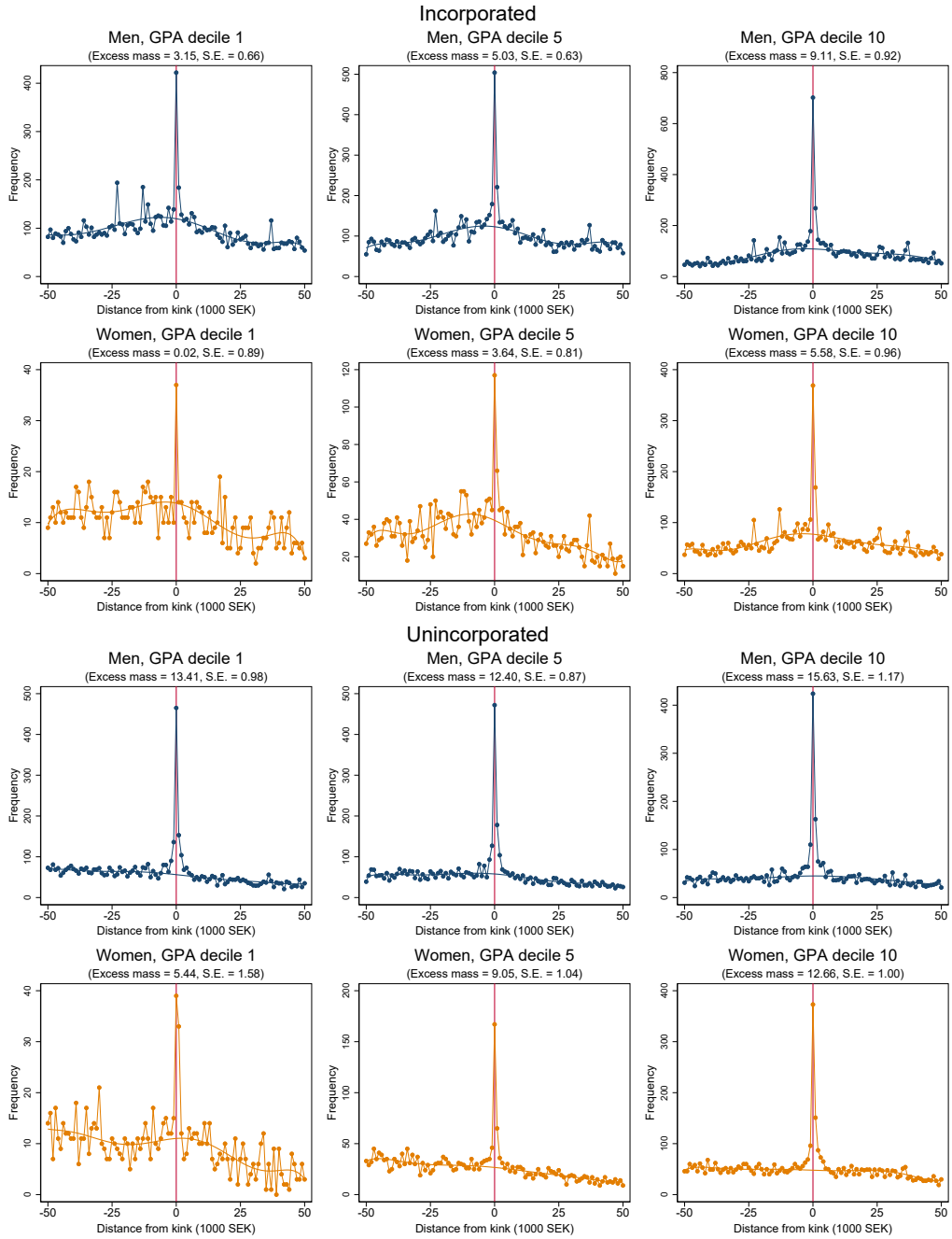


Figure C12: Bunching estimation and household ability

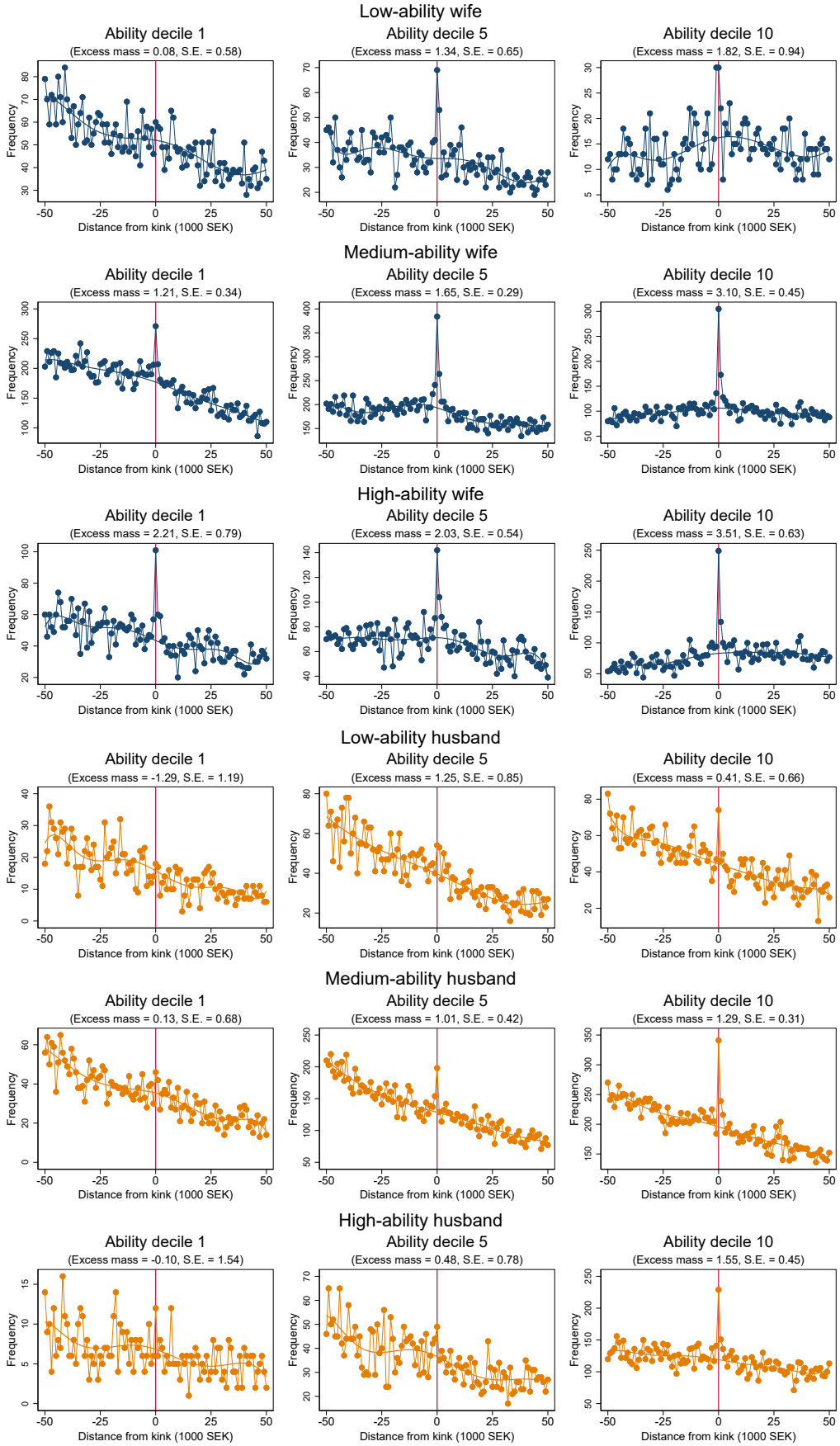


Figure C13: Gender differences in the ability gradient using GPA: Decomposition of the self-employed

Figure C14: Gender differences in the ability gradient using math grades

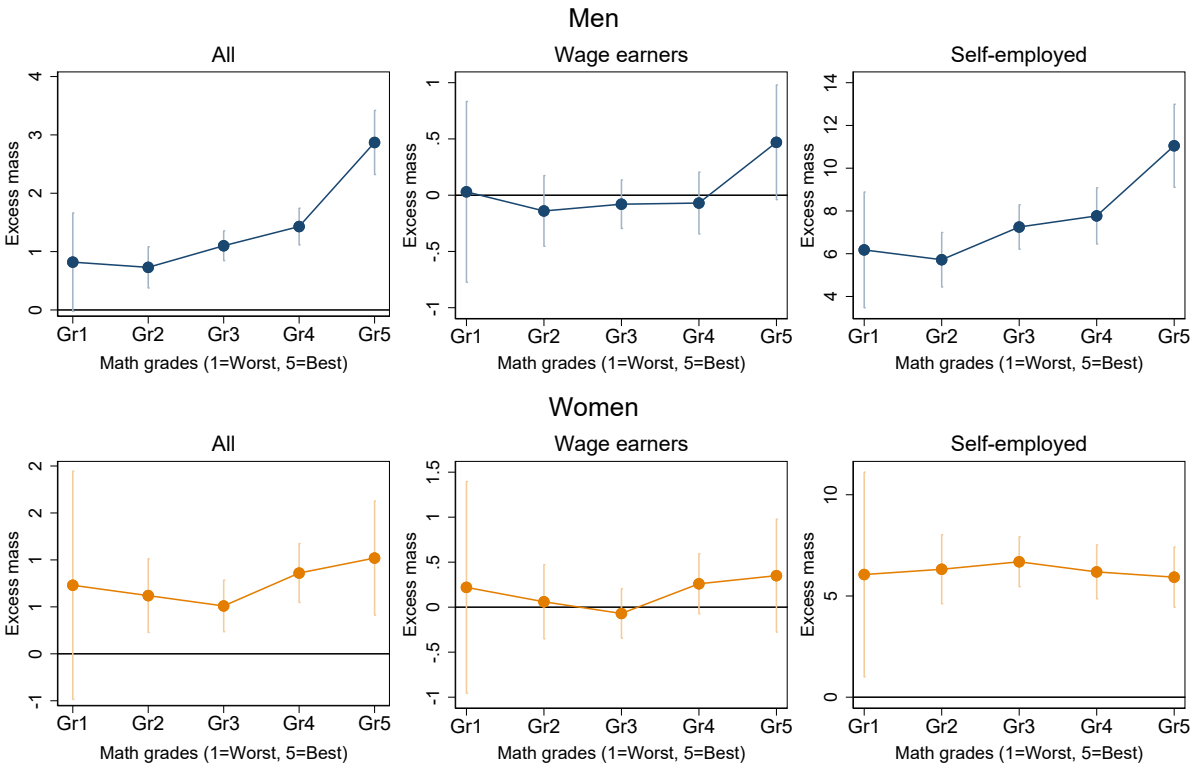


Figure C15: Gender differences in the ability gradient using math grades: Incorporated and unincorporated

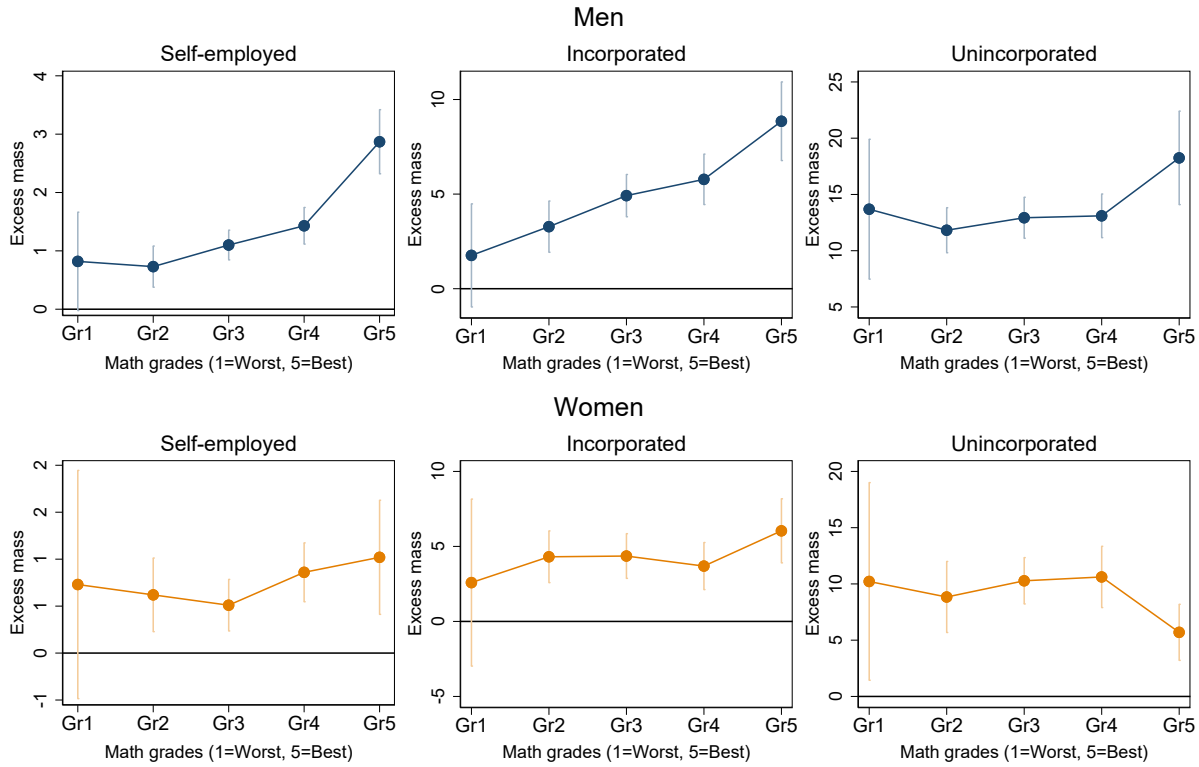
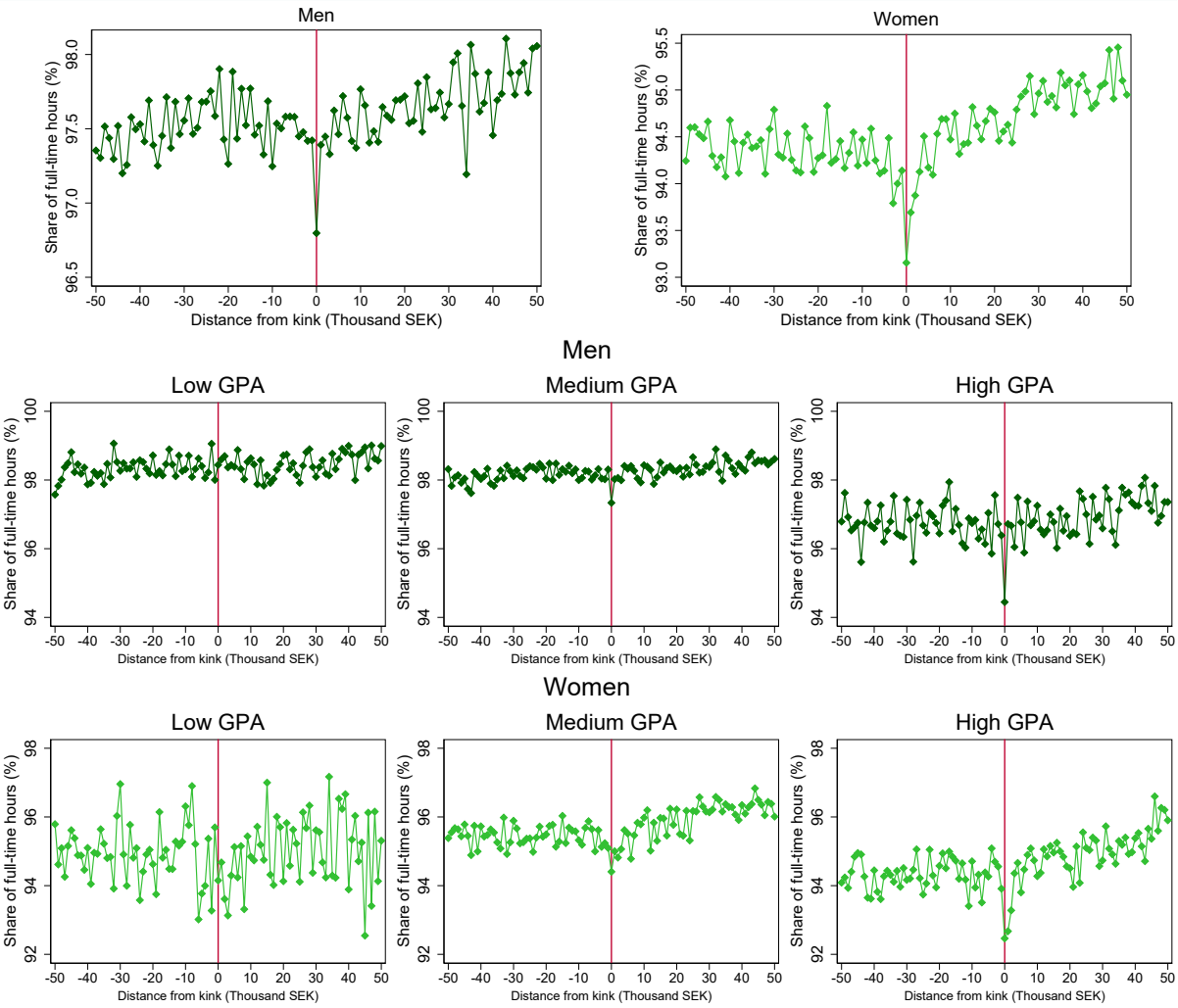


Figure C16: Labor supply for men and women around the kink



C.6 Additional bunching and other ability dimensions

Figure C17: Bunching gradient in non-cognitive traits

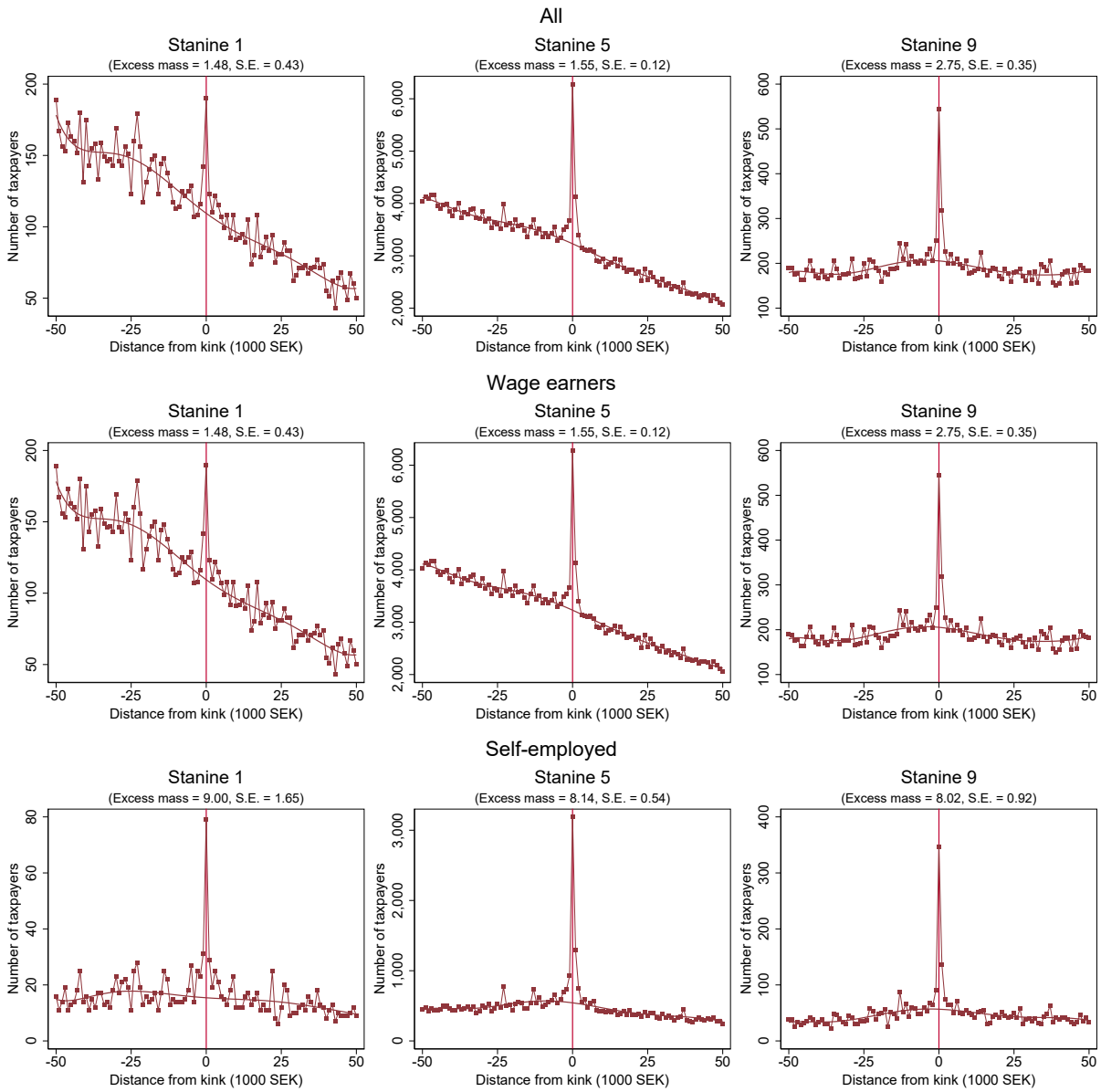


Figure C18: Bunching gradient in height

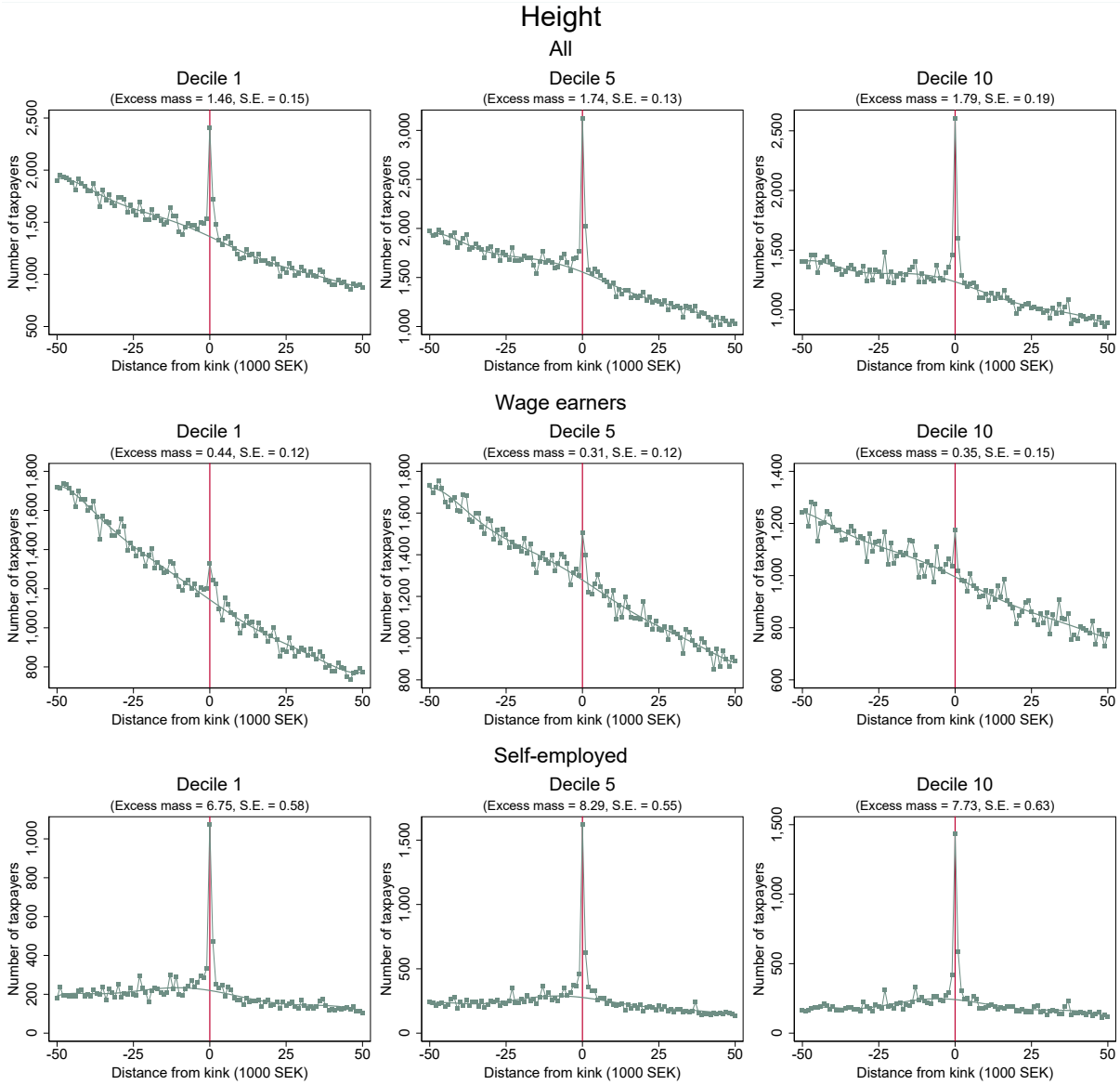


Figure C19: Bunching gradient in fitness

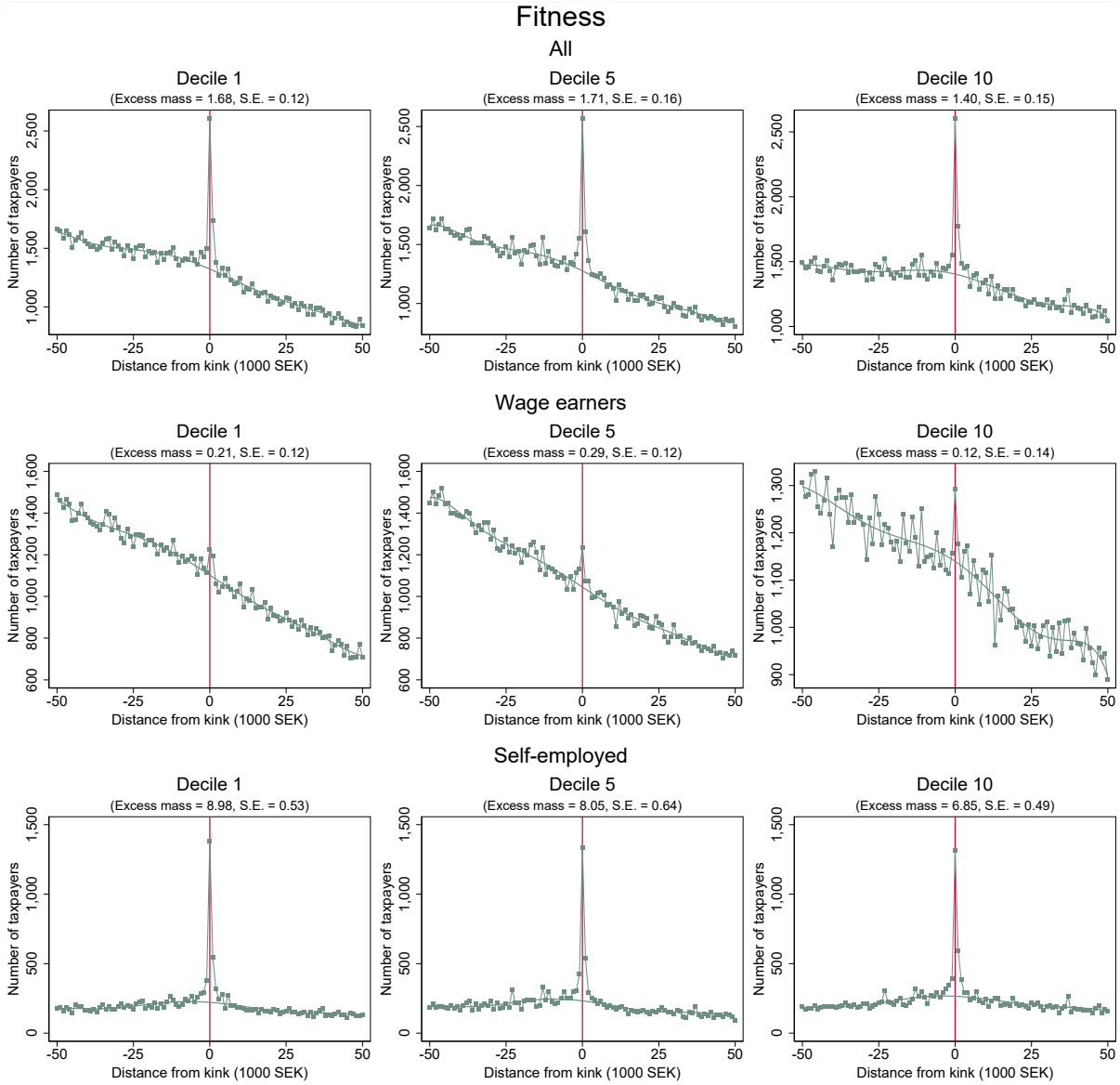


Figure C20: Bunching gradient in strength

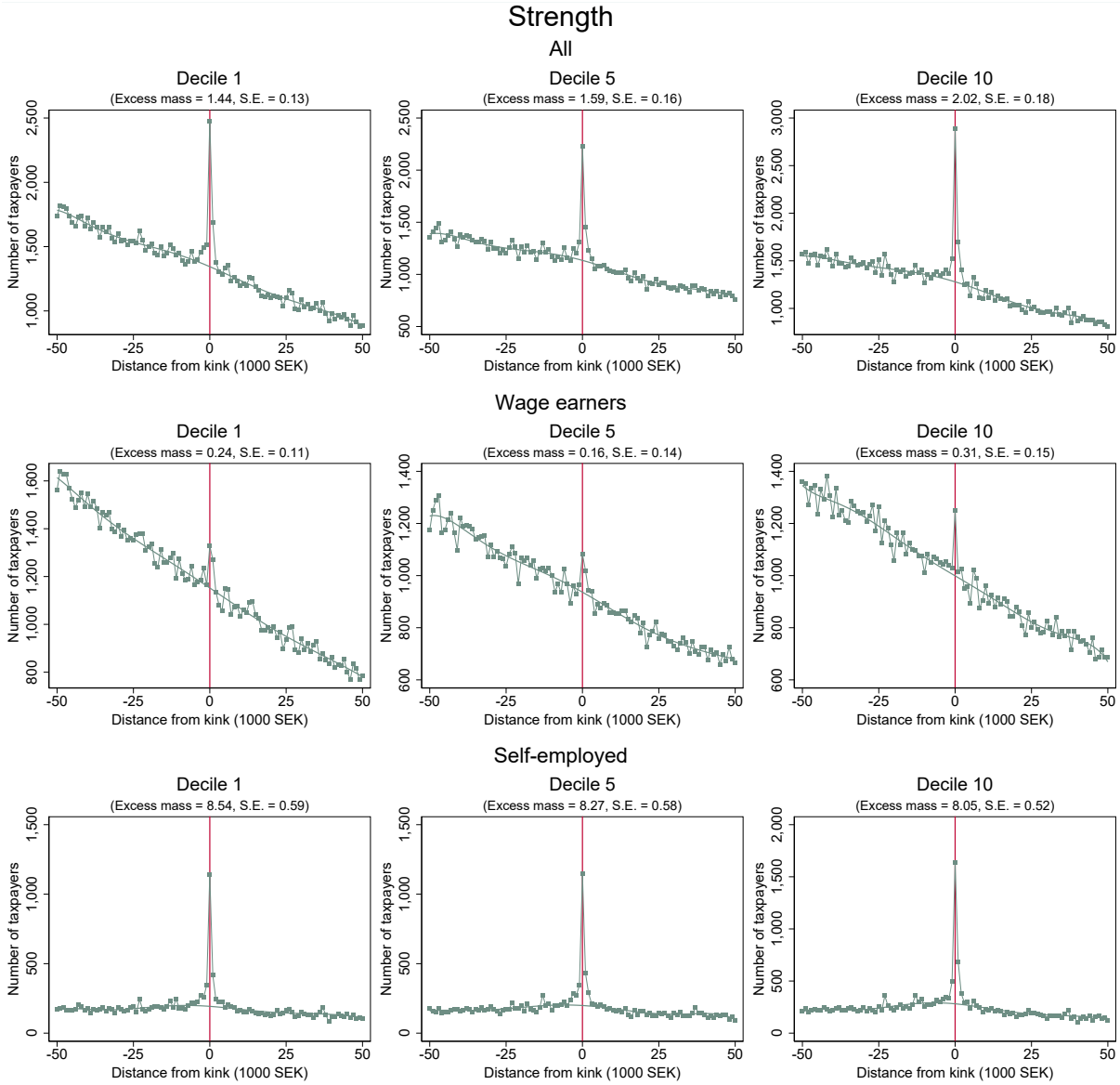
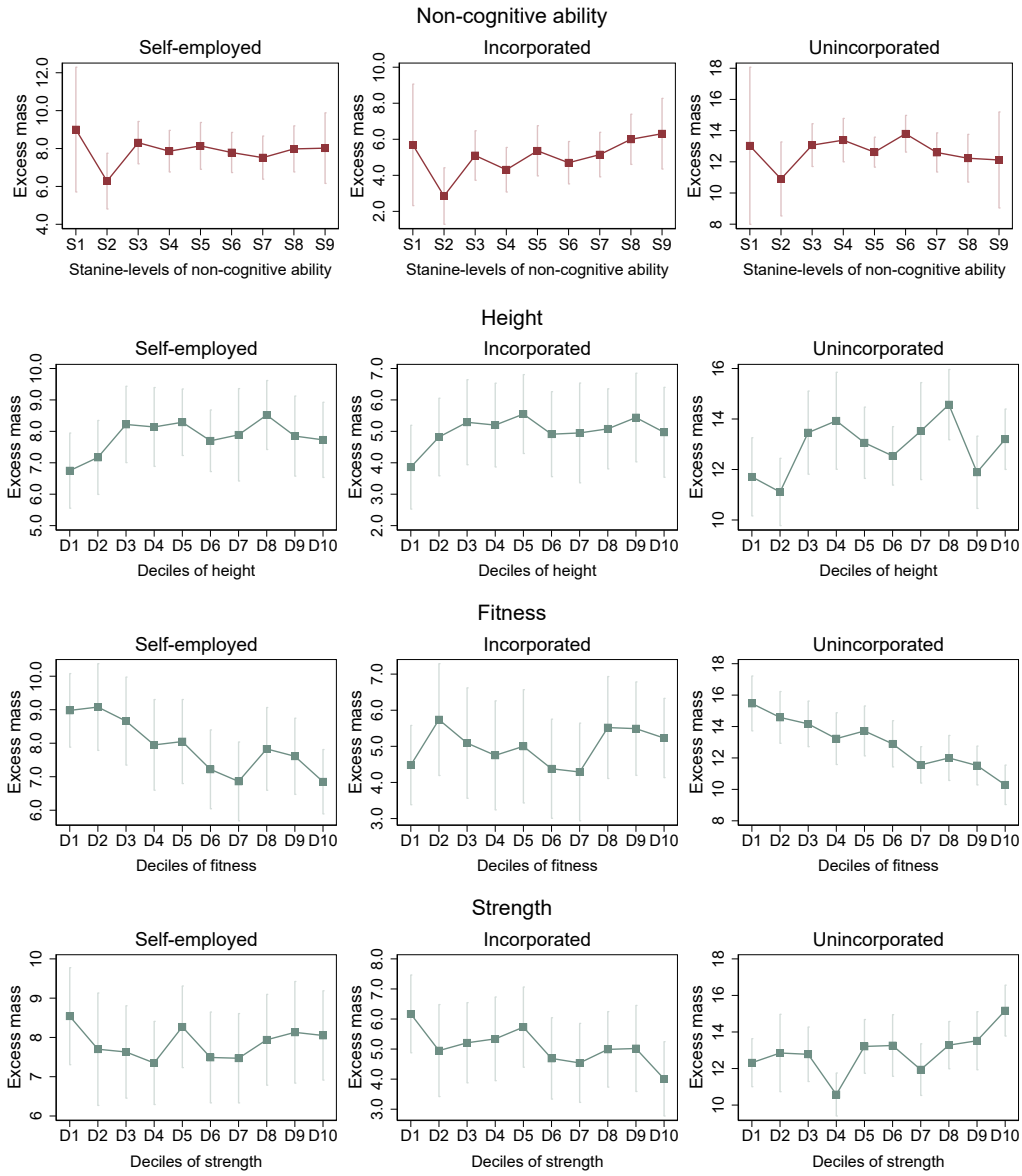


Figure C21: Ability gradients using non-cognitive traits and physical abilities: incorporated and unincorporated



Note: Analysis on incomes earned 2012-2016 by men born 1955-1975. Stanine levels show the population divided into fixed distribution shares with the following exact percentiles: S1 (P1-4, "the bottom 4%"), S2 (P5-11), S3 (P12-23), S4 (P24-40), S5 (P41-60, "the middle 20%"), S6 (P61-77), S7 (P78-89), S8 (P90-96), S9 (P97-100, "the top 4%").

C.7 Trends in bunching in Sweden, 2000-2016

This appendix section examines if bunching in our sample of middle-aged men has increased over time. In the study of bunching in Sweden during the period 1999–2005, Bastani and Selin (2014) found no evidence of bunching among wage-earners at the large kink in the government income tax schedule, in contrast to what we find in the 2012-2016 period. However, similar to us, they found significant bunching among the self-employed. Given the different results for wage earners, we are interested in examining if there are any *trends* in bunching at the large government income tax kink point

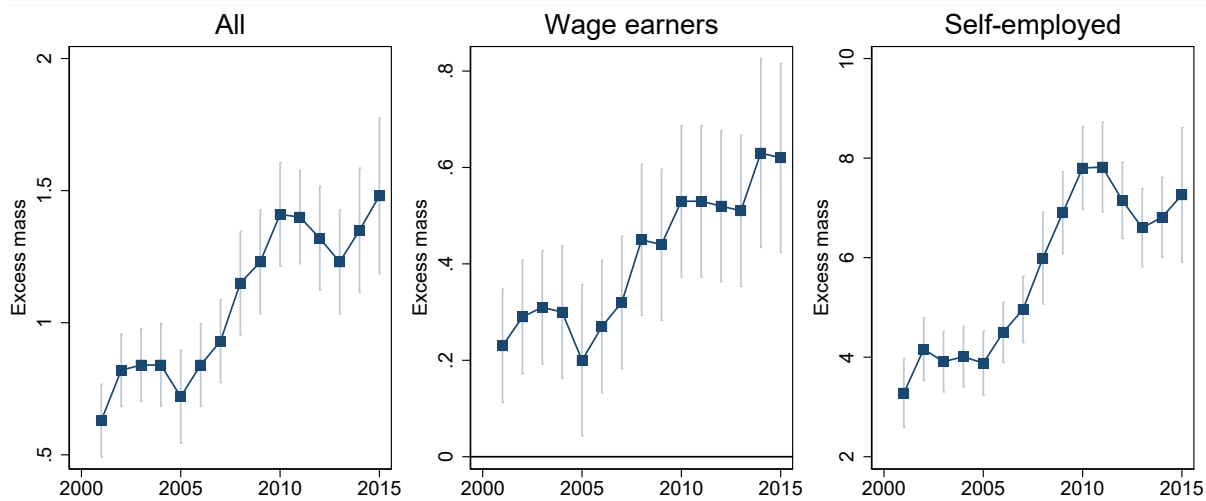
in Sweden over the past two decades.

To investigate the existence of trends in bunching, we focus on our sample of males registered in the military enlistment. We restrict the sample to men aged between 40 and 50 years in each income-year to mitigate long-term within-cohort learning effects that could drive up bunching over time. Since we keep the analysis within the main "military enlistment"–population studied in the rest of the paper, this will not be a complete analysis of bunching trends in Sweden. For that, one would need to include all age groups and also women, but such a broader investigation of trends in bunching (and its potential causes) lies beyond the scope of the present study.

The results are presented in figure C22 and show that bunching has increased notably from the early 2000s to the mid-2010s in our analyzed sample. Among all men, the excess mass has more than doubled, with the increase taking place in the late 2000s. Interestingly, the increase is notable among both wage earners and the self-employed. Although the level of bunching differs between these two groups, the relative increase in bunching over this period appears to be roughly similar at around a doubling of the excess mass. These results suggest that the differences between this paper's findings on bunching among wage earners and those of Bastani and Selin (2014) could partly be the result of an underlying trend in bunching over time.

We also would like to mention that we have analyzed how the ability gradient in tax responsiveness has evolved over time, exploiting the fact that the military enlistment records provide us with consistent measures of cognitive ability over several decades. This analysis shows that the ability gradient has been surprisingly stable over time. These results are available upon request.

Figure C22: Trends in bunching in Sweden, males 40-50 years old



Note: Bunching around the kink point in the government income tax schedule among men in our main sample, with the additional restriction that they are between 40 and 50 years old in each income-year to mitigate learning effects and other within-individual time trends. Three-year moving averages.

D For online publication: Additional regression results

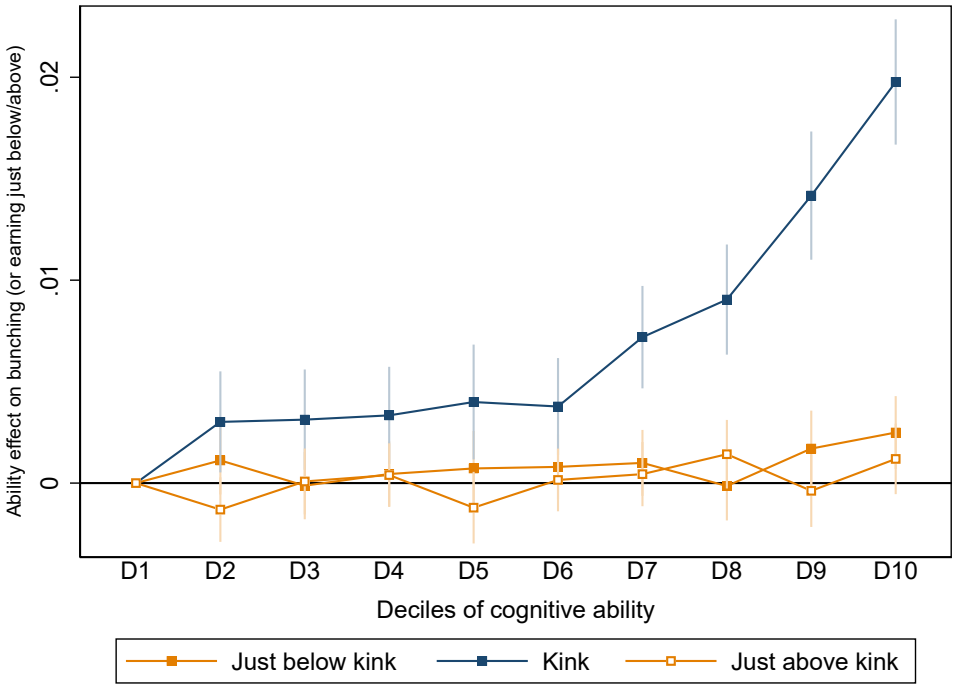
D.1 Regression-based ability gradients: Additional results

We generalize the analysis in section 6.1 by considering the effect of cognitive ability on the probability of locating in different income intervals $\mathscr{Y}$, by estimating the following equation:

$$\mathrm{I}\{y_{iat} \in \mathscr{Y}\} = \eta_a + \lambda_t + \sum_d \beta^d \cdot CogDec_i^d + \varepsilon_{iat}, \tag{20}$$

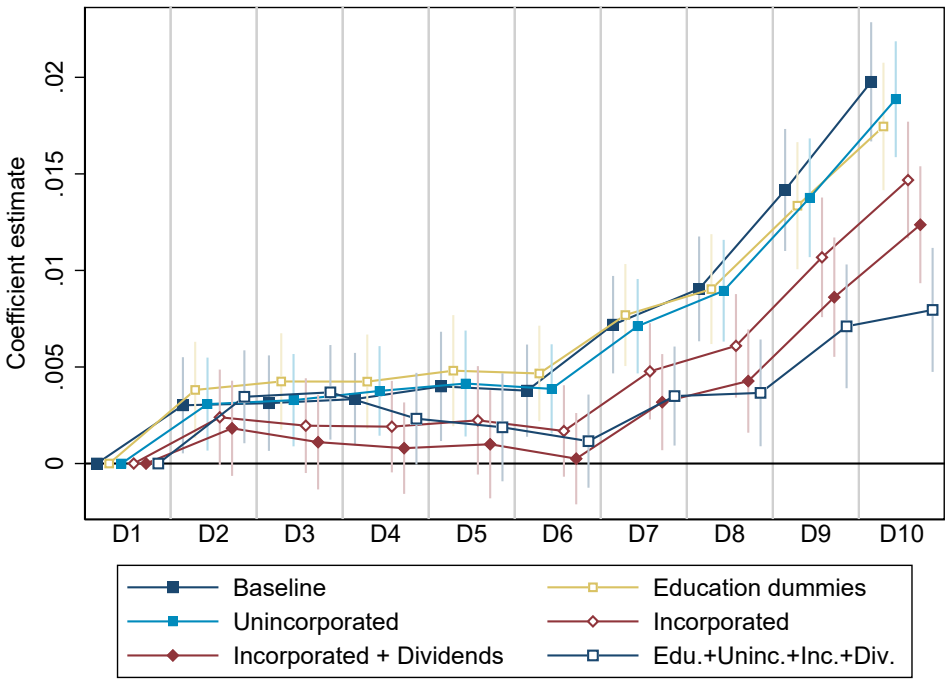
where $\mathrm{I}\{y_{iat} \in \mathscr{Y}\}$ is an indicator variable taking on the value of 1 if an individual's income y_{iat} falls within the income interval $\mathscr{Y}$, η_a represent cohort fixed effects, λ_t are year fixed effects, and ε_{iat} are error terms. Using a notation similar to that used in section 2.2, we let $\mathscr{Y}$ be equal either to a small window centered at $\hat{z}$, $[-\delta, \delta]$ (in which the case regression (20) becomes the same as the regression (18) in section 6.1 in the main text), an interval to the left of $\hat{z}$, $[-3\delta, -\delta]$, or an interval to the right of $\hat{z}$, $[\delta, +3\delta]$. The idea is to compare the ability gradient obtained from the estimates β^d, $d = 1, 2, \ldots, 10$, 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. As in the main text, we restrict the overall estimation sample to observations in the interval $[-\phi, \phi] = [-25k, 25k]$ and choose $\delta = 0.5k = 500SEK$. The results show that there is no ability gradient when considering individuals located in the two placebo-intervals.

Figure D1: Regression-based ability gradient and placebo-gradients



Note: Estimated coefficients from regressions in (20) for the three income intervals described in the text.

Figure D2: Regression-based ability gradient with controls, constant sample



Note: "Education dummies" controls for education length (in number of years) and field (3-digit ISCED 97 levels), "Unincorporated" and "Incorporated" are dummies for respective form of self-employment and "Dividends" denote earning any dividend income. Adding controls slightly re-shapes the sample (see figure 13), but here the sample is kept constant.

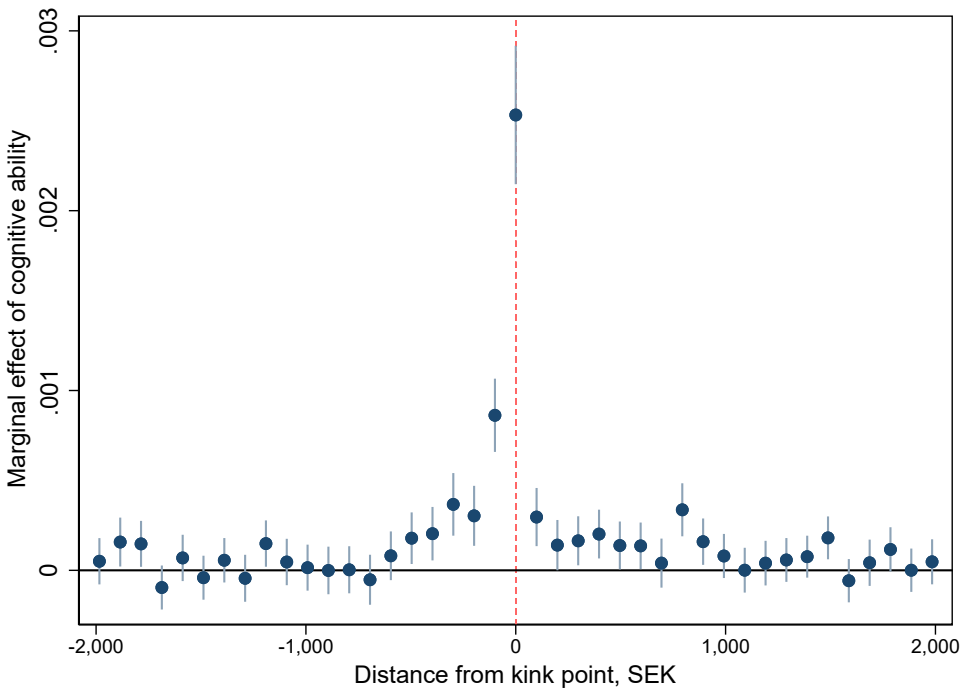
D.2 Zooming in on the kink

The purpose of this section is to characterize sharp bunching behavior *exactly* at the kink point. As in the analysis in section 6.1, 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.[53] Due to sample-size considerations, we focus on a linear specification with a continuous measure of ability rather than the flexible dummy-variable specification. Formally, we estimate:

$$\mathrm{I}\{y_{iat} = y\} = \eta_a + \lambda_t + \beta^y \cdot Cog_i + \varepsilon_{iat}, \quad y \in \{\hat{y} - \phi, \hat{y} - \phi + 100, \ldots, \hat{y} + \phi - 100, \hat{y} + \phi\}, \tag{21}$$

where Cog_i is a continuous measure of cognitive ability (the z-score) and $\hat{y}$ represents the exact location of the kink point. Notice that these are *rolling* regressions, with one separate regression for each value of $y \in \{\hat{y} - \phi, \hat{y} - \phi + 100, \ldots, \hat{y} + \phi - 100, \hat{y} + \phi\}$ with a corresponding estimate β^y. The results are shown in Figure D3.

Figure D3: 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 (21). 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

[53] 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.

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.

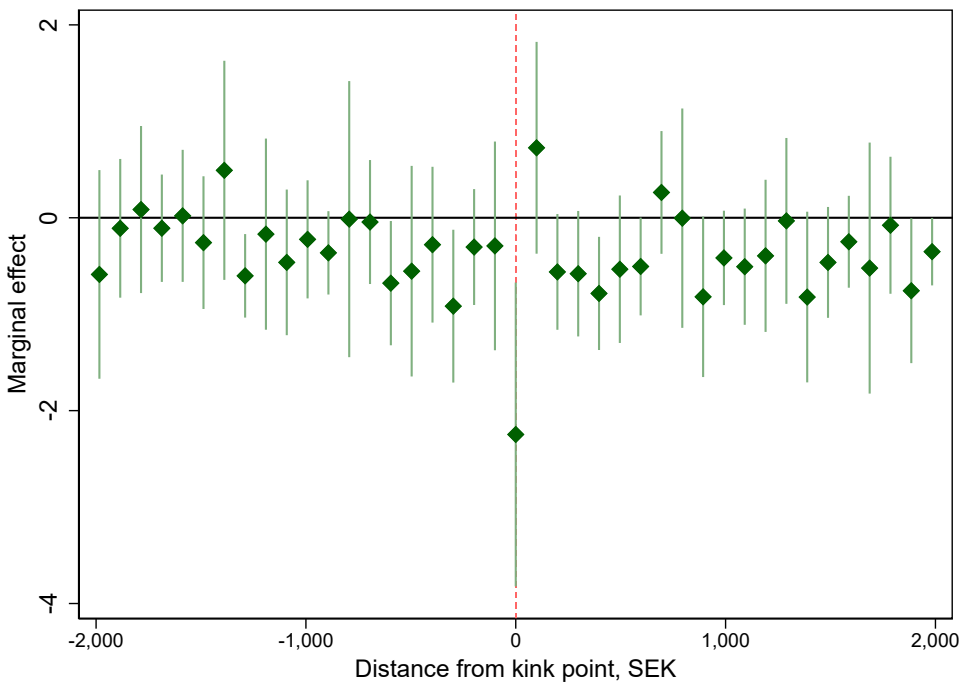
D.3 Zooming in on the kink: the role of labor supply

To shed further light on the role of labor supply adjustments for bunching behavior, we examine the conditional correlation between contracted hours and cognitive ability at different income levels y by running the following regression:

$$hours^{y}_{iat} = \eta_a + \lambda_t + \rho^y \cdot Cog_i + \varepsilon_{iat}. \tag{22}$$

Equation (22) bears similarity to the placebo regression defined in equation (21), which was visualized in figure D3. In the above equation, $hours^{y}_{iat}$ is contracted hours of individual i of age a in income year t, at income level y (measured as a percentage of full-time work), 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 D4 and demonstrate that the combination of having a high ability and low hours is especially prevalent at the kink, with no strong such indications at adjacent income levels.

Figure D4: Marginal effect of ability on hours at different income levels



D.4 Regression-based ability gradient in labor supply for men and women using GPA

In this section, we repeat the analysis of section 6.2 for men and women using GPA.

Figure D5: Ability gradient in labor supply reduction for men and women at the kink