Financial Frictions and the Non-distortionary Effects of Delayed Taxation

Andreas Fagereng and Marius A. K. Ring*

Preliminary

March 15, 2022

Abstract

Financially constrained agents discount future cash flows at above-market rates. In this paper, we introduce the hypothesis that delaying tax payments can materially reduce distortions when agents are financially constrained. We test this hypothesis in the context of the labor supply decisions of young workers in Norway, where a kinked income-contingent student-debt conversion scheme replicates an income tax with delayed payments. Bunching analyses reveal elasticities that are an order of magnitude below those we find at a regular income tax threshold, and which are increasing in ex-ante financial resources. These findings underline the potential for delayed taxation to be a powerful new component of optimal tax policy.

JEL: H21, G51, D15

Keywords: deferred taxation, delayed taxation, credit constraints, income taxation

---

*Andreas Fagereng is at BI Norwegian Business School. Marius Ring (corresponding author) is at the University of Texas at Austin. E-mail: mariuskallebergring[at]gmail.com. See www.mariusring.com for the most recent version. We thank seminar participants at MIT and UT Austin for helpful questions, and Anthony DeFusco, John Griffin, Sam Kruger, Michael Sockin, Andres Almazan, Daniel Neuhann, Will Fuchs, Clemens Sialm, Jonathan Parker, Lawrence Schmidt, Taha Choukhmane, Daniel Greenwald, and Eben Lazarus for helpful comments and discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 851891).

1 Introduction

The optimal design of any tax depends crucially on its distortionary effects. How responsive labor supply is to the net-of-tax wage has seen tremendous empirical attention. While there are offsetting substitution and income effects at play in the response to labor income taxes, it is generally considered that the substitution effect dominates: A reduction in take-home wages lowers labor supply. However, little attention has been given to the effects of introducing a substantial time delay between tax accrual and tax payment.[1] The purpose of this paper is to propose and test the hypothesis that delaying the payment of labor income taxes may reduce their distortionary effects.

The intuition for this hypothesis comes from basic finance theory. Financial frictions render agents with steep earnings profiles unable to borrow against higher future incomes at the market rate. This creates a wedge between the market rate and their personal discount rate (Carvalho, Meier, and Wang 2016, Epper 2017). Assuming that tax authorities may borrow at the market rate, financial frictions therefore introduce a wedge between the tax authorities and constrained agents' net present value (NPV) of a future given tax liabilities. By offering to finance income tax payments at the government rate, tax authorities may stimulate positive behavioral responses to lower taxation without (mechanically) foregoing revenues due to a lower tax rate. Put differently, tax authorities may reduce the distortionary effects of income taxation without foregoing tax revenues.

Our hypothesis is related to the notion that behavioral elasticities governing responses to taxation are not immutable parameters. Instead, they are affected by the the wide range of tax instruments available to the government. This implies that tax authorities should not only set the tax rate at its optimum, they should also set the behavioral elasticity at its (joint) optimum. This idea is formalized by Slemrod and Kopczuk (2002) who focus on tax enforcement as the instrument. Our paper stresses the timing of tax payments as an important instrument: by allowing for delayed tax payments, the behavioral labor supply elasticity goes down—and optimal tax rates go up. As with tax enforcement, this may have costs, but also benefits.

This paper focuses on empirically testing the core partial-equilibrium mechanism: In the presence of financial frictions, labor income taxes whose payments can be delayed become substantially less distortionary. Performing such a test is challenging, since there

[1] The concept that allowing tax payments (and not accrual) to be delayed may reduce the distortionary effects of labor income taxation in the presence of financial frictions is new, but the notion of delaying tax payments is not. For example, capital gains are typically only taxed at realization and retirement-savings schemes often allow for delayed taxation (see Amromin, Huang, and Sialm 2007 and Huang 2008 for more on tax-deferred savings accounts). Similarly, taxing entrepreneurial dividends rather than profits allow entrepreneurs to delay their tax payments (see Dávila and Hébert 2019).

is little variation in the timing of tax payments. Taxes are typically paid either immediately (through withholding) or a year later when tax returns are due. We overcome this challenge by studying the effects of a student debt forgiveness scheme in Norway. This scheme creates a large jump in the effective marginal income tax rate where marginally accrued taxes can be financed with the same generous terms as subsidized student loans. More specifically, the vast majority of Norwegian students receive a yearly loan of around \$13,000, of which roughly half may be forgiven. If the student has labor earnings above approximately \$17,000, each additional dollar of earnings reduces the amount forgiven by 50 cents.

This setting is well suited to investigate how financial frictions may render delayed taxation less distortionary. First, students are, almost by definition, highly constrained. Just a few years later they face significantly higher incomes against which it is hard to borrow. The dramatic increase in the effective tax rate at the earnings threshold is also more than significant enough for any student to be cognizant of it: At the threshold, the marginal net-of-tax (and debt increase) wage drops from 75 to 25 cents.[2] Despite this drastic reduction in the marginal (effective) wage, students are astoundingly irresponsive. While there is clear visual evidence that students do respond, these responses pale in comparison to the effective after-tax wage reduction that occurs.

Our bunching analysis offers an implied labor earnings elasticity to the after-tax wage of only 0.016. While this estimate is highly statistically significant, it is an order of magnitude below most existing estimates (Keane, 2011). Labor market frictions are unlikely to explain our relatively low elasticity. Our sample is limited to students near the debt-conversion threshold, which is substantially below a full-time salary in Norway. This ensures that students are part-time workers who likely face flexible work arrangements.

To shed light on the observed non-bunching behavior, we examine how student characteristics covary with their position relative to the debt-conversion threshold. These analyses suggest that non-bunchers (and their parents) have significantly less liquid wealth, but not lower future earnings. This is precisely what we would expect to see if irresponsiveness to the threshold is driven by financially constrained agents. We further find no evidence that the educational attainment of students' parents change in a manner consistent with these characteristics driving the differences in bunching behavior. Informed by these analyses, we study heterogeneity in bunching by the ex-ante financial situation of students and their parents. Students who have below median liquidity (and their parents as well) exhibit an implied labor earnings elasticity less than half as large as those

[2] The marginal tax rate around the threshold was approximately 25% during the sample period. This marginal tax applies to all marginal earnings regardless of the increase in student debt.

above the median. In a simple model, this heterogeneity can be rationalized by less-liquid students optimizing according to a 10 percentage points higher marginal borrowing rate.

We continue to examine student bunching behavior at a regular tax threshold. This allows us to compare the implied labor supply elasticities under different payment schemes but among a similar sample of individuals.[3] The tax threshold analyzed occurs at around \$6,000, where the marginal income tax rate goes from 0 to 25 percent. Using the same techniques as before, we estimate an implied labor supply elasticity of 0.13. This is about eight times higher than the elasticity inferred from the delayed-tax threshold, which is consistent with materially reduced distortions when payments are substantially delayed.

We further present a simple model that relates differential responses to regular and delayed taxation to the marginal rate of return on net savings of a life-cycle agent. Under the assumption of a homogenous structural labor supply elasticity, an annualized marginal interest rate of 23% can explain the relatively muted responses to delayed taxation.

The central contribution of this paper is to propose and test the hypothesis that delaying the payments of income taxes may substantially reduce its distortionary effects in the presence of financial frictions. To our knowledge, there exists neither theoretical nor empirical research on this topic. Our empirical setting fits the bill for testing this hypothesis due to three important features. (i) It effectively replicates a delayed income taxation system with a sizable hike in the marginal tax rate, where marginal taxes at this threshold are subject substantial delay in the payment. In addition, the sample consists of taxpayers where (ii) labor supply is highly flexible and (iii) borrowing constraints play an important role.

That delaying the payment of a tax reduces its distortionary effects is not too surprising. In the absence of severe debt aversion and in the presence of borrowing-constrained agents, this is what we would expect from economic theory. In some sense, this paper thus provides a test of whether life-cycle model reasoning is applicable to the study of constrained workers' labor supply decisions. Additionally, it provides empirical evidence on the potential economic magnitude of the effect, which in our setting is rather substantial. Both of these ingredients are necessary to assess the potential for delayed taxation as a new policy tool.

Related literature. On the conceptual front, this paper contributes to the literature on dynamic optimal taxation (see, e.g., Ndiaye 2020; Heathcote, Storesletten, and

[3] Ideally, this will keep unobservable factors causing frictions in labor supply optimization constant. An alternative would be to compare our elasticity under delayed taxation with elasticities from other research. However, this raises the concern that differences in labor market frictions are driving the differences in the elasticities.

Violante 2020; Yu 2021; and the surveys in Golosov and Tsyvinski 2015 and Stantcheva 2020). Most closely related is research that considers altering the timing of taxpayments or incorporating financial frictions.[4] The conceptual novelty of this paper lies in this intersection.

On the empirical front, this paper contributes to the growing literature studying bunching at tax thresholds (see, e.g., Saez 2010; Bastani and Waldenström 2020; Søgaard 2019; Seim 2017; and the review by Kleven 2016) or loan-term thresholds (see, e.g., Bachas, Kim, and Yannelis 2021; Bäckman, van Santen et al. 2020; DeFusco and Paciorek 2017; DeFusco, Johnson, and Mondragon 2020; and Best, Cloyne, Ilzetzki, and Kleven 2018). Our contribution is to study bunching at a threshold where the *payment* of marginally accrued taxes is substantially delayed. This adds an intertemporal dimension to bunching behavior not present in studies that consider the sensitivity to taxation.[5] We further add to the literature using income-contingent transfer schemes to identify labor supply elasticities (see, e.g., Ong 2020 who exploits the income contingency of child support.) Finally, this paper also relates to the emerging literature on the effects of debt on labor supply (see, e.g., Zator 2019; Bernstein 2021; Doornik, Gomes, Schoenherr, and Skrastins 2021; Brown and Matsa 2020; Donaldson, Piacentino, and Thakor 2019).

There is also related work considering how various tax instruments may affect behavioral elasticities. For example, Kostøl and Myhre (2020) consider how labor supply elasticities are affected by providing more information on kinks and notches, and for the price elasticity of giving, Fack and Landais (2016) consider the effect of changing documentation requirements and Ring and Thoresen (2021) consider the effect of wealth taxation.

This paper proceeds as follows. Section 2 describes the empirical setting. Section 3 presents the empirical analysis. Section 4 introduces a simple two-period life-cycle model with endogenous labor supply and financial frictions that formalizes some of the intuition introduced in this paper. Section 5 briefly discusses aspects related to the implementation and potential trade-offs associated with introducing delayed taxation.

[4] Lockwood (2020) theoretically examines how hyperbolic discounting affects the optimal timing of tax payments. Andreoni (1992) studies how financial frictions may affect tax policy, but the focus is on enforcement rather than timing. Lozachmeur (2006) studies optimal age-specific income taxation and finds that benefits from alleviating financial frictions lower the optimal tax rate for young (and more constrained) agents, but the analyses do not consider the potential optimality of delaying the payment of the tax (rather than lowering the rate itself) to achieve this benefit. Studying corporate taxation, Dávila and Hébert (2019) find that taxing payouts rather than profits is optimal in the presence of financial frictions. This essentially allows constrained firms with productive investment opportunities to delay when they pay taxes on their profits.

[5] A notable exception is Le Barbanchon (2020) who studies the response to an effective 100% *current* marginal tax that is offset by longer maximal duration of unemployment benefits.

2 Empirical Setting

The main years of study are 2004–2011. During these years, most Norwegian students faced an earnings threshold ranging from NOK 104,500 ($17,000) in 2004 to NOK 140,823 in 2011. The monthly transfers ranged from NOK 8000 ($1,300) in 2004 to NOK 9785 in 2011. These transfers are initially given as a loan, but 40% may be forgiven (converted to a stipend) to the extent that students pass classes and stay below the earnings thresholds above. Students are notified of the amount of transfers in the beginning of the academic year. These notification letters contain a breakdown of the transfers, noting the amount (40% of the total) that is given as a conversion loan, and stating that conversion from loan to stipend is contingent on incomes being below an income limit. The following year, students are notified how much of their loan was converted based on grades reported by educational institutions and income reported by the tax authorities. Loans must typically be paid off within 20 years following graduation. No interest is charged while still receiving support. Subsequently, interest rates are slightly above the risk-free rate and loan payments may be delayed at the (former) student's discretion for up to 3 years in total.[6]

This study is facilitated by administrative data hosted by Statistics Norway. The key data is derived from tax returns, including data on individuals' incomes, assets, and debts. The sample consists of students receiving standard student support for full-time studies for at least one full fiscal year during 2004–2011. We limit the sample to students who after conversion received a strictly positive stipend. This eliminates students who are ineligible for any debt-conversion due to, e.g., living at home with parents. This ensures that close to all students in our sample are subject to income-contingent debt-conversion.

Summary statistics are provided in Table 1. The average student is 23 years old. This is reasonable in light of high school graduation occurring at age 18 and that we condition on students being enrolled for higher education for both semesters within a given year. The summary statistics reveal a substantial spread in the amount of liquid assets available to students. While students at the 25th percentile only hold NOK 8,000 ($1,300) in liquid assets, students at the 75th percentile hold almost *ten* times more. A similar spread can be observed in the liquid assets of the students' parents. We further see that the average student earns around NOK 100,000 ($17,000), which is a direct consequence of our sample restrictions caused by focusing on students around the debt-conversion threshold. Four years later, the average student faces considerably higher earnings at around NOK 360,000 ($60,000).

[6] These generous terms differ from those offered in the U.S., where Gopalan, Hamilton, Sabat, and Sovich (2021) document debt responses to minimum-wage hikes that are consistent with either student debt aversion or very high perceived interest rates.

Table 1: Summary Statistics

This table provides summary statistics. The main sample period is 2004–2011. Financial variables are denominated in NOK. The USD/NOK exchange rate was around 6 in 2010. The main sample is restricted to students who had labor earnings within 50,000 of the debt-conversion threshold. Liquid Assets are made up of deposits, mutual funds, and ownership in public equity. Labor earnings are censored to be below NOK 1,000,000 in 2010 NOKs. The Bottom Tax Threshold is only considered for the years 2005–2011.

	N	Mean	p25	p50	p75
Liquid Assets$_{t-1}$	230,906	57,522	7,989	29,296	77,099
Liquid Assets$_{t-1}$ (Parents)	214,419	429,326	59,805	176,471	460,545
Age	231,036	23.4	22	23	25
Labor Earnings$_t$	231,036	101,394	81,156	98,536	118,966
Labor Earnings$_{t+4}$	229,027	357,506	226,244	372,615	464,829
Debt-Conversion Threshold$_t$	231,036	120,162	108,680	116,983	128,360
Bottom Tax Threshold$_t$	198,815	36,706	29,600	39,900	39,900

Salience. In order to meaningfully compare the implied elasticity from the debt-conversion threshold to those obtained at regular tax thresholds, the conversion threshold must be similarly salient. As past participants in this scheme, we certainly believe that to be the case. However, beyond anecdotal evidence, it is useful to consider the magnitude of the effective tax increase. A 50 percentage point reduction in the net-of-debt wage is unlikely to go unnoticed. In addition, students are informed of the presence of such a limit in a loan-agreement letter, which they must sign, and additionally receive letters informing them of any conversion that has taken place. Even if students are not expecting any debt-conversion reduction, students would want to read these letters to confirm that their educational institution has recorded and reported academic progress correctly. Non-passing grades in courses also reduce debt conversion. Students are also informed of their annual student debt balances when they receive their annual pre-filled tax returns that also contain information on their income tax liabilities.

2.1 Bunching methodology

The purpose of the bunching methodology is to estimate the earnings elasticity,

$$e = \frac{\Delta y^*/y^*}{\Delta\tau/(1-\tau)}, \tag{1}$$

where Δy^* is the reduction in earnings of the marginal buncher who is at an interior optimum at the debt-conversion threshold (i.e., the kink). The bunching mass is denoted B. By construction (see Saez 2010 and Kleven 2016 for graphical intuition), B equals $\int_{y^*}^{y^*+\Delta y^*} h_0(y)dy$, where $h_0(y)$ is the counter-factual (absent a kink) probability density function of earnings. We apply the standard approximation

$$B = \int_{y^*}^{y^*+\Delta y^*} h_0(y)dy \approx h_0(y^*)\Delta y^*. \tag{2}$$

Dividing through by y^*, we may write the (approximated) relative change in earnings of the marginal buncher as

$$\frac{\Delta y^*}{y^*} = \frac{B}{h_0(y^*)y^*} = \frac{b}{y^*}. \tag{3}$$

This is equation represents one of the central insights in the bunching literature, namely that the marginal buncher's earnings reduction caused by the kink is proportional to the excess mass at the kink.

We empirically estimate b, the relative excess mass at the threshold, using the methodology in Chetty et al. (2011),[7] which we call the bunching estimate. The empirical analogue of y^* is the (average) debt-conversion threshold denominated in the same units (thousands) as the empirical earnings bins.[8] We write our estimated compensated labor earnings elasticity as

$$\hat{e} = \frac{\hat{b}/y^*}{\widehat{\Delta\tau}/(1-\tau)}, \tag{4}$$

where $\widehat{\Delta\tau}$ is the estimated change in the effective nominal tax rate occurring at the debt-conversion threshold, and τ is the at-threshold after-tax keep rate of $1 - \tilde{\tau} = 0.75$.

3 Empirical Analysis

3.1 Bunching at the debt-conversion threshold

In the presence of delayed taxation, the tax rate $\tilde{\tau}$—according to which the agent optimizes—differs from the nominal tax rate, τ. This means that the standard result that e equals the Frisch elasticity of labor supply as in Saez (2010) does not necessarily hold. In Section 4, we outline a simple two-period model, in which the agent faces a labor income tax where only a fraction δ is payable in the current period. In this model, the agent behaves as if facing a standard (payable-today) labor income tax of

[7] See Seim (2017) for another application that combines the Saez (2010) and Chetty et al. (2011) approaches.

[8] Alternatively, we could multiply $\hat{B}$ and thus $\hat{b}$ by the width of the earnings bins (NOK 1,000), and let z^* equal the threshold in NOK.

$$\tilde{\tau} = \delta\tau + \frac{1-\delta}{\tilde{R}}\tau, \tag{5}$$

where $\tilde{R}$ is the relevant marginal (gross) cumulative interest rate faced by the agent, and τ is the nominal effective tax rate.

If taxes are payable today ($\delta = 1$), e may provide an estimate of the Frisch elasticity of labor supply. However, in the current setting, $\delta = 0$, as non-converted student loans are paid in the future. Thus, in order to relate our empirical estimate of e to the Frisch elasticity of labor supply, we need to know the applicable marginal gross interest rate of bunchers, $\tilde{R}$.

Since this is unobservable, we proceed as if $\delta = 1$. Then, in Section 3.4, we estimate the implied elasticity from bunching at a regular tax threshold (i.e., $\delta = 1$ and thus $d\tilde{\tau} = d\tau$) and compare the two elasticities to find the $\tilde{R}$ that would allow them to be consistent with the same *structural* labor supply elasticity.

Figure 1 shows some key details of the empirical analysis. Panel A verifies that earnings above the threshold lead to an increase in next period debt. Most students are on the expected kinked trajectory where each additional NOK of earnings increases debt by 0.50 NOK. The blue fitted line illustrates how we obtain our first-stage measure of the effect of excess earnings on debt accumulation. We find that the slope increases by 0.47. This is close to the nominal increase of 0.50 due to very few non-compliers.[9] Cast in terms of the previous notation, this implies that $d\tau = 0.47$.

In Panel (B), the yellow dotted line shows the distribution of students around the earnings threshold. The green line is the counter-factual distribution, which is a 5th-order polynomial fitted to the non-bunching region. We obtain a measure of the excess mass of individuals near the threshold by comparing the actual and counter-factual distributions. This offers a bunching estimate, b, of 1.21, which means that there are 121% more individuals around the threshold than the counter-factual distribution implies. Dividing 1.21 by the average threshold amount (120.162 in NOK 1,000s), per equation 4, we obtain a remarkably low elasticity of labor earnings to the net-of-tax (or net-of-debt-increase) wage of 0.0162.[10] The standard error is 0.0013.[11]

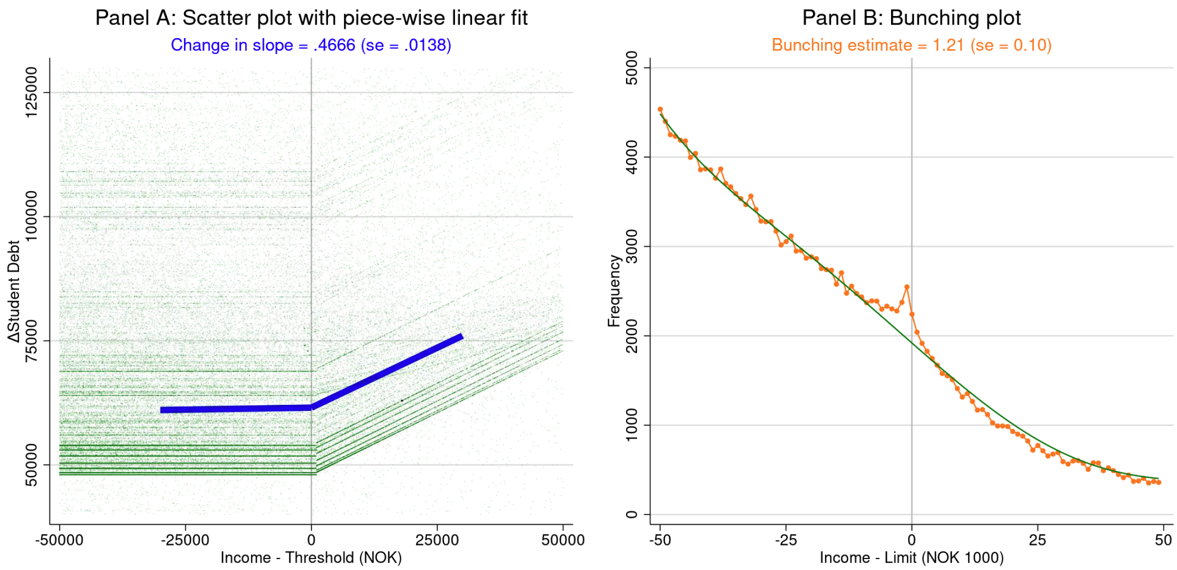
[9] Some non-compliers exist, for example, because they may have moved in with their parents during the fall semester, which would exclude them from receiving any conversion for fall semester loans. Such moves must be reported to the educational loan fund, but not to the tax authorities from which we receive address data.

[10] These calculations do not adjust for the fact that any accumulated debt is interest-free while in school. Adjusting for a 3-year 3%-interest discount would increase the elasticity by around 9%.

[11] We ignore the (very small) standard errors involved with estimating the change in debt per additional NOK earned above the threshold.

Figure 1: Verifying the Effect of Excess Earnings on Future Debt and Examining Bunching Responses

Panel (A) provides a scatter plot, in green, of the relationship betweeen debt accumulation and student earnings around the debt-conversion threshold. The fitted blue line illustrates the estimation of the effect of earnings in excess of the threshold and accumulated debt. Panel (B) provides a graphical illustration of how the bunching estimate. The orange connected line shows the actual distribution of students around the conversion threshold. The fitted green line shows the estimated counterfactual distribution. The bunching estimate provides the relative excess mass (actual versus counterfactual) of students near the threshold. This is done using the Stata .ado file provided by Chetty, Friedman, Olsen, and Pistaferri (2011). This program calculates the excess bunching between NOK -10,000 and NOK 6,000. Standard errors are computed from bootstrapping (1,000 reps). All plots represent statistics from the pooled sample years 2004–2011.



This analysis shows that students are remarkably irresponsive to de-facto delayed taxation. We show that the results are virtually identical when considering students employed in likely highly flexible hospitality and sales positions in Figure A.2. We further find qualitatively similar results when considering bunching around the conversion cap. This is where additional earnings no longer increase student debt because students are no longer eligible for *any* conversion from loans to stipends. We report these results in Figure A.1. We find that the bunching estimate is, in accordance with theory, negative, but statistically close to zero (t-stat=-1.64).

3.2 Determinants of non-bunching

We now investigate potential determinants of this (non-)bunching behavior. Our main approach is to plot student characteristics against their position relative to the conversion threshold.[12] This is a visual exercise where we attempt to draw conclusions from visual breaks in the relationship between a given characteristic and students' earnings occurring around the conversion threshold.

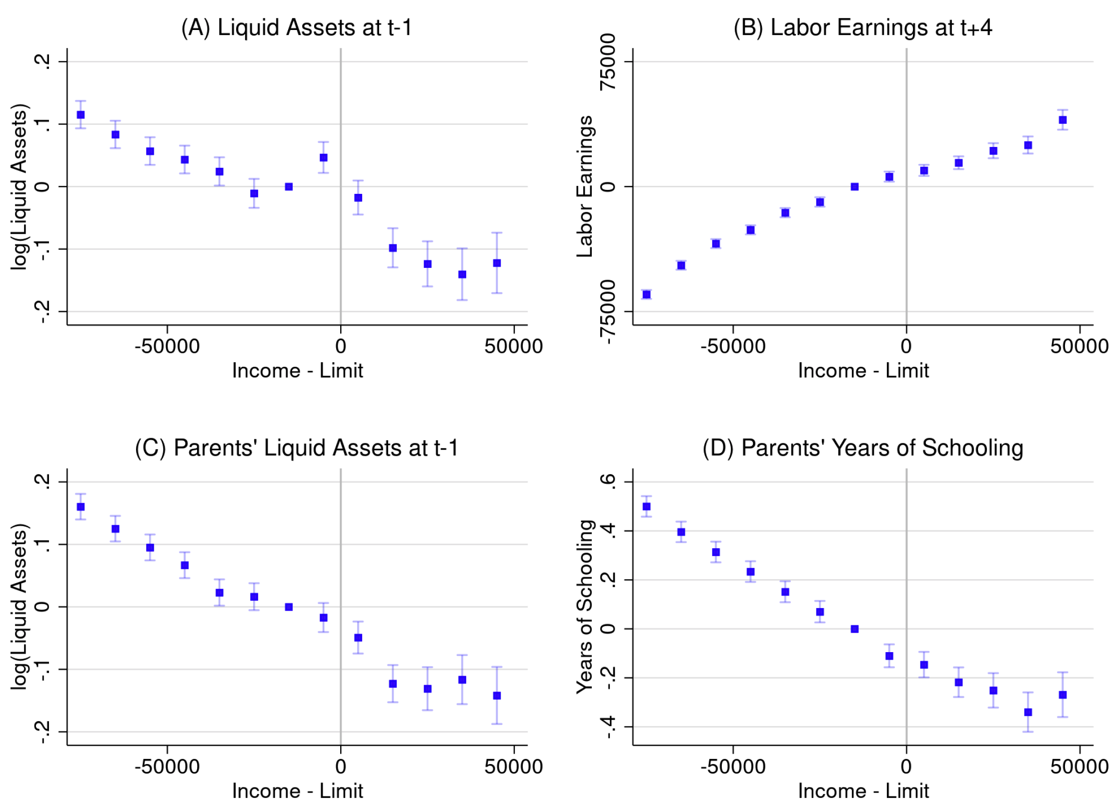
[12]Another application of this type of analysis can be found in concurrent work by Bastani and Waldenström (2020) who examine how ability covaries with taxpayers' position relative to a regular tax threshold to infer the ability gradient in tax responsiveness.

In Figure 2, Panel (A), we find that the amount of ex-ante liquid assets drops sharply right above the threshold. This suggests that non-bunchers have less liquid wealth, consistent with these students being financially constrained. Panel (B) of Figure 2 shows how future labor earnings vary with the student's position relative to the threshold. This reveals no sharp rise or decrease in realized future incomes above the threshold, which suggests that non-bunchers do not differ significantly in terms of medium-term earnings prospects.

Taken together, these findings emphasize financial frictions as a key channel in driving the insensitivity to the conversion threshold. Those who earn above the threshold have similar future earnings prospects, but have significantly less liquid assets. Holding less assets may both causally affect the extent to which the agents are constrained and be a proxy for financial frictions as it indicates a preference towards smoothing consumption toward the present.

Figure 2: Characteristics of Students Below and Above the Income-contingent Debt-conversion Threshold

The graphs below show the financial characteristics of students who are near the threshold. Panel A considers the liquid assets of students. These consist of deposits, stocks, bonds, and mutual fond holdings. Panel B shows future log labor earnings, measured 4 years later. Panel C shows the amount of liquid assets held by the student's parents. Panel D shows the educational attainment of the parents, measured as the maximum number of years of school among the set of parents. Standard errors used to provide 95% confidence intervals are clustered at the student level.



To investigate this liquidity channel further, we also show how *parents'* liquidity correlates with the students' earnings location in Panel (C) of Figure 2. This documents a noteworthy negative relationship between the parents' financial resources and the in-school labor earnings of the child. This suggests that parents play an important role in determining the amount of time students may dedicate to their studies. More relevant to the present study, is the finding that parents' assets drop shortly above the earnings threshold. This indicates that non-bunchers have access to fewer financial resources, which is consistent with financial frictions playing a key role in driving the observed non-responsiveness to the conversion threshold. However, wealth may proxy for human capital which influences tax responsiveness (Bastani and Waldenström, 2020). Therefore, we plot parental educational attainment on the y-axis in Panel (D). This shows that there is no break in the relationship between educational attainment, measured in the maximum years of schooling among the parents and the child's position relative to the conversion threshold. This addresses the hypothesis that less resources, in a human capital, rather than financial, sense can explain the irresponsiveness to the threshold. If anything, extrapolating from the below-threshold relationship, non-bunchers may have higher-educated parents. To the extent that this is correlated with the students' life-time wealth, this may explain some of the desire of students to front-load consumption through incurring higher student loans.

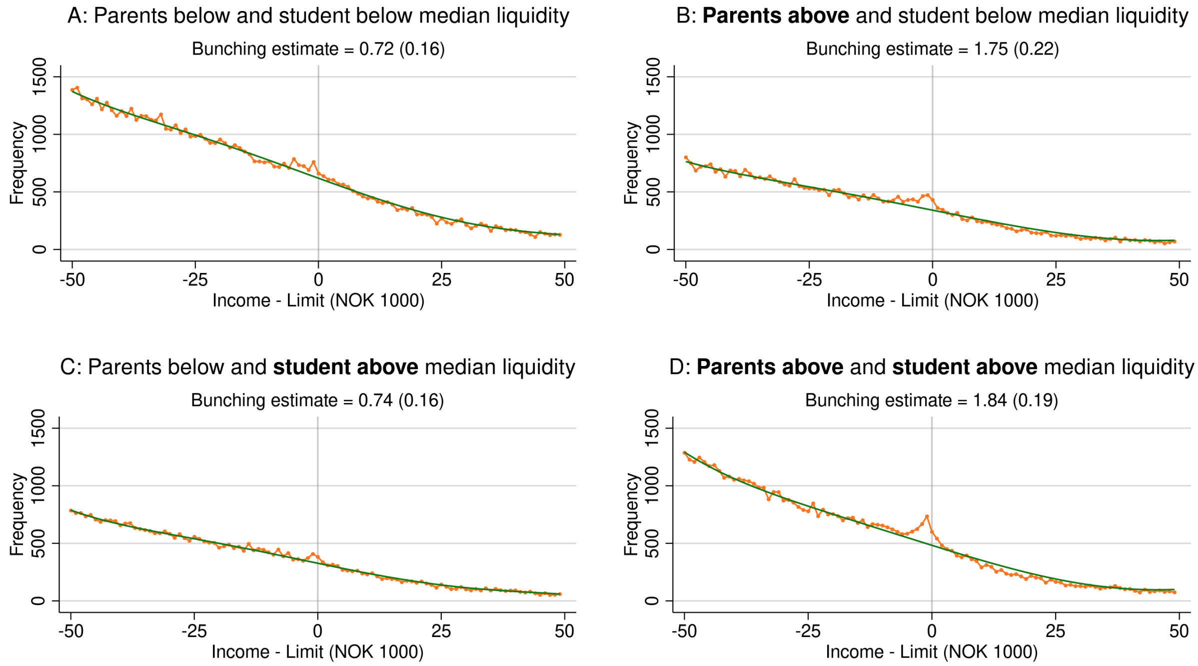
3.3 Bunching heterogeneity

We proceed with a supplementary, more standard approach of investigating heterogeneity in earnings sensitivity to the threshold in Figure 3. This approach splits the sample into subsets based on student and parental characteristic to compute heterogeneous bunching elasticities. We see that the largest contribution to the total excess mass in the preceeding Figure 1 is from students who themselves and their parents have above-median liquid assets. Figure 3 also suggests that the main driver of bunching responses is the parents' rather than the students' own liquid assets. Moving from the left to the right panels, which improves the parents' liquidity, more than doubles the bunching estimates.[13]

[13]In this case, it doesn't matter whether we compare excess mass in terms of students or earnings, since bin widths and thresholds are the same.

Figure 3: Heterogeneity in Bunching by Amount of Liquid Assets

These plots calculate the bunching elasticity for different subsamples. Students are split into four subsamples based on whether their and their parents' $\text{LiquidAssets}_{t-1}$ are below or above median. These medians are calculated separately for each year in the sample.



What can this heterogeneity tell us about how the severity of financial constraints vary with parents' liquidity? The source of variation in the implied labor supply elasticities across the liquidity subsamples are the bunching estimates provided in Figure 3. The ratio of bunching estimates therefore provide the relative implied elasticities. If we impose the same structural labor supply elasticity (e.g., same constant Frisch elasticity of labor supply) across the samples, differences can only be attributed to differences in the (gross) marginal rates of return on net saving, $\tilde{R}$ (see Proposition 1 in Section 4 for a theoretical example). This follows from replacing writing out $\tilde{\tau}$ (as defined in equation 5) in the expression for the $\hat{e}$ in equation 4, and setting the fraction payable today, δ, to zero.

In Figure 3, we see that the elasticity increases from 0.72 to 1.84 when moving from below to above the median in terms of both students' and their parents' resources. This thus implies that the doubly-below median students have on average have gross marginal rate that is 2.56 times larger. Annualizing this, assuming a 10-year horizon, implies a $2.56^{1/10}$=1.0986 times higher annualized gross marginal rate or approximately a 10 percentage point higher interest rate.[14]

[14] If the doubly-above median group has a baseline gross interest rate of 1.10, then the below-median group has a marginal rate that is (1.0986-1)*1.10 = 10.85 pp. higher.

3.4 Analysis of bunching at a regular tax threshold

In this section, we repeat the introductory analyses done in Figure 1 using a *tax* threshold rather than the debt-conversion threshold. The purpose of this exercise is to obtain a reference estimate of the implied labor earnings elasticity at a tax threshold where marginally accrued taxes are not delayed. We focus on the first tax threshold in the progressive income taxation system. This threshold as located at NOK 30,000 during 2005–06 and NOK 40,000 during 2007–2011.[15] At this threshold, the marginal income tax increases from 0 to around 25 percent for most taxpayers.

In Figure 4, we investigate this complementary empirical setting. Panel (A) provides a scatter-plot which verifies the presence of a rise in the marginal income tax rate by plotting total taxes accrued that year against incomes. It also provides the fitted kinked line, from which we infer an average increase in the marginal tax rate of 19 percentage points at the threshold. The coefficient is lower than the nominal increase of 25 percentage points since some individuals may be eligible for higher standard deductions.

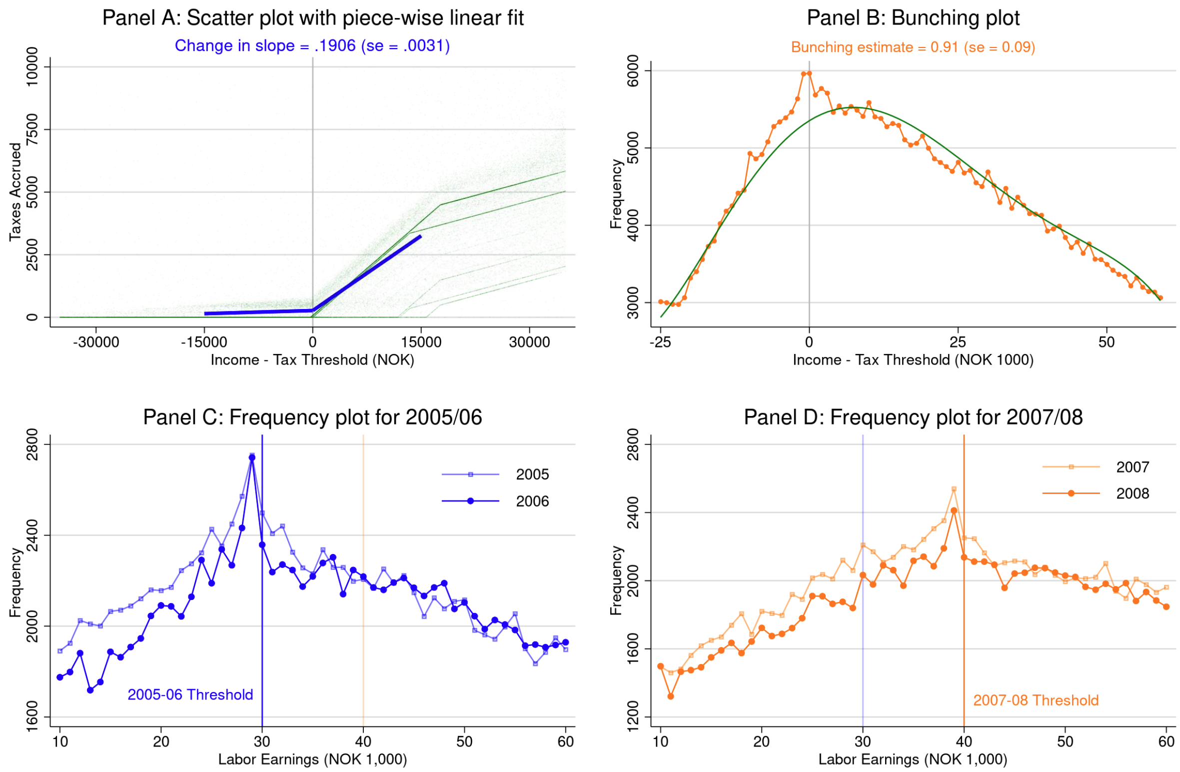
Panel (B) illustrates how the bunching estimate of $b = 0.91$ is calculated. While this bunching estimate is smaller than that found at the debt-conversion threshold, this one-to-one comparison is uninformative for two reasons. Firstly, we need to divide 0.91 by the threshold (36.706 in NOK 1,000s) to obtain a relative reduction in earnings for the marginal buncher of 2.48%. This is already larger than the reduction we found at the debt-conversion threshold of 1.00% (1.21 divided by 120.162). Secondly, we need to account for the fact that this is in response to a lower increase in the marginal (nominal) tax rate. Dividing 2.48% by the relative reduction in the after-tax keep rate of 19.6%/100%, we obtain a more substantial elasticity of 0.13

In Panel (B), we see that the bunching mass occurs at the mode of the distribution. If the location of the mode is not driven by students' responses to the tax threshold, then the co-location of the mode and threshold could lead to an upward bias in the bunching estimate. To address this concern, we show in Panels (C) and (D) that the location of the mode is driven by the location of the tax threshold. From 2005 to 2006 and from 2007 to 2008 there was no changes to the mode of the distribution. However, when the tax threshold rose from 2006 to 2007, the mode precisely followed. This reassures us that there is indeed substantial responsiveness to the tax threshold not driven by happenstance co-location of the mode and threshold.

[15] We omit 2004. During this year the threshold was only NOK 23,000, which substantially reduces how much of the left tail we can use to estimate a counterfactual distribution.

Figure 4: Bunching at a Regular Tax Threshold

The first and second plots shows the relationship between labor incomes ("pensionable income") and taxes accrued that year (payable same or next year) in the form of a scatter and binscatter plot, respectively. The third plot shows the distribution of students around the income tax threshold. The fourth plot calculates the bunching elasticity, in terms of the implied excess fraction of students located in the NOK 1,000 bin directly to the left of the threshold using the Stata .ado file provided by Chetty, Friedman, Olsen, and Pistaferri (2011). This program calculates the excess bunching between NOK -10,000 and NOK 6,000. Standard errors are computed from bootstrapping (N=1,000). All plots represent statistics from the pooled sample years.



The elasticity of 0.13 is eight times larger than the elasticity of 0.0162 found in when analyzing responsiveness to the debt-conversion threshold. For these to be consistent with the same structural labor supply elasticity, the average relevant cumulative gross marginal rate would have to be 8. Given that these loans have an average maturity of 10 years, this necessitates an annualized net discount rate of 23%=$8^{1/10}$-1. This number is comparable to average credit-card rates that lie slightly above 20%.[16] This implies that students are willing to borrow from the educational loan fund at a rate exceeding that offered by financial institutions. This may be driven in part by credit rationing, but likely primarily from the fact that the loan fund does not require payments while students are still in school and generally have a long maturity with the additional opportunity to delay payments for up to three years.

We can use this implied elasticity to get an idea of how much bunching would be caused by the debt-conversion threshold in the absence of financial frictions. In other

[16]Source: Statistics Norway's Statistics on Interest Rates in Banks and Credit Institutions, source table 12844, 2019Q4: 21.6%

words, how much bunching would there be in Figure 1 if students responded to the debt-conversion threshold as if it were a regular income tax threshold? To find this, we reverse the calculation used to infer labor supply elasticities from bunching estimates. This offers a counter-factual bunching estimate of 23.43.[17] This is considerably larger than the empirical bunching estimate of 1.21.

3.5 Regression-based approach to compare elasticities

In this section, we pool the samples used to examine bunching at the debt-conversion (delayed-tax) and regular-tax thresholds. We develop a regression-based approach that allows us to compare the underlying elasticities while keeping observables fixed.[18] This addresses the fact that higher-earning students in the debt-conversion sample may have different characteristics than those in the lower-earning regular-tax sample. We want to address the fact that differences in observable characteristics, such as occupation, may partially explain differences in bunching behavior.

We first define the individual-level elasticity as

$$\tilde{e}_i = \underbrace{\frac{1[y_i \in BR_s] - \hat{P}^{cf}(y_i \in BR_s)}{\hat{P}^{cf}(y_i \in BR_s)/N_s^{bins}}}_{\text{estimated } b_s} \cdot (\text{Bin width}_s/\text{Threshold}_s) \Big/ (\hat{d}\tau_s/(1-\tau_s)), \tag{6}$$

where $\hat{P}^{cf}$ denotes the estimated (counter-factual) probabilities of being in the bunching region absent any tax or debt-conversion kinks. This is estimated using the frequencies in the earnings bins around the bunching region as in Saez (2010). The s-sample mean, $\hat{E}$, of $\tilde{e}_i$ provides an estimate of the implied labor supply elasticity. For the delayed tax sample, this mean is around 0.0155, and thus very close to our baseline estimate of 0.0162.[19]

[17]=0.13*(120162/1000)*(75/50)

[18]See Ring and Thoresen (2021) for a related method, in which regressions of a bunching indicator on observables is used to infer bunching heterogeneity.

[19]The new estimate for the regular tax sample is about 0.2, which is larger than our baseline estimate for the regular tax threshold of 0.13. However, the graphical evidence in Figure 4 shows that this was likely a very conservative estimate. Differences arise because in the regression-based approach, we take the simpler approach of estimating the P^{cf}s using the observed number of observations in the two income bins right below and the two income bins right above the bunching region, BR_s (as in Saez 2010) rather than estimating a higher-order polynomial (as in Chetty et al. 2011).

Table 2: Regression-based Difference between Delayed-Taxation and Regular Labor Supply Elasticities

This table provides results from the regression-based approach to comparing the labor supply elasticities in the delayed-tax and regular-tax samples. The estimated relative elasticity difference is calculated as the coefficient on 1[regular tax sample] divided by $\hat{E}[\tilde{e}_i \mid s = delayed]$. We only keep observations for which we observe an employer-employee relationship, and thus can assign NACE and occupation codes based on the students' within-year highest-paid job spell. Standard errors are provided in parenthesis.

	(1)	(2)
Estimated Relative Difference in Elasticity		
$\widehat{\frac{e_{regular}-e_{delayed}}{e_{delayed}}}$	7.20	6.10
	(.59)	(.61)
Underlying Regression Coefficients		
1[regular tax sample]	0.0969***	0.0787***
	(0.0093)	(0.0094)
Male	0.0360***	0.0414***
	(0.0100)	(0.0102)
Age	-0.0434***	-0.0410***
	(0.0022)	(0.0022)
College, parents	0.0501**	0.0442**
	(0.0204)	(0.0205)
Years of schooling, parents	0.0056	0.0070**
	(0.0035)	(0.0036)
N	393443	390177
R2	0.01	0.02
$\widehat{E}[\tilde{e}_i \mid s = regular]$	0.2031	0.2032
$\widehat{E}[\tilde{e}_i \mid s = delayed]$	0.0156	0.0154
FEs	4-Digit Occ	4-Digit Occ × NACE2

We then estimate regression equations of the following form.

$$\tilde{e}_i = \alpha + \beta \mathbb{1}[\text{regular tax sample}]_i + \gamma' X_i + \varepsilon_i, \tag{7}$$

where y_i is individual labor earnings and X_i is a vector of individual-level observables, such as their 4-digit employee occupation code if available. We report the results from varying the contents of X_i in Table 2. To find the estimated relative increase in labor supply elasticities in the regular versus delayed-taxation samples, we divide $\hat{\beta}$ by the delayed-tax sample mean of $\tilde{e}_i$.

The main finding is that the relative difference in labor supply elasticities is about

6.1 once we address differences on observables. This is a little lower than what we found earlier, but the qualitative implications are the same: to rationalize a 6.1 times higher elasticity, we need an annualized net discount rate of 19.82%=$6.1^{1/10}$-1.

4 A simple model of labor supply under delayed taxation

4.1 Labor supply decision.

This section introduces a simple model that can guide the comparison of the implied labor supply elasticities. The central take-away is Proposition 1, which emphasizes the role of marginal rates of return on net saving in differential responses to regular and delayed taxation. However, we also use the simple model to study how a government may extract additional tax revenues through a non-distortionary tax reform, which results in Proposition 4.

Model environment. The agent works and consumes for two periods. A fraction, δ, of period 1 taxes taxes are paid in the first period. The remainder, $1-\delta$, is paid in the second. The period 2 wage, w_2, is net-of-tax, and payable in period 2. The agent faces the following maximization problem.

$$\begin{aligned} \max_{c_1,c_2,l_1,l_2,s} \quad & u(c_1,l_1)+\beta u(c_2,l_2), && (8)\\ \text{s.t.}\ \ c_1+s &= y_1+l_1w_1(1-\tau\delta) && (9)\\ \text{and } c_2 &= \bar{R}(s)+y_2+l_2w_2-l_1w_1\tau(1-\delta). && (10) \end{aligned}$$

Where w_1 is the first-period gross age, c_t is consumption, l_t is labor supply, l_t is exogenous income, s, and s is net savings. τ is the nominal tax rate for period 1 income.

The financial friction is the following. When agents save an amount s greater than $\bar{s}<0$, they face a gross interest rate of 1, which is the same gross interest rate that the tax authorities charge on delaying tax payments. When they save less than $\bar{s}$ (in general, borrow), they face a marginal gross interest rate $\tilde{R}>1$. Thus, an agent who saves s in period 1, enters period 2 with the following amount of net savings.

$$\begin{aligned} \bar{R}(s) &= s+(\tilde{R}-1)\mathbb{1}[s<\bar{s}](s-\bar{s}) && (11)\\ &\equiv s\tilde{R}-(R-1)\mathbb{1}[s<\bar{s}]\bar{s}, && (12) \end{aligned}$$

where $\tilde{R}=1+(R-1)\mathbb{1}[s<\bar{s}]$ is the agent's marginal gross interest rate.

Assume that we have additively separable (dis)preferences for consumption and labor supply, and that the per-period utility takes the following form:

$$u(c,l) = \frac{1}{1-\gamma}c^{1-\gamma} - \psi\frac{l^{1+\nu}}{1+\nu}. \tag{13}$$

We will focus on the cases where the first order conditions (FOCs) bind. In other words, we consider cases where the optimally chosen s is not at the kink point, $s = \bar{s}$, in the agent's budget constraint. We can think of this as focusing on either the unconstrained agent, where $s > \bar{s}$, or the highly constrained agent who chooses $s < \bar{s}$ even if this entails borrowing at $\tilde{R} = R > 1$. This could mimic a setting in which the only source of loans available is high-interest credit cards. We would conjecture that the responses of any agent who optimally chooses the kink point, $s = \bar{s}$, would be consistent with behavior "in the middle" of these two types of agents that we analyze.[20]

FOCs for these cases, where the optimal s is different from $\bar{s}$, are:

$$\begin{aligned}
s \;:& \quad c_1^{-\gamma} - \beta\tilde{R}c_2^{-\gamma} &= 0 \quad (14)\\
l_1 :& \quad w_1(1-\tau\delta)c_1^{-\gamma} - \psi l_1^{\nu} - w_1\tau(1-\delta)\beta c_2^{-\gamma} &= 0 \quad (15)\\
l_2 :& \quad w_2c_2^{-\gamma} - \psi l_2^{\nu} &= 0 \quad (16)
\end{aligned}$$

Wee see that the delayed tax scheme alters the standard optimization problem by introducing the third term in equation 15. Effectively, it adds an intertemporal component to the standard intratemporal trade-off between leisure and consumption.

To simplify the main proposition below, it is useful to define a somewhat stricter notion of not being at the kink point $s = \bar{s}$. This will allow us to not worry about agents hitting the kink point if we make changes to the tax environment.

Definition of IHS: The intertemporal first-order condition (14) holds strongly (**IHS**) whenever the the agent could increase saving by the delayed portion of period 1 income taxes without changing the marginal interest rate. More formally, this condition says that

$$s + l_1w_1\tau(1-\delta)/R < \bar{s} \quad \text{or} \quad s > \bar{s}. \tag{17}$$

This condition is weak in the sense that it can always be satisfied by considering a small enough fraction, $1-\delta$, of period 1 taxes that are delayed.

Proposition 1. It does not matter whether the agent faces a tax where a fraction,

[20] Think of the agents at the kink points as "moderately constrained" agents who would choose to borrow if the interest rate they faced were at least slightly lower than R.

$1-\delta$, is paid in period 2 or whether the agent faces a tax rate in period 1 of $\tilde{\tau} = \tau(\delta + (1-\delta)/\tilde{R})$, i.e., a discounted tax, whenever the IHS condition holds.

Proof: First, note that δ only enters in to the period 1 intratemporal first-order condition (15). Use equation 14 to substitute for $\beta c_2^{-\gamma}$ in equation 15. Then we see below that a tax of $\tilde{\tau}$, payable in period 1, is equivalent, in terms of the FOCs, to the case where δ of the tax is paid in period 2.

$$w_1\Big(1-\tau\{\delta+(1-\delta)/\tilde{R}\}\Big)c_1^{-\gamma}-\psi l_1^{\nu} \quad = 0 \tag{18}$$

By adding $l_1 w_1(1-\delta)\tau/\tilde{R}$ to both sides of the period-1 budget constraint, we see that the period-1 budget constraint still holds with equality when shifting from a $(1-\delta)$ delayed tax to a payable-today tax of $\tilde{\tau}$. This operation is equivalent to adding $l_1 w_1 \tau(1-\delta)/\tilde{R}$ to savings, s. Hence, we must add this increase to s in the period-2 constraint. This then cancels out the delayed tax payment term exactly when the IHS holds (we need $\tilde{R}$ to remain constant).

In other words, we can replace the delayed tax by a period 1 tax with the present value from the perspective of the agent, while leaving first-order conditions unaffected and still satisfy the budget constraints. The only thing that changes is that s increases. This holds true for any δ, but the IHS is easier to satisfy if δ is close to 1.

Corollary 1. If the elasticity of labor supply to the net of tax wage is positive (negative) and $\tau > 0$, then marginally decreasing δ – allowing more to be paid later – will strictly increase (decrease) period-1 labor supply for constrained agents, i.e., $s < \bar{s}$), whenever the IHS holds.

Proof: From Proposition 1, we know that the agent optimizes as if facing a regular income tax, payable during period 1, of $\tilde{\tau} = \tau(\delta + (1-\delta)/\tilde{R})$. A marginal decrease in δ (which delays taxation) will thus strictly decrease the effective tax rate, $\tilde{\tau}$, when $\tilde{R} > 1$, and thereby strictly increase labor supply whenever the labor supply is increasing in the net-of-tax wage is positive:

$$\frac{dl_1}{d\delta} = \frac{dl_1}{d\tau}\Big|_{\delta=1}\frac{d\tilde{\tau}}{d\delta} = \frac{dl_1}{d\tau}\Big|_{\delta=1}\tau(1-1/\tilde{R}). \tag{19}$$

4.2 Taxation from government's perspective.

Constant-Distortion Reform. Suppose there is currently a non-delayed tax regime with period-1 tax rate τ_0. There is only one taxpayer, and this taxpayer is constrained with $\tilde{R} > 1$. If the government introduces fully delayed taxation ($\delta = 0$), but wishes to keep the effective tax rate constant according to Proposition 1, the new tax rate must be $\tau_0\tilde{R}$. The government foregoes period-1 tax revenues of $\tau_0 l_1 w_1$ to obtain $\tau_0\tilde{R}l_1w_1$ in period 2, effectively earning a gross return of $\tilde{R}$. The only thing that changes in the economy is that household borrowing $(-s)$ increases by $\tau_0 l_1 w_1$, which is (potentially) offset by the same increase in government borrowing. In this scenario, the government does not increase over-all surplus, but rather extracts it from lenders.

Delayed taxation and government tax revenues. In the same economic environment as above, consider a simpler reform of $d\delta < 0$. We explore the effect on government revenues, assuming that the government's gross discount rate equals the market rate of 1. In Lemma 3 (Appendix A.1), we show that the NPV effect on government revenues is positive whenever the ratio of the Frisch elasticity, $\frac{1}{\nu}$, to the uncompensated elasticity, $-\frac{dl_1}{d\tilde{\tau}}\frac{1-\tilde{\tau}}{l_1}$ is higher than the ratio of life-time to period-1 taxes.

Since the compensated elasticity governs substitution effects, and income effects work in the other direction in determining the uncompensated effect, this is equivalent to saying that substitution effect must be larger than income effects. Roughly, if pre-reform tax revenues are equal in each period, we need substitution effects to be at least twice as large as income effects. The fact that they need to be twice as large, as opposed to just larger, is driven by intertemporal labor substitution.

The requirement that substitution effects must be relatively larger than income effects in order for delayed taxation to increase revenues is rather intuitive: income effects reduce the distortions from taxation. Hence it is when income effects are small relative to substitution effects that there is the most surplus obtainable from reducing distortions.

5 Discussion

This paper introduces the hypothesis that delaying labor income tax payments may reduce their distortionary effects in the presence of financially constrained agents. We exploit a unique setting in Norway that allow us to test this hypothesis empirically. Our results indicate that delaying the payment of taxes, while keeping time of accrual constant, materially reduces the distortionary effects of income taxation when agents are credit constrained. These findings highlight delayed taxation as a promising new tool in optimal taxation and a fertile ground for more theoretical and empirical research.

Delayed taxation amounts to government-provided income tax financing. Even if providing financing at the market rate, the government effectively earns an interest premium through higher tax revenues. Whether such a system of delayed taxation is optimal from a partial-equilibrium revenue-maximization perspective thus depends on whether the effective interest premium exceeds potential administrative and default costs.

The most feasible implementation of a system of delayed taxation is likely in connection with raising a given marginal income tax rate. Imagine an economy with a flat tax rate of 30%. Policy-makers are considering increasing the marginal rate to 40% at some threshold, say $100,000. Allowing marginal taxes (10% on the amount above $100,000) to be delayed is likely to reduce the distortionary effects among constrained agents. By construction, tax liabilities would only accrue to high-income individuals, which may limit the potential severity of issues such as adverse selection.

In a life-cycle model calibrated to U.S. workers, Scott, Shoven, Slavov, and Watson (2021) find that workers aged 25 would require a match rate above 1800% to participate in employer-sponsored retirement saving, which is driven by borrowing constraints and an upward-sloping earnings profile. This match rate decreases with age and approaches zero around age 45. This provides a useful statistic to explore the potential effects of delayed taxation: It indicates that the average 25 year old may be *eighteen times* less responsive to a labor income tax that can be paid at retirement and that incentives to delay accrued taxes would start vanishing at age 45. This suggests that a reasonable implementation of delayed taxation might involve age limits if policy-makers trade off tax-revenue effects and potential costs from mortality-induced non-payment.

There may be important costs associated with non-payment or debt-overhang (see, e.g., Donaldson, Piacentino, and Thakor 2019 and Cespedes, Parra, and Sialm 2020) induced by such a scheme, particularly if, e.g., hyperbolic discounting plays an important role in determining the extent to which the agent is constrained. These costs should be weighed against the potential benefits from reduced distortions while bearing in mind that the effect on over-all indebtedness may be limited due to substitution away from other sources of credit.

References

Amromin, G., J. Huang, and C. Sialm (2007): "The tradeoff between mortgage prepayments and tax-deferred retirement savings," *Journal of Public Economics*, 91, 2014–2040.

Andreoni, J. (1992): "IRS as loan shark tax compliance with borrowing constraints," *Journal of Public Economics*, 49, 35–46.

Bachas, N., O. S. Kim, and C. Yannelis (2021): "Loan guarantees and credit supply," *Journal of Financial Economics*, 139, 872–894.

Bäckman, C., P. van Santen, et al. (2020): *The Amortization Elasticity of Mortgage Demand*, Aarhus BSS, Aarhus University, Department of Economics and Business Economics.

Bastani, S. and D. Waldenström (2020): "The Ability Gradient in Tax Responsiveness," .

Bernstein, A. (2021): "Negative home equity and household labor supply," *Journal of Finance, forthcoming*.

Best, M. C., J. Cloyne, E. Ilzetzki, and H. Kleven (2018): "Estimating the elasticity of intertemporal substitution using mortgage notches," Tech. rep., National Bureau of Economic Research.

Brown, J. and D. A. Matsa (2020): "Locked in by leverage: Job search during the housing crisis," *Journal of Financial Economics*, 136, 623–648.

Carvalho, L. S., S. Meier, and S. W. Wang (2016): "Poverty and economic decision-making: Evidence from changes in financial resources at payday," *American Economic Review*, 106, 260–84.

Cespedes, J., C. Parra, and C. Sialm (2020): "The Effect of Principal Reduction on Household Distress: Evidence from Mortgage Cramdown," *Available at SSRN*.

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

Dávila, E. and B. M. Hébert (2019): "Optimal Corporate Taxation Under Financial Frictions," *NBER Working Paper*.

DeFusco, A. A., S. Johnson, and J. Mondragon (2020): "Regulating household leverage," *The Review of Economic Studies*, 87, 914–958.

DeFusco, A. A. and A. Paciorek (2017): "The interest rate elasticity of mortgage demand: Evidence from bunching at the conforming loan limit," *American Economic Journal: Economic Policy*, 9, 210–40.

Donaldson, J. R., G. Piacentino, and A. Thakor (2019): "Household debt overhang and unemployment," *The Journal of Finance*, 74, 1473–1502.

Doornik, B. F. N. V., A. R. Gomes, D. Schoenherr, and J. Skrastins (2021): "Financial Access and Labor Market Outcomes: Evidence from Credit Lotteries," *Available at SSRN 3800020*.

Epper, T. (2017): "Income Expectations, Limited Liquidity, and Anomilies in Intertemporal Choice," *Working Paper*.

Fack, G. and C. Landais (2016): "The effect of tax enforcement on tax elasticities: Evidence from charitable contributions in France," *Journal of Public Economics*, 133, 23–40.

Golosov, M. and A. Tsyvinski (2015): "Policy implications of dynamic public finance," *Annual Review of Economics*, 7, 147–171.

Gopalan, R., B. H. Hamilton, J. Sabat, and D. Sovich (2021): "Aversion to student debt? Evidence from low-wage workers," *Evidence from low-wage workers (May 11, 2021)*.

Heathcote, J., K. Storesletten, and G. L. Violante (2020): "Optimal progressivity with age-dependent taxation," *Journal of Public Economics*, 189, 104074.

Huang, J. (2008): "Taxable and tax-deferred investing: a tax-arbitrage approach," *The Review of Financial Studies*, 21, 2173–2207.

Keane, M. P. (2011): "Labor supply and taxes: A survey," *Journal of Economic Literature*, 49, 961–1075.

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

Kostøl, A. and A. Myhre (2020): "Labor supply responses to learning the tax and benefit schedule," *IZA Discussion Paper*.

Le Barbanchon, T. (2020): "Taxes Today, Benefits Tomorrow," Tech. rep., Working Paper.

Lockwood, B. B. (2020): "Optimal income taxation with present bias," *American Economic Journal: Economic Policy*, 12, 298–327.

Lozachmeur, J.-M. (2006): "Optimal age-specific income taxation," *Journal of Public Economic Theory*, 8, 697–711.

Ndiaye, A. (2020): "Flexible retirement and optimal taxation," *Working Paper*.

Ong, P. (2020): "The effect of child support on labor supply: An estimate of the Frisch Elasticity," *Working paper*.

Ring, M. and T. O. Thoresen (2021): "Wealth Taxation and Charitable Giving," *Working Paper*.

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

Scott, J., J. B. Shoven, S. Slavov, and J. G. Watson (2021): "Is Automatic Enrollment Consistent with a Life Cycle Model?" *NBER Working Paper*.

Seim, D. (2017): "Behavioral responses to wealth taxes: Evidence from Sweden," *American Economic Journal: Economic Policy*, 9, 395–421.

Slemrod, J. and W. Kopczuk (2002): "The optimal elasticity of taxable income," *Journal of Public*

Economics, 84, 91–112.

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

Stantcheva, S. (2020): "Dynamic Taxation," *Annual Review of Economics*, 12, 801–831.

Yu, P. C. (2021): "Optimal retirement policies with present-biased agents," *Journal of the European Economic Association*, 19, 2085–2130.

Zator, M. (2019): "Working More to Pay the Mortgage: Household Debt, Consumption Commitments, and Labor Supply," *Working paper*.

A Appendix

Figure A.1: Little Evidence of "Negative-Bunching" at Debt-Conversion-Cap Threshold

Panel (A) provides a scatter plot, in green, of the relationship between debt accumulation and student earnings around the debt-conversion-cap threshold. This is the threshold above which additional earnings do not increase future student debt because there is no more stipends to convert to debt. Panel (B) provides a graphical illustration of how the bunching estimate. See Figure 1 for further info on the methodology.

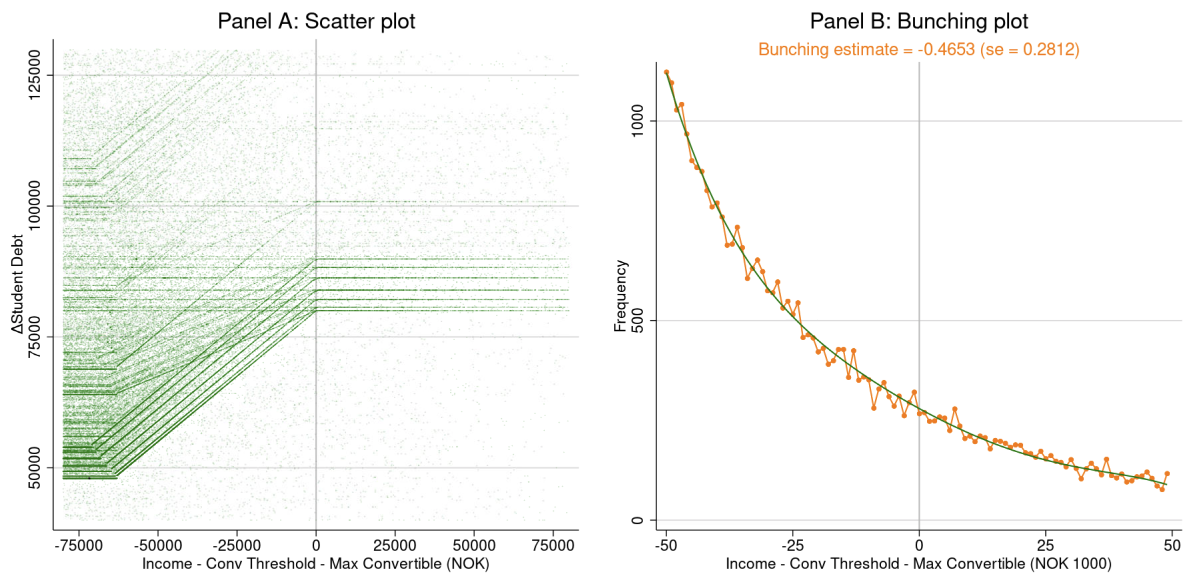
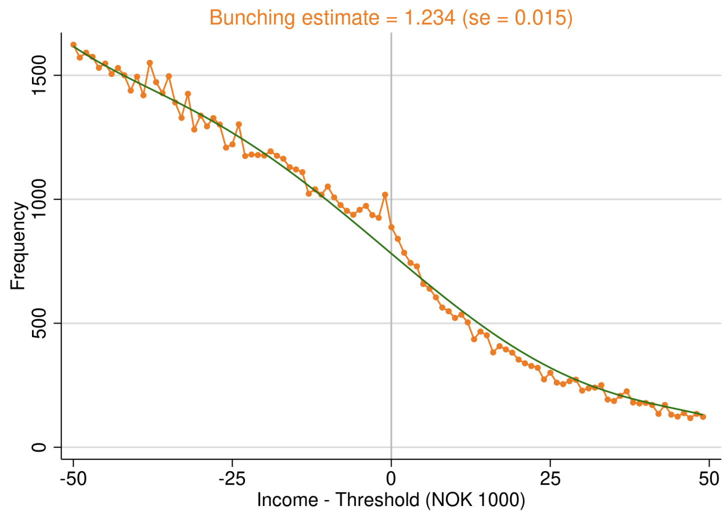


Figure A.2: Bunching at Debt-Conversion Threshold for Workers with Sales and Hospitality Occupations

We repeat the exercise in Panel (B) of Figure 1 on a subset of workers with hospitality (3-digit occupation code = 513, waiters and bartenders) and sales jobs (3-digit occupation codes = 513, 521, 522, 523, 524).



A.1 Additional Lemmas and Proofs

Lemma 1: The relationship between period-1 and period-2 labor supply responses is given by

$$-\frac{dl_2}{d\tilde{\tau}}\frac{1-\tilde{\tau}}{l_2} = \left(-\frac{dl_1}{d\tilde{\tau}}\frac{1-\tilde{\tau}}{l_1}\right) - \frac{1}{\nu}. \tag{20}$$

Proof: We use the intertemporal labor supply FOC, and differentiate it, allowing $\tilde{\tau}$

to change.

$$
\begin{aligned}
d\left[\frac{w_1}{w_2}(1-\tilde{\tau})\beta\tilde{R}\psi l_2^{\nu}-\psi l_1^{\nu}\right] &= 0 && (21)\\
\frac{w_1}{w_2}\beta\tilde{R}\left\{-d\tilde{\tau}l_2^{\nu}+(1-\tilde{\tau})\nu l_2^{\nu-1}dl_2\right\} &= \nu l_1^{\nu-1}dl_1 && (22)\\
\frac{w_1}{w_2}\beta\tilde{R}\left\{-d\tilde{\tau}+(1-\tilde{\tau})\nu l_2^{-1}dl_2\right\} &= \nu\left(\frac{l_1}{l_2}\right)^{\nu}l_1^{-1}dl_1 && (23)\\
\frac{w_1}{w_2}\beta\tilde{R}\left\{-d\tilde{\tau}+(1-\tilde{\tau})\nu l_2^{-1}dl_2\right\} &= \nu\left[\frac{w_1}{w_2}\beta\tilde{R}(1-\tilde{\tau})\right]l_1^{-1}dl_1 && (24)\\
-d\tilde{\tau}+(1-\tilde{\tau})\nu l_2^{-1}dl_2 &= \nu(1-\tilde{\tau})l_1^{-1}dl_1 && (25)\\
(1-\tilde{\tau})\nu l_2^{-1}dl_2 &= d\tilde{\tau}+\nu(1-\tilde{\tau})l_1^{-1}dl_1 && (26)\\
l_2^{-1}\frac{dl_2}{d\tilde{\tau}} &= \frac{1}{\nu}\frac{1}{1-\tilde{\tau}}+l_1^{-1}\frac{dl_1}{d\tilde{\tau}} && (27)\\
\frac{dl_2}{d\tilde{\tau}} &= l_2\frac{1}{\nu}\frac{1}{1-\tilde{\tau}}+\frac{l_2}{l_1}\frac{dl_1}{d\tilde{\tau}} && (28)\\
\frac{dl_2}{d\tilde{\tau}}\frac{1-\tilde{\tau}}{l_2} &= \frac{1}{\nu}+\frac{dl_1}{d\tilde{\tau}}\frac{1-\tilde{\tau}}{l_1}. && (29)
\end{aligned}
$$

This tells us that the elasticity of period-2 labor supply with respect to $1-\tilde{\tau}$ equals the difference between the uncompensated and compensated period-1 labor supply elasticities.

Lemma 2. Assume that the present value (PV) of government tax revenues is

$$PV = \tau w_1 l_1 \delta + \frac{1}{R^g}\left\{\tau w_1 l_1(1-\delta)+\tau_2 w_2 l_2\right\}, \tag{30}$$

where τ_2 is defined such that $\tau_2 w_2 l_2$ equals tax revenues from period-2 labor supply, then

$$dPV = \tau w_1 l_1\left(1-1/R^g\right)d\delta + \frac{1}{1-\tilde{\tau}}\left[\frac{1}{R^g}\tau_2 w_2 l_2\frac{1}{\nu}+\left\{\tau w_1 l_1+\frac{1}{R^g}\tau_2 w_2 l_2\right\}\left(\frac{dl_1}{d\tilde{\tau}}\frac{1-\tilde{\tau}}{l_1}\right)\right]d\tilde{\tau},$$

where $-\frac{1}{\nu}$ is the compensated (Frisch) labor supply elasticity and $-\frac{dl_1}{d\tilde{\tau}}\frac{1-\tilde{\tau}}{l_1}$ is the uncompensated labor supply elasticity.

Proof: Differentiate PV at $\delta = 1$ and use Lemma 1.

$$
\begin{aligned}
& \tau w_1 \left[l_1 d\delta + dl_1\right] + \frac{1}{R^g} \left\{\tau w_1 l_1 d\delta + \tau_2 w_2 dl_2\right\} && (31)\\
= \;& \tau w_1 l_1 \left(1 - 1/R^g\right) d\delta + \tau w_1 dl_1 + \tau_2 w_2 \frac{1}{R^g} dl_2 && (32)\\
= \;& \tau w_1 l_1 \left(1 - 1/R^g\right) d\delta + \tau w_1 dl_1 + \tau_2 w_2 \frac{1}{R^g} \left[l_2 \frac{1}{\nu} \frac{1}{1-\tilde{\tau}} d\tilde{\tau} + \frac{l_2}{l_1} dl_1\right] && (33)\\
= \;& \tau w_1 l_1 \left(1 - 1/R^g\right) d\delta + \tau w_1 dl_1 + \tau_2 w_2 l_2 \frac{1}{1-\tilde{\tau}} \frac{1}{R^g} \left[\frac{1}{\nu} + \frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1}\right] d\tilde{\tau} && (34)\\
= \;& \tau w_1 l_1 \left(1 - 1/R^g\right) d\delta + \tau w_1 dl_1 + \frac{1}{R^g} \tau_2 w_2 l_2 \frac{1}{1-\tilde{\tau}} \left[\frac{1}{\nu} + \frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1}\right] d\tilde{\tau} && (35)\\
= \;& \tau w_1 l_1 \left(1 - 1/R^g\right) d\delta + \tau w_1 \frac{dl_1}{d\tilde{\tau}} d\tilde{\tau} + \frac{1}{R^g} \tau_2 w_2 l_2 \frac{1}{1-\tilde{\tau}} \left[\frac{1}{\nu} + \frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1}\right] d\tilde{\tau} && (36)\\
= \;& \tau w_1 l_1 \left(1 - 1/R^g\right) d\delta + \frac{1}{R^g} \tau_2 w_2 l_2 \frac{1}{1-\tilde{\tau}} \frac{1}{\nu} d\tilde{\tau} + \tau w_1 l_1 \frac{1}{1-\tilde{\tau}} \frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1} d\tilde{\tau} + \frac{1}{R^g} \tau_2 w_2 l_2 \frac{1}{1-\tilde{\tau}} \frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1} d\tilde{\tau} && (37)\\
= \;& \tau w_1 l_1 \left(1 - 1/R^g\right) d\delta + \frac{1}{1-\tilde{\tau}} \left[\frac{1}{R^g} \tau_2 w_2 l_2 \frac{1}{\nu} + \left\{\tau w_1 l_1 + \frac{1}{R^g} \tau_2 w_2 l_2\right\} \left(\frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1}\right)\right] d\tilde{\tau} && (38)\\
& && (39)
\end{aligned}
$$

Note that in this setting it makes the most sense to set $R^g = 1$, as this is the rate at which delayed taxes accrue interest.

Lemma 3. Delayed taxation increases government tax revenues if the ratio of the Frisch elasticity to the uncompensated period-1 labor supply elasticity is below some threshold.

$$
\begin{aligned}
\frac{1}{\nu} \Big/ - \frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1} \;&<\; \frac{\tau w_1 l_1 + \tau_2 w_2 l_2}{\tau_2 w_2 l_2}. && (40)\\
& && (41)
\end{aligned}
$$

Proof: First divide through the expression for dPV in Lemma 2 by $d\delta$. Set $R^g = 1$. Since $d\tilde{\tau}/d\delta > 0$, we have that $dPV/d\delta > 0$ whenever the expression for dPV in Lemma 2 is positive.

$$
\begin{aligned}
\tau_2 w_2 l_2 \frac{1}{\nu} + \left\{\tau w_1 l_1 + \tau_2 w_2 l_2\right\} \left(\frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1}\right) \;&>\; 0 && (42)\\
\tau_2 w_2 l_2 \frac{1}{\nu} \;&>\; \left\{\tau w_1 l_1 + \tau_2 w_2 l_2\right\} \left(-\frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1}\right) && (43)\\
\frac{1}{\nu} \Big/ - \frac{dl_1}{d\tilde{\tau}} \frac{1-\tilde{\tau}}{l_1} \;&>\; \frac{\tau w_1 l_1 + \tau_2 w_2 l_2}{\tau_2 w_2 l_2}. && (44)\\
& && (45)
\end{aligned}
$$

Since delayed taxation implies $d\delta < 0$, we can reverse the inequalities above to find the cut-off for the elasticity ratio, above which $dPV/-d\delta > 0$.

Since the Frisch elasticity is the compensated elasticity, and the uncompensated elasticity equals the compensated elasticity plus an income effect (Slutsky equation), this says that we need income effects to be relatively smaller than the substitution effects: which is exactly the case in which the tax is the most distortionary.