Optimal Delayed Taxation in the Presence of Financial Frictions

Preliminary version, currently being revised

Spencer Bastani and Marius A. K. Ring*

March 18, 2024

Abstract

In the presence of financial frictions, the timing of cash flows matters. We apply this insight to optimal income taxation by studying a new policy: delayed taxation. Introducing a delay between the accrual and payment of income taxes provides two sources of welfare gains when some agents are borrowing constrained. First, it improves consumption smoothing for financially constrained agents. Second, it reduces the present value tax rate from the perspective of constrained agents, thereby reducing the distortionary effects of income taxation. We characterize the conditions under which marginally delayed taxation is welfare enhancing under different assumptions about the sophistication of the benchmark tax system, and we contrast the welfare gains with those achievable by offering low-interest loans or changing nominal tax rates. We then characterize optimal delayed taxation in a model calibrated to the Norwegian economy. This exercise reveals substantial welfare gains from delayed taxation. When limiting the amount the government may borrow to finance any given reform, delayed taxation materially outperforms age-dependent taxation and a policy in which the government offers subsidized loans. Finally, we empirically test the hypothesis that delayed taxation substantially reduces income tax distortions in the context 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 lower than those we find for a regular income tax threshold, and that increase with ex ante financial resources. Taken together, our results underscore 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

* Spencer Bastani is at the Institute for Evaluation of Labour Market and Education Policy (IFAU), Uppsala, Sweden; Uppsala University; IFN Stockholm. 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. This paper previously circulated as "Financial frictions and the Non-Distortionary Effects of Delayed Taxation," available on SSRN as of February 2021. We thank seminar participants at MIT, UT Austin, IFN, Washington University in St. Louis, LMU Munich, and Cornell University for helpful comments and 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, Eben Lazarus, Camille Landais, Daniel Waldenström, Andreas Kostøl, David Agrawal, Joel Slemrod, Andrey Ordin, Mindy Xiaolan, Sheridan Titman, Aaron Pancost, Andreas Haufler, Jonas Löbbing, Emanuel Hansen, Karl Schulz, Emmanuel Saez, Torstein Persson, Jakob Søgaard, Claus Kreiner, Arash Nekoei, Lukas Worku, Nikhilesh Prakash, Alexander Hilslov, and Ragnhild Schreiner for helpful feedback 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

One of the central questions in economics is how best to design the tax system. In most countries, the main source of tax revenue is taxes on labor income. This leads to efficiency losses, since it is generally believed that taxes on labor income reduce the amount of labor that taxpayers supply. The existence of these distortions has spawned an extensive literature in public finance. On the empirical side, researchers attempt to quantify the magnitude of the distortions in order to guide the calibration of optimal tax models. On the theoretical side, researchers attempt to model a tax system that minimizes distortions and maximizes overall welfare. However, these modeling efforts rarely take into account that many taxpayers face financial frictions. For example, taxpayers who expect higher wages in the future may want to borrow against higher future earnings, but are either unable to do so or face very high marginal borrowing rates.

In this paper, we point out that credit market imperfections imply that the *timing* of taxes is important. Importantly, one can change the timing of taxes in two ways. One way is age-dependent taxation (Lozachmeur 2006, Blomquist and Micheletto 2008, Weinzierl 2011), where tax rates vary with age. By taxing younger, more constrained households less, age-dependent taxation reduces the welfare losses caused by the inability to intertemporally smooth consumption. Our proposed path is to allow taxpayers to delay the *payment* of their taxes. This policy does not condition tax rates on age, but rather allows constrained taxpayers to delay their tax payments. Thus, delayed taxation imposes different tax regimes on different individuals based on their voluntary choices rather than on their exogenous characteristics (i.e., allowing individuals to ”self-select” into the delayed tax regime), thereby mitigating potential concerns about horizontal equity.

Delayed taxation generates welfare gains both by allowing agents to smooth consumption and by reducing the distortionary effects of income taxation. The latter effect arises because constrained taxpayers discount taxes payable in the future more than the government does. Thus, constrained taxpayers essentially behave as if the tax rate were lower, which implies lower welfare losses due to behavioral distortions. By reducing the “effective” tax rate, as perceived by agents, delayed taxation provides an instrument with which governments can reduce behavioral elasticities with respect to nominal tax rates. The notion that the government can use various tax instruments to lower behavioral elasticities is formalized by Slemrod and Kopczuk (2002), who focus on tax enforcement as an instrument. Our paper emphasizes the timing of tax payments as an important instrument: by allowing delayed tax payments, the behavioral labor supply elasticity with respect to the nominal tax rate falls.

Our study provides an in-depth examination of delayed taxation through a blend of theoretical analysis, numerical simulations, and empirical evidence. In the first part of our paper, we introduce delayed taxation into a dynamic optimal tax model with linear taxation. We consider heterogeneous workers who differ in their labor market productivity, which allows us to study the welfare gains from delayed taxation in the presence of distributional considerations. We formally show how optimal marginal tax rates depend on whether taxes can be delayed. We also show how the optimal delayed tax policy (i.e., the fraction of taxes that can be delayed) depends on standard behavioral elasticities of labor supply with respect to marginal tax rates and on the magnitude of financial frictions, measured by the difference between the interest rates faced by net borrowers and net savers.

To understand the welfare effects of delayed taxation, it is easiest to consider the case where an existing tax system is marginally complemented by delayed taxation. In this case, we provide a simple decomposition into four effects: (i) welfare gains from increased intertemporal consumption smoothing, (ii) a positive fiscal externality due to a positive substitution effect on the labor supply of young workers, which is (iii) partially offset by a negative intertemporal substitution effect on tax revenues of older workers, and (iv) a negative income effect on present value tax revenues as financially constrained workers become richer on a lifetime basis when some of their taxes are delayed (and repaid at a favorable interest rate from the perspective of financially constrained agents). Note that because those with higher earnings when young are effectively allowed to delay more taxes, the delayed tax scheme we propose introduces a history-dependent feature into the tax system.

If the government simply offered loans to financially constrained agents, the positive effects on labor supply would be absent. If the government were to engage in age-dependent taxation by offering tax breaks to young workers, this would affect all workers, not just those who are financially constrained. In addition, there would be a mechanical loss of tax revenue with age-based taxation. Thus, the favorable targeting properties of delayed taxation are an important source of its attractiveness as a policy tool.

The driving force behind the welfare gains from delayed taxation is financial frictions. It is widely recognized that financial frictions are relevant for many young people. For example, as argued by Herbst and Hendren (2023), students possess substantial private knowledge about their future earnings, academic persistence, employment, and likelihood of loan repayment beyond what is captured by observable characteristics. This makes it difficult for lenders to offer fair loans to students who need money for college, and causes the market for student loans to collapse. Thus, the government has a central role to play in mitigating financial frictions, and our paper argues that we can do so in a welfare-enhancing way that also increases tax revenues through positive effects on labor supply.

In the second part of our paper, we provide numerical solutions by calibrating the optimal tax model to the Norwegian economy. The Norwegian register data, which cover the entire population, allow us to compute realized earnings trajectories for a large sample of young workers. We consider all workers between the ages of 20 and 30 in 1990, for whom we can compute a measure of effective wages in both 1990 and 2011. This allows for considerable variation in initial earnings and wage trajectories, and thus in the degree of financial frictions.[1] We then consider the optimal tax schedule that the government would impose on these workers to maximize welfare. We consider three different policies: (i) age-independent linear taxation, (ii) allowing for age-dependent marginal tax rates, (iii) allowing for age-dependent marginal tax rates as well as lump-sum transfers, and (iv) delayed taxation. We also examine the welfare effects of letting

[1] That is, those with low initial earnings and high future earnings will optimally want to borrow but face high interest rates, but those with flat or declining earnings will want to save.

the government lend directly to households at the government interest rate. In the linear age-independent tax scheme, the government implements high marginal taxes of about 60% in part to redistribute across workers and in part to provide lump-sum transfers that help intertemporal consumption smoothing.

As in Weinzierl (2011), we find that age-dependent taxation increases welfare even in the absence of financial frictions. As financial frictions increase in severity, the welfare gains become larger. For delayed taxation, of course, there are no welfare gains when borrowing and saving rates are equal. However, as the marginal borrowing rate increases, the welfare gains of delayed taxation eventually exceed those of age-dependent taxation, and equal those of a fully-age dependent tax *and* lump-sum transfer scheme. To illustrate, we consider the welfare implications of an exogenous increase in the severity of financial frictions. We find that increasing the borrowing rate from 3 to 10 percent leads to a large welfare loss. This welfare reduction is equivalent to the government experiencing a negative revenue shock equal to 9.8% of GDP. If the government uses age-dependent taxation and is allowed to re-optimize to this shock, this welfare loss is reduced to 5.4% of GDP. If the government engages in delayed taxation, the welfare loss is only 0.8% of GDP. Thus, in the money-metric sense, age-dependent taxation reduces the utility cost of financial frictions by a sizable 44%, but delayed taxation reduces it by an even more substantial 92%.

Both age-dependent taxation and delayed taxation are policies that help consumers smooth intertemporal consumption. Behind the scenes, the government must increase its own borrowing to achieve these welfare gains. With delayed and age-dependent taxation, the government borrows to offset the effects of postponed tax revenues. In practice, the government's ability to borrow may be limited by political (e.g., debt ceilings) or financial (e.g., credit ratings) constraints. Therefore, we investigate what welfare gains can be achieved while restricting the government's ability to borrow. We first consider the case where the government cannot increase its borrowing at all. Interestingly, this still allows for meaningful welfare gains of about 0.5% of benchmark GDP under both age-dependent and delayed taxation. Relaxing the government's borrowing constraint increases the potential welfare gains, but more so for delayed taxation. If the government is allowed to increase borrowing by 50%, the welfare gains from age-dependent taxation are about 2.2% and the gains from delayed taxation are 2.6%. For age-dependent taxation, whether the government can also implement age-dependent lump-sum transfers plays an immaterial role. This is an interesting finding since delayed taxation only involves a single additional policy variable (share of taxes that are delayed) while the fully age-dependent tax and transfer schedule involves two. The favorable targeting properties of delayed taxation implies that it can achieve higher welfare gains with fewer policy parameters

We also study the welfare effects of a government lending program, wherein the government offers loans at the government interest rate up to a uniform loan limit. When the government cannot increase borrowing by more than 50%, this lending policy produces welfare gains of about 1.37% of GDP, which is barely half that of delayed taxation. The relatively low gains from government lending is consistent with our theoretical findings. By simply offering loans, the government does not exploit the opportunity to reduce the distortionary effects of income taxation.

In the third and final part of the paper, we empirically test whether delayed taxation in fact reduces income tax distortions when taxpayers are financially constrained. Conducting such a test is challenging because few settings allow for substantial variation in the timing of tax payments. Taxes are typically paid either immediately (through withholding) or one year later when tax returns are due. We overcome this challenge by studying the effects of a student debt conversion 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. Specifically, the vast majority of Norwegian students receive an annual loan of about $13,000, about half of which is typically forgiven at the end of the year. However, if the student earns more than about $17,000, each additional dollar of earnings reduces the amount forgiven by 50 cents.

This quasi-experimental setting is well suited to examine how financial frictions can make delayed taxation less distortionary. First, students are almost by definition highly constrained. Only a few years later, they face significantly higher incomes against which it is difficult 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 of bunching, indicating that students do respond, these responses pale in comparison to the effective after-tax wage reduction that occurs.

Our bunching analysis provides an implied elasticity of labor earnings to after-tax wages of only 0.016. While this estimate is highly statistically significant, it is an order of magnitude smaller than most existing estimates (Keane, 2011). A fair comparison of our estimates would be with existing bunching estimates of labor earnings elasticities, which tend to be lower than those from other regression-based methods. However, these earnings elasticities typically suffer from a downward bias caused by labor market and optimization frictions. Our main approach to still obtain some qualitative insights is to consider a homogeneous group of workers (e.g., students) and to compare the bunching elasticities implied by a (de facto) delayed tax threshold with those obtained from a regular tax threshold.

To shed light on the observed non-bunching behavior at the delayed tax threshold, we examine how students' characteristics covary with their position relative to the debt conversion threshold. These analyses suggest that non-bunchers (and their parents) have significantly lower liquid assets, but not lower future earnings. This is exactly what we would expect to see if irresponsiveness to the threshold is driven by financially constrained agents. We also find no evidence that the educational attainment of students' parents changes in a manner consistent with these characteristics driving differences in bunching behavior. Building on these analyses, we examine heterogeneity in bunching by the ex ante financial situation of students and their parents. Students with liquidity below the median (and their parents as well) have an implied labor earnings elasticity less than half as large as those above the median. In our modeling framework, this

[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.

heterogeneity can be rationalized by the fact that less liquid students optimize to a 10 percentage point higher marginal borrowing rate.

We also examine the bunching behavior of students at a regular tax threshold. This allows us to compare implied labor supply elasticities under different tax regimes but among a similar sample of individuals.[3] The tax threshold analyzed occurs around $6,000, where the marginal income tax rate ranges 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 implied by the delayed tax threshold. Under some simplifying assumptions about the exact elasticities measured by our bunching framework, these large elasticity differences can be rationalized by the fact that students face an average marginal borrowing rate of more than 20 percent and thus are considerably less responsive to the de facto delayed tax scheme created by the student loan program. We argue that it is unlikely that this difference can be explained by the fact that the kink occurs at different income levels. Using a regression-based approach that controls for differences on observables, such as occupation codes and age, we find a qualitatively similar elasticity difference.

That delaying the payment of a tax reduces its distortionary effects is not too surprising. In the absence of strong debt aversion and in the presence of borrowing-constrained agents, this is what we would expect from economic theory. In a sense, then, our bunching analysis provides a test of the applicability of life-cycle model reasoning to the study of the labor supply decisions of constrained workers. In addition, it provides empirical evidence on the potential economic magnitude of the effect, which is quite substantial in our setting. Both are necessary to assess the potential of delayed taxation as a new policy tool.

The central contribution of this paper is to propose and study the welfare implications of the simple idea that —in the presence of credit market imperfections —altering the timing of income tax payments may offer material welfare gains, in large part by reducing the distortionary effects of income taxation. We strengthen this contribution by providing quasi-experimental evidence that delayed taxation in fact reduces income tax distortions. To our knowledge, this idea has not been previously explored, either theoretically or empirically.

Related literature. On the conceptual front, this paper contributes to the literature on dynamic optimal taxation (see, e.g., Ndiaye 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 tax payments or incorporating financial frictions.[4] The conceptual novelty of our paper

[3] Ideally, this will control for unobservable factors that influence labor supply optimization. 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 or financial frictions, or differences in structural elasticities, are driving the differences in elasticities.

[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.

lies in this intersection.

Our paper is also related to the optimal tax literature that allows tax rates to depend on taxpayer characteristics, i.e., tagging (Akerlof, 1978). Our work is perhaps closest to the literature on age-dependent taxation (Weinzierl 2011; Bastani, Blomquist, and Micheletto 2013; Heathcote, Storesletten, and Violante 2020; Gervais 2012): by introducing delayed taxation, financially constrained taxpayers see a reduction in effective (present-value) tax rates. Once taxpayers age and borrowing constraints no longer bind, effective tax rates equal the higher nominal rate. In that sense, delayed taxation has a strong element of age-dependent taxation. However, the key differences is that (i) delayed taxation does not necessarily require the government to condition tax rates on taxpayer characteristics (which may likely be controversial) and (ii) it does not rely on using age as a proxy for liquidity constraints. Instead, delayed taxation allows constrained borrowers, who face a high marginal borrowing rate, to self-select into the scheme. Optimal delayed taxation also does not necessarily imply changing statutory tax rates.[5]

On the empirical front, this paper contributes to the growing literature studying bunching at tax thresholds (see, e.g., Saez 2010; Bastani and Selin 2014; Seim 2017; Søgaard 2019; and the review by Kleven 2016), 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) or student-loan repayment thresholds (de Silva 2023). 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.[6] We further add to the literature using income-contingent transfer schemes to identify labor supply elasticities (see, e.g., Ong 2020). 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 studies delayed taxation in a dynamic optimal tax framework. Section 3 provides numerical solutions to the optimal delayed tax problem. Section 4 discusses whether there are existing tax regimes that are similar to delayed taxation. Section 5 uses a de-facto delayed tax scheme in Norway to test some of the behavioral implications of our theoretical framework. Section 6 briefly discusses aspects related to the implementation and unmodeled trade-offs associated with introducing delayed taxation.

[5] In our calibration, for example, the optimal linear tax rate only changes by 1 percentage point when delayed taxation is introduced.

[6] 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 The welfare gains of delayed taxation

We consider an economy consisting of heterogeneous agents who live and work for two periods and differ in their exogenous lifetime labor productivity profiles (w_1^i, w_2^i), where w_t^i denotes the market productivity of agent i in period $t = 1, 2$. In each period, an agent earns an income of $y_t^i = w_t^i \ell_t^i$, where ℓ_t^i is the labor supply. For tractability reasons, we focus on a dynamic extension of the linear (progressive) taxation framework (Sheshinski 1972). This setting allows us to study the welfare gains from delayed taxation in the presence of distributional considerations. The tax schedules in periods 1 and 2 are given by

$$T_1(y_1) = -G_1 + \delta\tau_1 y_1,$$
$$T_2(y_1, y_2) = -G_2 + \tau_2 y_2 + (1 + r)(1 - \delta)\tau_1 y_1,$$

where

- τ_t denotes the nominal (statutory) marginal tax rate in period t.
- $G_t \geq 0$ denotes the age-dependent lump-sum transfer in period t.[7]
- $\delta \in [0, 1]$ is the fraction of the period 1 tax that must be paid in period 1 while the fraction $1 - \delta$ must be paid in period 2. The government charges an interest on the delayed tax payment equal to r.

Agents can save between periods, and the saving technology of private agents is given by the function $R(s)$, which is the amount by which disposable income in the second period is increased if the *individual* saves (or borrows) the amount s.[8] Consumption c_t^i is given by

$$c_1^i = w_1^i \ell_1 [1 - \delta\tau_1] + x_1^i + G_1 - s^i, \tag{1}$$
$$c_2^i = w_2^i \ell_2 [1 - \tau_2] + x_2^i + G_2 - (1 + r)[1 - \delta]\tau_1 w_1^i \ell_1^i + R(s^i), \tag{2}$$

where x_t^i is non-labor income (such as wealth or partner income). Individual preferences are represented by the utility function:

$$u_1(c_1^i) - v(\ell_1^i) + \beta[u_2(c_2^i) - v(\ell_2^i)], \tag{3}$$

where u is increasing, twice differentiable and strictly concave, and v is increasing, twice differentiable, and strictly convex.

The individual's problem The problem solved by an individual i with a lifetime wage profile of (w_1^i, w_2^i) is to choose ℓ_1^i, ℓ_2^i, s^i in order to maximize (3) subject to constraints (1) and (2). The

[7] We realistically rule out lump-sum taxes. On the infeasibility of lump-sum taxes, see Smith (1991) for a discussion of Margaret Thatcher's disastrous attempt to introduce a poll tax in the United Kingdom between 1989 and 1990.

[8] We exclude taxes on capital income. Including a proportional tax on capital income would not affect the qualitative nature of our results.

first-order conditions are:

$$(\ell_1): u_1'(c_1^i)w_1^i[1-\delta\tau_1] - \beta(1+r)u_2'(c_2^i)[1-\delta]\tau_1 w_1^i = v'(\ell_1^i), \tag{4}$$

$$(\ell_2): u_2'(c_2^i)w_2^i[1-\tau_2] = v'(\ell_2), \tag{5}$$

$$(s): u_1'(c_1^i) = \beta u_2'(c_2^i)R'(s^i). \tag{6}$$

The first order condition for ℓ_1 (equation 4) can be written:

$$u_1'(c_1^i)w_1^i\left(1-\tau_1\Big[\delta+[1-\delta]\theta\Big]\right) = v'(\ell_1^i) \quad \text{where} \quad \theta^i = \frac{\beta(1+r)u'(c_2^i)}{u'(c_1^i)}. \tag{7}$$

Thus, the labor supply in period 1 depends on the extent to which taxes are delayed (as reflected by δ) and the wedge θ between the marginal utility of consumption in period 1 and the discounted marginal utility of consumption in period 2. When agents are free to save and borrow at the government interest rate, consumption is perfectly smoothed across periods, $\theta^i = 1$, which implies that the first-order condition is independent of δ.

However, in the presence of financial frictions, we have that $\theta^i < 1$, which implies that when $0 < \delta < 1$, the effective marginal tax rate faced by agents is lower than in an economy without delayed taxation.

For the second period, the first-order condition for ℓ_2 (equation 5) is the same as in a standard model without delayed taxation.

As can be seen from (7), the effects of delayed taxation depend on the degree of consumption smoothing. Assuming an interior solution for s^i such that the FOC (6) holds, and insertion of this condition into (4) yields:

$$u_1'(c_1^i)w_1^i\left(1-\tau_1\left[\delta+[1-\delta]\frac{1+r}{R'(s^i)}\right]\right) = v'(\ell_1^i). \tag{8}$$

Equation (8) illustrates that increasing the fraction of period 1 taxes that are delayed (increasing $1-\delta$) reduces the distortion on period 1 labor supply at a rate determined by the difference in interest rates between government and private agents, which captures the "strength" of the financial friction. Note that this is a testable implication: (compensated) behavioral responses to a delayed tax are smaller than to a regular tax when agents face financial frictions. We define:

$$\Delta_r^i = 1 - \frac{1+r}{R'(s^i)}, \tag{9}$$

$$\tilde{\tau}_1^i = \tau_1\left[\delta+[1-\delta]\frac{1+r}{R'(s)}\right] = \tau_1\left[1-(1-\delta)\Delta_r^i\right], \tag{10}$$

$$\tilde{w}_1^i = w_1^i\left(1-\tilde{\tau}_1^i\right) \quad \text{and} \quad \tilde{w}_2^i(1-\tau_2), \tag{11}$$

where (9) is the interest wedge, (10) defines the effective tax rate for period 1, and (11) defines the "net" wage rate relevant to the labor supply decision in periods 1 and 2, respectively. The definition of the effective tax rate emphasizes the fact that the effective tax rate is equal to the nominal tax rate if there is no delayed taxation ($1-\delta=0$) or no financial frictions ($\Delta_r^i=0$).

Note that period-1 and period-2 labor supply can be related by substituting (6) into (8) and then substituting in (5):

$$\beta R'(s^i)\frac{\tilde{w}_1^i}{\tilde{w}_2^i}=\frac{v'(\ell_1^i)}{v'(\ell_2^i)}. \tag{12}$$

Intuitively, borrowing constrained individuals smooth their consumption by increasing their labor supply in period 1, making it less elastic to taxes.

In some parts of our analysis, we will consider a specific piece-wise linear saving technology:

$$R(s)=\begin{cases}(1+r)s & \text{if } s\geq 0,\\ (1+r_b)s & \text{if } s<0,\end{cases} \tag{13}$$

where $r_b-r>0$ reflects the "credit penalty" faced by financially constrained agents. An interior solution $s<0$ is optimal if r_b-r is sufficiently close to zero and the wage profile is sufficiently steep. Focusing on such interior solutions, we have $\frac{dR(s)}{ds}=1+r_b$. This implies that for agents with $s^i<0$ we have $\theta^i=\frac{1+r}{1+r_b}<1$. Note that (13) can be seen as a generalization of the general "no-borrowing constraint" in macro (obtained when $r_b\to\infty$).

The government's problem Assuming that the government is utilitarian (for notational simplicity, all our results generalize to the case where the government weights the welfare of different individuals differently), and denoting by π^i the proportion of agents of type i in the population, the government's problem is:

$$\max_{G_1,G_2,\tau_1,\tau_2,\delta}\sum_i \pi^i V^i, \tag{14}$$

subject to:

$$\sum_i \pi^i\left(\tau_1 w_1^i\ell_1^i+\frac{\tau_2 w_2^i\ell_2^i}{1+r}\right)\geq G_1+\frac{G_2}{1+r}+M, \tag{15}$$

$$0\leq\delta\leq 1, \tag{16}$$

where M is an exogenous revenue requirement that is not refunded to agents. Note that δ does not enter (15) because the government is indifferent between receiving tax revenue in period 1 or period 2. This is because the government charges an interest rate of r on delayed taxes, which is the same interest rate the government faces. If the private borrowing rate exceeds the government's rate, we implicitly assume that there are financial frictions between workers and private lenders, but not between workers and the government. For example, the government typically has more effective means to collect delinquent taxes compared to the ability of private lenders to collect unsecured debt. As a result, the actuarially fair interest rates applied to delayed taxes differ from those applied to private loans. A case in point is the United States, where the interest rate on delinquent taxes is set at the federal short-term rate plus 3 percent. This rate is significantly lower than the average interest rate on unsecured personal loans.

Before proceeding, two comments are in order regarding the government optimization problem.

First, dynamic models of optimal taxation typically assume that the budget constraint is satisfied in the expectation, as we have formulated our government budget constraint (15). This implicitly assumes that the government is free to borrow if, for example, it runs a deficit in period 1. In reality, however, governments often face constraints on the amount of borrowing they can undertake. In our numerical simulations, we go beyond the standard model by shedding light on the consequences of imposing constraints on the amount of such implicit borrowing (see section 3.4).

Second, both age-dependent and delayed taxation have in common that they allow significant redistribution from period 2 to period 1 (achieved through lower average tax rates in period 1). For example, with age-dependent taxation, we are likely to have $G_1 > G_2$ and $\tau_2 > \tau_1$. This may imply incentives to move abroad or to engage in other forms of tax evasion. Assumption 1 makes it clear that we assume perfect enforcement of both age-dependent taxation and delayed taxation.

Assumption 1 (No Default Assumption) *Workers cannot default on taxes or government loans, nor can they evade taxes or move abroad. In other words, there's perfect enforcement in all dimensions.*

In our particular setting, the no-default assumption is not very restrictive. The reason is that we model rational agents who borrow only if they expect higher future wages, so they are not really liquidity constrained as they age. Central to our discussion is the introduction of delayed taxation as an innovative policy instrument that may be more politically feasible than age-dependent taxation. We argue that concerns about default apply equally to both delayed and age-dependent taxation.

Now we want to characterize the optimal solution to the government's problem. For this purpose, let λ denote the Lagrange multiplier associated with (15). The Lagrangian is

$$W = \sum_i \pi^i V^i - \lambda \left(- \sum_i \pi^i \left(\tau_1 w_1^i \ell_1^i + \frac{\tau_2 w_2^i \ell_2^i}{1+r} \right) + G_1 + \frac{G_2}{1+r} + M \right).$$

We first characterize the optimal age-independent tax system, given by the solution to the above optimization problem, assuming $\tau_1 = \tau_2$, $G_1 = G_2$. For this purpose, we define:

$$g_1^i = \frac{u_1'(\cdot)}{\lambda}, \tag{17}$$

$$g_2^i = \beta(1+r)\frac{u_2'(\cdot)}{\lambda}, \tag{18}$$

$$\varepsilon_{ts}^i = \frac{1-\tau_s}{y_t^i}\frac{dy_t^i}{d(1-\tau_s)}, \tag{19}$$

$$\varepsilon_t^i = \frac{1-\tau}{y_t^i}\frac{dy_t^i}{d(1-\tau)}, \tag{20}$$

$$\rho^i = \frac{d}{dG_1}\left(\tau_1 y_1^i + \frac{\tau_2 y_2^i}{1+r}\right) \leq 0, \tag{21}$$

$$\eta^i = \frac{d}{dG}\left(\tau_1 y_1^i + \frac{\tau_2 y_2^i}{1+r}\right) \leq 0, \tag{22}$$

where (17)–(18) defines the social value of giving an additional dollar to an agent of type i in period $s = 1, 2$ (in money metric terms) and (19) is the elasticity of period t income with respect to the period s net-of-tax rate. This elasticity allows to capture labor supply adjustments to within-period tax changes as well as across-period tax changes (intertemporal labor substitution effects). Equation (20) is the elasticity of period t income with respect to a change in $1 - \tau$ (a change in the net-of-tax rate in both periods). Equations (21) and (22) define income effect parameters that represent the reduction in present value taxes caused by an increase in lump-sum income in period 1 and lump-sum income in both periods, respectively. Finally, we also define $\mathbb{E}_t[x] = \sum_i \pi^i y_t^i x$ as the period-s-income and population-weighted summation operator.

Proposition 1 (Benchmark Linear Tax Scheme) *Consider the optimization program* (14) *with $\tau_1 = \tau_2 = \tau$ and $G_1 = G_2 = G$, and some fixed value of $1 - \delta$.*

(i) The optimal marginal tax rate τ satisfies

$$\mathbb{E}_1\left[\delta g_1^i + (1-\delta) g_2^i\right] + \frac{1}{1+r}\mathbb{E}_2\left[g_2^i\right] = \mathbb{E}_1\left[1 + \frac{\tau}{1-\tau}\varepsilon_1^i\right] + \frac{1}{1+r}\mathbb{E}_2\left[1 + \frac{\tau}{1-\tau}\varepsilon_2^i\right]. \tag{23}$$

(ii) The optimal per-period transfer G satisfies

$$\sum_i \pi^i \left(g_1^i + \frac{1}{1+r} g_2^i\right) = \sum_i \pi^i \left(1 + \frac{1}{1+r} - \eta^i\right). \tag{24}$$

Proof. See Appendix A.1. ■

The formulations in equations (23) and (24) reflect a dynamic extension of the seminal work on optimal taxation by Atkinson and Stiglitz (1980), among others. The determination of the optimal linear (progressive) tax rate involves balancing equity, as reflected on the left-hand side of the equation (23), with efficiency, as reflected on the right-hand side of the same equation. Similarly, the optimal level of the transfer depends on a balance between equity, shown on the LHS of equation (24), and costs, shown on the RHS of equation (24). These costs are the direct costs of providing the transfer, adjusted for the resulting negative impact on the tax base due to income effects.[9]

We then turn to the age-dependent tax system, allowing $\tau_1 \neq \tau_2$ and $G_1 \neq G_2$, while considering a fixed amount of delayed taxation $1-\delta$.

Proposition 2 (Optimal Age-Dependent Taxation) *Consider the optimization program (14), with some fixed value of $1-\delta$.*

(i) The optimal marginal tax rates (τ_1, τ_2) satisfy:

$$\mathbb{E}_1\left[\delta g_1^i + [1-\delta] g_2^i\right] = \mathbb{E}_1\left[1 - \frac{\tau_1}{1-\tau_1}\left(\varepsilon_{11}^i + \frac{1}{1+r}\frac{\tau_2}{\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{21}^i\right)\right], \tag{25}$$

$$\mathbb{E}_2\left[g_2^i\right] = \mathbb{E}_2\left[1 - \frac{\tau_2}{1-\tau_2}\left(\varepsilon_{22}^i + (1+r)\frac{\tau_1}{\tau_2}\frac{y_1^i}{y_2^i}\varepsilon_{12}^i\right)\right]. \tag{26}$$

(i) The optimal transfers G_1 and G_2 satisfy:

$$\sum_i \pi^i g_1^i = \sum_i \pi^i \left[1-\rho^i\right], \tag{27}$$

$$\sum_i \pi^i g_2^i = \sum_i \pi^i \left[1-\rho^i \frac{1+r}{R'(s^i)}\right]. \tag{28}$$

Proof. See Appendix A.2. ■

The weighted average of the social weights on the LHS of (25) reflects that the burden of a marginal increase in τ_1 is borne partly in period 1 and partly in period 2 (when $\delta \in (0,1)$). The timing of taxes is important because agents are financially constrained. Note that $g_2^i = \frac{1+r}{R'(s^i)} g_1^i$ by virtue of (6). The RHS of (25) and (26) reflect that a marginal change in a tax rate in one period affects labor supply in *both* periods. The intertemporal substitution of labor supply in response to the age-dependent tax rates τ_1 and τ_2 is an important feature of our framework.

Equations (27) and (28) require that G_t, $j = 1,2$ are set so that the average social value of giving everyone an additional dollar in period j ($\sum_i \pi^i g_t^i$) is exactly equal to the resource cost of an additional dollar ($\sum_i \pi^i = 1$) minus the loss of tax revenue due to fiscal externalities (individuals reduce their labor supply when transfers are increased). Note again that $g_2^i = \frac{1+r}{R'(s^i)} g_1^i$. If $R'(s) = 1+r$ when $s > 0$ and $R'(s) > 1+r$ when $s < 0$, then if at least one agent borrows, we have $\sum_i \pi^i g_1^i > \sum_i \pi g_2^i$, working towards $G_1 > G_2$. Note that since $\frac{1+r}{R'(s^i)} \leq 1$, the negative income effects on tax revenue are generally less severe for period 2 labor supply than they are for period 1 labor supply. This is because financially constrained agents discount future cash flows at above-market rates.

We now turn to the optimal amount of delayed taxation, $1-\delta$. To set the stage for our next proposition, we first derive a lemma showing that *marginally* increasing $1-\delta$ has effects on ℓ_t, $t = 1,2$, that are proportional to the effect of marginally changing $1-\tau_1$. This result allows us to characterize optimal delayed taxation in terms of standard labor supply elasticities.

[9] In Marginal Value of Public Funds (MVPF) language, the optimality conditions can be stated as $MVPF_\tau = MVPF_G = 1$.

Lemma 1 *When $s^i \neq 0$, then*

$$\frac{d\ell_1^i}{d(1-\delta)} = \tau_1 \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i} \right] \frac{d\ell_1^i}{d(1-\tau_1)}, \tag{29}$$

where $\bar{\delta}^i = \delta + [1-\delta]\frac{1+r}{R'(s^i)}$ *and* $\bar{\delta}^i \in [0,1]$ *when* $\delta \in [0,1]$, *and*

$$\frac{d\ell_2^i}{d(1-\delta)} = \tau_1 \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i} \right] \frac{d\ell_2^i}{d(1-\tau_1)}. \tag{30}$$

Proof. See Appendix A.3. ■

Proposition 3 characterizes the optimal amount of delayed taxation $1-\delta$. To better convey the effects of delayed taxation, we express this proposition in terms of compensated tax elasticities (indicated by the superscript c).

Proposition 3 (Optimal Delayed Taxation) *Consider the optimization program (14), given some fixed values of τ_1, τ_2, G (not necessarily optimal). Assuming an interior solution for δ, the optimal share of delayed taxation, $1-\delta$, satisfies:*

$$\tau_1 \cdot \mathbb{E}_1\left(g_1^i - g_2^i\right) = -\mathbb{E}_1\left[\tau_1 \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\left(\frac{\tau_1}{1-\tau_1}\varepsilon_{11}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{21}^{i,c} + \rho^i\right)\right], \tag{31}$$

Proof. See Appendix A.4. ■

The LHS is the welfare effect of increased consumption smoothing, and the RHS captures the fiscal externalities of marginally delayed taxation. If there were no financial constraints, both the left and right sides of (31) would equal zero. In other words, if no one is financially constrained, it does not matter whether the government delays taxation. There are no consumption-smoothing benefits, since agents can already borrow freely at the government rate, and on the right-hand side there are no fiscal externalities, since the present value of the delayed tax from the perspective of an unconstrained agent is equal to the nominal tax. However, in the presence of at least one borrowing-constrained agent, that is, an agent who borrows at some $R'(s) = 1+r_b > 1+r$, the left-hand side of (31) is positive and equal to the marginal welfare gains from improved consumption smoothing. The right-hand side is also nonzero due to fiscal externalities caused by constrained agents who now face a lower effective tax rate in period 1.

2.1 Welfare effects of marginal reforms

An interesting question is under what conditions the introduction of delayed taxation increases welfare in an economy without delayed taxation. We start with our benchmark economy, which is characterized by a tax system with $\tau_1 = \tau_2 = \tau$ and no delayed taxation $1-\delta = 0$. We then consider a marginal delay in taxation (i.e., a marginal increase in $1-\delta$). Note that this reform has no mechanical cost to the government. To illustrate the economic forces at play, consider $\frac{dW}{d(1-\delta)}\big|_{1-\delta=0}$, based on equation (A24) in Appendix A.4:

$$\tau \sum_i \pi^i y_1^i \Delta_r^i \Bigg[\underbrace{\lambda \cdot u'(c_1^i)}_{\substack{\text{Welfare gains} \\ \text{from intertemporal} \\ \text{consumption smoothing}}} + \underbrace{\frac{\tau}{1-\tau}\varepsilon_{1,1-\tau_1}^{i,c}}_{\substack{\text{Increase in period-1} \\ \text{tax revenues} \\ \text{(substitution)}}} + \underbrace{\frac{1}{1+r}\frac{\tau}{1-\tau}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c}}_{\substack{\text{Decrease in period-2} \\ \text{tax revenues} \\ \text{(intertemp. substitution)}}} + \underbrace{\rho^i}_{\substack{\text{Decrease} \\ \text{in PV taxes} \\ \text{(income effect)}}} \Bigg], \tag{32}$$

which includes (i) positive welfare effects from increasing intertemporal consumption smoothing, (ii) a positive fiscal externality from increasing tax revenues in period 1 through a substitution effect, (iii) a partially offsetting negative intertemporal substitution effect on tax revenues due to a decrease in labor supply in period 2, (iv) a negative income effect on present value tax revenues.

Lemma 2 below formally presents the welfare effects of three marginal reforms: (i) delaying taxation, (ii) offering a uniform loan, and (iii) lowering the marginal tax rate in period 1. This will allow us to later establish Proposition 4, which relates the welfare effects of delayed taxation to loans and age-dependent taxation.

Lemma 2 *Assume that $R(s)$ is piecewise linear around $s = 0$, $s^i \neq 0$ for all i, and consider an initial benchmark economy without delayed taxation ($\delta = 1$) and age-independent taxation.*

(i) The money-metric welfare effect of a marginal introduction of delayed taxation is:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \tau\Delta_r \left(\sum_{i:s^i<0} \pi^i y_1^i (g_1^i + \rho^i) + \sum_{i:s^i<0} \pi^i \mathcal{X}^i \right). \tag{33}$$

(ii) The money metric welfare effect of the government offering a marginal loan, $dx > 0$, at an interest rate of r, can be written as:

$$\frac{1}{\lambda}\frac{dW}{dx}\Big|_{\delta=1} = \sum_i \pi^i \left(g_1^i - g_2^i\right) + \sum_i \pi^i \Delta_r^i \rho^i = \Delta_r \sum_{i:s^i<0} \pi^i (g_1^i + \rho^i). \tag{34}$$

(iii) The money metric welfare effect of a marginal increase in $1-\tau_1$, while keeping $1-\tau_2$ fixed at $1-\tau$, is:

$$\frac{1}{\lambda}\frac{dW}{d(1-\tau_1)}\Big|_{\delta=1} = \sum_i \pi^i y_1^i (g_1^i + \rho^i) + \sum_i \pi^i \mathcal{X}^i - \sum_i \pi^i y_1^i, \tag{35}$$

where $\Delta_r = 1 - \frac{1+r}{1+r_b} > 0$ and $\mathcal{X}^i = \left(\frac{\tau}{1-\tau} y_1^i \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau}{1-\tau} y_2^i \varepsilon_{2,1-\tau_1}^{i,c}\right)$ denotes the (income-weighted) substitution effects of the tax change for individual i.

Proof. See Appendix A.5 ■

Using Lemma 2, we can now establish Proposition 4.

Proposition 4 (Decomposing Delayed Taxation) *Assume that $R(s)$ is piecewise linear around $s = 0$, $s^i \neq 0$ for all i, and consider an initial benchmark economy with neither delayed taxation nor age-dependent taxation. Then, the marginal welfare effect of introducing delayed taxation can* *be written in terms of either a uniform loan or an age-dependent tax change as follows:*

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \underbrace{\tau\bar{y}_1\frac{1}{\lambda}\frac{dW}{dx}}_{Loan} + \tau_1\Delta_r\Big[\underbrace{\sum_{i:s^i<0}\pi^i(y_1^i-\bar{y}_1)(g_1^i+\rho^i)}_{=Cov(y_1,g_1+\rho|s<0)} + \sum_{i:s^i<0}\pi^i\mathcal{X}^i\Big] \tag{36}$$

$$= \tau\Delta_r\Big[\underbrace{\frac{1}{\lambda}\frac{dW}{d(1-\tau_1)}}_{AD\ tax\ change} - \sum_{i:s^i>0}\pi^i y_1^i(g_1^i+\rho^i) - \sum_{i:s^i>0}\pi^i\mathcal{X}^i + \sum_i \pi^i y_1^i\Big], \tag{37}$$

where $\bar{y}_1 = \sum_{i:s^i<0}\pi^i y_1^i$ *and* $\tau_1\Delta_r = \tau_1\left(1-\frac{1+r}{1+r_b}\right)$.

Proof. See Appendix A.6. ■

Proposition 4 highlights that the marginal introduction of delayed taxation affects only borrowers and shows that the total welfare effects can be decomposed into the effect of a uniform loan of $\tau_1\bar{y}_1$, which has a welfare effect of $\tau_1\bar{y}_1\frac{1}{\lambda}\frac{dW}{dx}$, and an age-dependent reduction in the marginal tax change in period 1 of magnitude $\tau_1\Delta_r > 0$. This marginal tax reduction has redistributive, income, and substitution effects, which are captured by the bracketed terms on the RHS of equation (36). Compared to offering loans, the delayed tax has positive substitution effects on labor supply in period 1, because it implies an effective increase in the marginal return to work for financially constrained agents, along with partially offsetting negative substitution effects in period 2 (as part of the increase in labor supply in period 1 is just a substitution away from labor supply in period 2). In addition, the delayed tax has direct welfare and income effects that are income-dependent, as those with higher period 1 incomes get to delay more taxes, while everyone's disposable income is raised by the same amount in the case of a (uniform) loan.

Proposition 4 also compares the marginal introduction of delayed taxation with a cut in the period 1 marginal tax rate (age-dependent taxation). From (37), we see that the effects of delayed taxation differ from those of a period 1 tax cut of the magnitude $\tau_1\Delta_r$ in the sense that delayed taxation does not affect savers, and it has a mechanical effect on tax revenue.[10]

Note that if everyone borrows, it follows from Proposition 4 that:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \tau_1\Delta_r\left(\frac{1}{\lambda}\frac{dW}{d(1-\tau_1)} + \sum_i \pi^i y_1^i\right) = \tau_1\Delta_r\sum_i \pi^i y_1^i,$$

where the last equality follows because $\frac{1}{\lambda}\frac{dW}{d(1-\tau_1)} = 0$ when the income tax is optimally age-dependent. Thus, we can establish Corollary 1 which discusses what happens when the income tax is age-dependent and everyone borrows.

Corollary 1 *When all agents borrow ($s^i < 0$), starting from no delayed taxation ($\delta = 1$) but optimal age-dependent tax rates (τ_1, τ_2), the money-metric welfare effects of marginally delaying taxation, $d(1-\delta) > 0$ equals*

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \tau_1\Delta_r\sum_i \pi^i y_1^i, \tag{38}$$

[10] Note that when $\delta = 1$, lowering τ_1 does not cause any direct redistribution between periods.

where λ is the Lagrangian of the optimal age-dependent tax rate problem.

The intuition for (38) is that if you delay taxation slightly, you can afford a slightly higher tax rate in period 1 while leaving the effective wedge in period 1 unaffected. This increases the total tax revenue. Note that since marginal taxes are optimally age-dependent in the pre-reform situation, there is no welfare gain from changing the wedge in period 1 by introducing a small amount of delayed taxation.

Before moving on to the quantitative analysis, it is important to clarify a point from our theoretical framework. We set the interest rate on delayed taxes, r_{dtax}, equal to the net saving rate, r. This assumption implies two things: first, for workers whose marginal borrowing rate is equal to r, delayed taxation does not affect their behavior (since for them delayed taxation leaves the present value of the tax rate unchanged); second, the government incurs no additional costs by delaying tax collection, since r is also assumed to be the government's interest rate. This means that delayed taxation does not affect the behavior of net savers in partial equilibrium, nor does it alter outcomes in a frictionless financial environment where $r_b = r$.

In Appendix B, we explore a variation of our model that allows r_{dtax} to deviate from the government's borrowing rate, $r_{gov} = r$, and consider scenarios without financial frictions where $r_b = r$. If the government has the flexibility to set any $r_{dtax} \in \mathbb{R}$, it could emulate any age-based tax system characterized by $(\tau_1, \tau_2) \in \mathbb{R}^2_+$ by selecting $\tau = \tau_2$, assigning $r_{dtax} = -1$, and adjusting $\delta = \tau_1/\tau_2$.[11] Nevertheless, this approach overlooks the heterogeneity of individual borrowing rates, $R'(s^i)$, and fails to exploit the strategic potential of delayed taxation —its ability to tailor tax rates to an individual's borrowing status. In section 3.5 of our quantitative analysis, we explore the implications of treating r_{dtax} as a policy instrument.

3 A quantitative investigation of delayed taxation

3.1 Calibration

We assume that the utility function has the form

$$u(c) - v(\ell) = \frac{c^{1-\sigma} - 1}{1-\sigma} - \xi \frac{\ell^{1+\frac{1}{k}}}{1+\frac{1}{k}}, \tag{39}$$

which implies that σ is the inverse of the elasticity of intertemporal substitution and k is the constant consumption elasticity of labor supply, while ξ is a scaling parameter reflecting the intensity of the disutility of labor. In the simulations below, we choose $\sigma = 5$ and $k = 0.5$. We also set $\xi = 1$. Our baseline analyses consider constant social welfare weights, = 1.

We calibrate our model to Norwegian workers between the ages of 20 and 30 in 1990 who are employed and not in school. There are 100 agents. For each decile in the 1990 wage distribution, there are 10 agents corresponding to their decile in the 2011 wage distribution. We set π_i, $i = 1, \ldots, 100$ equal to the population fraction of each of these types. Our calibration assumes perfect foresight to obtain variation in expected inflation-adjusted wage trajectories. We set

[11] If $\tau_2 = 0$, δ becomes undefined, but this is inconsequential since δ loses relevance if $\tau = \tau_2 = 0$.

the exogenous revenue requirement, M, equal to 15% of "GDP", which is the sum of labor income.

In our calibration, there are no exogenous non-labor income or endowments. Accordingly, all variation in savings incentives (and thus the degree of financial friction) comes from differences in earnings trajectories. For example, those who start in the bottom decile and end in the top decile will want to borrow the most.

We set the base interest rate (faced by the government) to 3%. Since we are modeling periods that are 21 years apart, the cumulative interest rates enter the budget constraints. That is, the present value in period 1 of \$ in period 2 is 1.03^{-21}. Accordingly, we set $\beta = 1.03^{-21}$ so that a net saver facing an interest rate of 3% would choose the same amount of consumption in the two periods.

Policies. For all policies, lump-sum transfers must be equal across periods ($G_1 = G_2$). For age-dependent taxation (AD), the government may choose different tax rates, i.e. we allow $\tau_1 \neq \tau_2$, but we require $\tau_t \geq 0$. Allowing $\tau_1 \neq \tau_2$ is not an option under either the benchmark policy or delayed taxation. If the government can do delayed taxation, we restrict the fraction of period 1 taxes payable in period 1 to be within $[0, 1]$. That is, you cannot borrow from the government in excess of the amount of taxes you accrue, which is typically binding when financial frictions are severe. However, imposing $\delta \leq 1$ is not a binding constraint. In the presence of financial frictions, the optimizing government will not force workers to save an amount proportional to their accrued taxes. In other words, "social security" contributions would not arise in our model.

Individual-level delayed taxes. We further impose that agents dissave rather than delay taxes if they weakly prefer to. This occurs when $R'(s^i) \leq 1 + r_{dtax}$.

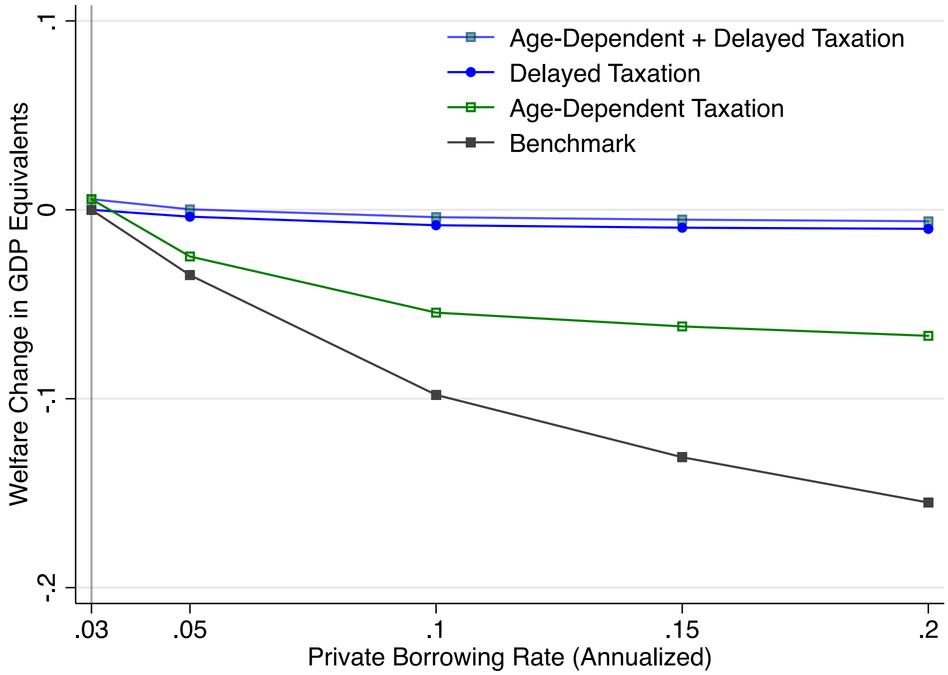
3.2 Main numerical results

Figure 1 summarizes our main results. The bottom black line shows that welfare is lower when the private borrowing rate, r_b, is higher. Going from a private borrowing rate equal to that of the government (3%) to 10% reduces welfare by as much as an exogenous negative GDP shock of 15%. We find that allowing either delayed taxation or age-dependent taxation substantially attenuates this negative welfare effect. With age-dependent taxation, the welfare loss is only 10%, and with delayed taxation, the loss is about 2%. Thus, in a money-metric sense, AD removes about 67% of the welfare cost of financial frictions and DT removes 87%. We interpret this to mean that there are substantial welfare gains from age-dependent taxation —but even more so from delayed taxation.

In addition to contrasting the two policies, we also simulate the effects of combining them. The top line in Figure 1 shows the welfare effects when the government engages in both AD and DT. Contrasting these effects with the stand-alone effects of either AD or DT, we see that age-dependent taxation provides almost no welfare gains on top of a delayed tax policy, while delayed taxation adds substantial welfare gains on top of an age-dependent tax regime.

Figure 1: Welfare Effects from Implementing Delayed and Age-Dependent Taxation under Different Assumptions on Marginal Borrowing Rates

This figure plots the monetary-equivalent reduction in welfare from raising the private borrowing rate above the government rate ($r_{gov} = 0.03$). We consider the lifetime welfare of agents starting in period 1. The black squares show the effect of increasing the borrowing rate in the benchmark economy, where marginal tax rates and transfers are equal across periods. The blue circles show the effect when only the delayed tax policy is implemented: that is, transfers and marginal tax rates are equal across periods, but the government optimally chooses a fraction of the taxes incurred in period 1 to be paid in period 2, that is, $1-\delta$. The green hollow squares show the effect when we instead have age-dependent taxation: both marginal tax rates and transfers may differ across periods. The blue-green squares show the effect of allowing both delayed and age-dependent taxation.



We provide more details in Table 1. Panel A shows summary statistics for the wage trajectories we use for calibration. We see that the median cumulative real wage growth is 0.88. Annualizing this over a 21-year period yields annual real wage growth of 5.16%. This masks considerable heterogeneity. The fifth percentile of cumulative real wage growth is -28% and the ninety-fifth percentile is 464%.

Panel B shows the optimal tax policies and allocations for the case where the private borrowing rate, r_b, is 10%. We see that allowing age-dependent taxation (different τ_1 and τ_2) leads to a corner solution where $\tau_1 = 0$. The intuition for this is that workers' utility is very sensitive to period 1 disposable income, and thus it is optimal to help agents smooth their consumption by taxing them very little in period 1. An interesting observation is that allowing both age-dependent and delayed taxation moves the optimal tax scheme out of this corner solution, with both τ_1 and τ_2 being positive and quite large.

We further see that, whenever we allow for delayed taxation, the government optimally chooses to delay 100% of period-1 taxes. Defining average tax rates (ATRs) as the ratio of taxes minus transfers to labor earnings, we see that delayed taxation produces the smallest (most negative) period-1 ATR.

Another observation is that the present value of taxes is slightly lower with delayed taxation than in the benchmark economy, due to strong income effects. However, we see that the present value of taxes is about 46% higher with delayed taxation than with age-dependent taxation.

Table 1: Optimal Taxation with Financial Frictions

This table provides summary statistics for the calibrated economy when $r_b = 10\%$. The present value function calculates present values according to the government's discount rate. The (money-metric) welfare gap is the exogenous shock to revenue that the government must experience in the baseline case ($r_b = r_{gov} = 3\%$) to be equally worse off as in the benchmark economy (neither delayed nor age-dependent taxation) when the borrowing rate, $r_b = 10\%$. This number is measured as a fraction of the baseline economy's GDP.

	Panel A: Exogenous wage heterogeneity				
	p5	p25	p50	p75	p95
w_1	0.53	0.80	0.98	1.20	1.75
w_2	0.76	1.56	1.83	2.55	5.13
w_2/w_1	0.72	1.29	1.88	2.92	5.64

	Panel B: Tax schedule and allocations with $r_b = 10\%$					
	Benchmark	Delayed Taxation	Age-Dependent	AD & DT	AD+	DT+
τ_1	0.60	0.57	0.25	0.49	0.48	0.62
τ_2	0.60	0.57	0.69	0.62	0.66	0.62
G_1	0.51	0.49	0.37	0.48	0.75	0.48
G_2	0.51	0.49	0.37	0.48	0.00	0.48
$1-\delta$		1.00		1.00		1.00
r_{dtax}		0.03		0.03		0.02
Δ Welfare (% GDP) rel to Benchmark ($r_b = 10\%$)	0.00	8.70	4.22	9.11	8.70	9.11
Δ Welfare (% GDP) rel to Benchmark ($r_b = 3\%$)	-9.80	-0.82	-5.44	-0.39	-0.82	-0.39
means						
$l_1 w_1$	0.93	0.73	0.96	0.79	0.81	0.79
$l_2 w_2$	2.78	3.46	2.82	3.33	3.19	3.33
PV($l_1 w_1$,$l_2 w_2$)	2.43	2.59	2.48	2.58	2.53	2.58
PV($l_1 w_1 \tau_1$,$l_2 w_2 \tau_1$)	1.46	1.48	1.30	1.50	1.52	1.60
l_1	0.88	0.71	0.93	0.76	0.78	0.76
l_2	0.76	0.97	0.79	0.92	0.88	0.92
$(1-\delta^i) l_1 w_1 \tau_1$	0.00	0.22	0.00	0.15	0.00	0.49
s	-0.01	0.00	0.07	0.00	0.18	0.24

3.3 Heterogeneity in welfare effects of increasing financial frictions

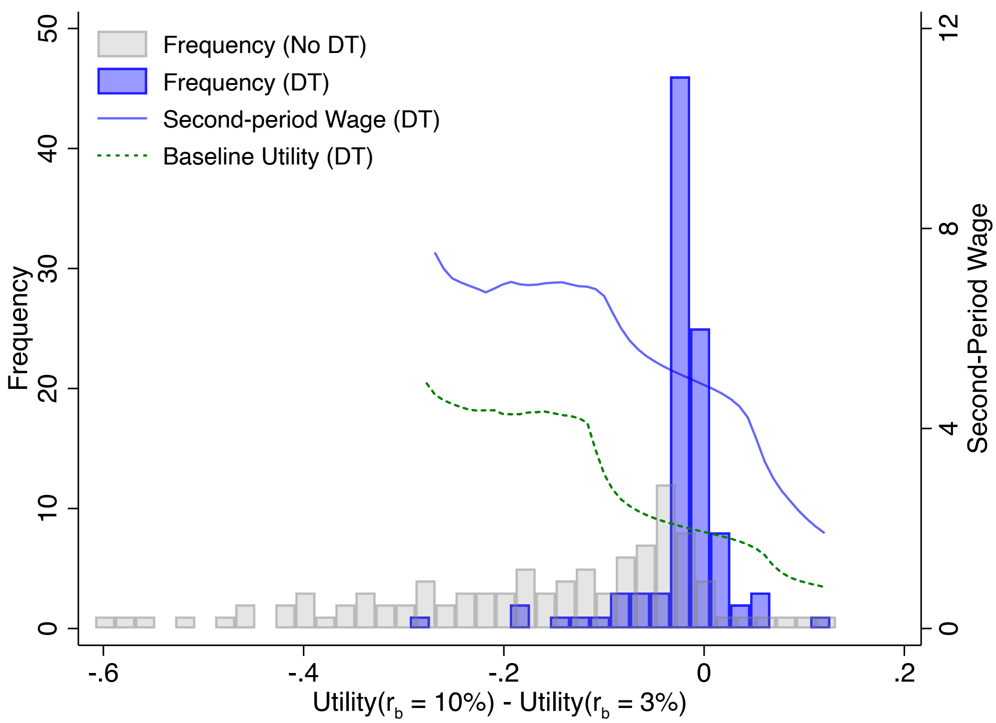
In the previous section we saw, not surprisingly, that increasing financial frictions by raising the private borrowing rate reduces welfare. This is true regardless of whether the government uses age-dependent or delayed taxation. However, as we will show, in the presence of delayed taxation, there are indeed many agents who become *better off* once marginal borrowing rates rise.

Figure 2 plots the distribution of changes in lifetime utility caused by increasing the marginal borrowing rate from 3% to 10%. While most agents experience a decrease in utility, many agents are actually better off. To provide some intuition for this finding, we also superimpose a plot of the mean period 2 wage. This shows that the agents who are better off are those with low period-2 earnings. These households are better off because they were net savers to begin with. Given their low future earnings, they have little future earnings against which to borrow. However, they benefit from the fact that higher borrowing rates imply that delayed taxation reduces the distortionary effects of income taxation more, leading to higher tax revenues and thus larger transfers.

By also superimposing the baseline utility, we see that those who benefit from higher borrowing rates are those with the lowest utility in the baseline case (no financial frictions). Thus, while increasing financial frictions reduces overall welfare, it increases welfare for those who were worse off to begin with.

Figure 2: Delayed Taxation and Heterogeneity in Welfare Effects of Increasing Financial Frictions

This figure plots the distribution of lifetime utility changes from a 3% borrowing rate to a 10% borrowing rate. The gray bars in the background provide the distribution without delayed taxation (DT). The blue bars in the foreground are the distribution with delayed taxation. We also plot a local polynomial fit of second period wages (right-hand y-axis). We also add (solid blue line) the baseline utility, transformed as $10\exp(u)$ to be visible on the second y-axis.



3.4 Reforms when the government is borrowing constrained

An important source of welfare gains from both age-dependent and delayed taxation is improved consumption smoothing, which is facilitated by government borrowing. In our model, the government can borrow an unlimited amount at r_{gov}, and thus is not concerned with the amount it must borrow to finance the various reforms. In practice, this may not be the case.

We proceed by directly addressing the question of optimal tax policy in the presence of government borrowing constraints. When implementing any tax reform, the government can now only borrow .

$$B \leq (1+b)B^{*}(r_b), \tag{40}$$

where $B^{*}(r_b)$ is the amount of government borrowing in the benchmark economy with a borrowing rate of r_b. We consider moderate values of the parameter b from 0 to 1.0, where a value of 0.5 implies that the government can increase borrowing by 50%.

We present our main results for the case where $b = 0.5$ and $r_b = 10\%$ in Table 2. We see that the government now optimally chooses to allow only about 16% of taxes to be delayed. Relative to the unconstrained scenario, the government now faces more pressing tradeoffs in implementing delayed taxation: once the borrowing limit is reached, any increase in $1-\delta$ must be financed either by higher tax rates in period 1 (more distortions) or by lower transfers, G (less redistribution). Similar trade-offs apply to age-dependent taxation. The only way to finance a reduction in τ_1 is to reduce G.

We find that delayed taxation offers the highest welfare gains, equal to about 2.58% of the GDP of the benchmark economy. Interestingly, this welfare effect is also larger than for *AD T&T*, which is a more flexible age-dependent tax system that also allows the government to choose age-dependent lump-sum transfers in addition to age-dependent (marginal) tax rates. This is all the more surprising given that delayed taxation, in which the government chooses only three parameters (δ, τ, G), outperforms a scenario in which the government chooses four parameters $(\tau_1, \tau_2, G_1, G_2)$.

We illustrate how welfare varies with the degree of government borrowing constraints, b, in Figure 3 for the case where $r_b = 10\%$. We see that delayed taxation offers larger welfare gains than all the other reform options whenever $b < \infty$. Interestingly, all policies provide non-negligible welfare gains of about 0.5% of GDP even when the government cannot increase borrowing at all.

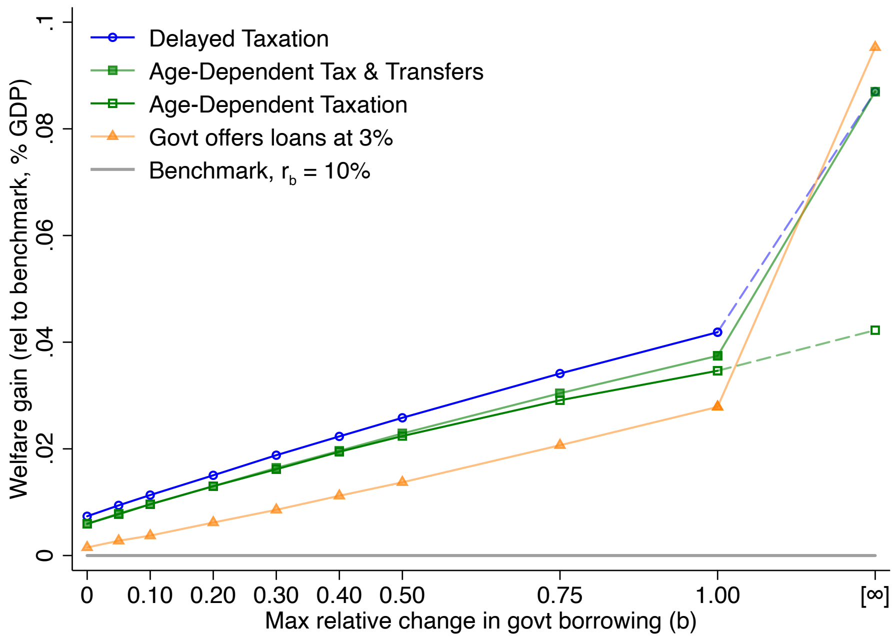
Table 2: Optimal Taxation with Financial Frictions when Government Faces Borrowing Constraints

This table provides summary statistics for the calibrated economy when $r_b = 10\%$ and the government borrowing limit, b, is 50%. The present value function calculates present values according to the government's discount rate. For a given policy reform, the money-metric welfare measure is the exogenous change in government tax revenue that would be required for the government of the benchmark economy to be as well off as it would be with the reform. It is expressed as a percentage of the benchmark economy's GDP (PV labor earnings).

	Benchmark	Delayed Taxation	Age-Dependent	AD (T&T)	Lending
τ_1	0.60	0.51	0.44	0.46	0.55
τ_2	0.60	0.51	0.61	0.57	0.55
G_1	0.51	0.42	0.43	0.44	0.48
G_2	0.51	0.42	0.43	0.38	0.48
$1-\delta$	0	0.16	0	0	0
Govt loan limit	0	0	0	0	0.04
Δ Welfare, % GDP (Benchmark with $r_b = 10\%$)	0.00	2.58	2.24	2.29	1.37
means					
$l_1 w_1$	0.89	0.89	0.92	0.91	0.88
$l_2 w_2$	1.70	1.77	1.68	1.71	1.83
PV($l_1 w_1$,$l_2 w_2$)	1.81	1.84	1.82	1.83	1.86
PV($l_1 w_1 \tau_1$,$l_2 w_2 \tau_1$)	1.09	0.94	0.95	0.94	1.03
ATR_1	-0.00	-0.06	-0.04	-0.04	-0.02
ATR_2	0.24	0.28	0.32	0.31	0.25
l_1	0.88	0.88	0.91	0.90	0.86
l_2	0.76	0.81	0.77	0.79	0.83
$(1-\delta^i) l_1 w_1 \tau_1$	0	0.06	0	0	0
NPV DT	0	0.03	0.00	0	0
s	-0.01	-0.01	0.01	0.00	0.04

Figure 3: Welfare Effects of Delayed and Age-Dependent Taxation With Government Borrowing Constraints

This figure plots the money-metric welfare effects (measured in terms of the GDP of the benchmark economy) of implementing either delayed taxation or age-dependent taxation. We do this for different values of b, which is defined as the maximum relative increase in borrowing relative to the benchmark economy. For example, if $b = 0$, the government can introduce delayed taxation but cannot itself borrow more than it did before implementing delayed taxation.

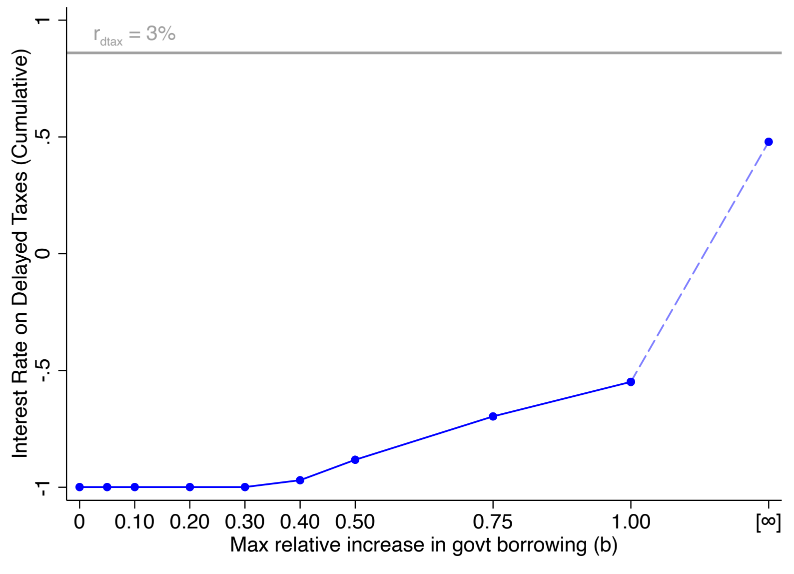


3.5 Unmodeled aspects of the various policies

Our analyses highlight the potential benefits of delayed taxation and show that, in most cases, it generates more welfare than age-dependent taxation. Our modeling does not take into account the possibility that borrowers may default. One concern might be that agents who acquire a significant amount of delayed taxation might choose to evade tax payment by moving abroad and refusing to pay tax liabilities. This problem also applies to delayed taxation: those who benefit from young tax rates when they are young may choose to move to a tax jurisdiction with age-independent tax rates when they are old. In our framework, the possibility of default could *increase* the welfare gains from delayed taxation. This happens because the government, if it could, would like to charge a lower interest rate on delayed taxes than the government interest rate. We show this when we consider "DT+", which is a more flexible form of delayed taxation in which the government is free to choose the interest rate on delayed taxes. When given this freedom, the government chooses a lower interest rate, and when the government is subject to tight borrowing constraints, it chooses to set $r_{dtax} = -1$ (see figure 4), which is equivalent to ex ante loan forgiveness.

Figure 4: Optimal interest rate on delayed taxes

This figure shows the optimally chosen $r_{dtax} \geq -1$ in the flexible delayed tax version. We fix $r_b = 10\%$ and vary the government borrowing constraint parameter, b. The y-axis gives the cumulative interest rate, $(1 + r_{dtax})^{21} - 1$, where a value of -1 implies immediate debt forgiveness.



4 Does delayed taxation already exist?

We define delayed taxation as a scheme in which tax payment, but not tax accrual, is (substantially) delayed. As we show theoretically, the benefits of delayed taxation arise when this wedge is large enough to substantially reduce marginal borrowing rates between the time of tax accrual and tax payment. While the use of this time wedge as a policy tool is novel to our paper, there are several tax systems in place today that also introduce such a wedge. Below, we discuss some of these tax schemes and contrast them with our notion of delayed taxation.

Non-withheld taxes. In the absence of employer withholding, most employees would face a modest tax delay. If taxes are due in, say, April of the following tax year, then taxes on earnings in, say, January are delayed by about 15 months. While most developed countries require employers to withhold taxes, not all countries require the entire tax to be withheld. In Sweden, for example, the progressive portion of the income tax is sometimes not withheld, meaning that some high-income Swedish workers face a modest lag in taxation on marginal earnings.

Installment plans. While many countries allow for the use of installment plans to delay the payment of tax liabilities, these tax deferral systems are typically not designed in a way that resembles our concept of delayed taxation. In the U.S., for example, employers are required to withhold federal income taxes, which means that taxes are paid immediately. The IRS does offer (up to) 10-year installment plans for taxpayers in adverse financial circumstances, but only for the balance due, i.e., the taxes owed less what has already been withheld. Thus, unless a taxpayer expects tax liabilities to substantially exceed withholding amounts, the option to enter into an installment plan will not dampen behavioral responses to labor income taxes.

Social Security. Mandating workers to save a portion of their income for retirement effectively amounts to a negative tax deferral. If we assume that Social Security contributions are about 8% of pre-tax income and taxes are about 24%, then workers are effectively paying 133% of their taxes today and getting 33% back (with interest) in retirement.

The idea that social security contributions can reduce welfare in the presence of financial frictions is not new. Hubbard et al. (1986) argue that when there are liquidity constraints, social security contributions lead to reduced welfare and excessive saving. In discussing the paper, Larry Summers notes that "reversing the direction of transfers" under social security would seem a natural way to reduce welfare losses from financial frictions. However, the key point that negative Social Security taxes (i.e., delayed taxation) can also reduce welfare losses from the distortionary effects of income taxation is missing.[12]

Capital gains. In most jurisdictions, capital gains are not taxed until realized. This effectively allows taxpayers to delay their taxes indefinitely, especially if there is a step-up in basis at death, as in the U.S. However, the underlying mechanism through which delayed capital gains taxation affects behavior is very different from delayed income taxation. Since unrealized capital gains are, in principle, difficult to consume, capital gains deferral does not facilitate intertemporal consumption smoothing in the same way as delayed income taxation.

Student stipends. A case of delayed taxation is when student financial aid is based on income level. In the U.S., for example, the generosity of financial aid generally depends on parental income. Thus, if parents work and earn more, the financial aid package may consist to a greater extent of student loans. If we consider the family as a single economic unit, this is essentially delayed taxation: higher earned income results in the accrual of a (de facto) tax to be paid in the future. In Norway, the mix of financial support provided by the government does not depend on parental income, but rather, mechanically, on the student's own earnings while in school. We discuss this in the next section.

Retirement savings accounts. Tax-deductible contributions to retirement accounts where distributions are taxed as income (such as traditional IRAs in the U.S) also lower effective tax rates.[13] However, these savings incentives are often capped, which means that many workers (who contribute the maximal amount) see no effect on their effective marginal tax rate. In addition, they may be irrelevant for the labor supply decisions of constrained workers whose marginal saving likely go towards reduced borrowing as opposed to IRA contributions.

[12] It is also useful to note that the presence of an income limit ("maximum taxable earnings") for Social Security taxes limits the labor supply distortions for high-income earners. Social security contributions do not affect the marginal effective tax rates of these earners, but only serve to exacerbate financial frictions that may *increase* labor supply through an income effect.

[13] In the U.S., contributions to traditional IRAs are income-tax deductible and distributions are later taxable as income. Under the assumption that marginal savings go into a traditional IRA account, this savings scheme lowers the effective marginal tax rate in two ways. First, the nominal rate changes to the rate that workers expect to face during retirement when the savings are distributed. Since retirement incomes are generally lower than working-age incomes and the tax system is progressive, this channel lowers the nominal tax rate. Second, the effective tax rate is reduced because savings in IRA accounts are not subject to capital taxation and thus grow at a higher (pre-tax) rate of return.

5 De facto Delayed Taxation in Norway

5.1 Empirical setting

Norwegian students receive monthly transfers from the government to pay for housing and other consumption while pursuing higher education. Importantly, these transfers are a mix of stipends and loans. If students earn above a certain threshold, each additional NOK of earnings causes a reduction in the stipend amount which is offset by an equal increase in the student's loan balance.

Our empirical study focuses on the years 2004 to 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. 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% can be forgiven (converted to a stipend) as long as students pass their courses and stay below the earnings thresholds mentioned above. Students are notified of the amount of the transfer at the beginning of the academic year. These notification letters include a breakdown of the transfers, noting the amount (40% of the total) that will be given as a conversion loan, and stating that the conversion of the loan to a stipend is contingent upon income being below an income threshold. The following year, students are notified how much of their loan has been converted, based on grades reported by the educational institutions and income reported to the tax authorities. Loans must generally be repaid within 20 years of graduation. No interest is charged while the student is receiving aid. Thereafter, interest rates are slightly above the risk-free rate and loan payments can be delayed at the (former) student's discretion for up to a total of 3 years.[14]

This study is facilitated by administrative data hosted by Statistics Norway. The key data are derived from tax returns, including data on income, assets and debts of individuals. The sample consists of students who received the standard student support for full-time studies for at least one full financial year during 2004-2011. We restrict the sample to students who received a strictly positive grant after conversion. This excludes students who are ineligible for debt conversion because, for example, they live at home with their parents. This ensures that nearly all students in our sample are subject to income-contingent debt conversion.

Summary statistics are given in Table 3. The average student is 23 years old. This is reasonable given that high school graduation occurs at age 18 and that we condition on students being enrolled in higher education for both semesters within a given year. The summary statistics reveal a significant spread in the amount of liquid assets available to students. While students at the 25th percentile have only NOK 8,000 ($1,300) in liquid assets, students at the 75th percentile have almost *ten* times more. A similar spread can be observed in the liquid assets of the students' parents. We also 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 has a much higher income of around

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

NOK 360,000 ($60,000).

TABLE 3: SUMMARY STATISTICS

This table presents summary statistics. The main sample period is 2004-2011. The 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 earned income within 50,000 of the debt conversion threshold. Liquid assets consist of deposits, investment funds and ownership of public shares. Labor earnings are censored to be below NOK 1,000,000 in 2010 NOK. 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 of the debt conversion threshold with that of the regular tax thresholds, the conversion threshold must be similarly salient. As one of the authors is a former participant in this program, we certainly believe this to be the case. Beyond anecdotal evidence, however, 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. Moreover, students are informed of the existence of such a cap in a loan agreement letter that they must sign, and they also receive letters informing them of any conversion that has occurred. Even if students do not expect to receive a reduction in their debt through conversion, they will want to read these letters to confirm that their institution has accurately recorded and reported their academic progress. Non-passing grades in courses also reduce debt conversion. Students are also informed of their annual student debt balances when they receive their annual prefilled tax returns, which also include information about their income tax liabilities.

5.2 Bunching methodology

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

$$e = \frac{\Delta y^*/y^*}{\Delta\tau/(1-\tau)}, \tag{41}$$

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 counterfactual (absent a kink) probability density function of earnings. We apply the standard approximation

$$B = \int_{y^*}^{y^*+\Delta y^*} h_0(y)dy \approx h_0(y^*)\Delta y^*. \tag{42}$$

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

$$\frac{\Delta y^*}{y^*} = \frac{B}{h_0(y^*)y^*} = \frac{b}{y^*}. \tag{43}$$

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), which we call the bunching estimate. The empirical analog of y^* is the (average) debt conversion threshold, expressed in the same units (thousands) as the empirical earnings bins.[15] We write our estimated compensated labor earnings elasticity as

$$\hat{e} = \frac{\hat{b}/y^*}{\widehat{\Delta\tau}/(1-\tau)}, \tag{44}$$

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 $1 - \tilde{\tau} = 0.75$.

In a standard model without adjustment frictions, the estimator $\hat{e}$ is considered to estimate the Frisch elasticity (Saez 2010 and Kleven 2016). When preferences are additively separable as in our calibration (equation 39), this implies that $\hat{e}$ identifies the structural Frisch elasticity, κ. However, this is not true in the presence of financial frictions and delayed taxation.

In our two-period model, the FOC for period 1 labor supply from equation (8) can be written as:

$$u'(c_1^i) \cdot w_1^i(1 - \tau_1(1-\delta)\Delta_r^i) = v'(\ell_1^i). \tag{45}$$

Differentiating this expression with respect to τ_1, keeping $u'(c_1)$ constant and value it at the threshold where $1 - \delta = 0$, we obtain

$$\varepsilon^{i,F}_{\ell_1, 1-\tau_1} = \left(1 - (1-\delta)\Delta_r^i\right) \frac{v'(\ell_1^i)}{\ell_1^i v''(\ell_1^i)} = \left(1 - (1-\delta)\Delta_r^i\right) \kappa, \tag{46}$$

where $\Delta_r^i = 1 - \frac{1+r}{R'(s^i)}$ is the interest rate wedge and κ is the "structural" Frisch elasticity. In our empirical setting, the marginal tax is fully delayed, i.e., $1 - \delta = 1$, and hence,

$$\varepsilon^{i,F}_{\ell_1, 1-\tau_1} = \left(1 - \Delta_r^i\right) \kappa. \tag{47}$$

Our estimator estimates a scaled-down structural Frisch elasticity, where the scaling depends on local average marginal interest rates. Furthermore, we allow our estimator to be biased downward by a factor of ζ due to, for example, labor supply adjustment frictions. We denote these factors

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

as

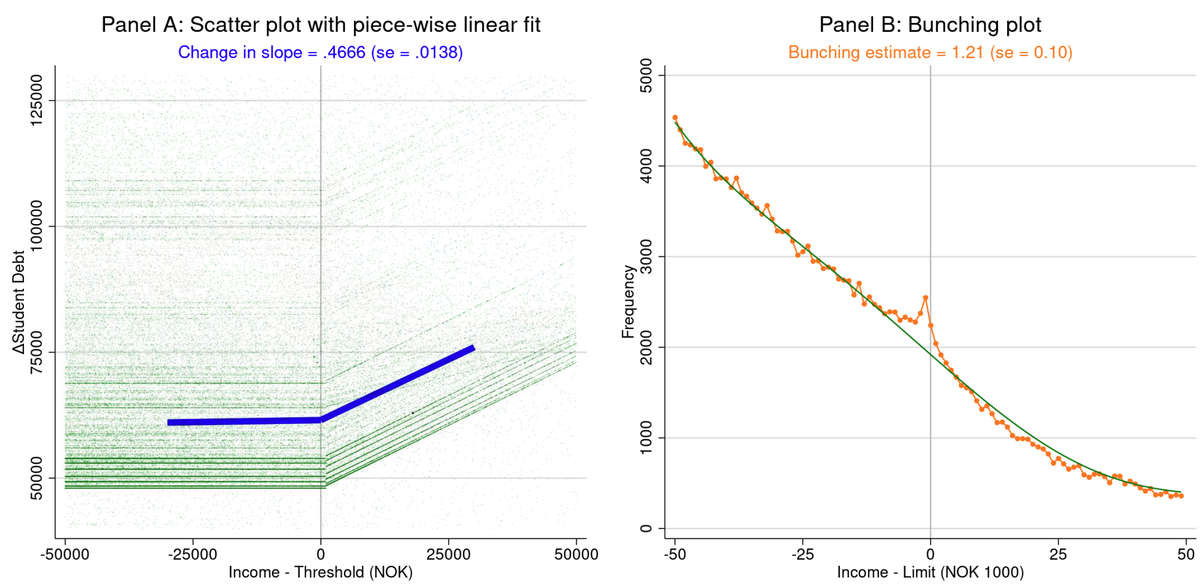
$$E[\hat{e}] = E[\zeta \varepsilon^{i,F}_{\ell_1, 1-\tau_1}] = \underbrace{E\left[1 - \Delta^i_r\right]}_{\text{Delayed taxation effect}} \cdot \underbrace{\zeta}_{\text{bias}} \cdot \underbrace{\kappa.}_{\text{structural Frisch}} \quad . \tag{48}$$

5.3 Bunching at the debt-conversion threshold

Figure 5 summarizes the empirical analysis. Panel A verifies that earnings above the threshold lead to an increase in debt in the next period. Most students are on the expected kinked trajectory, where each additional NOK of earnings increases debt by NOK 0.50. 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 because there are very few non-compliers.[16] In terms of the previous notation, this means that $\widehat{\Delta\tau} = 0.47$.

Figure 5: Verifying the Effect of Excess Earnings on Future Debt and Examining Bunching Responses

Panel (A) shows a scatterplot in green of the relationship between debt accumulation and student earnings around the debt conversion threshold. The fitted blue line illustrates the estimate of the effect of earnings above the threshold and accumulated debt. Panel (B) provides a graphical illustration of the bunching estimate. The orange fitted 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 calculated from bootstrapping (1,000 replications). All plots show statistics from the pooled sample years 2004-2011.



In Panel (B), the yellow dotted line shows the distribution of students around the earnings threshold. The green line is the counterfactual distribution, which is a 5th order polynomial fitted to the non-bunching region. By comparing the actual and counterfactual distributions, we obtain a measure of the excess mass of individuals near the threshold. This provides a bunching

[16] 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 a 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.

estimate, b, of 1.21, which means that there are 121% more individuals around the threshold than the counterfactual distribution implies. Dividing 1.21 by the average threshold amount (120.162 in NOK 1,000s), per equation 44, we obtain a remarkably low elasticity of labor earnings to the net-of-tax (or net-of-debt-increase) wage of 0.0162.[17] The standard error is 0.0013.[18]

This analysis shows that students are remarkably unresponsive 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 also find qualitatively similar results when we consider bunching around the conversion cap. Here, 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 negative, in line with theory, but statistically close to zero (t-stat=-1.64).

5.4 Determinants of non-bunching

We now examine potential determinants of this (non-)bunching behavior. Our main approach is to plot students' characteristics against their position relative to the conversion threshold.[19] This is a visual exercise in which we try to draw conclusions from visual breaks in the relationship between a given characteristic and students' earnings that occur around the conversion threshold.

In Figure 6, Panel (A), we see that the amount of ex-ante liquid wealth drops sharply just above the threshold. This suggests that non-bunchers have less liquid assets, which is consistent with these students being financially constrained. Panel (B) of Figure 6 shows how future labor earnings vary with the student's position relative to the threshold. This shows no sharp increase or decrease in realized future earnings above the threshold, suggesting that non-bunchers do not differ significantly in terms of medium-term earnings prospects.

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

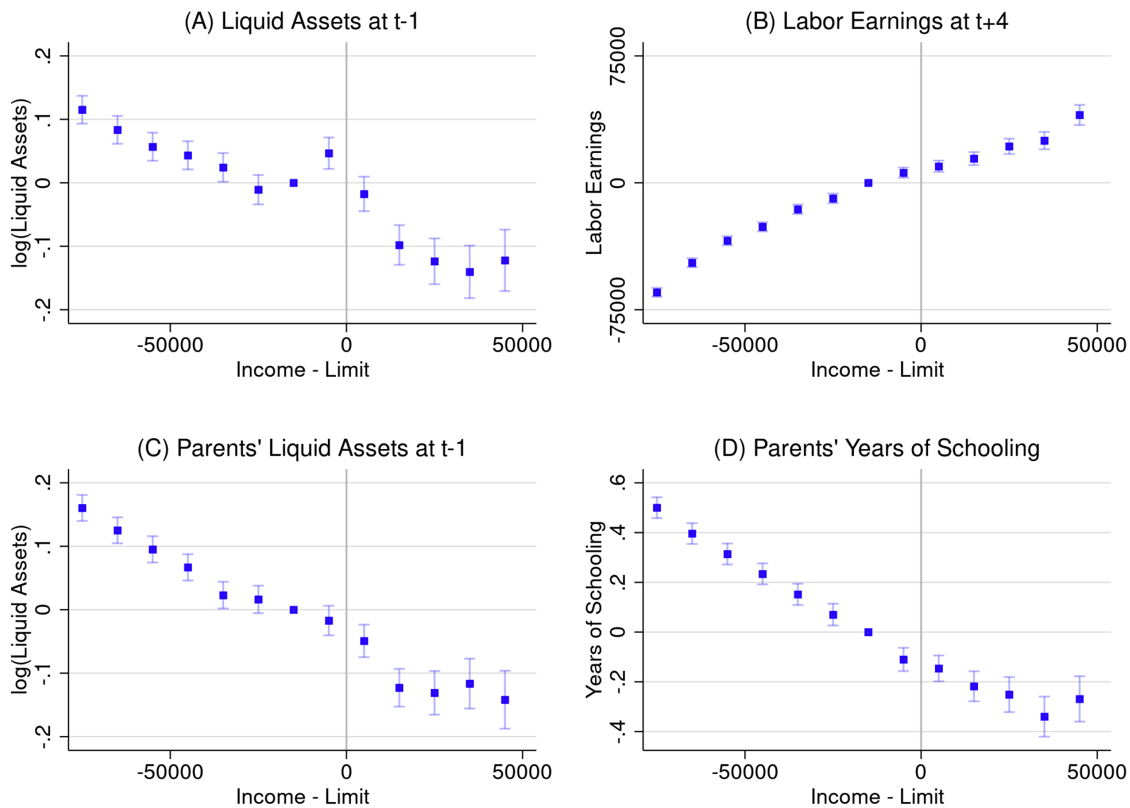
[17]These calculations do not adjust for the fact that any debt accumulated while in school is interest-free. Adjusting for a 3-year 3% interest discount would increase the elasticity by about 9%.

[18]We ignore the (very small) estimation error in $\widehat{\Delta\tau}$.

[19]Another application of this type of analysis can be found in the concurrent work of Bastani and Waldenström (2021), who examine how ability covaries with taxpayers' position relative to a regular tax threshold to infer the ability gradient in tax responsiveness.

Figure 6: Characteristics of Students Below and Above the Income-contingent Debt-conversion Threshold

The graphs below show the financial characteristics of students near the threshold. Panel A looks at students' liquid assets. These consist of deposits, stocks, bonds, and mutual fund 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 schooling among the set of parents. Standard errors, used to provide 95% confidence intervals, are clustered at the student level.



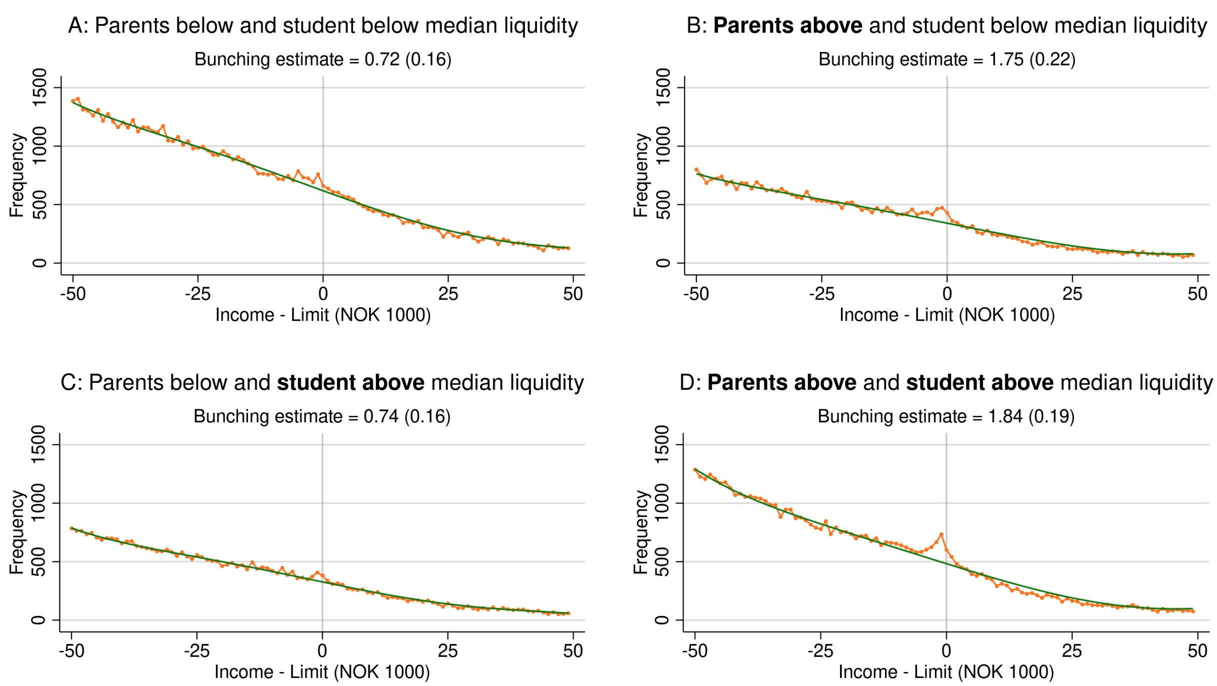
To investigate this liquidity channel further, in panel (C) of figure 6 we also show how *parents'* liquidity correlates with the student's earnings location. This documents a notable negative relationship between parents' financial resources and the child's in-school labor earnings. This suggests that parents play an important role in determining the amount of time students are able to devote to their studies. More relevant to the present study is the finding that parental wealth declines just above the earnings threshold. This suggests that non-bunchers have access to fewer financial resources, which is consistent with financial frictions playing a key role in the observed non-responsiveness to the conversion threshold. However, wealth may be a proxy for human capital, which influences tax responsiveness (Bastani and Waldenström, 2021). Therefore, we plot parental education on the y-axis in panel (D). This shows that there is no break in the relationship between educational attainment, measured by the maximum number of years of schooling of the parents, and the position of the child relative to the conversion threshold. This argues against the hypothesis that fewer resources, in a human capital rather than a financial sense, can explain the irresponsiveness to the threshold. If anything, extrapolating from the relationship below the threshold, non-bunchers may have more highly educated parents. To the extent that this is correlated with students' lifetime wealth, it may explain some of the students' desire to front-load consumption by taking on higher student loans.

5.5 Bunching heterogeneity

We proceed with a complementary, more standard approach to examine the heterogeneity in the earnings sensitivity to the threshold in Figure 7. This approach splits the sample into subsets based on student and parental characteristics to compute heterogeneous bunching elasticities. We see that the largest contribution to the total excess mass in the previous figure 5 comes from students who themselves and their parents have above-average liquid assets. Figure 7 also suggests that the main driver of bunching responses is parents' rather than students' own liquid assets. Moving from the left to the right panel, which improves parental liquidity, more than doubles the bunching estimates.[20]

FIGURE 7: HETEROGENEITY IN BUNCHING BY AMOUNT OF LIQUID ASSETS

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



What can this heterogeneity tell us about how the severity of financial constraints varies with liquidity? In Figure 7, we see that the elasticity increases from 0.72 to 1.84 as we move from below to above the median for both student and parent resources. Using our expression (48) for the expectation of the estimator $\hat{e}$, we can write

$$\frac{1.84}{0.72} = \frac{\mathbb{E}[\hat{e}_{\text{above}}]}{\mathbb{E}[\hat{e}_{\text{below}}]} = \frac{E_{\text{below}}\left[(1-\tau_1)(1-\Delta_r^i)\right] \quad \cdot \quad \zeta \quad \cdot \kappa}{E_{\text{below}}\left[(1-\tau_1)(1-\Delta_r^i)\right] \quad \cdot \quad \zeta \quad \cdot \kappa}. \tag{49}$$

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

From this expression, we get that

$$\frac{1.84}{0.72} = \frac{E_{\text{below}}[1-\Delta_r^i]}{E_{\text{below}}[1-\Delta_r^i]} = \frac{E_{\text{below}}[R'(s^i)]}{E_{\text{below}}[R'(s^i)]}, \tag{50}$$

Given an average maturity for these loans of about 10 years, we find that the gross annual interest rate $(1+r_b)$ is $(1.84/0.72)^{\frac{1}{10}} = 1.0984$ times greater for the below-median liquidity group, roughly a 10 percentage point difference. This is a substantial difference in marginal borrowing rates. While our theoretical framework allows us to calculate implied differences in marginal interest rates, we cannot derive them directly from the data. First, while we can calculate average interest rates on debt, we do not observe marginal interest rates. For example, in practice, students may have a choice between not borrowing and accumulating credit card debt at interest rates close to 20%. If the marginal rate at which they would borrow is 10%, these students would borrow 0, and thus we would not observe any (realized) interest rates for them.

5.6 Analysis of bunching at a regular tax threshold

In this section, we repeat the introductory analyses performed in Figure 5 using a *tax* threshold instead of 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 at which marginally accrued taxes are not delayed. We focus on the first tax threshold in the progressive income tax system. This threshold was NOK 30,000 in 2005-06 and NOK 40,000 in 2007-2011.[21] At this threshold, the marginal income tax increases from 0 to about 25 percent for most taxpayers.

In Figure 8 we examine this complementary empirical setting. Panel (A) provides a scatterplot that verifies the presence of an increase in the marginal income tax rate by plotting total taxes accrued in the year against income. It also provides the fitted 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 because 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. First, we must divide 0.91 by the threshold (36.706 in NOK 1,000) 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). Second, we need to take into account the fact that this is in response to a smaller increase in the marginal (nominal) tax rate. Dividing 2.48% by the relative reduction in the after-tax rate of 19.6%/100%, we get a more substantial elasticity of 0.13.

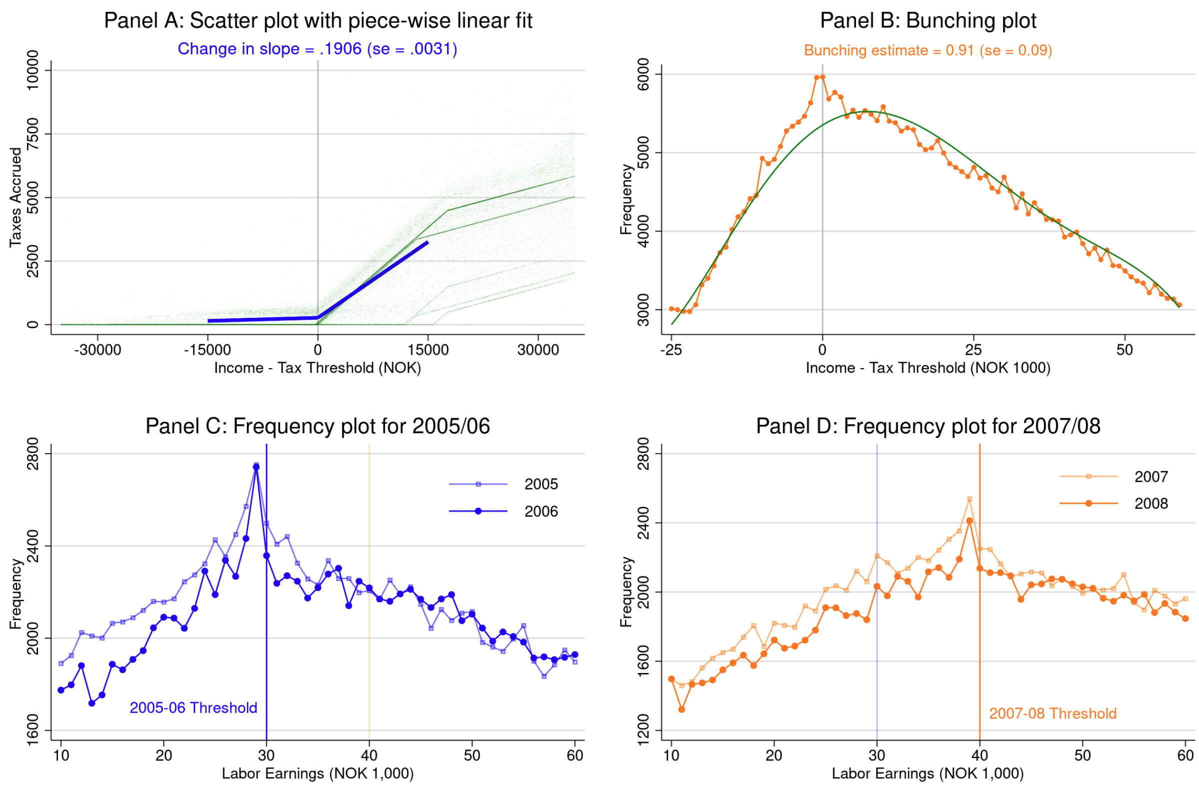
In Panel (B), we see that the bunching mass occurs at the mode of the distribution. If the location of the mode is not driven by students' responses to the tax threshold, then the co-location of the mode and the threshold could lead to an upward bias in the bunching estimate. To address

[21] We omit 2004. In that year the threshold was only NOK 23,000, which significantly reduces the size of the left tail we can use to estimate a counterfactual distribution.

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 change in the mode of the distribution. However, when the tax threshold increased from 2006 to 2007, the mode followed exactly. This reassures us that there is indeed substantial responsiveness to the tax threshold that is not driven by a coincidental co-location of the mode and the threshold.

Figure 8: Bunching at a Regular Tax Threshold

The first and second plots show the relationship between labor income ("pensionable income") and taxes accrued in that year (payable in the same or next year) in the form of a scatterplot and binscatterplot, 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 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 calculated from bootstrapping (N=1,000). All plots are statistics from the pooled sample years.



The elasticity of 0.13 is eight times greater than the elasticity of 0.0162 found in the analysis of responsiveness to the debt conversion threshold. For these differences to be consistent with the same structural Frisch elasticity, κ, we need

$$\frac{0.13}{0.0162} = \frac{E_{\text{delayed}}\left[(1-(1-0)\Delta_r^i)\right]\ \zeta\ \kappa}{E_{\text{regular}}\left[(1-(1-1)\Delta_r^i)\right]\ \zeta\ \kappa} = (1 - E_{\text{delayed}}[\Delta_r^i]). \tag{51}$$

which implies an average annual interest rate over 10 years of $\left(\frac{0.13}{0.162}\right)^{\frac{1}{10}} - 1 = 23\%$. This figure is comparable to average credit card rates, which are slightly above 20%.[22] This suggests that some

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

students are willing to borrow from the Education Loan Fund at an interest rate higher than that offered by financial institutions. This may be partly due to credit rationing, but probably mainly due to the fact that the loan fund does not require payments while students are still in school, and generally has a long maturity, with the additional option of delaying 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 words, how much bunching would there be in Figure 5 if students responded to the debt conversion threshold as if it were a regular income tax threshold? To find out, we reverse the calculation used to derive labor supply elasticities from the bunching estimates. This yields a counterfactual bunching estimate of 23.43.[23] This is considerably larger than the empirical bunching estimate of 1.21.

Figure 9: Contrasting Characteristics of Bunchers at the Delayed Tax and Regular Tax Thresholds

This figure shows how financial characteristics vary across the delayed tax threshold (blue squares) and the regular tax threshold (orange triangles). Panel A looks at the propensity to take out unsecured loans, defined as having interest expenses of more than NOK 1,000. For this sample, we exclude students who, according to the tax authorities, own a house, car or boat. Panel B considers log liquid assets, where liquid assets consist mainly of deposits, but also stocks and bonds. For each threshold, using observations for 2007-2011, we regress the y-variable on earnings-bin fixed effects. We control for year fixed effects and third-order polynomials in the student's age and (max) years of parental education. The bin width is NOK 2000 and is based on the distance between labor earnings and the applicable threshold.

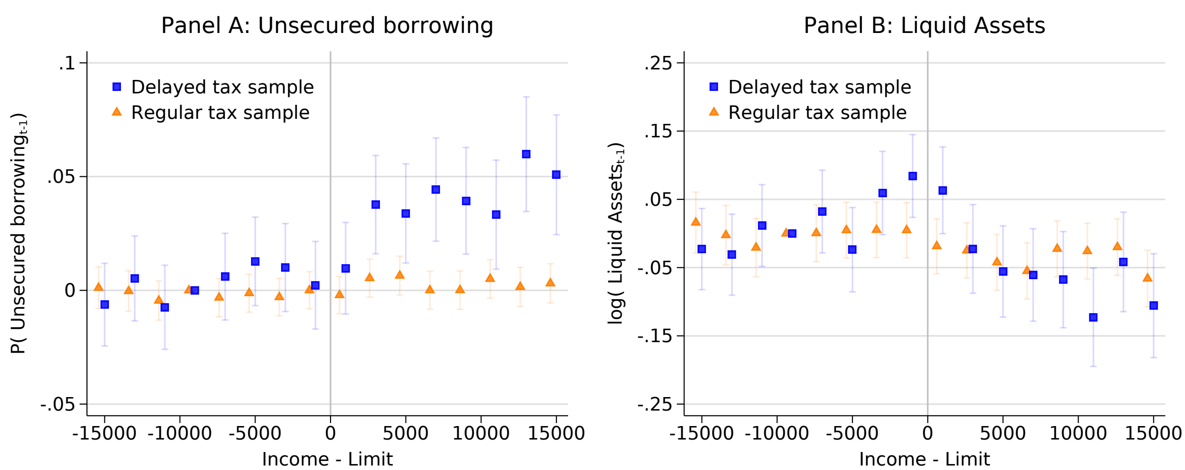


Figure 6 shows that those who bunch at the *delayed tax* threshold have more liquid wealth. This is intuitive because more constrained students have higher marginal borrowing rates on average and are thus less sensitive to a tax that is payable in the future. However, it is useful to show that proxies for financial constraints matter more for delayed tax bunching than for regular tax bunching to rule out the possibility that financial frictions are simply correlated with labor supply adjustment frictions. Accordingly, we empirically investigate whether regular-tax bunchers also appear to be more or less constrained, and we contrast this with the characteristics of bunchers at the delayed-tax threshold in Figure 9.

Panel A shows that the delayed tax non-bunchers are significantly more likely to take on

[23]=0.13*(120162/1000)*(75/50)

unsecured debt, such as credit cards or consumer loans. In the regular tax sample, however, there does not appear to be a systematic relationship between whether someone bunches and their unsecured borrowing. Panel B looks at liquid assets. In the delayed tax sample, we see that bunchers have more liquidity and that those earning above the threshold have significantly less liquid assets. In the regular tax sample, we do not see strong deviations for bunchers, but moving to the right, we see that non-bunchers appear to have less liquid assets.

5.7 Accounting for differences in observables when comparing 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 holding observables fixed.[24] 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 due to, for example, differences in labor supply adjustment frictions.

We first define the individual-level elasticity as

$$\tilde{e}_i = \underbrace{\frac{1[y_i \in BR_s] - \hat{P}^{cf}(y_i \in BR_s)}{\hat{P}^{cf}(y_i \in BR_s)/N_s^{bins}}}_{\text{estimated } b_s} \cdot (\text{Bin width}_s/\text{Threshold}_s) \Big/ (\hat{d}\tau_s/(1-\tau_s)), \tag{52}$$

where $\hat{P}^{cf}$ denotes the estimated (counterfactual) probabilities of being in the bunching region in the absence of 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 about 0.0155, which is very close to our baseline estimate of 0.0162.[25]

We then estimate regression equations of the following form.

$$\tilde{e}_i = \alpha + \beta \mathbb{1}[\text{regular tax sample}]_i + \gamma' X_i + \varepsilon_i, \tag{53}$$

where y_i is the individual's labor earnings and X_i is a vector of individual-level observables, such as the worker's 4-digit occupation code, if available. We report the results of varying the contents of X_i in Table 4. To find the estimated relative increase in labor supply elasticities in the regular versus delayed tax 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 7.20 (CI

[24] See Ring and Thoresen (2021) for a related method that uses regressions of a bunching indicator on observables to infer bunching heterogeneity.

[25] 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 8 shows that this is likely to be a very conservative estimate. Differences arise because in the regression-based approach we take the simpler approach of estimating P^{cf}s using the observed number of observations in the two income bins just below and the two income bins just above the bunching region, BR_s (as in Saez 2010), rather than estimating a higher order polynomial (as in Chetty et al. 2011).

= [6.04, 8.36]) once we control for sex, age, parental education, and 4-digit occupation fixed effects. When we include narrower 2-digit industry interacted with 4-digit occupation code fixed effects, the relative difference is 6.10 (CI = [4.90, 7.30]). This is slightly less than the relative difference obtained by simply contrasting the implied elasticities from the bunching analyses, but the qualitative implications are the same: to rationalize a 6.1 times higher elasticity, we need an average marginal interest rate of 19.82%=$6.1^{1/10}$-1.

Table 4: Regression-based Approach to Account for Differences on Observables in Delayed and Regular Tax Samples

This table presents the results of the regression-based approach to comparing labor supply elasticities in the delayed 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 student's highest paid job within the year. Standard errors are shown in parentheses.

	(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

6 Discussion

This paper introduces the hypothesis that delaying the payment of labor income taxes can reduce their distortionary effects in the presence of financially constrained agents. We exploit a unique setting in Norway that allows us to test this hypothesis empirically. Our results indicate that delaying tax payments, holding the accrual period constant, significantly reduces the distortionary effects of income taxation when agents are credit constrained. We further study delayed taxation in a dynamic optimal tax framework, which we then calibrate to the Norwegian economy in order to numerically assess the welfare effects and contrast them with other policies, such as age-dependent taxation. Our results highlight delayed taxation as a promising new tool in optimal taxation and a fertile ground for further theoretical and empirical research.

There may be important costs associated with nonpayment or debt overhang (see, for example, Donaldson, Piacentino, and Thakor 2019 and Cespedes, Parra, and Sialm 2020) induced by such a system. In our theoretical and calibration exercises, we implicitly assumed that while financial frictions may plague private credit markets, there are no frictions in the effective lending relationship between taxpayers and the government when there is delayed taxation. One possible justification for this would be that the government has a strong advantage in collecting debts. However, even in the absence of such an advantage, we expect welfare gains from delayed taxation. This is because even if the government lends at a "too low" rate, this is compensated for in two ways. The first is, of course, reduced interest costs and increased intertemporal consumption smoothing by agents. The second is reduced distortions from income taxation.

References

Akerlof, G. A. (1978): "The economics of" tagging" as applied to the optimal income tax, welfare programs, and manpower planning," *The American economic review*, 68, 8–19.

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

Atkinson, A. B. and J. E. Stiglitz (1980): *Lectures on Public Economics*, McGraw-Hill Book Company.

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., S. Blomquist, and L. Micheletto (2013): "The Welfare Gains of Age-Related Optimal Income Taxation," *International Economic Review*, 54, 1219–1249.

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

Bastani, S. and D. Waldenström (2021): "The ability gradient in tax responsiveness," *Journal of Public Economics Plus*, 2, 100007.

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.

Blomquist, S. and L. Micheletto (2008): "Age-Related Optimal Income Taxation," *The Scandinavian Journal of Economics*, 110, 45–71.

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

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.*

de Silva, T. (2023): "Insurance versus Moral Hazard in Income-Contingent Student Loan Repayment," *Available at SSRN 4614108.*

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.*

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.

Gervais, M. (2012): "On the optimality of age-dependent taxes and the progressive US tax system," *Journal of Economic Dynamics and Control*, 36, 682–691.

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.

Herbst, D. and N. Hendren (2023): "Opportunity Unraveled: Private Information and the Missing Markets for Financing Human Capital," *NBER Working Paper 29214. Updated May 2023.*

Hubbard, R. G., K. L. Judd, R. E. Hall, and L. Summers (1986): "Liquidity constraints, fiscal policy, and consumption," *Brookings Papers on Economic Activity*, 1986, 1–59.

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.

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

Sheshinski, E. (1972): "The Optimal Linear Income-tax," *The Review of Economic Studies*, 39, 297–302.

Slemrod, J. and W. Kopczuk (2002): "The optimal elasticity of taxable income," *Journal of Public Economics*, 84, 91–112.

Smith, P. (1991): "Lessons from the British Poll Tax Disaster," *National Tax Journal*, 44, 421–436.

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.

Weinzierl, M. (2011): "The surprising power of age-dependent taxes," *The Review of Economic Studies*, 78, 1490–1518.

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 Proofs

A.1 Proof of Proposition 1

Assuming $\tau_1 = \tau_2 = \tau$, differentiating the Lagrangian of the government's optimization problem with respect to τ and setting it equal to zero, yields

$$-\sum_i \pi^i \Big(\delta u_1'(\cdot) y_1^i + \Big([1-\delta](1+r)y_1^i + y_2^i\Big) u_2'(\cdot)\beta\Big) + \lambda \sum_i \pi^i \left(y_1^i + \tau \frac{dy_1^i}{d\tau} + \frac{1}{1+r}\left[y_2^i + \tau \frac{dy_2^i}{d\tau}\right]\right) = 0. \tag{A1}$$

Substituting in for g_t^i and rearranging yields

$$-\sum_i \pi^i \Big(\delta g_1^i y_1^i + \Big([1-\delta]y_1^i + \frac{1}{1+r} y_2^i\Big) g_2^i\Big) + \sum_i \pi^i \left(y_1^i + \tau \frac{dy_1^i}{d\tau} + \frac{1}{1+r}\left[y_2^i + \tau \frac{dy_2^i}{d\tau}\right]\right) = 0. \tag{A2}$$

Setting $\varepsilon_{t,\tau}^i = \frac{1-\tau}{y_t^i}\frac{dy_t^i}{d(1-\tau)} = -\frac{1-\tau}{y_t^i}\frac{dy_t^i}{d\tau}$, we may rewrite as

$$\sum_i \pi^i \Big(\delta g_1^i y_1^i + \Big([1-\delta]y_1^i + \frac{1}{1+r} y_2^i\Big) g_2^i\Big) = \sum_i \pi^i \left(y_1^i + \frac{1}{1+r} y_2^i - \frac{\tau}{1-\tau} y_1^i \varepsilon_{1,1-\tau}^i - \frac{1}{1+r}\frac{\tau}{1-\tau} y_2^i \varepsilon_{2,1-\tau}^i \right). \tag{A3}$$

Equation (23) follows by re-arrangement and using the operator $\mathbb{E}_t[x] = \sum_i \pi^i y_t^i x$, $t = 1, 2$. Condition (24) follows by noting that the the FOC for G can be written as:

$$\sum_i \pi^i \Big(u'(c_1) + \beta u'(c_2)\Big) = \lambda \sum_i \pi^i \left(1 + \frac{1}{1+r} - \tau \frac{dy_1^i}{dG} - \tau \frac{1}{1+r}\frac{dy_2^i}{dG}\right). \tag{A4}$$

A.2 Proof of Proposition 2

Forming the Lagrangian expression of the government optimization problem defined above and letting λ denote the multiplier attached to the government's budget constraint, the first-order condition with respect to τ_1 is

$$-\sum_i \pi^i y_1^i \Big(\delta u_1'(\cdot) + [1-\delta] u_2'(\cdot)\beta(1+r)\Big) + \lambda \sum_i \pi^i \left[y_1^i + \tau_1 \frac{dy_1^i}{d\tau_1} + \frac{\tau_2}{1+r}\frac{dy_2^i}{d\tau_1}\right] = 0, \tag{A5}$$

where the envelope theorem is invoked on the utility terms, V_i. Let $\varepsilon_1^i = \frac{1-\tau_1}{y_1^i}\frac{dy_1^i}{d(1-\tau_1)}$ and $\varepsilon_{2,1}^i = \frac{1-\tau_1}{y_2^i}\frac{dy_2^i}{d(1-\tau_1)}$. We can then write:

$$-\sum_i \pi^i y_1^i \Big(\delta u_1'(\cdot) + [1-\delta]u_2'(\cdot)\beta(1+r)\Big) + \lambda \sum_i \pi^i y_1^i \Big[1 - \frac{\tau_1}{1-\tau_1}\varepsilon_1^i - \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^i\Big] = 0. \tag{A6}$$

Reorganizing and using $\mathbb{E}_1[x] = \sum_i \pi^i y_1^i x$ and the definition of g_t^i, $t = 1, 2$, we can rewrite as

$$-\mathbb{E}_1\left[\delta g_1^i + [1-\delta]g_2^i\right] + \mathbb{E}_1\left[1 - \frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^i - \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^i\right] = 0. \tag{A7}$$

Alternatively, we may write it as

$$\frac{\tau_1}{1-\tau_1} = \frac{\mathbb{E}_1\left[1-\delta g_1^i + [1-\delta]g_2^i\right]}{\mathbb{E}_1\left[\varepsilon_1^i\right]} - \frac{\tau_2}{1-\tau_1}\frac{\mathbb{E}_1\left[\frac{y_2^i}{y_1^i}\varepsilon_{2,1}^i\right]}{\mathbb{E}_1\left[\varepsilon_1^i\right]} = 0. \tag{A8}$$

The formula for τ_2 is directly derived from the first order condition:

$$-\sum_i \pi^i y_2^i \beta u_2'(\cdot) + \lambda \sum_i \pi^i \left[\tau_1 \frac{dy_1}{d\tau_2} + \frac{1}{1+r}\left(y_2^i + \tau_2 \frac{dy_2^i}{d\tau_2}\right)\right] = 0, \tag{A9}$$

which we may rewrite as

$$-\frac{1}{1+r}\mathbb{E}_1\left[\frac{y_2^i}{y_1^i}g_2^i\right] + \mathbb{E}_1\left[-\frac{\tau_1}{1-\tau_2}\varepsilon_{1,1-\tau_2}^i + \frac{1}{1+r}\left(\frac{y_2^i}{y_1^i} - \frac{\tau_2}{1-\tau_2}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_2}^i\right)\right] = 0. \tag{A10}$$

The conditions for G_1 and G_2 follow from the first-order conditions for G_1 and G_2 which are:

$$\sum_i \pi^i u_1'(\cdot) = \lambda \sum_i \pi^i \left[1 - \tau_1 \frac{dy_1^i}{dG_1} - \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_1}\right], \tag{A11}$$

$$\beta \sum_i \pi^i u_2'(\cdot) = \lambda \sum_i \pi^i \left[\frac{1}{1+r} - \tau_1 \frac{dy_1^i}{dG_2} - \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_2}\right]. \tag{A12}$$

These expressions may be further simplified under the assumption that $R'(s^i)$ is well defined, i.e., that $s^i \neq 0$. In that case, changing G_2 by dG_2 is equivalent to changing G_1 by the agent's present value of dG_2. Hence, $R'(s^i)\frac{dy_t^i}{dG_2} = \frac{dy_t^i}{dG_1}$. Hence, the LHS of the last equation becomes

$$\beta \sum_i \pi^i u_2'(\cdot) = \lambda \sum_i \pi^i \left[\frac{1}{1+r} - \left(\tau_1 \frac{dy_1^i}{dG_1} + \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_1}\right)\frac{1}{R'(s^i)}\right]. \tag{A13}$$

Now define $\rho^i = \frac{d}{dG_1}\left(\tau_1 y_1^i + \frac{\tau_2 y_2^i}{1+r}\right)$ as an income effect parameter that provides the change in present-value tax revenues from increasing period-1 unearned income.

A.3 Proof of Lemma 1

Assuming that all $s^i \neq 0$, then $R'(s^i)$ is well-defined and constant for marginal changes in economic incentives. Then, substituting the Euler equation into the FOC for ℓ_1 and re-organizing the life-time budget constraint reveals that δ and τ only enter in a multiplicative manner, which implies that their effect on labor supply is closely related. More formally, we start with the FOC for ℓ_1, which holds when $s^i \neq 0$. Using equation (D11), we can write:

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-\ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\tilde{\tau}_1^i\right)}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-u_1''(c_1^i)w_1^{i2}\left(1-\tilde{\tau}_1^i\right)^2}. \tag{A14}$$

From this, we see that marginal changes in δ and τ_1 only affect l_1 through the term $\tilde{\tau}_1^i = \tau_1[1-(1-\delta)\Delta_r^i]$. By allowing *either* δ or τ_1 to vary, we obtain expressions for $\frac{d\ell_1}{d(1-\tau_1)}$ and $\frac{d\ell_1}{d(1-\delta)}$ that directly lead to (29).

The proof of (30) has the following steps. We substitute the Euler equation (6) into the period-1 intratemporal FOC (8) to obtain an expression that relates ℓ_1 and ℓ_2. We differentiate this to obtain an expression that relates $d\ell_1$ and $d\ell_2$. This allows us to substitute out $d\ell_1$ in (A14) and replace it with an expression for $d\ell_2$. Then the above logic applies, since marginal changes in δ and τ_1 only affect l_2 through the term $\tilde{\tau}_1^i$.

We may also note that the optimal solution $(c_1, c_2, \ell_1, \ell_2)$ to the individual's problem is given by the solution to the following set of equations:

$$u_1'(c_1)w_1^i\left(1-\tau_1\bar{\delta}^i\right) = v'(\ell_1) \tag{A15}$$

$$\frac{u_1'(c_1)}{\beta(1+r_b)}w_2^i[1-\tau_2] = v'(\ell_2) \tag{A16}$$

$$c_1 + \frac{c_2}{1+r_b} = w_1\ell_1(1-\tau_1\bar{\delta}^i) + \frac{w_2\ell_2(1-\tau_2)+G_2}{1+r_b}. \tag{A17}$$

The first condition is just (8), obtained by inserting the intertemporal FOC (6) into (4), the second condition is obtained by inserting (6) into (5), and the third constraint is just the life-time budget constraint.[26] Thus, the optimal individual allocation (and any comparative statics exercise) only depends on τ_1 and δ through the term $\tau_1\bar{\delta}^i$. Note that $\frac{d(\tau_1\bar{\delta}^i)}{d\tau_1} = \bar{\delta}^i$ and $\frac{d(\tau_1\bar{\delta}^i)}{d\delta} = \tau_1\left(1-\frac{1+r}{R'(s)}\right)$.

[26] The latter is derived by noticing that

$$\begin{aligned} c_1 + \frac{c_2}{1+r_b} &= y_1(1-\delta\tau_1) + G_1 + \frac{y_2(1-\tau_2)+G_2}{1+r_b} - \frac{1+r}{1+r_b}(1-\delta)\tau_1 y_1 \\ &= y_1 - \delta\tau_1 y_1 - \frac{1+r}{1+r_b}(1-\delta)\tau_1 y_1 + \frac{y_2(1-\tau_2)+G_2}{1+r_b} \\ &= y_1 - \bar{\delta}^i\tau_1 y_1 + \frac{y_2(1-\tau_2)+G_2}{1+r_b} \\ &= y_1(1-\bar{\delta}^i\tau_1) + \frac{y_2(1-\tau_2)+G_2}{1+r_b} \end{aligned}$$

Thus, we have that

$$\frac{d\ell_1}{d\tau_1} = \frac{d\ell_1}{d(\tau_1\bar{\delta}^i)}\frac{d(\tau_1\bar{\delta}^i)}{d\tau_1} = \bar{\delta}^i\frac{d\ell_1}{d(\tau_1\bar{\delta}^i)} \tag{A18}$$

$$\frac{d\ell_1}{d\delta} = \frac{d\ell_1}{d(\tau_1\bar{\delta}^i)}\frac{d(\tau_1\bar{\delta}^i)}{d\delta} = \tau_1\left(1-\frac{1+r}{R'(s)}\right)\frac{d\ell_1}{d(\tau_1\bar{\delta}^i)} \tag{A19}$$

Substituting from (A18) into (A19), we get:

$$\frac{d\ell_1}{d\delta} = \tau_1\left(1-\frac{1+r}{R'(s)}\right)\frac{1}{\bar{\delta}^i}\frac{d\ell_1}{d\tau_1}.$$

The proof for ℓ_2 is analogous.

A.4 Proof of Proposition 3

We first differentiate the Lagrangian of the government's optimization problem with respect to $1-\delta$ and invoke the envelope theorem on the V_i terms.

$$\sum_i \pi^i\left(u'(c_1)\tau_1 y_1 - \beta u'(c_2)(1+r)\tau_1 y_1\right) + \lambda\sum_i \pi^i\left(\tau_1\frac{dy_1^i}{d(1-\delta)} + \tau_2\frac{1}{1+r}\frac{dy_2^i}{d(1-\delta)}\right) = 0. \tag{A20}$$

We then assume $s^i \neq 0$ and use Lemma 1 to modify the terms in the parenthesis in the second summation term.

$$\left(\tau_1^2\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\frac{dy_1^i}{d(1-\tau_1)} + \tau_2\tau_1\frac{1}{1+r}\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\frac{dy_2^i}{d(1-\tau_1)}\right). \tag{A21}$$

We use the Slutsky equation to rewrite $\frac{dy_1^i}{d(1-\tau_1)} = \left(\frac{dy_1^i}{d(1-\tau_1)}\right)^c + y_1\frac{dy_1}{dG_1}$ and the cross-price Slutsky equation to rewrite $\frac{dy_2^i}{d(1-\tau_1)} = \left(\frac{dy_2^i}{d(1-\tau_1)}\right)^c + y_1\frac{dy_2}{dG_1}$. The expression above becomes

$$\left(\tau_1^2\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\left\{\left(\frac{dy_1^i}{d(1-\tau_1)}\right)^c + y_1\frac{dy_1}{dG_1}\right\} + \tau_2\tau_1\frac{1}{1+r}\left[1-\frac{1+r}{R'(s^i)}\right]\frac{1}{\bar{\delta}^i}\left\{\left(\frac{dy_2^i}{d(1-\tau_1)}\right)^c + y_1\frac{dy_2}{dG_1}\right\}\right). \tag{A22}$$

Further rearranging and using elasticity notation yields

$$y_1^i\frac{\tau_1}{1-\tau_1}\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\left(\left\{\tau_1\varepsilon_{1,1-\tau_1}^{i,c} + (1-\tau_1)\tau_1\frac{dy_1}{dG_1}\right\} + \frac{1}{1+r}\left\{\tau_2\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + (1-\tau_1)\tau_2\frac{dy_2}{dG_1}\right\}\right). \tag{A23}$$

Further using the definitions $\rho^i = \frac{d}{dG_1}\left(\tau_1 y_1^i + \frac{\tau_2 y_2^i}{1+r}\right)$, we may now rewrite the government's FOC with respect to $1-\delta$ as

$$\tau_1 \sum_i \pi^i y_1^i \left(\frac{}{\lambda} \Delta_r^i \cdot u'(c_1^i) \quad + \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i \right)\right) = 0. \tag{A24}$$

Using the definitions of g_1^i and g_2^i and re-arranging yields:

$$\tau_1 \mathbb{E}_1\left(g_1^i - g_2^i\right) = -\mathbb{E}_1\left[\tau_1 \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i \right)\right]. \tag{A25}$$

Using Lemma 1 yields (31).

A.5 Proof of Lemma 2

Part i) Using the derivations from the proof of Proposition 3 (see Appendix A.4), we have that:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)} = \tau_1 \mathbb{E}_1\left(g_1^i - g_2^i\right) + \mathbb{E}_1\left[\tau_1 \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i \right)\right]. \tag{A26}$$

Now divide by τ_1 and use the fact that that $g_2^i = \frac{\beta(1+r)u'(c_2^i)}{\lambda} = \frac{u'(c_1^i)}{\lambda}\frac{1+r}{R'(s^i)} = g_1^i \frac{1+r}{R'(s^i)}$:

$$\frac{dW}{\lambda d(1-\delta)}\frac{1}{\tau_1} = \mathbb{E}_1\left(g_1^i\left[1-\frac{1+r}{R'(s^i)}\right]\right) + \mathbb{E}_1\left[\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i \right)\right]. \tag{A27}$$

Setting $\delta = 1$ yields:

$$\frac{dW}{\lambda d(1-\delta)}\frac{1}{\tau_1} = \mathbb{E}_1\left(g_1^i\left[1-\frac{1+r}{R'(s^i)}\right]\right) + \mathbb{E}_1\left[\left(1-\frac{1+r}{R'(s^i)}\right) \left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i \right)\right]. \tag{A28}$$

Given the piece-wise linear return technology, we have that $R'(s^i) = 1+r$ when $s^i > 0$ and $R'(s) = 1 + r_b$ when $s^i < 0$. Remember $s^i \neq 0$ by assumption. Letting $\Delta_r = 1 - \frac{1+r}{1+r_b} > 0$ we get:

$$\frac{dW}{\lambda d(1-\delta)}\frac{1}{\tau_1\Delta_r} = \sum_{i:s^i<0}\pi^i y_1^i g_1^i + \sum_{i:s^i<0}\pi^i y_1^i \rho^i + \sum_{i:s^i<0}\pi^i y_1^i\left(\frac{\tau_1}{1-\tau_1}\varepsilon^{i,c}_{1,1-\tau_1} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon^{i,c}_{2,1-\tau_1}\right), \tag{A29}$$

$$= \sum_{i:s^i<0}\pi^i y_1^i (g_1^i + \rho^i) + \sum_{i:s^i<0}\pi^i y_1^i\left(\frac{\tau_1}{1-\tau_1}\varepsilon^{i,c}_{1,1-\tau_1} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon^{i,c}_{2,1-\tau_1}\right). \tag{A30}$$

Part ii) We calculate the welfare effect as if everyone accepts the loan. Because, if $s^i < 0$, then agents would strictly prefer to accept, if $s^i > 0$, then agents are indifferent.

$$\frac{dW}{dx} = \sum_i \pi^i\left(u'(c_1^i) - (1+r)\beta u'(c_2^i)\right) + \lambda\sum_i \pi^i\left(\frac{dy_1^i}{dx} + \frac{1}{1+r}\frac{dy_2^i}{dx}\right). \tag{A31}$$

From the agent's perspective, as long as $s^i \neq 0$, a loan of dx is equivalent to $dG_1 = \left(1 - \frac{1+r}{R'(s^i)}\right)dx$. This follows from using the period-2 budget constraint to replace s^i in the period-1 budget constraint. Therefore, we can rewrite the above expression as

$$\frac{dW}{dx} = \sum_i \pi^i\left(u'(c_1^i) - (1+r)\beta u'(c_2^i)\right) + \lambda\sum_i \pi^i\left(1 - \frac{1+r}{R'(s^i)}\right)\left(\tau_1\frac{dy_1^i}{dG_1} + \frac{1}{1+r}\tau_2\frac{dy_2^i}{dG_1}\right). \tag{A32}$$

Further rearranging and using the definition of $\rho^i = \tau_1\frac{dy_1^i}{dG_1} + \frac{1}{1+r}\tau_2\frac{dy_2^i}{dG_1}$,

$$\frac{1}{\lambda}\frac{dW}{dx} = \sum_i \pi^i\left(g_1^i - g_2^i\right) + \sum_i \pi^i\left(1 - \frac{1+r}{R'(s^i)}\right)\rho^i. \tag{A33}$$

The last step follows by realizing that $g_2^i = \frac{\beta(1+r)u'(c_2^i)}{\lambda} = \frac{u'(c_1^i)}{\lambda}\frac{1+r}{R'(s^i)} = g_1^i\frac{1+r}{R'(s^i)}$ and that $R'(s) = 1+r$ when $s > 0$.

Part iii) Equation (A7) in the proof of Proposition 2 provides the money-metric welfare effect of a marginal increase in $1-\tau_1$:

$$\frac{dW}{d(1-\tau_1)\lambda} = \mathbb{E}_1\left[\delta g_1^i + [1-\delta]g_2^i\right] - \mathbb{E}_1\left[1 - \frac{\tau_1}{1-\tau_1}\varepsilon^i_{1,1-\tau_1} - \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon^i_{2,1-\tau_1}\right]. \tag{A34}$$

This holds at any baseline tax system, including an age-independent one where $\tau_1 = \tau_2$. When $\delta = 1$, the expression above simplifies to:

$$\frac{dW}{d(1-\tau_1)\lambda} = \sum_i \pi^i y_1^i g_1^i + \sum_i \pi^i y_1^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right) + \sum_i \pi^i y_1^i \rho^i - \sum_i \pi^i y_1^i, \tag{A35}$$

$$= \sum_i \pi^i y_1^i (g_1^i + \rho^i) + \sum_i \pi^i y_1^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right) - \sum_i \pi^i y_1^i. \tag{A36}$$

A.6 Proof of Proposition 4

Multiplying (34) in Lemma 2 by $\bar{y}_1 = \sum_{i:s^i<0} \pi^i y_1^i$ yields:

$$\frac{\bar{y}_1}{\Delta_r} \frac{dW}{\lambda dx} = \bar{y}_1 \sum_{i:s^i<0} \pi^i (g_1^i + \rho^i). \tag{A37}$$

Taking the difference between (33) in Lemma 2 and (A37) yields:

$$\frac{dW}{\lambda d(1-\delta)} \frac{1}{\tau_1 \Delta_r} - \frac{\bar{y}_1}{\Delta_r} \frac{dW}{\lambda dx} = \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1)(g_1^i + \rho^i) + \sum_{i:s^i<0} \pi^i y_1^i \left(\frac{\tau}{1-\tau} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau}{1-\tau} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right),$$

which can be written

$$\frac{dW}{\lambda d(1-\delta)} = \frac{\tau_1 \bar{y}_1}{\lambda} \frac{dW}{dx} + \tau_1 \Delta_r \left[\sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1)(g_1^i + \rho^i) + \sum_{i:s^i<0} \pi^i y_1^i \left(\frac{\tau}{1-\tau} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau}{1-\tau} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right) \right].$$

This establishes (36). To establish (37), we take the difference between (33) and (35) in Lemma 2 to obtain:

$$\frac{dW}{\lambda d(1-\delta)} \frac{1}{\tau_1 \Delta_r} - \frac{dW}{d(1-\tau_1)\lambda} = \sum_{i:s^i<0} \pi^i y_1^i (g_1^i + \rho^i) + \sum_{i:s^i<0} \pi^i y_1^i \left(\frac{\tau}{1-\tau} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau}{1-\tau} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right) - \left(\sum_i \pi^i y_1^i (g_1^i + \rho^i) + \sum_i \pi^i y_1^i \left(\frac{\tau}{1-\tau} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau}{1-\tau} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right) - \sum_i \pi^i y_1^i \right).$$

Re-arranging we get:

$$\frac{dW}{\lambda d(1-\delta)} = \tau_1 \Delta_r \left(\frac{dW}{d(1-\tau_1)\lambda} - \sum_{i:s^i>0} \pi^i y_1^i (g_1^i + \rho^i) - \sum_{i:s^i>0} \pi^i y_1^i \left(\frac{\tau}{1-\tau} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau}{1-\tau} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right) + \sum_i \pi^i y_1^i \right).$$

A.7 Proof of Proposition 5

Suppose there exists an optimal age-dependent tax scheme characterized by G, τ_1, and τ_2, where $\tau_1 < \tau_2$. We consider an alternative delayed tax scheme on top of the benchmark linear tax policy. We want to show that we can make savers just as well off and borrowers strictly better off while leaving slack in the delayed-tax policy's government budget constraint.

Consider a delayed tax scheme with $1-\delta > 0$, where the age-independent (nominal) tax rate $\tau^* = \tau_2$. Given this $\delta \neq 1$, we set $r_{dtax} \leq r$ such that the effective marginal period-1 tax rate for workers with $R'(s^i) = r$ equals τ_1 from the AD scheme.

$$\tau_1^{i,*} = \tau_2 \left[1 - (1-\delta)\left(1 - \frac{1+r_{dtax}}{1+r}\right)\right] = \tau_1 \quad \text{for all } i \text{ s.t. } R'(s^i) = r \tag{A38}$$

We also set $G^* = G$. This ensures that choices of workers is the same under AD and the new DT policies $\forall i : s^i > 0$. Hence V^i are the same under the AD and DT schemes $\forall i : R'(s^i) = r$. Importantly, (A38) also ensures that the PV tax revenues obtained under AD and DT from all i such that $R'(s^i) = r$ are the same.

We need to ensure that $r_{dtax} > \underline{r_{dtax}}$. From (A38), we that this is equivalent to

$$\frac{\tau_1}{\tau_2} \geq \frac{1+\underline{r_{dtax}}}{1+r}. \tag{A39}$$

Hence if $\underline{r_{dtax}} > -1$, we need $\tau > 0$.

If $R'(s^i) = 1+r$ for all i, then the proof is complete because we have exactly replicated the AD policy. Hence, now we assume that there exists at least one i for which $R'(s^i) = 1 + r_b < 1 + r_{gov}$. We proceed to ensure that V^i increases for those with $s^i < 0$ and that their contribution to tax revenues does not decrease. We want to show that borrowers can be made better off while not violating the government budget constraint. Under the DT policy, net borrowers face an effective tax rate of

$$\tilde{\tau}_1^{i,*} = \tau_2 \left[1 - (1-\delta)\left(1 - \frac{1+r_{dtax}}{1+r}\right)\right] \quad \text{for all } i \text{ s.t. } R'(s^i) = 1 + r_b. \tag{A40}$$

Since $G^* = G$ and $\tilde{\tau}_1^{i,*} < \tau_1$, they are strictly better off under DT than AD, i.e., $V^{i,*} > V^i$ for all i such that $s^i < 0$. We next explore feasibility. The change to PV tax revenues is

$$\sum_{s^i<0} \left(\ell_1^{i,*} w_1^i \tau_2 \left[1 - (1-\delta)\left(1 - \frac{1+r_{dtax}}{1+r}\right)\right] + \frac{\ell_2^{i,*} w_2^i \tau_2}{1+r}\right) - \sum_{s^i<0} \left(\ell_1^i w_1^i \tau_1 + \frac{\ell_2^i w_2^i \tau_2}{1+r}\right). \tag{A41}$$

By virtue of how r_{dtax} is set (equation A38), this revenue change may be rewritten as

$$\sum_{i:s^i<0} \ell_1^{i,*} w_1^i \tau_1 + \frac{1}{1+r} \sum_{i:s^i<0} \ell_2^{i,*} w_2^i \tau_2 - \left(\sum_{i:s^i<0} \ell_1^i w_1^i \tau_1 + \frac{1}{1+r} \sum_{i:s^i<0} \ell_2^i w_2^i \tau_2\right). \tag{A42}$$

or

$$\sum_{i:s^i<0}\left((\ell_1^{i,*}-\ell_1^i)w_1^i\tau_1+\frac{1}{1+r}(\ell_2^{i,*}-\ell_2^i)w_2^i\tau_2\right), \tag{A43}$$

which is equivalent to (B1). Since the government perceives the tax rates in the same way as financially unconstrained individuals, the only thing that matters is the change in labor supply in periods 1 and 2 by the constrained agents who experience the tax rate as $\tilde{\tau}_1^* < \tau_1$ (see equation A40). The slack in the budget constraint will materialize as long as the direct substitution effect on labor supply in period 1 is not offset by income effects (individuals become wealthier over their lifetimes, which may reduce labor supply in both periods) and intertemporal substitution effects (the tax rate cut in period 1 may be accompanied by a labor supply increase in period 1 and a labor supply reduction in period 2).

B Extension: Letting the interest rate on delayed taxes be a policy tool

If the interest rate on delayed taxes, r_{dtax}, is constrained to equal the interest rate on net saving, r, delayed taxation has no effect on the behavior of workers whose marginal borrowing rate is r. Delayed taxation does not change the present value of the tax rate. In our framework, this implies that delayed taxation does not affect the behavior of net savers in partial equilibrium, nor does it have any effect in the absence of financial frictions (when $r_b = r$).

As an extension of our framework, we now allow the interest rate on delayed taxes to be a policy tool. That is, we let r_{dtax} differ from $r_{gov} = r$. We also allow for the possibility that there are no financial frictions, i.e. $r_b = r$. A trivial result if the government can choose any $r_{dtax} \in \mathbb{R}$ is that it can replicate any age-dependent marginal tax system characterized by $(\tau_1, \tau_2) \in \mathbb{R}_+^2$ by choosing $\tau = \tau_2$, setting $r_{dtax} = -1$, and choosing $\delta = \tau_1/\tau_2$. [27] However, such a policy does not exploit the fact that there is heterogeneity in marginal borrowing rates, $R'(s^i)$.

Our next proposition examines whether a flexible delayed taxation policy can replicate (and improve upon) an age-dependent marginal tax schedule under lower bound restrictions on the interest rate.

Proposition 5 (When Delayed Taxation Pareto dominates Age-Dependent Taxation)

Assume that policymakers can choose an interest rate on delayed taxes, $\underline{r_{dtax}} < r_{dtax} \leq r$. *Then any optimal age-dependent marginal tax scheme characterized by* $G_1 = G_2 = G$ *and* $1 > \frac{\tau_1}{\tau_2} \geq \frac{1+\underline{r_{dtax}}}{1+r}$ *may be weakly Pareto dominated by a (not necessarily optimal) delayed tax policy with* $1 - \delta < 1$*, which leaves the following slack in the government budget constraint.*

[27] If $\tau_2 = 0$, δ is not well defined, but it does not matter since δ becomes irrelevant when $\tau = \tau_2 = 0$.

$$\frac{1}{1+r}\sum_{i:s^i<0}\Big((1+r_{dtax})(\ell_1^{i,*}-\ell_1^i)w_1^i\tau_1+(\ell_2^{i,*}-\ell_2^i)w_2^i\tau_2\Big)\,, \tag{B1}$$

where ℓ_t^i *is the labor supply under the AD scheme and* $\ell_t^{i,*}$ *is the labor supply under the DT scheme.* $\ell_t^{i,*}=\ell_t^i\ \forall i: s^i>0$ *and* $\ell_t^{i,*}$*, for* i*, such that* $s^i<0$ *differs from* ℓ_t^i *only because of a lower effective tax rate in period 1 among borrowers,*

$$\tilde{\tau}_1^*=\tau_2\left[1-(1-\delta)\left(1-\frac{1+r_{dtax}}{1+r}\right)\right]\le\tau_1 \quad \textit{for } i \textit{ s.t. } s^i<0. \tag{B2}$$

Proof. See Appendix A.7. ■

Note that (i) when there are no financial frictions (i.e., $r_b=r=r_{gov}$), the delayed tax policy in Proposition 5 exactly replicates the allocations under the age-dependent policy and satisfies the budget constraint with equality.

We also note that (ii), for $r_b>r$, this specific Pareto-dominant delayed tax policy does not exist if the behavioral response to a decrease in the effective tax rate is sufficiently negative to cause tax revenue to decrease. However, we do not consider the assumption of a non-negative revenue effect (B1) to be particularly strong. This is because the relevant revenue effect only includes behavioral responses to a tax cut —and not the mechanical negative effects typically caused by a tax cut. For example, if the age-dependent optimal tax rates of the existing tax system coincide with the revenue-maximizing rates, then the marginal behavioral responses would be strictly positive and equal in magnitude to the negative mechanical effects of a marginal tax rate reduction.

C Bunching Analysis 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 5 for further info on the methodology.

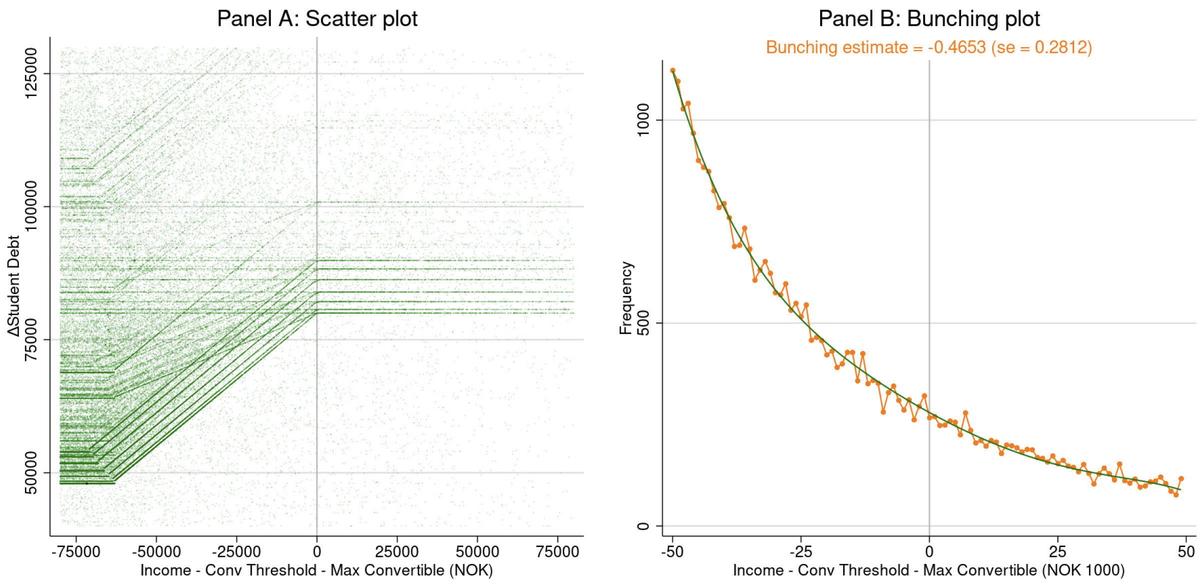
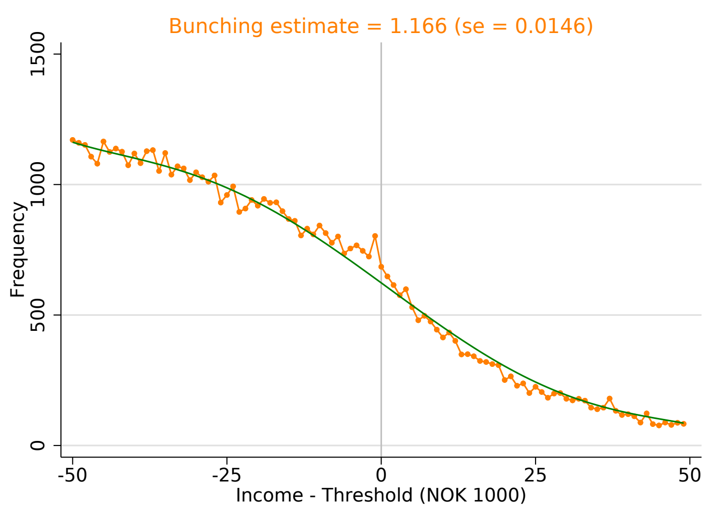


Figure A.2: Bunching at Debt-Conversion Threshold for Workers with Sales and Hospitality Occupations

We repeat the exercise in Panel (B) of Figure 5 on a subset of workers with hospitality (4-digit "STYRK-98"occupation code = 5123, waiters and bartenders) and store sales/clerk jobs (4-digit occupation code = 5221.



D Dynamic Uncompensated and Compensated Elasticities

In dynamic economies, Frisch elasticities impose restrictions that are helpful in obtaining simple elasticity expressions in cases where accounting for the full range of substitution effects across periods would be intractable. In this section, we derive unrestricted elasticities that allow for intertemporal substitution in the context of our two-period framework.

D.1 Derivative of period-1 labor supply w.r.t. $\bar{\delta}^i \tau_1$

We first differentiate c_1 using the first-period budget constraint of agent i, allowing τ_1 and δ^i to vary.

$$dc_1 = d\ell_1 w_1(1-\delta\tau_1) - \ell_1 w_1 d(\delta\tau_1) - ds. \tag{D1}$$

Since $s^i \neq 0$, we can use the period-2 budget constraint to obtain an expression for s^i and differentiate it to obtain

$$ds = \frac{1}{R'(s)}\left[dc_2 + \ell_1 w_1 d\left\{(1-\delta)(1+r)\tau_1\right\} + (1-\delta)(1+r)w_1\tau_1 d\ell_1 - (1-\tau_2)w_2 d\ell_2\right]. \tag{D2}$$

Substituting (D2) into (D1), and using the expression for $\bar{\delta}^i$, yields

$$dc_1 = d\ell_1 w_1(1-\bar{\delta}^i\tau_1) - \ell_1 w_1 d(\bar{\delta}^i\tau_1) - \frac{1}{R'(s^i)}dc_2 + \frac{1}{R'(s^i)}(1-\tau_2)w_2 d\ell_2. \tag{D3}$$

Now we can differentiate the second period intratemporal FOC (5) to get $d\ell_2 = \frac{u''(c_2)}{v''(l_2)} w_2^i (1-\tau_2) dc_2$ and substitute this into (D3) and collect multiplicative terms on dc_2 to get

$$dc_1 = d\ell_1 w_1(1-\bar{\delta}^i\tau_1) - \ell_1 w_1 d(\bar{\delta}^i\tau_1) - \frac{1}{R'(s^i)}\left(1 - [(1-\tau_2)w_2]^2 \frac{u''(c_2)}{v''(l_2)}\right) dc_2. \tag{D4}$$

Similarly, we use the differentiated intertemporal FOC, $u''(c_1)dc_1 = \beta u''(c_2)R'(s)dc_2$, to replace dc_2 with an expression that includes dc_1.

$$dc_1\left[1 + \left(1 - [(1-\tau_2)w_2]^2 \frac{u''(c_2)}{v''(l_2)}\right)\frac{u''(c_1)}{\beta u''(c_2)}\frac{1}{R'(s)^2}\right] = d\ell_1 w_1(1-\bar{\delta}^i\tau_1) - \ell_1 w_1 d(\bar{\delta}^i\tau_1). \tag{D5}$$

Now, we wish to substitute out dc_1 with a term that only contains dl_1. We obtain this by differentiating the FOC for l_1 (equation 8, which relies on the Euler equation 6):

$$u_1''(c_1^i)w_1^i\left(1-\tau_1\bar{\delta}^i\tau_1\right)dc_1 + u'(c_1^i)w_1^i d\left(\tau_1\bar{\delta}^i\right) \quad = \quad v''(\ell_1^i)d\ell_1^i. \tag{D6}$$

We then substitute (D6) into (D5). We denote the term in the brackets in (D5) as ι^i.

$$\left[\frac{v''(\ell_1^i)d\ell_1^i}{u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)} - \frac{u'(c_1^i)w_1^i d\left(\tau\bar{\delta}^i\tau_1\right)}{u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}\right]\iota^i = d\ell_1 w_1(1-\bar{\delta}^i\tau_1) - \ell_1 w_1 d(\bar{\delta}^i\tau_1). \tag{D7}$$

Then we re-arrange to obtain

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\iota^i - w_1^i(1-\bar{\delta}^i\tau_1)u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}. \tag{D8}$$

Then we re-arrange to obtain

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\iota^i - w_1^i(1-\bar{\delta}^i\tau_1)u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}. \tag{D9}$$

This may also be written as

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i u_1''(c_1^i)\tilde{w}_1^i}. \tag{D10}$$

Writing out the ι^i and $\bar{w}_1^i$ terms, we get

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2 u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right] - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2 u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right] - u_1''(c_1^i)w_1^{i2}\left(1-\bar{\delta}^i\tau_1\right)^2}, \tag{D11}$$

which depends explicitly on an individual's marginal interest rate, $R'(s^i)$.

D.2 Period 1 income effects

We want to have an expression for $\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)}$. All derivations assume $s^i \neq 0$.

(a) We use the period-2 budget constraint to substitute in for s in the period-1 budget constraint to get an expression for c_1 and differentiate.

$$dc_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)}\left((1-\tau_2)w_2 d\ell_2 - dc_2\right) \tag{D12}$$

(b) Now find an expression for $d\ell_2$. We use the implied intratemporal FOC for labor (12), and differentiate it to get

$$d\ell_2 = \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)}d\ell_1. \tag{D13}$$

(c) Now find an expression for dc_2. We differentiate the Euler equation (6).

$$dc_2 = \frac{u''(c_1)}{\beta R'(s)u''(c_2)}dc_1 \tag{D14}$$

(d) Now find an expression for dc_1. We differentiate the intratemporal FOC for ℓ_1, (8), which relies on the intertemporal FOC, (6).

$$dc_1 = \frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1. \tag{D15}$$

(e) Now substitute the expressions found in steps (b) and (c) into (a).

$$dc_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)} \left((1-\tau_2) w_2 \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} d\ell_1 - \frac{u''(c_1)}{\beta R'(s) u''(c_2)} dc_1 \right) \tag{D16}$$

(f) Now substitute in the expression for dc_1 found in step (d).

$$\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)} \left((1-\tau_2) w_2 \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} d\ell_1 - \frac{u''(c_1)}{\beta R'(s) u''(c_2)} \frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1 \right) \tag{D17}$$

(g) Reorganize by collecting terms on $d\ell_1$.

$$\left(\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} - \tilde{w}_1 - \frac{1}{R'(s)} \left[\frac{[w_2(1-\tau_2)]^2 v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} - \frac{1}{\beta R'(s) u''(c_2)} \frac{v''(\ell_1)}{\tilde{w}_1} \right] \right) d\ell_1 = dG_1 \tag{D18}$$

(g) Reorganize by collecting terms on $d\ell_1$.

$$\left(\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} - \frac{1}{R'(s)} \left[\frac{[w_2(1-\tau_2)]^2 v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} - \frac{1}{\beta R'(s) u''(c_2)} \frac{v''(\ell_1)}{\tilde{w}_1} \right] - \tilde{w}_1 \right) d\ell_1 = dG_1 \tag{D19}$$

$$\left(v''(l_1) \left[1 - u''(c_1) \frac{1}{\beta R'(s)^2} \left[\frac{[w_2(1-\tau_2)]^2}{v''(\ell_2)} - \frac{1}{u''(c_2)} \right] \right] - \tilde{w}_1 u''(c_1) \tilde{w}_1 \right) d\ell_1 = u''(c_1)\tilde{w}_1 dG_1 \tag{D20}$$

$$\left(v''(l_1) \left[1 + \left(1 - \frac{[w_2(1-\tau_2)]^2 u''(c_2)}{v''(\ell_2)} \right) \frac{u''(c_1)}{u''(c_2)} \frac{1}{\beta R'(s)^2} \right] - \tilde{w}_1 u''(c_1) \tilde{w}_1 \right) d\ell_1 = u''(c_1)\tilde{w}_1 dG_1 \tag{D21}$$

(i) Finally, we may write the income effect term as

$$\frac{d\ell_1^i}{dG_1} = \frac{u''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i) \left[1 + \left(1 - \frac{[w_2^i(1-\tau_2)]^2 u''(c_2^i)}{v''(\ell_2^i)} \right) \frac{u''(c_1^i)}{u''(c_2^i)} \frac{1}{\beta R'(s^i)^2} \right] - \tilde{w}_1^i u''(c_1^i) \tilde{w}_1^i}, \tag{D22}$$

or using the definition of ι^i,

$$\frac{d\ell_1^i}{dG_1} = \frac{u''(c_1^i)\tilde{w}_1}{v''(\ell_1^i)\iota^i - \tilde{w}_1 u''(c_1^i)\tilde{w}_1^i}, \tag{D23}$$

D.3 Slutsky application to obtain period-1 compensated labor supply elasticity

We may use the results in the previous two subsection to get an expression for the compensated period-1 labor supply elasticity, $\varepsilon^i_{1,1-\tau_1}$ as follows.

By Slutsky,

$$\frac{d\ell^i_1}{d\tilde{w}^i_1} = \left(\frac{d\ell^i_1}{d\tilde{w}^i_1)}\right)^c + \frac{d\ell^i_1}{dG_1}\ell^i_1. \tag{D24}$$

If we keep $\bar{\delta}^i$ fixed, $d\tilde{w}^i_1 = -\bar{\delta}^i w_1 d\tau_1 = \bar{\delta}^i w_1 d(1-\tau_1)$. Hence, the relevant Slutsky equation becomes

$$\frac{d\ell^i_1}{d(1-\tau_1)} = \left(\frac{d\ell^i_1}{d(1-\tau_1)}\right)^c + \frac{d\ell^i_1}{dG_1}\ell^i_1\bar{\delta}w^i_1. \tag{D25}$$

Substituting in for the LHS using equation (D11) and the second-term on the RHS using (D23), we obtain

$$-\bar{\delta}^i\frac{u'(c^i_1)w^i_1\iota^i - \ell^i_1 w^i_1 u''_1(c^i_1)\tilde{w}^i_1}{v''(\ell^i_1)\iota^i - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1} = \left(\frac{d\ell^i_1}{d(1-\tau_1)}\right)^c + \frac{u''(c^i_1)\tilde{w}_1}{v''(\ell^i_1)\iota^i - \tilde{w}_1 u''(c^i_1)\tilde{w}^i_1}\ell^i_1 w^i_1\bar{\delta}^i. \tag{D26}$$

Rearrange and cancel out to get

$$\left(\frac{d\ell^i_1}{d(1-\tau_1)}\right)^c = -\bar{\delta}^i\frac{u'(c^i_1)w^i_1\iota^i}{v''(\ell^i_1)\iota^i - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1}. \tag{D27}$$

In terms of an elasticity, we can write it as

$$\varepsilon^c_{1,1-\tau_1} = -\bar{\delta}^i\frac{1-\tau_1}{\ell^i_1}\frac{u'(c^i_1)w^i_1\iota^i}{v''(\ell^i_1)\iota^i - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1}. \tag{D28}$$

We may rewrite this using the intratemporal FOC (8), which says that $u'(c^i_1)\tilde{w}^i_1 = v'(\ell_1)$ and thus $u'(c^i_1)w^i_1 = \frac{1}{1-\bar{\delta}^i\tau 1}v'(\ell_1)$ to get

$$\varepsilon^c_{1,1-\tau_1} = -\frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\frac{1-\tau_1}{\ell^i_1}\frac{v'(\ell_1)\iota^i}{v''(\ell^i_1)\iota^i - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1}. \tag{D29}$$

Writing out the ι^i terms, we get

$$\varepsilon^c_{1,1-\tau_1} = -\frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\frac{1-\tau_1}{\ell^i_1}\frac{v'(\ell_1)\left[1+\left(1-\frac{[w^i_2(1-\tau_2)]^2u''(c^i_2)}{v''(\ell^i_2)}\right)\frac{u''(c^i_1)}{u''(c^i_2)}\frac{1}{\beta R'(s^i)^2}\right]}{v''(\ell^i_1)\left[1+\left(1-\frac{[w^i_2(1-\tau_2)]^2u''(c^i_2)}{v''(\ell^i_2)}\right)\frac{u''(c^i_1)}{u''(c^i_2)}\frac{1}{\beta R'(s^i)^2}\right] - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1}. \tag{D30}$$

Writing out the $\tilde{w}_1^i$ and $\bar{\delta}^i$ terms, we get

$$\varepsilon^c_{1,1-\tau_1} = -\frac{\frac{\delta+(1-\delta)\frac{1+r}{R'(s^i)}}{1-\left[\delta+(1-\delta)\frac{1+r}{R'(s^i)}\right]\tau_1}\frac{1-\tau_1}{\ell_1^i}v'(\ell_1)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-\left[w_1^i\right]^2\left(1-\tau_1\left[\delta+(1-\delta)\frac{1+r}{R'(s^i)}\right]\right)^2u_1''(c_1^i)}. \tag{D31}$$

E Data for calibration

We first use microdata from the 1990 and 2011 censuses to compute micro-level data on (proxies for) effective hourly wages.

Wages in 1990.

- To construct a measure of effective wages in 1990, we first construct a measure of hours worked. The 1990 census ("folke- og boligtellingen") contains categories for usual weekly hours worked, [1,10], [11,20], [20,30], [30,35], or full time (37 hours). We assign numeric values of 5, 15, 25, 32.5, and 37 to these categories. This variable is defined as *typical hours worked.*
- We calculate a measure of *minimum hours worked* as 37.5 hours × 47 weeks × number of reported months of full time work / 12 months.
- The survey also includes information on the number of months worked full-time and part-time. We sum the number of months to get a measure of *total months worked.*
- Our final measure of weekly hours worked is then max(*typical hours worked* × 47 weeks × *total months worked* / 12, *minimum hours worked*).
- WWe replace hours worked as missing if either the number of months worked is less than 2 or the total hours worked is less than 47 × 7.5 (i.e., we require an average of 7.5 hours worked per week).
- We calculate the 1990 wage as annual labor income divided by the number of hours worked. To avoid low wage outliers, we replace as missing whenever the wage is below 25% of the median wage. This drops about 3.5% of the observations. To avoid high wage outliers, we winsorize at the 99.9th percentile.

Wages in 2011

- Using the 2011 Census, we calculate hours as the number of contractual weekly hours × 47 weeks.
- We then calculate wages as annual labor earnings / hours worked. We take the same approach to dealing with outliers as for 1990 wages.
- Finally, we deflate the 2011 wages by 1.5576, which is the ratio of the 2011 to 1990 Consumer Price Index from Statistics Norway.

We then keep only observations for which the age in 1990 was between 20 and 30. For individuals under the age of 25, we further require that their age exceeds 7 (first year of education) plus the number of years of education reported in the 2011 census by one year. We also require that we observe effective wages for the individual in both 1990 and 2011.

Wage trajectories for calibration.

Using the microdata above, we calculate the median wage within each 1990 decile and each 2011 decile. This gives us 100 different (w_1, w_2) combinations: $i = 1, ..., 100$. For each wage combination, we assign π_i using the empirical probabilities in the microdata above.