Optimal Delayed Taxation in the Presence of Financial Frictions

Preliminary version, currently being revised

Spencer Bastani and Marius A. K. Ring*

February 26, 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. We thank seminar participants at MIT, UT Austin, IFN, Washington University in St. Louis, and LMU Munich 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 comes from taxes on labor income. This leads to efficiency losses, since labor income taxes are generally believed to 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 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 thus 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.[1] This 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 lower behavioral elasticities with respect to nominal tax rates. The notion that the government may 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 paper conducts a comprehensive study of delayed taxation. 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 interest rates faced by net borrowers and net savers.

[1] 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.

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, (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 since 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-based taxation by offering tax breaks to young workers, this would affect all workers, not just those who are financially constrained. In addition, with age-dependent taxation, there would be a mechanical loss of tax revenue. 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 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.[2] 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 (iii) delayed taxation. We also examine the welfare effects of letting the government lend directly to households at the lower 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

[2] 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.

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 substantial 44%, but delayed taxation reduces it by an even more dramatic 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 delayed 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 delayed taxation are about 2.6% and the gains from age-dependent taxation are about 2.2%. Interestingly, under government borrowing constraints, 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.

Finally, when government borrowing is restricted, we also examine the extent to which a government lending program, wherein the government offers loans at the government interest rate up to a uniform loan limit. This policy increases produces welfare gains of about 1.37% of GDP, which is barely half that of delayed taxation. The low relatively low gains from government lending is consistent with our theoretical findings. By simply offering loans, the government does not exploit the possibility of also reducing the distortionary effects of income taxation.

The final part of the paper empirically tests the intuitive implication of a simple life-cycle model with borrowing constraints and endogenous labor supply—namely, that delayed taxation reduces the distortionary effects of income taxes when taxpayers face borrowing constraints. 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.[3] 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 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

[3] 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.

us to compare implied labor supply elasticities under different tax regimes but among a similar sample of individuals.[4] 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, this paper 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, test, and study the optimal tax policy implications of the simple idea that —in the presence of credit market imperfections —altering the timing of income tax payments may reduce the distortionary effects of taxation. To our knowledge, there is no theoretical or empirical research on this topic.

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.[5] The conceptual novelty of our paper 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 con-

[4] 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.

[5] 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.

strained 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.[6]

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) or loan-term thresholds (see, e.g., Bachas, Kim, and Yannelis 2021; Bäckman, van Santen et al. 2020; DeFusco and Paciorek 2017; DeFusco, Johnson, and Mondragon 2020; and Best, Cloyne, Ilzetzki, and Kleven 2018). Our contribution is to study bunching at a threshold where the *payment* of marginally accrued taxes is substantially delayed. This adds an intertemporal dimension to bunching behavior not present in studies that consider the sensitivity to taxation.[7] 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.

2 The welfare gains of delayed taxation

2.1 A simple illustration

Consider a two-period model where a government imposes proportional tax rates on labor income of τ_1 and τ_2 in periods 1 and 2, respectively, and where a fraction $1-\delta$ of the tax payment

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

[7]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.

in period 1 is deferred to period 2. The government charges interest on the delayed tax payment equal to r. The saving technology of private agents is given by the function $R(s)$, which is the amount by which the disposable income in the second period is increased if the *individual* saves (or borrows) the amount s.

The individual maximization problem is:

$$\begin{aligned} \max_{\ell_1,\ell_2,s} \quad & u(c_1) - v(\ell_1) + \beta[u(c_2) - v(\ell_2)], && (1)\\ \text{s.t. } c_1 &= w_1\ell_1[1-\delta\tau_1] + x_1 - s, && (2)\\ \text{and } c_2 &= w_2\ell_2[1-\tau_2] + x_2 - (1+r)[1-\delta]\tau_1 w_1\ell_1 + R(s), && (3)\end{aligned}$$

where u is increasing, twice differentiable and strictly concave, and v is increasing, twice differentiable, and strictly convex. For $t = 1, 2$, c_t is the consumption, w_t is the wage rate, ℓ_t is the labor supply, and x_t the non-labor income in period t.

The first order condition for ℓ_1 can be written:

$$u_1'(c_1)w_1\left(1 - \tau_1\Big[\delta + [1-\delta]\theta\Big]\right) = v'(\ell_1) \quad \text{where} \quad \theta = \frac{\beta(1+r)u'(c_2)}{u'(c_1)}. \tag{4}$$

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, $\theta = 1$, which implies that the first-order condition is independent of δ. However, in the presence of financial frictions, we have that $\theta < 1$, which implies that when $0 < \delta < 1$, the effective marginal tax rate faced by agents is lower than in an economy without delayed taxation.[8]

As can be seen from (4), the effects of delayed taxation depend on the degree of consumption smoothing. To gain insight into the latter and to make further progress, we consider a specific saving technology:

$$R(s) = \begin{cases} (1+r)s & \text{if } s \geq 0 \\ (1+r_b)s & \text{if } s < 0, \end{cases} \tag{5}$$

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 $\frac{dR(s)}{ds} = 1 + r_b$ and the first-order condition for savings is:

$$u'(c_1) = (1+r_b)\beta u'(c_2). \tag{6}$$

[8] For the second period, the first order condition for ℓ_2 is $w_2(1-\tau_2)u'(c_2) = v'(\ell_2)$. Thus, the optimality condition for ℓ_2 is the same as in a standard model without delayed taxation.

This implies that $\theta = \frac{1+r}{1+r_b} < 1$ and (4) has the form:

$$u_1'(c_1)w_1\left(1 - \tau_1\Big[\delta + [1-\delta]\frac{1+r}{1+r_b}\Big]\right) = v'(\ell_1) \tag{7}$$

Equation (7) illustrates that increasing the share of period 1 taxes that are delayed (increasing $1-\delta$) reduces the distortion on labor supply in period 1 at a rate determined by the difference in interest rates between government and private agents, which captures the "strength" of the financial friction. Intuitively, borrowing constrained individuals smooth their consumption by increasing their period 1 labor supply, which makes it less elastic to taxes.

2.2 Optimal delayed taxation

We now formally consider delayed taxation in an optimal tax framework. We consider heterogeneous households that differ in their labor market productivity, which allows us to study the welfare gains from delayed taxation in the presence of distributional considerations. For tractability, we focus on a dynamic extension of the linear (progressive) taxation framework (Sheshinski 1972). The tax schedules in periods 1 and 2 are given by:

$$\begin{aligned} T_1(y_1) &= -G_1 + \tau_1 y_1 \\ T_2(y_2) &= -G_2 + \tau_2 y_2, \end{aligned}$$

where G_t, $t = 1, 2$ are age-dependent lump-sum taxes/transfers.[9] We focus on the case $G_t \geq 0$, $t = 1, 2$, which realistically precludes lump-sum taxation but allows for age-dependent lump-sum transfers.[10] The linear tax system with age-dependent lump-sum transfers and delayed marginal taxation has the form

$$\begin{aligned} \tilde{T}_1(y_1) &= -G_1 + \delta\tau_1 y_1 \\ \tilde{T}_2(y_1, y_2) &= -G_2 + \tau_2 y_2 + (1+r)(1-\delta)\tau_1 y_1, \end{aligned}$$

where $\delta \in [0, 1]$ is the fraction of the period 1 tax that has to be paid in period 1.

The individual's problem The problem solved by an individual i with a lifetime wage profile of (w_1^i, w_2^i) is:

$$V^i = \max_{\ell_1^i, \ell_2^i, s^i} \Big(u_1(c_1^i) - v(\ell_1^i) + \beta[u_2(c_2^i) - v(\ell_2^i)]\Big), \tag{8}$$

[9] In the model examined in section 2.1, lump-sum taxes/transfers had to be excluded because they would imply that a first-best allocation could be achieved (the social planner would raise the required tax revenue without imposing any labor distortions and would eliminate the financial friction). This is a well-known limitation of optimal tax analysis in representative agent settings. In the more realistic setting considered here, uniform age-dependent lump-sum taxes/transfers are generally not sufficient to implement a first-best allocation, since the revenue-raising potential of lump-sum taxation must be weighted against its distributional incidence, and different agents may face different intertemporal consumption trade-offs (in the sense that θ in equation (4) may differ across individuals).

[10] 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.

subject to:

$$c_1^i = w_1^i \ell_1[1 - \delta\tau_1] + G_1 - s^i, \tag{9}$$

$$c_2^i = w_2^i \ell_2[1 - \tau_2] + G_2 - (1 + r)[1 - \delta]\tau_1 w_1^i \ell_1^i + R(s^i), \tag{10}$$

with associated first-order conditions:

$$(\ell_1) : u_1'(c_1^i) w_1^i [1 - \delta\tau_1] - \beta(1 + r) u_2'(c_2^i)[1 - \delta]\tau_1 w_1^i = v'(\ell_1^i), \tag{11}$$

$$(\ell_2) : u_2'(c_2^i) w_2^i [1 - \tau_2] = v'(\ell_2), \tag{12}$$

$$(s) : u_1'(c_1^i) = \beta u_2'(c_2^i) R'(s^i). \tag{13}$$

Insertion of (13) into (11) yields:

$$u_1'(c_1^i) w_1^i \left(1 - \tau_1 \left[\delta + [1 - \delta]\frac{1 + r}{R'(s^i)}\right]\right) = v'(\ell_1^i). \tag{14}$$

This is just a slightly more general version of (7). For future use we introduce the notation $\tilde{w}^i = w_1^i \left(1 - \tau_1 \left[\delta + [1 - \delta]\frac{1+r}{R'(s^i)}\right]\right)$ to denote the "net" wage rate relevant for the period 1 labor supply decision for agent i.

Substituting (13) into (14) and then substituting in (12), we obtain

$$\beta R'(s^i) \frac{\tilde{w}^i}{w_2^i(1 - \tau_2)} \cdot v'(\ell_2^i) = v'(\ell_1^i), \tag{15}$$

which relates period-1 and period-2 labor supply.

To simplify notation, we define agent i's interest wedge to be

$$\Delta_r^i = 1 - \frac{1 + r}{R'(s^i)}. \tag{16}$$

Note that this implies that the effective tax rate, $\tilde{\tau}_1^i$, can be written

$$\tilde{\tau}_1^i = \tau_1 \left[\delta + [1 - \delta]\frac{1 + r}{R'(s)}\right] = \tau_1 \left[1 - (1 - \delta)\Delta_r^i\right], \tag{17}$$

which emphasizes the fact that the effective tax rate equals the nominal tax rate when there is no delayed taxation ($1 - \delta = 0$) or there are no financial frictions ($\Delta_r^i = 0$).

The government's problem Assuming that the government assigns a welfare weight of α^i to agents of type i, and denoting by π^i the proportion of agents of type i in the population, the government's problem in the presence of age-dependent taxation and the potential for delayed taxation can be expressed as

$$\max_{G_1, G_2, \tau_1, \tau_2, \delta} \sum_i \alpha^i \pi^i V^i, \tag{18}$$

subject to:

$$\sum_i \pi^i \left(\tau_1 w_1^i \ell_1^i + \frac{\tau_2 w_2^i \ell_2^i}{1+r}\right) \geq G_1 + \frac{G_2}{1+r} + M, \tag{19}$$

$$0 \leq \delta \leq 1, \tag{20}$$

where M is an exogenous revenue requirement that is not refunded to agents. Note that δ does not enter (19) 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 that the government faces. When the private borrowing rate, r_b, exceeds r, we are implicitly assuming that there are financial frictions between workers and private lenders but not between workers and the government.[11] To simplify notation, we henceforth set $y_t^i = w_t^i \ell_t^i$, $t = 1, 2$.

Let λ denote the Lagrange multiplier associated with (19). The Lagrangian is:

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

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

Define $g_1^i = \frac{\alpha^i u_1'(\cdot)}{\lambda}$ as the social value of giving an additional dollar to an agent of type i in period 1 (in money metric terms) and $g_2^i = \beta(1+r)\frac{\alpha^i u_2'(\cdot)}{\lambda}$ as the social value (in money-metric terms) of giving an additional dollar to an agent of type i in period 2. Furthermore, let $\varepsilon_t^i = \frac{1-\tau_t}{y_t^i}\frac{dy_t^i}{d(1-\tau_t)}$ be the elasticity of period j income with respect to the net-of-tax rate $1-\tau_t$. We can then establish the following.

Proposition 1 (Benchmark Linear Tax Scheme) *Consider the solution of the optimization program (18), which imposes $\tau_1 = \tau_2 = \tau$ and $G_1 = G_2 = G$, while considering some fixed value of $1-\delta$.*

(i) The optimal marginal tax rate τ satisfies

$$\mathbb{E}_1\left[\delta g_1^i + (1-\delta) g_2^i\right] + \frac{1}{1+r}\mathbb{E}_2\left[g_2^i\right] = \mathbb{E}_1\left[1 + \frac{\tau}{1-\tau}\varepsilon_{1,1-\tau}^i\right] + \frac{1}{1+r}\mathbb{E}_2\left[1 + \frac{\tau}{1-\tau}\varepsilon_{2,1-\tau}^i\right], \tag{21}$$

where $\mathbb{E}_t[x] = \sum_i \pi^i y_t^i x$ is the period-t-income and population-weighted summation operator and $\varepsilon_{t,1-\tau}^i = \frac{1-\tau}{y_t^i}\frac{dy_t^i}{d(1-\tau)}$ is the elasticity of period-t labor earnings with respect to $1-\tau$.

[11] For example, the government may be better suited to collect on taxes owed than private lenders are on unsecured debt. This would naturally cause a gap between the actuarially fair interest rates offered on delayed taxes versus that on private loans. This is the case in, e.g., the U.S., where the government charges an interest rate equal to the federal short-term rate plus 3% on past-due taxes. This rate is substantially lower than typical interest rates on unsecured credit.

(ii) The optimal per-period transfer G satisfies

$$\sum_i \pi^i \Big(g_1^i + \frac{1}{1+r} g_2^i \Big) = \sum_i \pi^i \left(1 + \frac{1}{1+r} - \eta^i \right). \tag{22}$$

where $\eta^i = \frac{d}{dG}\left(\tau_1 y_1^i + \frac{\tau_2 y_2^i}{1+r}\right) \leq 0$ is an income effect parameter that provides the reduction in present-value taxes caused by an increase in G.

Proof. 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

$$-\sum_i \alpha^i \pi^i \Big(\delta u_1'(\cdot) y_1^i + \Big([1-\delta](1+r) y_1^i + y_2^i \Big) u_2'(\cdot)\beta \Big) + \lambda \sum_i \pi^i \left(y_1^i + \tau \frac{dy_1^i}{d\tau} + \frac{1}{1+r} \left[y_2^i + \tau \frac{dy_2^i}{d\tau} \right] \right) = 0. \tag{23}$$

Substituting in for g_t^i and rearranging yields

$$-\sum_i \pi^i \Big(\delta g_1^i y_1^i + \Big([1-\delta] y_1^i + \frac{1}{1+r} y_2^i \Big) g_2^i \Big) + \sum_i \pi^i \left(y_1^i + \tau \frac{dy_1^i}{d\tau} + \frac{1}{1+r} \left[y_2^i + \tau \frac{dy_2^i}{d\tau} \right] \right) = 0. \tag{24}$$

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

$$\sum_i \pi^i \Big(\delta g_1^i y_1^i + \Big([1-\delta] y_1^i + \frac{1}{1+r} y_2^i \Big) g_2^i \Big) = \sum_i \pi^i \left(y_1^i + \frac{1}{1+r} y_2^i - \frac{\tau}{1-\tau} y_1^i \varepsilon_{1,1-\tau}^i - \frac{1}{1+r} \frac{\tau}{1-\tau} y_2^i \varepsilon_{2,1-\tau}^i \right). \tag{25}$$

Equation (21) follows by re-arrangement and using the operator $\mathbb{E}_t[x] = \sum_i \pi^i y_t^i x,\ t = 1, 2$. Condition (22) follows by noting that the the FOC for G can be written as:

$$\sum_i \alpha_i \pi^i \Big(u'(c_1) + \beta u'(c_2) \Big) = \lambda \sum_i \pi^i \left(1 + \frac{1}{1+r} - \tau \frac{dy_1^i}{dG} - \tau \frac{1}{1+r} \frac{dy_2^i}{dG} \right). \tag{26}$$

■

We then turn to the age-dependent tax system, allowing $\tau_1 \neq \tau_2$ and $G_1 \neq G_2$, while considering some fixed extent of delayed taxation $1-\delta$. To facilitate our exposition, we introduce the cross-period elasticity of period-t earnings with respect to period-s taxes, $\varepsilon_{t,s}^i = \frac{1-\tau_s}{y_t^i} \frac{dy_t^i}{d(1-\tau_s)}$, which captures intertemporal labor substitution effects.

Proposition 2 (Optimal Age-Dependent Taxation) *Consider the solution to the problem described in equation 18, with some fixed value of $1-\delta$. The optimal age-dependent tax scheme is characterized by the following.*

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

$$\mathbb{E}_1\left[\delta g_1^i + [1-\delta] g_2^i\right] = \mathbb{E}_1\left[1 - \frac{\tau_1}{1-\tau_1}\left(\varepsilon_{1,1-\tau_1}^i + \frac{1}{1+r}\frac{\tau_2}{\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^i\right)\right], \tag{27}$$

$$\mathbb{E}_2\left[g_2^i\right] = \mathbb{E}_2\left[1 - \frac{\tau_2}{1-\tau_2}\left(\varepsilon_{2,1-\tau_2}^i + (1+r)\frac{\tau_1}{\tau_2}\frac{y_1^i}{y_2^i}\varepsilon_{1,1-\tau_2}^i\right)\right]. \tag{28}$$

(i) The optimal transfers G_1 and G_2 are set so that

$$\sum_i \pi^i g_1^i = \sum_i \pi^i \left[1 - \rho^i\right], \tag{29}$$

$$\sum_i \pi^i g_2^i = \sum_i \pi^i \left[1 - \rho^i \frac{1+r}{R'(s^i)}\right], \tag{30}$$

where $\rho^i = \frac{d}{dG_1}\left(\tau_1 y_1^i + \frac{\tau_2 y_2^i}{1+r}\right) \leq 0$ is an income effect parameter that provides the reduction in present-value taxes caused by increases in period-1 unearned income.

Proof. 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

$$-\sum_i \alpha^i \pi^i y_1^i \left(\delta u_1'(\cdot) + [1-\delta] u_2'(\cdot)\beta(1+r)\right) + \lambda \sum_i \pi^i \left[y_1^i + \tau_1 \frac{dy_1^i}{d\tau_1} + \frac{\tau_2}{1+r}\frac{dy_2^i}{d\tau_1}\right] = 0, \tag{31}$$

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:

$$-\sum_i \pi^i \alpha^i y_1^i \left(\delta u_1'(\cdot) + [1-\delta] u_2'(\cdot)\beta(1+r)\right) + \lambda \sum_i \pi^i y_1^i \left[1 - \frac{\tau_1}{1-\tau_1}\varepsilon_1^i - \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^i\right] = 0. \tag{32}$$

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

$$-\mathbb{E}_1\left[\delta g_1^i + [1-\delta] g_2^i\right] + \mathbb{E}_1\left[1 - \frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^i - \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^i\right] = 0. \tag{33}$$

Alternatively, we may write it as

$$\frac{\tau_1}{1-\tau_1} = \frac{\mathbb{E}_1\left[1 - \delta g_1^i + [1-\delta] g_2^i\right]}{\mathbb{E}_1\left[\varepsilon_1^i\right]} - \frac{\tau_2}{1-\tau_1}\frac{\mathbb{E}_1\left[\frac{y_2^i}{y_1^i}\varepsilon_{2,1}^i\right]}{\mathbb{E}_1\left[\varepsilon_1^i\right]} = 0. \tag{34}$$

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

$$-\sum_i \pi^i \alpha^i y_2^i \beta u_2'(\cdot) + \lambda \sum_i \pi^i \left[\tau_1 \frac{dy_1}{d\tau_2} + \frac{1}{1+r}\left(y_2^i + \tau_2 \frac{dy_2^i}{d\tau_2}\right)\right] = 0, \tag{35}$$

which we may rewrite as

$$-\frac{1}{1+r}\mathbb{E}_1\left[\frac{y_2^i}{y_1^i} g_2^i\right] + \mathbb{E}_1\left[-\frac{\tau_1}{1-\tau_2}\varepsilon^i_{1,1-\tau_2} + \frac{1}{1+r}\left(\frac{y_2^i}{y_1^i} - \frac{\tau_2}{1-\tau_2}\frac{y_2^i}{y_1^i}\varepsilon^i_{2,1-\tau_2}\right)\right] = 0. \tag{36}$$

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

$$\sum_i \pi^i \alpha^i u_1'(\cdot) = \lambda \sum_i \pi^i \left[1 - \tau_1 \frac{dy_1^i}{dG_1} - \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_1}\right], \tag{37}$$

$$\beta \sum_i \pi^i \alpha^i u_2'(\cdot) = \lambda \sum_i \pi^i \left[\frac{1}{1+r} - \tau_1 \frac{dy_1^i}{dG_2} - \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_2}\right]. \tag{38}$$

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

$$\beta \sum_i \pi^i \alpha^i u_2'(\cdot) = \lambda \sum_i \pi^i \left[\frac{1}{1+r} - \left(\tau_1 \frac{dy_1^i}{dG_1} + \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_1}\right)\frac{1}{R'(s^i)}\right]. \tag{39}$$

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. ■

Equations (27) and (27) are reminiscent of the optimal tax analysis presented by Atkinson and Stiglitz (1980) and others, linear (progressive) tax rates are determined as a trade-off between equity and efficiency considerations. One non-standard element of (27) is that the weighted average of the social weights in the numerator of the RHS reflects that the burden of a marginal increase in τ_1 is borne partly in period 1 and partly in period 2. 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 (13). Another non-standard element is that we consider optimal linear (progressive) age-dependent in a dynamic setting, and take into account intertemporal substitution of labor supply in response to the age-dependent tax system.

Equations (29) and (30) prescribe 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 in the direction of $G_1 > G_2$. 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.

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

$$\frac{d\ell_1^i}{d(1-\delta)} = \tau_1 \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \frac{d\ell_1^i}{d(1-\tau_1)}, \tag{40}$$

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

$$\frac{d\ell_2^i}{d(1-\delta)} = \tau_1 \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \frac{d\ell_2^i}{d(1-\tau_1)}. \tag{41}$$

Proof. See Appendix C. ■

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.

Proposition 3 (Optimal Delayed Taxation) *Consider the problem described in equation 18, 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:*

$$\tau_1 \cdot \mathbb{E}_1\left(g_1^i - g_2^i\right) = -\mathbb{E}_1\left[\tau_1\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i\right)\right], \tag{42}$$

where the LHS is the welfare effect due to enhanced consumption smoothing, and the RHS captures the fiscal externalities of marginally delaying taxation, where $\rho^i = \frac{d}{dG_1}\left(\tau_1 y_1^i + \frac{\tau_2 y_2^i}{1+r}\right)$.

Proof. 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.

$$\sum_i \pi^i \alpha_i \left(u'(c_1)\tau_1 y_1 - \beta u'(c_2)(1+r)\tau_1 y_1\right) + \lambda \sum_i \pi^i \left(\tau_1 \frac{dy_1^i}{d(1-\delta)} + \tau_2 \frac{1}{1+r}\frac{dy_2^i}{d(1-\delta)}\right) = 0. \tag{43}$$

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

$$\left(\tau_1^2\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \frac{dy_1^i}{d(1-\tau_1)} + \tau_2\tau_1\frac{1}{1+r}\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \frac{dy_2^i}{d(1-\tau_1)}\right). \tag{44}$$

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

$$\left(\tau_1^2 \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \left\{\left(\frac{dy_1^i}{d(1-\tau_1)}\right)^c + y_1 \frac{dy_1}{dG_1}\right\} + \tau_2\tau_1 \frac{1}{1+r}\left[1 - \frac{1+r}{R'(s^i)}\right] \frac{1}{\delta^i} \left\{\left(\frac{dy_2^i}{d(1-\tau_1)}\right)^c + y_1 \frac{dy_2}{dG_1}\right\}\right). \tag{45}$$

Further rearranging and using elasticity notation yields

$$y_1^i \frac{\tau_1}{1-\tau_1}\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right] \left(\left\{\tau_1 \varepsilon_{1,1-\tau_1}^{i,c} + (1-\tau_1)\tau_1 \frac{dy_1}{dG_1}\right\} + \frac{1}{1+r}\left\{\tau_2 \frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + (1-\tau_1)\tau_2 \frac{dy_2}{dG_1}\right\}\right). \tag{46}$$

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

$$\tau_1 \sum_i \pi^i y_1^i \left(\frac{\alpha_i}{\lambda}\ \Delta_r^i \cdot u'(c_1^i) \ + \left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i\right)\right) = 0. \tag{47}$$

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

$$\tau_1 \mathbb{E}_1\left(g_1^i - g_2^i\right) = -\mathbb{E}_1\left[\tau_1\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i\right)\right]. \tag{48}$$

Using Lemma 1 yields (42) ■

If there were no financial constraints, both the left-hand and right-hand sides of (42) 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 an unconstrained agent's perspective equals 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 (42) is positive and equal to the marginal welfare gains from enhanced consumption smoothing. The right-hand-side is also non-zero due to fiscal externalities caused by constrained agents who now face a lower effective period-1 tax rate.

2.3 Welfare effects of marginal reforms

An interesting question is under what conditions the introduction of delayed taxation might increase welfare. Starting from our benchmark economy characterized by a tax regime with with $\tau_1 = \tau_2 = \tau$ and no delayed taxation $1 - \delta = 0$, we ask whether marginally delaying taxation (i.e., marginally increasing $1 - \delta$) increases tax revenue. Note that this reform has no mechanical cost to the government.

To illustrate the economic forces at play, we consider the welfare effects of marginally delaying taxation, $\frac{dW}{d(1-\delta)}\big|_{1-\delta=0}$. This can be summarized as follows based on equation (47):

$$\tau \sum_i \pi^i y_1^i \Delta_r^i \Bigg[\underbrace{\frac{\alpha_i}{\lambda} \cdot u'(c_1^i)}_{\substack{\text{Welfare gains}\\ \text{from intertemporal}\\ \text{consumption smoothing}}} + \underbrace{\frac{\tau}{1-\tau}\varepsilon_{1,1-\tau_1}^{i,c}}_{\substack{\text{Increase in period-1}\\ \text{tax revenues}\\ \text{(substitution)}}} + \underbrace{\frac{1}{1+r}\frac{\tau}{1-\tau}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c}}_{\substack{\text{Decrease in period-2}\\ \text{tax revenues}\\ \text{(intertemp. substitution)}}} + \underbrace{\rho^i}_{\substack{\text{Decrease}\\ \text{in PV taxes}\\ \text{(income effect)}}} \Bigg], \tag{49}$$

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

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \tau\Delta_r \left(\sum_{i:s^i<0} \pi^i y_1^i (g_1^i + \rho^i) + \sum_{i:s^i<0} \pi^i \mathcal{X}^i \right). \tag{50}$$

(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

$$\frac{1}{\lambda}\frac{dW}{dx}\Big|_{\delta=1} = \sum_i \pi^i \left(g_1^i - g_2^i\right) + \sum_i \pi^i \Delta_r^i \rho^i = \Delta_r \sum_{i:s^i<0} \pi^i (g_1^i + \rho^i) \tag{51}$$

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

$$\frac{1}{\lambda}\frac{dW}{d(1-\tau_1)}\Big|_{\delta=1} = \sum_i \pi^i y_1^i (g_1^i + \rho^i) + \sum_i \pi^i \mathcal{X}^i - \sum_i \pi^i y_1^i. \tag{52}$$

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 D ∎

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 without delayed taxation ($\delta = 1$) and age-independent taxation. Then, the marginal welfare effect of delayed taxation can be written* *in terms of either a uniform loan or an age-dependent tax change as follows:*

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \underbrace{\tau\bar{y}_1\frac{1}{\lambda}\frac{dW}{dx}}_{Loan} + \tau_1\Delta_r\Big[\underbrace{\sum_{i:s^i<0}\pi^i(y_1^i-\bar{y}_1)(g_1^i+\rho^i)}_{=Cov(y_1,g_1+\rho|s<0)} + \sum_{i:s^i<0}\pi^i\mathcal{X}^i\Big] \tag{53}$$

$$= \tau\Delta_r\Big[\underbrace{\frac{1}{\lambda}\frac{dW}{d(1-\tau_1)}}_{AD\ tax\ change} - \sum_{i:s^i>0}\pi^i y_1^i(g_1^i+\rho^i) - \sum_{i:s^i>0}\pi^i\mathcal{X}^i + \sum_i \pi^i y_1^i\Big], \tag{54}$$

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. Multiplying (51) in Lemma 2 by $\bar{y}_1 = \sum_{i:s^i<0}\pi^i y_1^i$ yields:

$$\frac{\bar{y}_1}{\Delta_r}\frac{dW}{\lambda dx} = \bar{y}_1\sum_{i:s^i<0}\pi^i(g_1^i+\rho^i). \tag{55}$$

Taking the difference between (50) in Lemma 2 and (55) yields:

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

which can be written

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

This establishes (53). To establish (54), we take the difference between (50) and (52) in Lemma 2 to obtain:

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

Re-arranging we get:

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

■

Proposition 4 highlights that the marginal introduction of delayed taxation affects only borrowers and shows that the overall 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 effects, income effects, and substitution effects, captured by the terms in brackets on the RHS of equation (53). 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 a partially offsetting negative substitution effects in period 2 (as part of the increase in period-1 labor supply is just a substitution away from period-2 labor supply). 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 (54), 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.[12]

Corollary 1 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*

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \tau_1 \Delta_r \sum_i \pi^i y_1^i, \tag{56}$$

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

Proof. This follows from Proposition 4 because, if everyone borrows, we have:

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

where the last equality follows because $\frac{1}{\lambda}\frac{dW}{d(1-\tau_1)} = 0$ when the income tax is optimally age-dependent. ■

The intuition for (56) is that if you marginally delay taxation, you can afford a marginally higher tax rate in period 1 while leaving the effective distortion (wedge) in period 1 unaffected (note that since marginal taxes are optimally age-dependent in the pre-reform situation, there is no welfare gain from changing the period 1 wedge by introducing a small amount of delayed taxation).

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

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

When the interest rate on delayed taxes, r_{dtax}, is constrained to equal the interest rate on net savings, r, delayed taxation has no effect on the behavior of workers whose marginal borrowing rate equals r. Delayed taxation does not alter the present-value 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 an effect in the absence of financial frictions (when $r_b = r$).

As an extension to 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 further allow for the possibility that there are no financial frictions, i.e., $r_b = r$. A trivial result when the government can choose any $r_{dtax} \in \mathbb{R}$ is that it may replicate any age-dependent marginal tax scheme characterized by $(\tau_1, \tau_2) \in \mathbb{R}^2_+$ by choosing $\tau = \tau_2$, setting $r_{dtax} = -1$, and choosing $\delta = \tau_1/\tau_2$.[13] Such a policy, however, does not exploit the fact the heterogeneity in marginal borrowing rates, $R'(s^i)$.

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

$$\frac{1}{1+r} \sum_{i:s^i<0} \left((1 + r_{dtax})(\ell_1^{i,*} - \ell_1^i) w_1^i \tau_1 + (\ell_2^{i,*} - \ell_2^i) w_2^i \tau_2 \right), \tag{57}$$

where ℓ_t^i is labor supply under the AD scheme and $\ell_t^{i,*}$ is 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 $s^i < 0$ only differs from ℓ_t^i due to a lower effective period-1 tax rate among borrowers,

$$\tilde{\tau}_1^* = \tau_2 \left[1 - (1-\delta) \left(1 - \frac{1 + r_{dtax}}{1+r} \right) \right] \leq \tau_1 \quad \textit{for } i \textit{ s.t. } s^i < 0. \tag{58}$$

Proof. 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

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

workers with $R'(s^i) = r$ equals τ_1 from the AD scheme.

$$\tau_1^{i,*} = \tau_2 \left[1 - (1-\delta)\left(1 - \frac{1 + r_{dtax}}{1+r}\right)\right] = \tau_1 \quad \text{for all } i \text{ s.t. } R'(s^i) = r \tag{59}$$

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, (59) 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 (59), we that this is equivalent to

$$\frac{\tau_1}{\tau_2} \geq \frac{1 + \underline{r_{dtax}}}{1+r}. \tag{60}$$

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

$$\tilde{\tau}_1^{i,*} = \tau_2 \left[1 - (1-\delta)\left(1 - \frac{1 + r_{dtax}}{1+r}\right)\right] \quad \text{for all } i \text{ s.t. } R'(s^i) = 1 + r_b. \tag{61}$$

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

$$\sum_{s^i<0} \left(\ell_1^{i,*} w_1^i \tau_2 \left[1 - (1-\delta)\left(1 - \frac{1 + r_{dtax}}{1+r}\right)\right] + \frac{\ell_2^{i,*} w_2^i \tau_2}{1+r}\right) - \sum_{s^i<0} \left(\ell_1^i w_1^i \tau_1 + \frac{\ell_2^i w_2^i \tau_2}{1+r}\right). \tag{62}$$

By virtue of how r_{dtax} is set (equation 59), this revenue change may be rewritten as

$$\sum_{i:s^i<0} \ell_1^{i,*} w_1^i \tau_1 + \frac{1}{1+r} \sum_{i:s^i<0} \ell_2^{i,*} w_2^i \tau_2 - \left(\sum_{i:s^i<0} \ell_1^i w_1^i \tau_1 + \frac{1}{1+r} \sum_{i:s^i<0} \ell_2^i w_2^i \tau_2\right). \tag{63}$$

or

$$\sum_{i:s^i<0} \left((\ell_1^{i,*} - \ell_1^i) w_1^i \tau_1 + \frac{1}{1+r} (\ell_2^{i,*} - \ell_2^i) w_2^i \tau_2\right), \tag{64}$$

which is equivalent to (57). 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 61). 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). ■

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-dominating delayed tax policy does not exist if the behavioral response to a decrease in the effective tax rate is sufficiently negative to cause tax revenues to decrease. However, we do not consider the assumption of a nonnegative revenue effect (57) to be particularly strong. That is because the relevant revenue effect only contains behavioral responses to a tax reduction— and not the mechanical negative effects typically caused by of lowering tax rates. If, for example, the existing tax system's age-dependent optimal tax rates coincides 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 decrease.

3 A quantitative investigation of delayed taxation

3.1 Calibration

We assume the utility function takes the form:

$$u(c) - v(\ell) = \frac{c^{1-\sigma} - 1}{1-\sigma} - \xi \frac{\ell^{1+\frac{1}{k}}}{1+\frac{1}{k}}, \tag{65}$$

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 work. 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, $\alpha^i = 1$.

We calibrate our model to Norwegian workers between the age 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 share of each of these types. Our calibration assumes perfect foresight in order to obtain variation in expected inflation-adjusted wage trajectories. We set the exogenous revenue requirement, M equal to 15% of "GDP', which is the sum of labor earnings.

In our calibration, there is no exogenous income or endowments. Accordingly, all of the variation in saving incentives (and thus the degrees 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 baseline interest rate (faced by the government) to 3%. Since we are modeling periods that are 21 years apart, cumulative interest rates enter the budget constraints. That is, the present value in period 1 of $ in period 2 is 1.03^{-21}. We accordingly set $\beta = 1.03^{-21}$ such that a net-saving agent facing an interest rate of 3% would choose an equal amount of consumption in the two periods.

Policies. Under 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, that is, we allow $\tau_1 \neq \tau_2$, but we require $\tau_t \geq 0$. Letting $\tau_1 \neq \tau_2$ is not an option under either the benchmark policy or delayed taxation. When the government can do delayed taxation, we restrict the share of period-1 taxes payable in period 1 to be inside $[0, 1]$. That is, you may not borrow from the government in excess of the amount of taxes you accrue, which typically binds when financial frictions are severe. Imposing $\delta \leq 1$, however, 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 $R'(s^i) \leq 1 + r_{dtax}$.

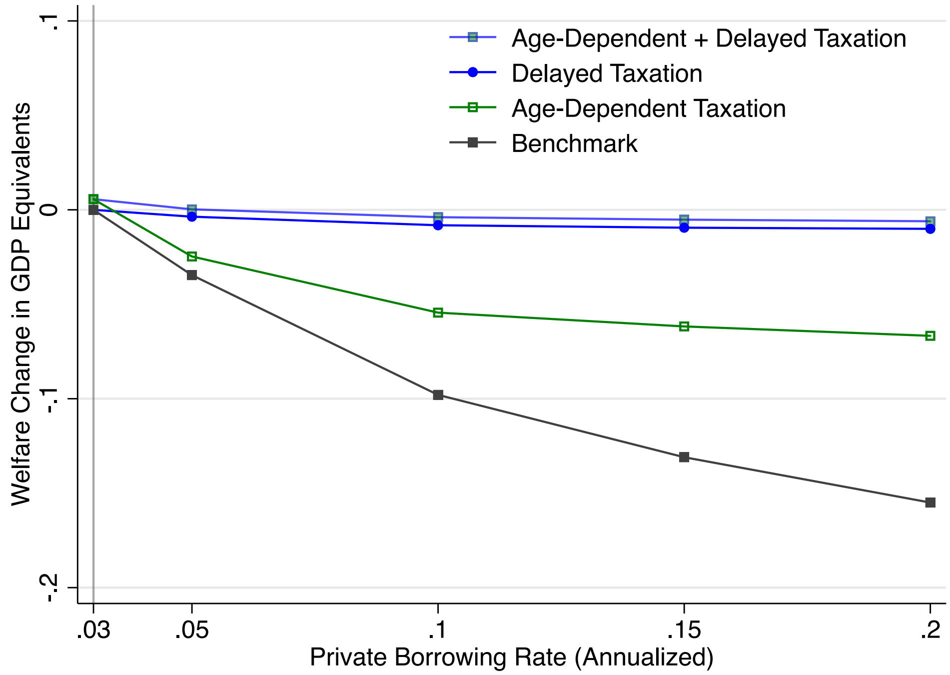
3.2 Main numerical results

Figure 1 summarizes our main findings. 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 governmente (3%) to 10% reduces welfare by as much as an exogenous negative GDP shock of 15%. We find that allowing for either delayed taxation or age-dependent taxation substantially weakens this adverse effect on welfare. With age-dependent taxation, the welfare loss is only 10%, and with delayed taxation the loss is about 2%. Hence, in a money-metric sense, AD removes about 67% of the welfare cost of financial frictions and DT removes 87%. We interpret this as there being substantial welfare gains from age-dependent taxation–but even more so by delayed taxation.

Beyond 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. By contrasting these effects to the stand-alone effects of either AD or DT, we see that age-dependent taxation provides almost no welfare gains on top to a delayed tax policy, while delayed taxation adds considerable welfare 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 increasing the private borrowing rate above the government rate ($r_{gov} = 0.03$). We consider the life-time welfare of agents as of period 1. The black squares shows the effect of increasing the borrowing rate in the benchmark economy in which marginal tax rates and transfers are equal across periods. The blue circles provide the effect when only the delayed taxation policy is implemented: that is, transfers and marginal tax rates are equal across periods but the government optimally chooses a fraction of taxes incurred in period 1 to be paid in period 2, that is, $1-\delta$. The green hollow squares provide the effect when we instead do age-dependent taxation: both marginal tax rates and transfers may differ across periods. The blue-green squares provide the effect when both delayed taxation and age-dependent taxation are allowed.



We provide more detail in Table 1. Panel A shows summary statistics for the wage trajectories we use for calibration. We see that median real cumulative wage growth is 0.88. Annualizing this over a 21-year period gives an 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 provides the optimal tax policies and allocations for the case when the private borrowing rate, r_b equals 10%. We see that allowing age-dependent taxation (differing τ_1 and τ_2) leads to a corner solution in which $\tau_1 = 0$. The intuition for this is that the workers utility is very sensitive to period-1 disposable income, and hence 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 fairly 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 modestly lower for delayed taxation than in the benchmark economy, due to strong income effects. However, we see that present-value taxes are about 46% higher with delayed taxation than 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 discount rate. The (money-metric) welfare differential is the exogenous shock to revenues 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 or 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

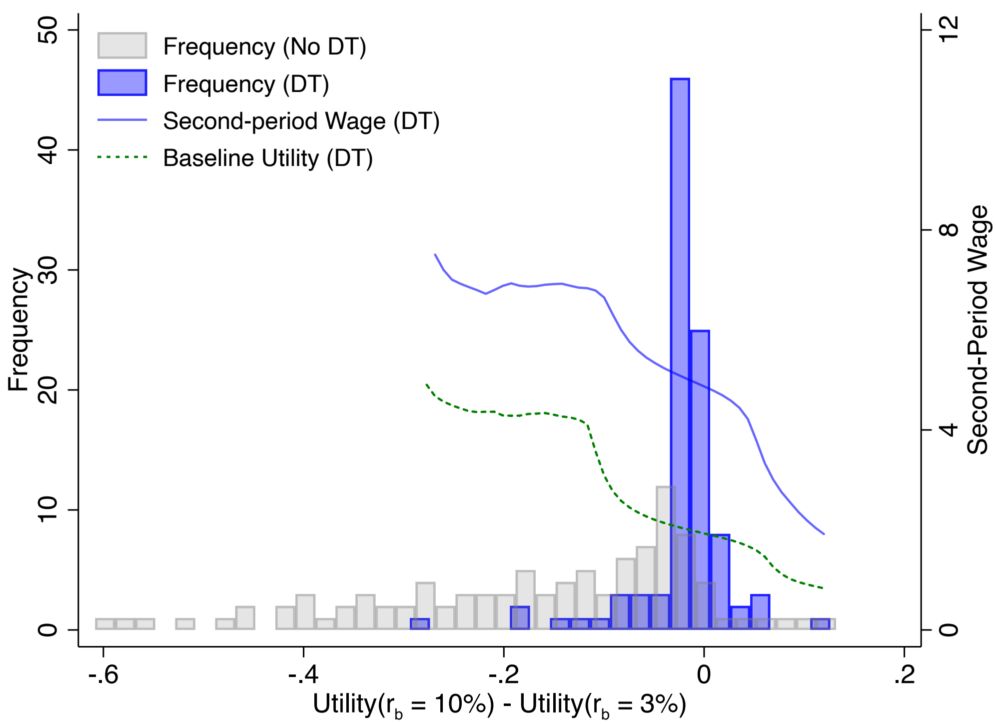
In the previous section, we see, not surprisingly, that increasing financial frictions by increasing the private borrowing rate decreases welfare. This holds true regardless of whether the government engages in age-dependent or delayed taxation. However, as we will show, in the presence of delayed taxation, there are in fact many agents who become *better off* once marginal borrowing rates increase.

Figure 2 plots the distribution of changes in life-time utility caused by increasing the marginal borrowing rate from 3% to 10%. While most agents see reduced utility, many agents in fact are better off. To provide some intuition for this finding, we also overlay a plot of the mean period-2 wage. This shows that the agents who are better off are the agents with low period-2 earnings. These households are better off because they were net savers in the first place. Given their low future earnings, there is little future earnings against which they want to borrow. However, they benefit from the fact that higher borrowing rates implies that delayed taxation decreases the distortionary effects of income taxation more, which causes higher tax revenues and thus larger transfers.

By overlaying the baseline utility as well, we see that those who benefit from higher borrowing rates are those with the lowest utility in the baseline (no financial friction) case. Hence, while increasing financial frictions does reduce overall welfare, it increases it for those who were worse off in the first place.

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

This figure plots the distribution of life-time utility changes from going from a 3% borrowing rate to a 10% borrowing rate. The gray bars in the background provide the distribution in absence of delayed taxation (DT). The blue bars in the foreground provide the distribution with delayed taxation. On top of it, we plot a local polynomial fit of the second-period wages (right-hand-side y axis). We also add (solid blue line) the baseline utility, transformed as $10\exp(u)$ in order 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 enhanced consumption smoothing, which is facilitated by government borrowing. In our model, the government may borrow an unlimited amount at r_{gov} and thus is not concerned with the amount it must borrow to finance the different 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

$$B \leq (1+b)B^*(r_b), \tag{66}$$

where $B^*(r_b)$ is the amount of government borrowing in the benchmark economy with a borrowing rate of r_b. We consider modest 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 provide our main results for the case when $b = 0.5$ 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 $1 - \delta$ has to be financed by either higher period-1 tax rates (more distortions) or lower transfers, G (less redistribution). Similar tradeoffs affect 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 benchmark economy's GDP. Interestingly, this welfare effect is also larger than for *AD T&T*, which is a more flexible age-dependent tax scheme that also allows the government to chose age-dependent lump-sum transfers on top of age-dependent (marginal) tax rates. This is more surprising given that delayed taxation, where 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 when $r_b = 10\%$. We see that delayed taxation offers larger welfare gains than all the other reform options whenever $b < \infty$. Interestingly, all policies deliver non-neglible welfare gains of about 0.5% of GDP even when the government cannot increase borrowing at all.

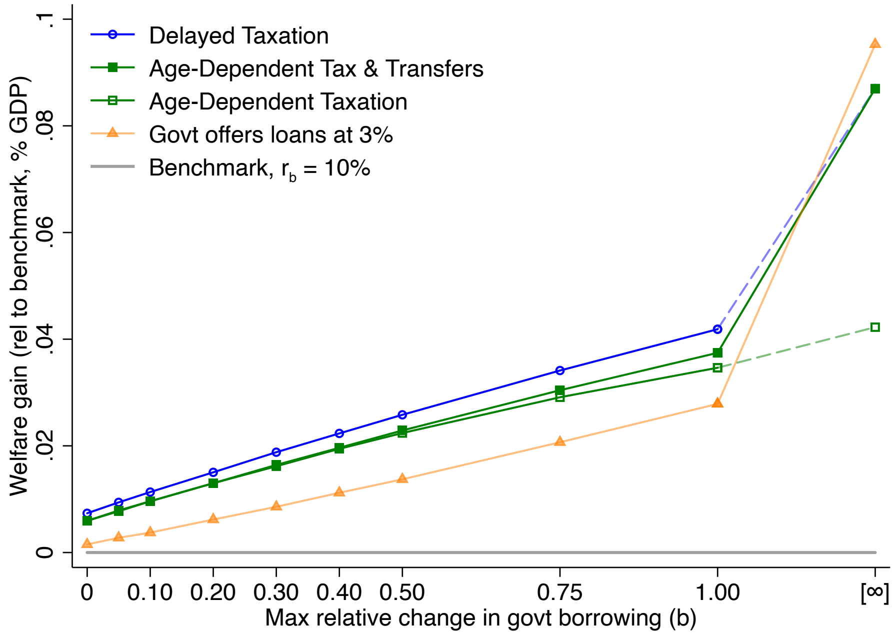
TABLE 2: OPTIMAL TAXATION WITH FINANCIAL FRICTIONS WHEN GOVERNMENT FACES BORROWING CONSTRAINTS

This table provides summary statistics for the calibrated economy when $r_b = 10\%$ and the government borrowing limit, b, equals 50%. The present-value function calculates present values according to the government discount rate. For a given policy reform, the money-metric welfare measure is the exogenous change in government tax revenues that the Benchmark economy government would need to be equally well off as with the reform. The number is expressed as a percent 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 money-metric welfare effects (measured in terms of the benchmark economy's GDP) of implementing delayed taxation or age-dependent taxation. We do this for different values for b, which is defined as the maximal relative increase in borrowing relative to the benchmark economy. For example, when $b = 0$, the government may implement delayed taxation but may itself not borrow more than it did before implementing delayed taxation.

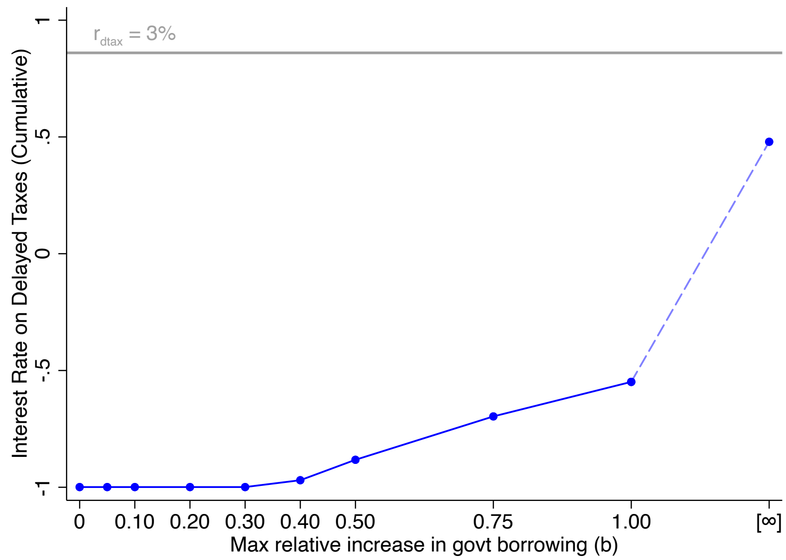


3.5 Unmodeled aspects of the different policies

Our analyses highlight the potential benefit of delayed taxation and shows that it in most cases produces more welfare than age-dependent taxation. Our modeling exercise does not account for the possibility that borrowers may default on their debt. One concern could be that agents who acquire a considerable amount of delayed taxes choose to evade the tax payment by moving abroad and refusing to pay tax liabilities. This issue also applies to delayed taxation: those who benefit from young tax rates and while young may choose to relocate to a tax jurisdiction with age-independent tax rates when old. In our framework, the possibility of default might *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 considering "DT+", which is a more flexible type of delayed taxation in which the government is free to chose 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 equates to ex-ante loan forgiveness.

Figure 4: Optimal interest rate on delayed taxes

This figure plots the optimally chosen $r_{dtax} \geq -1$ in the flexible delayed-tax version. We fix $r_b = 10\%$ and vary the govt borrowing constraint parameter, b. The y-axis provides 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 the tax payment but not tax accrual is (substantially) delayed. As we we show theoretically, the benefits of delayed taxation arise when this wedge is large enough for marginal borrowing rates to materially decrease between the time of tax accrual and payment. While exploiting this time wedge as a policy tool is novel to our paper, there exists various tax schemes today that also introduce such a wedge. Below we discuss some of these tax schemes and contrast them to our notion of delayed taxation.

Non-withheld taxes. In the absence of employer income-tax withholding, most workers would face modest delayed taxation. If taxes are due in, e.g., April, the following tax year, then taxes on earnings in, e.g., January is 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 not withheld, implying that high-income Swedish workers face modest delayed taxation on marginal earnings.

Installment plans. While the option to enter into installment plans to delay the payment tax liabilities exists in many countries, these tax deferral schemes are typically not set up in a way that resembles our notion of delayed taxation. In the U.S. for example, employers must withhold federal income taxes, implying that taxes are paid immediately. The IRS does offer options to enter into (up to) 10-year installment plans for taxpayers facing adverse financial circumstances, but only on the balance due, that is, any taxes owed net of what is already withheld. Hence, unless a taxpayer expects tax liabilities to substantially exceed withholding amounts, the option to enter into an installment plan will not mute behavioral responses to labor income taxes.

Social Security. Mandating that employees save a portion of their income towards retirement effectively amounts to negatively-delayed taxation. If we assume that social security contributions equal about 8% of pre-tax income and taxes equal about 24%, then workers are effectively paying 133% of their taxes today and receiving 33% back (with interest) in retirement.

The notion that social security contributions may 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 oversaving. In discussing the paper, Larry Summers notes that "reversing the direction of transfers" under social security seems would be a natural way to reduce welfare losses from financial frictions. However, the key point that negative social security contributions (i.e., delayed taxation) may reduce also the welfare losses from the distortionary effects of income taxation is missing.[14]

Capital gains. In most tax jurisdictions, capital gains are only taxed upon realization. This effectively allows taxpayers to delay their taxes indefinitely, especially when there is step-up in basis at death, such 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, hard to consume, delaying capital gains 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 support depends on income levels. In the U.S., for example, the generosity of financial aid generally depends on parental income. Hence, if parents work and earn more, then the financial aid package may to a larger degree consist of student loans. If we consider the family as a single economic unit, this is essentially delayed taxation: higher labor earnings cause the accrual of a (de-facto) tax that is payable in the future. In Norway, the mix of government provided financial support package does not depend on parental income but rather, mechanically, on the students' own labor earnings while in school. We expand on this in the next section.

Retirement savings accounts. Traditional IRAs allow you to save pre-tax income, and any withdrawals are subsequently taxable as income. This affects the effective (present-value) tax rate that workers face in two ways. First, the nominal tax rate is change to the nominal tax rate during retirement. This is rate is likely lower than the concurrent rate due to tax progressivity and the fact that most households have lower incomes during retirement. Second, it reduces the effective tax rate because the IRA savings cumulate at the pre-tax rate of return while other savings grow at the post-capital-income-tax rate of return. Hence, IRAs have a clear element of delayed taxation in that effective period-1 tax rates are lowered. However, there are two important differences relative to delayed taxation. (i) In practice, IRA contributions are subject to fairly low limits (e.g., $6,500 in 2023), implying that for workers who'd optimally contribute more than the limit, marginal labor supply is not increased—instead, there's a negative income effect on labor supply. (ii) In addition, while the positive present-value effect of saving at the after-tax rate

[14] It is also useful to note that the presence of an income limit ("maximum taxable earnings") for social security contributions limits the labor supply distortions on high income earner. Social security contributions do not affect these earners' marginal effective tax rates, instead it only works to exacerbate financial frictions which may *increase* labor supply through an income effect.

mirrors the present-value effect that constrained borrowers see from delaying taxes at a rate lower than their borrowing rate, this is driven by a tax policy. The government is strictly worse off in pure revenue sense by allowing deferred IRA taxation since they are foregoing capital income taxes. With delayed taxation, on the other hand, the idea is that (net-borrowing) households and governments may face different interest rates due to financial frictions.

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.

We focus our empirical study 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. The monthly transfers ranged from NOK 8000 ($1,300) in 2004 to NOK 9785 in 2011. These transfers are initially given as a loan, but 40% may be forgiven (converted to a stipend) to the extent that students pass classes and stay below the earnings thresholds above. Students are notified of the amount of transfers in the beginning of the academic year. These notification letters contain a breakdown of the transfers, noting the amount (40% of the total) that is given as a conversion loan, and stating that conversion from loan to stipend is contingent on incomes being below an income limit. The following year, students are notified how much of their loan was converted based on grades reported by educational institutions and income reported by the tax authorities. Loans must typically be paid off within 20 years following graduation. No interest is charged while still receiving support. Subsequently, interest rates are slightly above the risk-free rate and loan payments may be delayed at the (former) student's discretion for up to 3 years in total.[15]

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

Summary statistics are provided in Table 3. The average student is 23 years old. This is reasonable in light of high school graduation occurring at age 18 and that we condition on students being enrolled for higher education for both semesters within a given year. The summary statistics

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

reveal a substantial spread in the amount of liquid assets available to students. While students at the 25th percentile only hold NOK 8,000 ($1,300) in liquid assets, students at the 75th percentile hold almost *ten* times more. A similar spread can be observed in the liquid assets of the students' parents. We further see that the average student earns around NOK 100,000 ($17,000), which is a direct consequence of our sample restrictions caused by focusing on students around the debt-conversion threshold. Four years later, the average student faces considerably higher earnings at around NOK 360,000 ($60,000).

Table 3: Summary Statistics

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

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

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

5.2 Bunching methodology

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

$$e = \frac{\Delta y^*/y^*}{\Delta \tau/(1-\tau)}, \tag{67}$$

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

$$B = \int_{y^*}^{y^*+\Delta y^*} h_0(y)dy \approx h_0(y^*)\Delta y^*. \tag{68}$$

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

$$\frac{\Delta y^*}{y^*} = \frac{B}{h_0(y^*)y^*} = \frac{b}{y^*}. \tag{69}$$

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 analogue of y^* is the (average) debt-conversion threshold denominated in the same units (thousands) as the empirical earnings bins.[16] We write our estimated compensated labor earnings elasticity as

$$\hat{e} = \frac{\hat{b}/y^*}{\widehat{\Delta\tau}/(1-\tau)}, \tag{70}$$

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 with no 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 65), 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 (14) can be written as:

$$u'(c_1^i) \cdot w_1^i(1-\tau_1(1-\delta)\Delta_r^i) = v'(\ell_1^i). \tag{71}$$

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

$$\varepsilon_{\ell_1,1-\tau_1}^{i,F} = \left(1-(1-\delta)\Delta_r^i\right) \frac{v'(\ell_1^i)}{\ell_1^i v''(\ell_1^i)} = \left(1-(1-\delta)\Delta_r^i\right) \kappa, \tag{72}$$

[16] 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.

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,

$$\varepsilon_{\ell_1, 1-\tau_1}^{i,F} = \left(1 - \Delta_r^i\right) \kappa. \tag{73}$$

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

$$E[\hat{e}] = E[\zeta \varepsilon_{\ell_1, 1-\tau_1}^{i,F}] = \underbrace{E\left[1 - \Delta_r^i\right]}_{\text{Delayed taxation effect}} \cdot \underbrace{\zeta}_{\text{bias}} \underbrace{\cdot \kappa.}_{\text{structural Frisch}} \tag{74}$$

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 next period debt. Most students are on the expected kinked trajectory where each additional NOK of earnings increases debt by 0.50 NOK. The blue fitted line illustrates how we obtain our first-stage measure of the effect of excess earnings on debt accumulation. We find that the slope increases by 0.47. This is close to the nominal increase of 0.50 due to very few non-compliers.[17] Cast in terms of the previous notation, this implies that $\widehat{\Delta\tau} = 0.47$.

In Panel (B), the yellow dotted line shows the distribution of students around the earnings threshold. The green line is the counter-factual distribution, which is a 5th-order polynomial fitted to the non-bunching region. We obtain a measure of the excess mass of individuals near the threshold by comparing the actual and counter-factual distributions. This offers a bunching estimate, b, of 1.21, which means that there are 121% more individuals around the threshold than the counter-factual distribution implies. Dividing 1.21 by the average threshold amount (120.162 in NOK 1,000s), per equation 70, we obtain a remarkably low elasticity of labor earnings to the net-of-tax (or net-of-debt-increase) wage of 0.0162.[18] The standard error is 0.0013.[19]

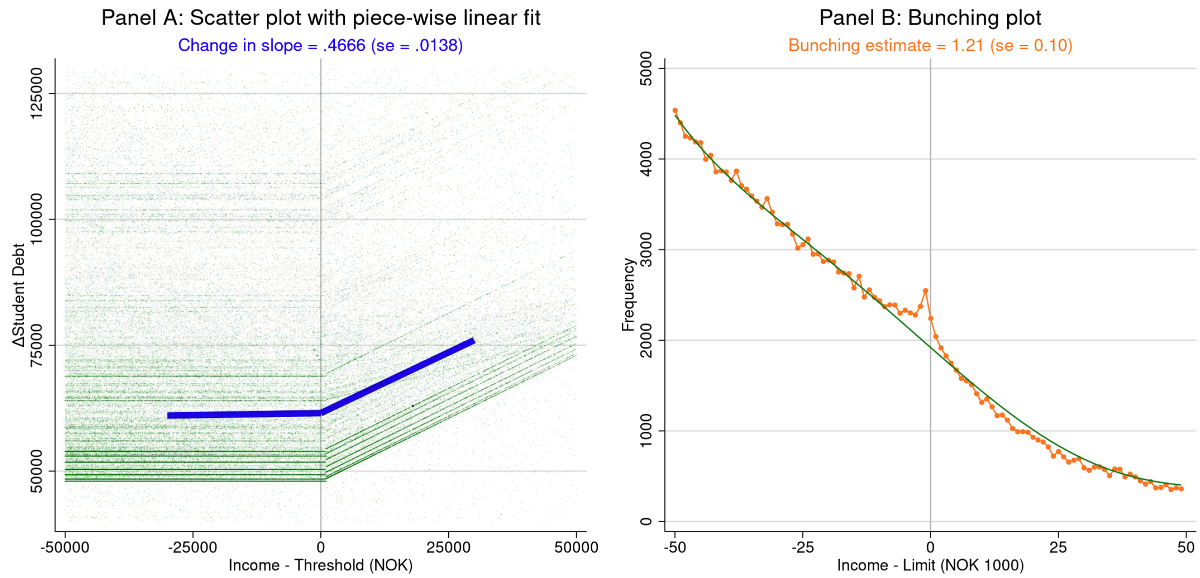
[17] Some non-compliers exist, for example, because they may have moved in with their parents during the fall semester, which would exclude them from receiving any conversion for fall semester loans. Such moves must be reported to the educational loan fund, but not to the tax authorities from which we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

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

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

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

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



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

5.4 Determinants of non-bunching

We now investigate potential determinants of this (non-)bunching behavior. Our main approach is to plot student characteristics against their position relative to the conversion threshold.[20] This is a visual exercise where we attempt to draw conclusions from visual breaks in the relationship between a given characteristic and students' earnings occurring around the conversion threshold.

In Figure 6, Panel (A), we find that the amount of ex-ante liquid assets drops sharply right above the threshold. This suggests that non-bunchers have less liquid wealth, consistent with these

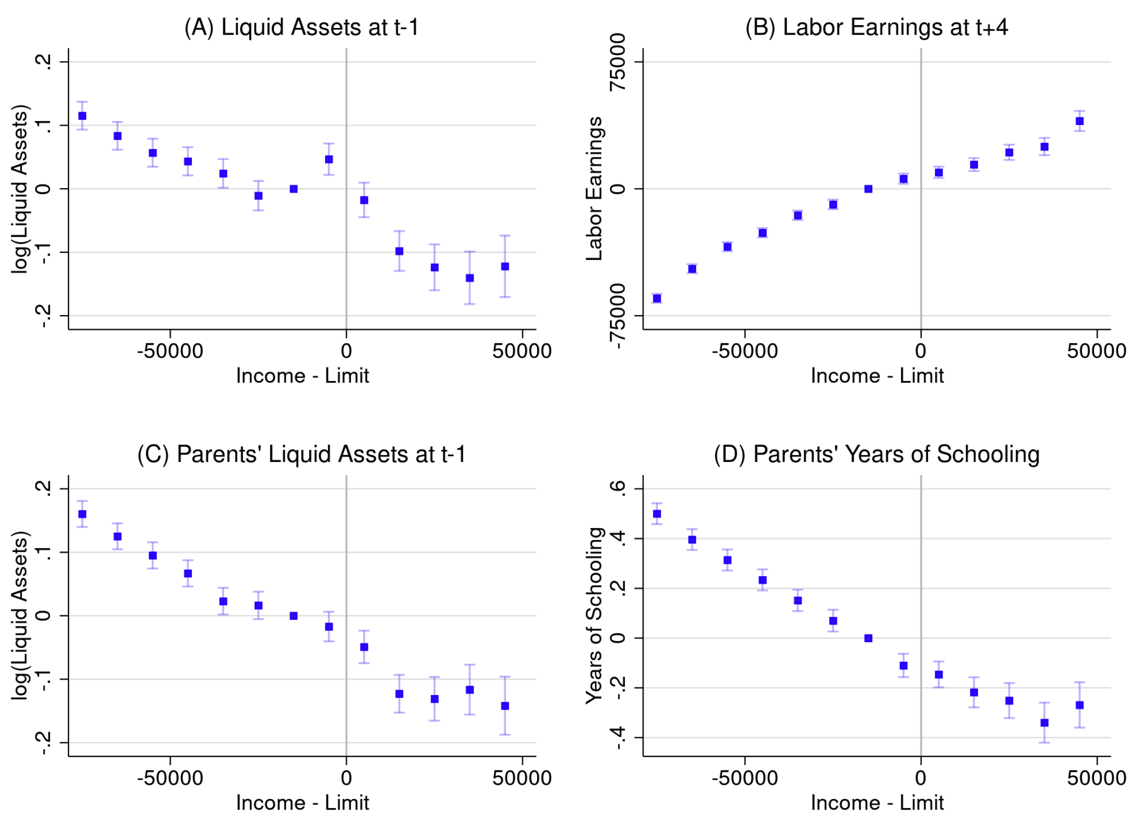
[20] Another application of this type of analysis can be found in concurrent work by Bastani and Waldenström (2021) who examine how ability covaries with taxpayers' position relative to a regular tax threshold to infer the ability gradient in tax responsiveness.

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 reveals no sharp rise or decrease in realized future incomes above the threshold, which suggests that non-bunchers do not differ significantly in terms of medium-term earnings prospects.

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

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

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



To investigate this liquidity channel further, we also show how *parents'* liquidity correlates with the students' earnings location in Panel (C) of Figure 6. This documents a noteworthy negative relationship between the parents' financial resources and the in-school labor earnings of the child. This suggests that parents play an important role in determining the amount of time students may dedicate to their studies. More relevant to the present study, is the finding that parents' assets drop shortly above the earnings threshold. This indicates that non-bunchers have access to fewer financial resources, which is consistent with financial frictions playing a key role in driving the observed non-responsiveness to the conversion threshold. However, wealth may proxy for human capital which influences tax responsiveness (Bastani and Waldenström, 2021). Therefore, we plot parental educational attainment on the y-axis in Panel (D). This shows that there is no break in the relationship between educational attainment, measured in the maximum years of schooling among the parents and the child's position relative to the conversion threshold. This addresses the hypothesis that less resources, in a human capital, rather than financial, sense can explain the irresponsiveness to the threshold. If anything, extrapolating from the below-threshold relationship, non-bunchers may have higher-educated parents. To the extent that this is correlated with the students' life-time wealth, this may explain some of the desire of students to front-load consumption through incurring higher student loans.

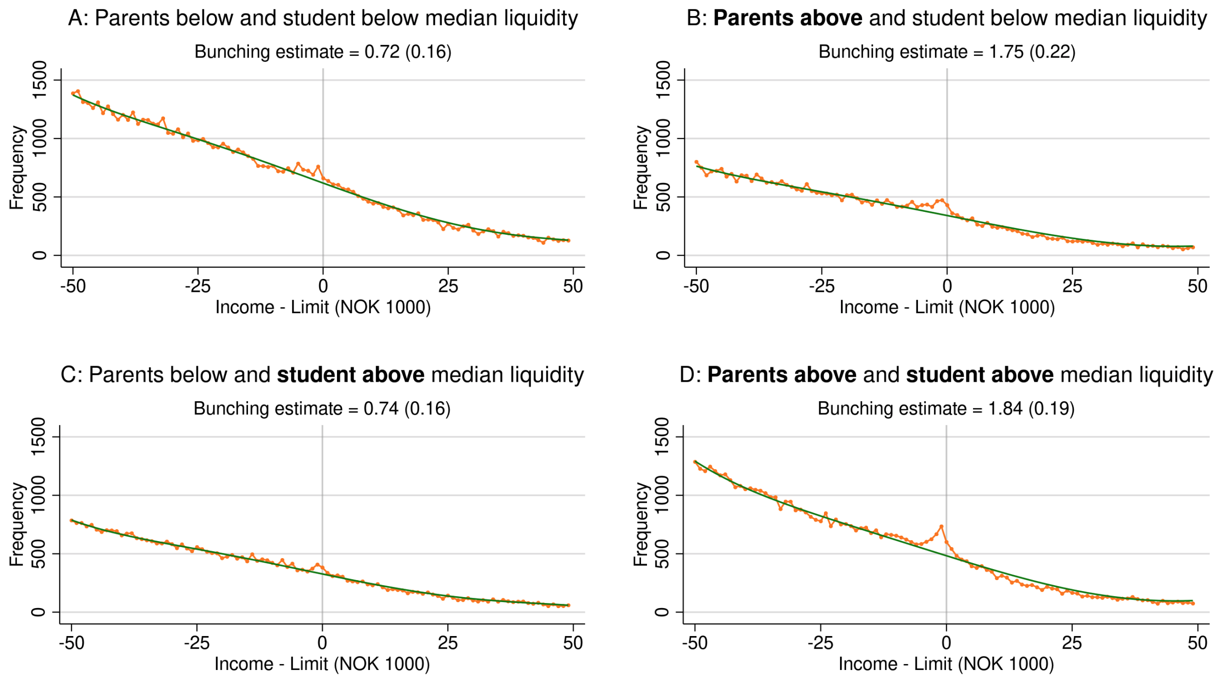
5.5 Bunching heterogeneity

We proceed with a supplementary, more standard approach of investigating heterogeneity in earnings sensitivity to the threshold in Figure 7. This approach splits the sample into subsets based on student and parental characteristic to compute heterogeneous bunching elasticities. We see that the largest contribution to the total excess mass in the preceeding Figure 5 is from students who themselves and their parents have above-median liquid assets. Figure 7 also suggests that the main driver of bunching responses is the parents' rather than the students' own liquid assets. Moving from the left to the right panels, which improves the parents' liquidity, more than doubles the bunching estimates.[21]

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

Figure 7: Heterogeneity in Bunching by Amount of Liquid Assets

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



What can this heterogeneity tell us about how the severity of financial constraints vary with liquidity? In Figure 7, we see that the elasticity increases from 0.72 to 1.84 when moving from below to above the median in terms of both students' and their parents' resources. By using our expression (74) for the expectation of the estimator $\hat{e}$, we may write

$$\frac{1.84}{0.72} = \frac{\mathbb{E}[\hat{e}_{\text{above}}]}{\mathbb{E}[\hat{e}_{\text{below}}]} = \frac{E_{\text{below}}\left[(1-\tau_1)(1-\Delta_r^i)\right] \quad \cdot \quad \zeta \quad \cdot \kappa}{E_{\text{below}}\left[(1-\tau_1)(1-\Delta_r^i)\right] \quad \cdot \quad \zeta \quad \cdot \kappa}. \tag{75}$$

From this expression, we get that

$$\frac{1.84}{0.72} = \frac{E_{\text{below}}[1-\Delta_r^i]}{E_{\text{below}}[1-\Delta_r^i]} = \frac{E_{\text{below}}[R'(s^i)]}{E_{\text{below}}[R'(s^i)]}, \tag{76}$$

Given an average maturity for these loans of about 10 years, we find that the annual gross interest rate $(1 + r_b)$ is $(1.84/0.72)^{\frac{1}{10}} = 1.0984$ times larger for the below-median-liquidity group, which is roughly equal to a 10 percentage point difference. This is a material difference in marginal borrowing rates. While our theoretical framework allows us to calculate implied differences in marginal interest rates, we cannot back this out directly from the data. Firstly, while we can calculate average interest rates on debt, we do not observe marginal interest rates. For example, in practice, students may perhaps be choosing from not borrowing and accruing 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 done in Figure 5 using a *tax* threshold rather than the debt-conversion threshold. The purpose of this exercise is to obtain a reference estimate of the implied labor earnings elasticity at a tax threshold where marginally accrued taxes are not delayed. We focus on the first tax threshold in the progressive income taxation system. This threshold as located at NOK 30,000 during 2005–06 and NOK 40,000 during 2007–2011.[22] At this threshold, the marginal income tax increases from 0 to around 25 percent for most taxpayers.

In Figure 8, we investigate this complementary empirical setting. Panel (A) provides a scatter-plot which verifies the presence of a rise in the marginal income tax rate by plotting total taxes accrued that year against incomes. It also provides the fitted kinked line, from which we infer an average increase in the marginal tax rate of 19 percentage points at the threshold. The coefficient is lower than the nominal increase of 25 percentage points since some individuals may be eligible for higher standard deductions.

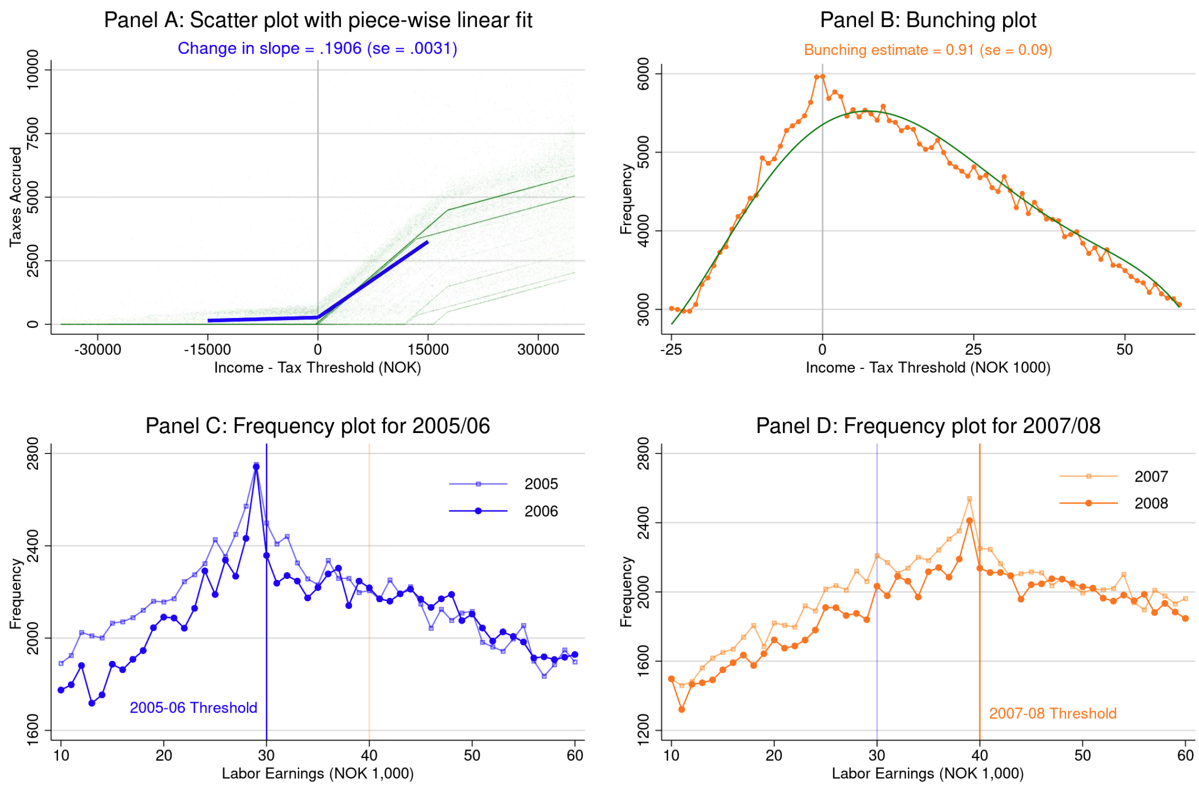
Panel (B) illustrates how the bunching estimate of $b = 0.91$ is calculated. While this bunching estimate is smaller than that found at the debt-conversion threshold, this one-to-one comparison is uninformative for two reasons. Firstly, we need to divide 0.91 by the threshold (36.706 in NOK 1,000s) to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

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

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

Figure 8: Bunching at a Regular Tax Threshold

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



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

$$\frac{0.13}{0.0162} = \frac{E_{\text{delayed}}\left[(1-(1-0)\Delta_r^i)\right]\ \zeta\ \kappa}{E_{\text{regular}}\left[(1-(1-1)\Delta_r^i)\right]\ \zeta\ \kappa} = (1 - E_{\text{delayed}}[\Delta_r^i]). \tag{77}$$

which implies an an average annual interest rate over 10 years of $\left(\frac{0.13}{0.162}\right)^{\frac{1}{10}} - 1 = 23\%$. This number is comparable to average credit-card rates that lie slightly above 20%.[23] This implies that some students are willing to borrow from the educational loan fund at a rate exceeding that offered by financial institutions. This may be driven in part by credit rationing, but likely primarily from the fact that the loan fund does not require payments while students are still in school and generally have a long maturity with the additional opportunity to delay payments for up to three years.

We can use this implied elasticity to get an idea of how much bunching would be caused

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

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 this, we reverse the calculation used to infer labor supply elasticities from bunching estimates. This offers a counter-factual bunching estimate of 23.43.[24] 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 varies across the delayed tax (blue squares) and regular tax threshold (orange triangles). Panel A considers the propensity to engage in unsecured borrowing, defined as having interest expenses in excess of NOK 1,000. For this sample, we omit students who own a house, car, or boat according to the tax authorities. Panel B considers log liquid assets, where liquid assets consist of mainly 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 age of the student and the (max) years of education of the parents. The bin width is NOK 2000 and is based on the distance between labor earnings and the applicable threshold.

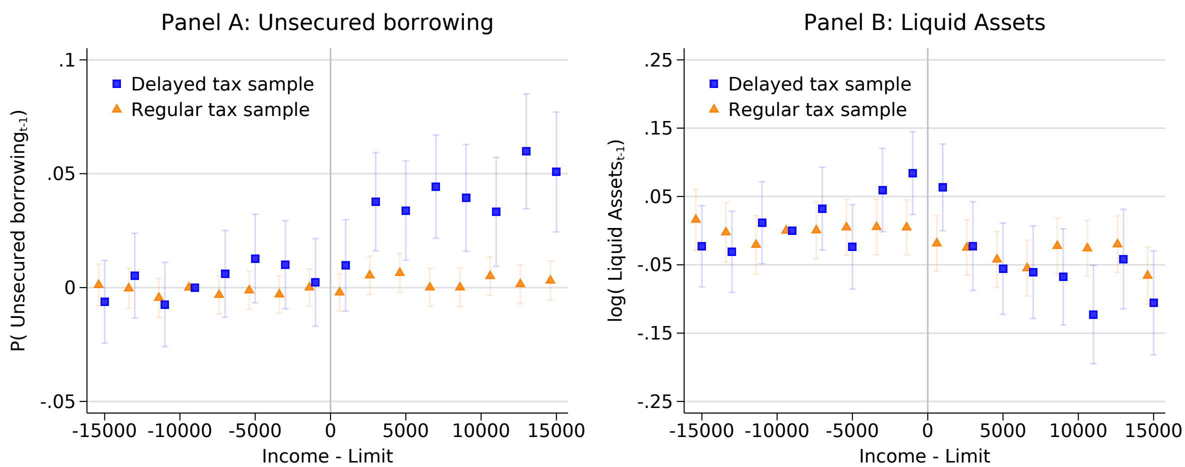


Figure 6 shows that those who bunch at the *delayed tax* threshold have more liquid wealth. This is intuitive because more constrained students will on average have higher marginal borrowing rates and thus be less sensitive to a tax that is payable in the future. However, it is useful to show that proxies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

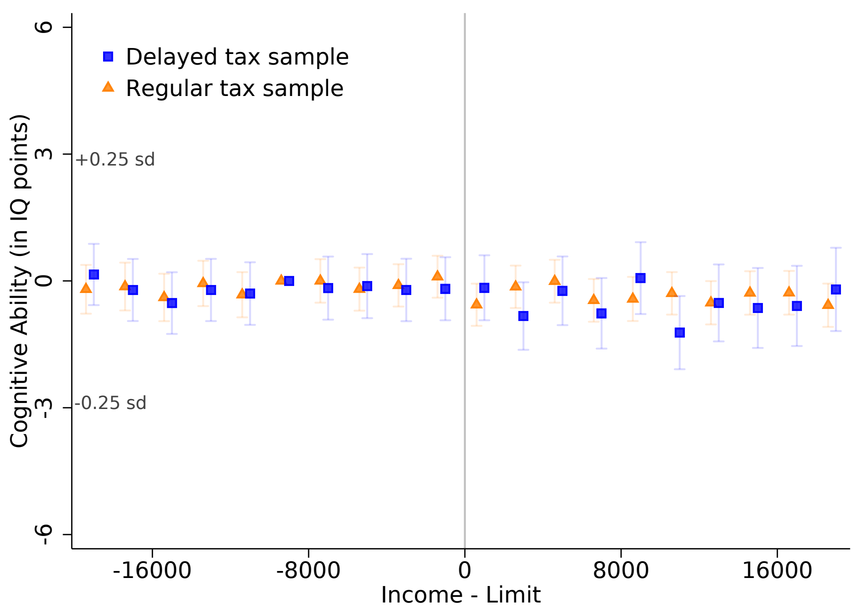
Panel A shows that the delayed-tax non-bunchers are significantly more likely to take take on unsecured debt, such as credit cards or consumption loans. In the regular tax sample, however, there does not seem to be a systematic relationship between whether someone bunches and their unsecured borrowing. Panel B considers liquid assets. In the delayed tax sample, we see that bunchers have more liquidity, and that those who earn above the threshold have significantly less liquid assets. In the regular tax sample, we do not see any strong deviations for bunchers, but moving toward the right, we see that nonbunchers do seem to have less liquid assets.

[24]=0.13*(120162/1000)*(75/50)

We also examine whether cognitive ability (as measured during mandatory military enrollment tests for males) varies with students' position relative to the delayed-tax and regular tax thresholds in Figure 10. This does not reveal any systematic relationship between the income location and ability, which speaks against the concern that students may irresponsive to delayed taxation due particularly low cognitive ability as opposed to financial constraints.

Figure 10: Cognitive Ability of Bunchers at the Delayed Tax and Regular Tax Thresholds

This figure shows how cognitive ability scores varies across the delayed tax (blue squares) and regular tax threshold (orange triangles). Ability scores are measured during military enrollment testing, which is mandatory for all males aged 18–19. The raw scores are provided on a 1–9 (stanine) scale, which we convert to IQ points using the standard formula: $IQ = (\text{raw score} - 5) \times 7.5$. We control for year fixed effects and third-order polynomials in age of the student. The bin width is NOK 2000 and is based on the distance between labor earnings and the applicable threshold.



5.7 Accounting for differences on 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 keeping observables fixed.[25] 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, e.g., differences in labor supply adjustment frictions.

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

We first define the individual-level elasticity as

$$\tilde{e}_i = \underbrace{\frac{1[y_i \in BR_s] - \hat{P}^{cf}(y_i \in BR_s)}{\hat{P}^{cf}(y_i \in BR_s)/N_s^{bins}}}_{\text{estimated } b_s} \cdot (\text{Bin width}_s/\text{Threshold}_s) \Big/ (\hat{d}\tau_s/(1-\tau_s)), \tag{78}$$

where $\hat{P}^{cf}$ denotes the estimated (counter-factual) probabilities of being in the bunching region absent any tax or debt-conversion kinks. This is estimated using the frequencies in the earnings bins around the bunching region as in Saez (2010). The s-sample mean, $\hat{E}$, of $\tilde{e}_i$ provides an estimate of the implied labor supply elasticity. For the delayed tax sample, this mean is around 0.0155, and thus very close to our baseline estimate of 0.0162.[26]

We then estimate regression equations of the following form.

$$\tilde{e}_i = \alpha + \beta \mathbb{1}[\text{regular tax sample}]_i + \gamma' X_i + \varepsilon_i, \tag{79}$$

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

The main finding is that the relative difference in labor supply elasticities is about 7.20 (CI = [6.04, 8.36]) once we control for gender, age, parental educational attainment, and 4-digit occupation fixed effects. Once we include more narrow 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 lower than the relative difference obtained by simply contrasting implied elasticities from the bunching analyses, but the qualitative implications are the same: to rationalize a 6.1 times higher elasticity, we require an average marginal interest rate of 19.82%=$6.1^{1/10}$-1.

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

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

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

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

6 Discussion

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

There may be important costs associated with non-payment or debt-overhang (see, e.g., Donaldson, Piacentino, and Thakor 2019 and Cespedes, Parra, and Sialm 2020) induced by such a scheme. 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 potential justification for this would be that the government has a strong advantage in collecting debts. However, even absent such an advantage, we anticipate 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 naturally the reduced interest costs and increased intertemporal consumption smoothing of the 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.*

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.

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

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

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 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.

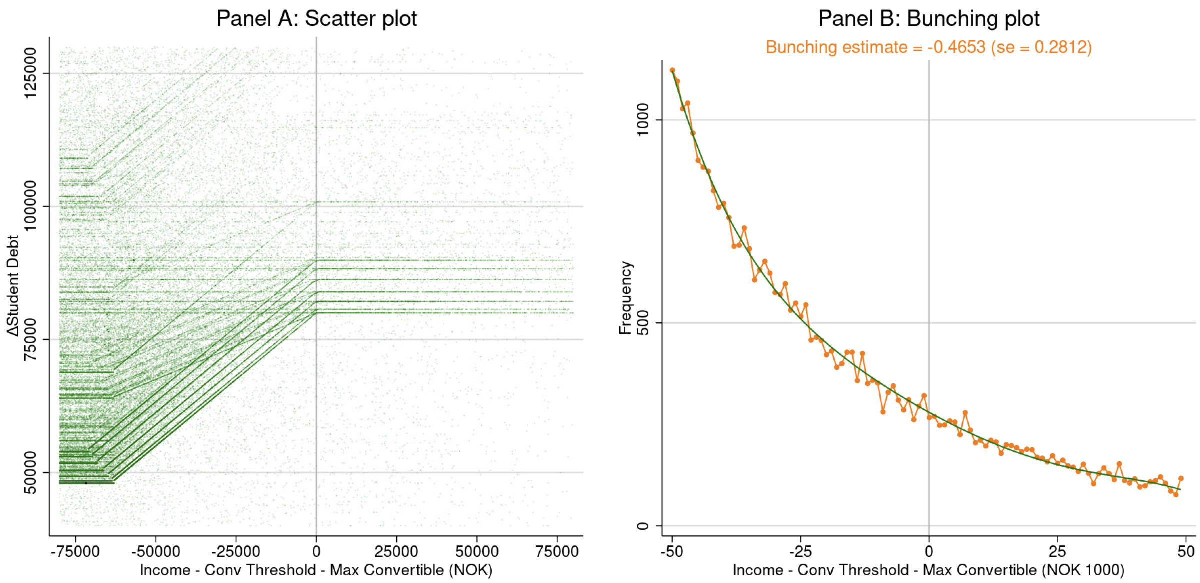
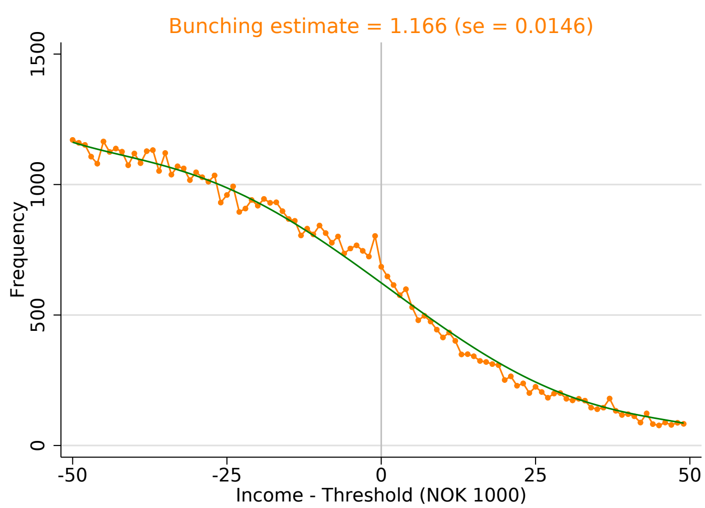


Figure A.2: Bunching at Debt-Conversion Threshold for Workers with Sales and Hospitality Occupations

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.



B 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.

B.0.1 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.

$$dc_1 = d\ell_1 w_1 (1 - \delta\tau_1) - \ell_1 w_1 d(\delta\tau_1) - ds. \tag{B1}$$

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

$$ds = \frac{1}{R'(s)} \left[dc_2 + \ell_1 w_1 d\left\{(1-\delta)(1+r)\tau_1\right\} + (1-\delta)(1+r) w_1 \tau_1 d\ell_1 - (1-\tau_2) w_2 d\ell_2 \right]. \tag{B2}$$

Substituting (B2) into (B1), and using the expression for $\bar{\delta}^i$, yields

$$dc_1 = d\ell_1 w_1 (1 - \bar{\delta}^i \tau_1) - \ell_1 w_1 d(\bar{\delta}^i \tau_1) - \frac{1}{R'(s^i)} dc_2 + \frac{1}{R'(s^i)} (1-\tau_2) w_2 d\ell_2. \tag{B3}$$

Now we can differentiate the second period intratemporal FOC (12) to get $d\ell_2 = \frac{u''(c_2)}{v''(l_2)} w_2^i (1 - \tau_2) dc_2$ and substitute this into (B3) and collect multiplicative terms on dc_2 to get

$$dc_1 = d\ell_1 w_1 (1 - \bar{\delta}^i \tau_1) - \ell_1 w_1 d(\bar{\delta}^i \tau_1) - \frac{1}{R'(s^i)} \left(1 - \left[(1-\tau_2) w_2\right]^2 \frac{u''(c_2)}{v''(l_2)} \right) dc_2. \tag{B4}$$

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.

$$dc_1 \left[1 + \left(1 - \left[(1-\tau_2) w_2\right]^2 \frac{u''(c_2)}{v''(l_2)} \right) \frac{u''(c_1)}{\beta u''(c_2)} \frac{1}{R'(s)^2} \right] = d\ell_1 w_1 (1 - \bar{\delta}^i \tau_1) - \ell_1 w_1 d(\bar{\delta}^i \tau_1). \tag{B5}$$

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 14, which relies on the Euler equation 13):

$$u_1''(c_1^i) w_1^i \left(1 - \tau_1 \bar{\delta}^i \tau_1\right) dc_1 + u'(c_1^i) w_1^i d\left(\tau_1 \bar{\delta}^i\right) \quad = \quad v''(\ell_1^i) d\ell_1^i. \tag{B6}$$

We then substitute (B6) into (B5). We denote the term in the brackets in (B5) as ι^i.

$$\left[\frac{v''(\ell_1^i) d\ell_1^i}{u_1''(c_1^i) w_1^i \left(1 - \bar{\delta}^i \tau_1\right)} - \frac{u'(c_1^i) w_1^i d\left(\tau \bar{\delta}^i \tau_1\right)}{u_1''(c_1^i) w_1^i \left(1 - \bar{\delta}^i \tau_1\right)} \right] \iota^i = d\ell_1 w_1 (1 - \bar{\delta}^i \tau_1) - \ell_1 w_1 d(\bar{\delta}^i \tau_1). \tag{B7}$$

Then we re-arrange to obtain

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\iota^i - w_1^i(1-\bar{\delta}^i\tau_1)u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}. \tag{B8}$$

Then we re-arrange to obtain

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\iota^i - w_1^i(1-\bar{\delta}^i\tau_1)u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}. \tag{B9}$$

This may also be written as

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i u_1''(c_1^i)\tilde{w}_1^i}. \tag{B10}$$

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

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right] - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right] - u_1''(c_1^i)w_1^{i2}\left(1-\bar{\delta}^i\tau_1\right)^2}, \tag{B11}$$

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

B.0.2 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.

$$dc_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)}\left((1-\tau_2)w_2 d\ell_2 - dc_2\right) \tag{B12}$$

(b) Now find an expression for $d\ell_2$. We use the implied intratemporal FOC for labor (15), and differentiate it to get

$$d\ell_2 = \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)}d\ell_1. \tag{B13}$$

(c) Now find an expression for dc_2. We differentiate the Euler equation (13).

$$dc_2 = \frac{u''(c_1)}{\beta R'(s)u''(c_2)}dc_1 \tag{B14}$$

(d) Now find an expression for dc_1. We differentiate the intratemporal FOC for ℓ_1, (14), which relies on the intertemporal FOC, (13).

$$dc_1 = \frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1. \tag{B15}$$

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

$$dc_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)}\left((1-\tau_2)w_2 \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} d\ell_1 - \frac{u''(c_1)}{\beta R'(s)u''(c_2)} dc_1\right) \tag{B16}$$

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

$$\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)}\left((1-\tau_2)w_2 \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} d\ell_1 - \frac{u''(c_1)}{\beta R'(s)u''(c_2)} \frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1\right) \tag{B17}$$

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

$$\left(\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} - \tilde{w}_1 - \frac{1}{R'(s)}\left[\frac{[w_2(1-\tau_2)]^2 v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} - \frac{1}{\beta R'(s)u''(c_2)}\frac{v''(\ell_1)}{\tilde{w}_1}\right]\right) d\ell_1 = dG_1 \tag{B18}$$

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

$$\left(\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} - \frac{1}{R'(s)}\left[\frac{[w_2(1-\tau_2)]^2 v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} - \frac{1}{\beta R'(s)u''(c_2)}\frac{v''(\ell_1)}{\tilde{w}_1}\right] - \tilde{w}_1\right) d\ell_1 = dG_1 \tag{B19}$$

$$\left(v''(l_1)\left[1 - u''(c_1)\frac{1}{\beta R'(s)^2}\left[\frac{[w_2(1-\tau_2)]^2}{v''(\ell_2)} - \frac{1}{u''(c_2)}\right]\right] - \tilde{w}_1 u''(c_1)\tilde{w}_1\right) d\ell_1 = u''(c_1)\tilde{w}_1 dG_1 \tag{B20}$$

$$\left(v''(l_1)\left[1 + \left(1 - \frac{[w_2(1-\tau_2)]^2 u''(c_2)}{v''(\ell_2)}\right)\frac{u''(c_1)}{u''(c_2)}\frac{1}{\beta R'(s)^2}\right] - \tilde{w}_1 u''(c_1)\tilde{w}_1\right) d\ell_1 = u''(c_1)\tilde{w}_1 dG_1 \tag{B21}$$

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

$$\frac{d\ell_1^i}{dG_1} = \frac{u''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i)\left[1 + \left(1 - \frac{[w_2^i(1-\tau_2)]^2 u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right] - \tilde{w}_1^i u''(c_1^i)\tilde{w}_1^i}, \tag{B22}$$

or using the definition of ι^i,

$$\frac{d\ell_1^i}{dG_1} = \frac{u''(c_1^i)\tilde{w}_1}{v''(\ell_1^i)\iota^i - \tilde{w}_1 u''(c_1^i)\tilde{w}_1^i}, \tag{B23}$$

B.0.3 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,

$$\frac{d\ell^i_1}{d\tilde{w}^i_1} = \left(\frac{d\ell^i_1}{d\tilde{w}^i_1)}\right)^c + \frac{d\ell^i_1}{dG_1}\ell^i_1. \tag{B24}$$

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

$$\frac{d\ell^i_1}{d(1-\tau_1)} = \left(\frac{d\ell^i_1}{d(1-\tau_1)}\right)^c + \frac{d\ell^i_1}{dG_1}\ell^i_1\bar{\delta}w^i_1. \tag{B25}$$

Substituting in for the LHS using equation (B11) and the second-term on the RHS using (B23), we obtain

$$-\bar{\delta}^i\frac{u'(c^i_1)w^i_1\iota^i - \ell^i_1 w^i_1 u''_1(c^i_1)\tilde{w}^i_1}{v''(\ell^i_1)\iota^i - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1} = \left(\frac{d\ell^i_1}{d(1-\tau_1)}\right)^c + \frac{u''(c^i_1)\tilde{w}_1}{v''(\ell^i_1)\iota^i - \tilde{w}_1 u''(c^i_1)\tilde{w}^i_1}\ell^i_1 w^i_1\bar{\delta}^i. \tag{B26}$$

Rearrange and cancel out to get

$$\left(\frac{d\ell^i_1}{d(1-\tau_1)}\right)^c = -\bar{\delta}^i\frac{u'(c^i_1)w^i_1\iota^i}{v''(\ell^i_1)\iota^i - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1}. \tag{B27}$$

In terms of an elasticity, we can write it as

$$\varepsilon^c_{1,1-\tau_1} = -\bar{\delta}^i\frac{1-\tau_1}{\ell^i_1}\frac{u'(c^i_1)w^i_1\iota^i}{v''(\ell^i_1)\iota^i - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1}. \tag{B28}$$

We may rewrite this using the intratemporal FOC (14), 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

$$\varepsilon^c_{1,1-\tau_1} = -\frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\frac{1-\tau_1}{\ell^i_1}\frac{v'(\ell_1)\iota^i}{v''(\ell^i_1)\iota^i - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1}. \tag{B29}$$

Writing out the ι^i terms, we get

$$\varepsilon^c_{1,1-\tau_1} = -\frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\frac{1-\tau_1}{\ell^i_1}\frac{v'(\ell_1)\left[1+\left(1-\frac{[w^i_2(1-\tau_2)]^2u''(c^i_2)}{v''(\ell^i_2)}\right)\frac{u''(c^i_1)}{u''(c^i_2)}\frac{1}{\beta R'(s^i)^2}\right]}{v''(\ell^i_1)\left[1+\left(1-\frac{[w^i_2(1-\tau_2)]^2u''(c^i_2)}{v''(\ell^i_2)}\right)\frac{u''(c^i_1)}{u''(c^i_2)}\frac{1}{\beta R'(s^i)^2}\right] - \tilde{w}^i_1 u''_1(c^i_1)\tilde{w}^i_1}. \tag{B30}$$

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

$$\varepsilon^c_{1,1-\tau_1} = -\frac{\frac{\delta+(1-\delta)\frac{1+r}{R'(s^i)}}{1-\left[\delta+(1-\delta)\frac{1+r}{R'(s^i)}\right]\tau_1}\frac{1-\tau_1}{\ell_1^i}v'(\ell_1)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-\left[w_1^i\right]^2\left(1-\tau_1\left[\delta+(1-\delta)\frac{1+r}{R'(s^i)}\right]\right)^2u_1''(c_1^i)}. \tag{B31}$$

C 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 (B11), we can write:

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-\ell_1^iw_1^iu_1''(c_1^i)w_1^i\left(1-\tilde{\tau}_1^i\right)}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-u_1''(c_1^i)w_1^{i2}\left(1-\tilde{\tau}_1^i\right)^2}. \tag{C1}$$

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 (40).

The proof of (41) has the following steps. We substitute the Euler equation (13) into the period-1 intratemporal FOC (14) 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 (C1) 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:

$$u_1'(c_1)w_1^i\left(1-\tau_1\bar{\delta}^i\right) = v'(\ell_1) \tag{C2}$$

$$\frac{u_1'(c_1)}{\beta(1+r_b)}w_2^i[1-\tau_2] = v'(\ell_2) \tag{C3}$$

$$c_1 + \frac{c_2}{1+r_b} = w_1\ell_1(1-\tau_1\bar{\delta}^i) + \frac{w_2\ell_2(1-\tau_2)+G_2}{1+r_b}. \tag{C4}$$

The first condition is just (14), obtained by inserting the intertemporal FOC (13) into (11), the second condition is obtained by inserting (13) into (12), and the third constraint is just the life-time budget constraint.[27] 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 $\frac{d(\tau_1\bar{\delta}^i)}{d\delta} = \tau_1\left(1 - \frac{1+r}{R'(s)}\right)$. Thus, we have that

$$\frac{d\ell_1}{d\tau_1} = \frac{d\ell_1}{d(\tau_1\bar{\delta}^i)}\frac{d(\tau_1\bar{\delta}^i)}{d\tau_1} = \bar{\delta}^i\frac{d\ell_1}{d(\tau_1\bar{\delta}^i)} \tag{C5}$$

$$\frac{d\ell_1}{d\delta} = \frac{d\ell_1}{d(\tau_1\bar{\delta}^i)}\frac{d(\tau_1\bar{\delta}^i)}{d\delta} = \tau_1\left(1 - \frac{1+r}{R'(s)}\right)\frac{d\ell_1}{d(\tau_1\bar{\delta}^i)} \tag{C6}$$

Substituting from (C5) into (C6), 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.

D Proof of Lemma 2

Part i) Using the derivations from Proposition 3, we have that:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)} = \tau_1\mathbb{E}_1\left(g_1^i - g_2^i\right) + \mathbb{E}_1\left[\tau_1\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i\right)\right]. \tag{D1}$$

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

$$\frac{dW}{\lambda d(1-\delta)}\frac{1}{\tau_1} = \mathbb{E}_1\left(g_1^i\left[1 - \frac{1+r}{R'(s^i)}\right]\right) + \mathbb{E}_1\left[\left[\frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}\right]\left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i\right)\right]. \tag{D2}$$

Setting $\delta = 1$ yields:

$$\frac{dW}{\lambda d(1-\delta)}\frac{1}{\tau_1} = \mathbb{E}_1\left(g_1^i\left[1 - \frac{1+r}{R'(s^i)}\right]\right) + \mathbb{E}_1\left[\left(1 - \frac{1+r}{R'(s^i)}\right)\left(\frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i\right)\right]. \tag{D3}$$

[27] The latter is derived by noticing that

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

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:

$$\frac{dW}{\lambda d(1-\delta)} \frac{1}{\tau_1 \Delta_r} = \sum_{i:s^i<0} \pi^i y_1^i g_1^i + \sum_{i:s^i<0} \pi^i y_1^i \rho^i + \sum_{i:s^i<0} \pi^i y_1^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right), \tag{D4}$$

$$= \sum_{i:s^i<0} \pi^i y_1^i (g_1^i + \rho^i) + \sum_{i:s^i<0} \pi^i y_1^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right). \tag{D5}$$

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.

$$\frac{dW}{dx} = \sum_i \pi^i \alpha_i \left(u'(c_1^i) - (1+r)\beta u'(c_2^i) \right) + \lambda \sum_i \pi^i \left(\frac{dy_1^i}{dx} + \frac{1}{1+r} \frac{dy_2^i}{dx} \right). \tag{D6}$$

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

$$\frac{dW}{dx} = \sum_i \pi^i \alpha_i \left(u'(c_1^i) - (1+r)\beta u'(c_2^i) \right) + \lambda \sum_i \pi^i \left(1 - \frac{1+r}{R'(s^i)} \right) \left(\tau_1 \frac{dy_1^i}{dG_1} + \frac{1}{1+r} \tau_2 \frac{dy_2^i}{dG_1} \right). \tag{D7}$$

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

$$\frac{1}{\lambda} \frac{dW}{dx} = \sum_i \pi^i \left(g_1^i - g_2^i \right) + \sum_i \pi^i \left(1 - \frac{1+r}{R'(s^i)} \right) \rho^i. \tag{D8}$$

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

Part iii) Equation (33) in the proof of Proposition 2 provides the money-metric welfare effect of a marginal increase in $1 - \tau_1$:

$$\frac{dW}{d(1-\tau_1)\lambda} = \mathbb{E}_1 \left[\delta g_1^i + [1-\delta] g_2^i \right] - \mathbb{E}_1 \left[1 - \frac{\tau_1}{1-\tau_1} \varepsilon_{1,1-\tau_1}^i - \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^i \right]. \tag{D9}$$

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:

$$\frac{dW}{d(1-\tau_1)\lambda} = \sum_i \pi^i y_1^i g_1^i + \sum_i \pi^i y_1^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right) + \sum_i \pi^i y_1^i \rho^i - \sum_i \pi^i y_1^i, \tag{D10}$$

$$= \sum_i \pi^i y_1^i (g_1^i + \rho^i) + \sum_i \pi^i y_1^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{2,1-\tau_1}^{i,c} \right) - \sum_i \pi^i y_1^i. \tag{D11}$$

E Data for calibration

We first use microdata from the censuses in 1990 and 2011 to calculate 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 the usual weekly hours worked, (1,10], (11,20], (20,30], (30,35], or full time (37 hours). For these categories we assign numerical values of 5, 15, 25, 32.5, and 37. 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 contains 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*).
- We replace hours worked as missing if either the number of months worked is below 2 or the total hours worked is below 47 × 7.5 (i.e., we require an average of 7.5 hours worked per week).
- We the 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 hours as the number of contractual weekly hours × 47 weeks
- We then calculate the wage as annual labor earnings / hours worked. We use the same approach to address outliers as for 1990 wages.
- Finally, we deflate the 2011 wages by 1.5576 which is the ratio of the 2011 to the 1990 Statistics Norway consumption price index

We then only keep observations for which the age in 1990 was between 20 and 30. For individuals aged below 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 further require that we observe effective wages for the individual in both 1990 and in 2011.

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