Optimal Delayed Taxation in the Presence of Financial Frictions

Spencer Bastani and Marius A. K. Ring*

November 13, 2025 (First version: February 23, 2021)

Abstract

In the presence of financial frictions, the timing of cash flows matters. We apply this insight to optimal income taxation by proposing a new policy: delayed taxation. Introducing a delay between the accrual and payment of income taxes yields two sources of welfare gains when some agents are borrowing constrained. First, it improves consumption smoothing by enabling constrained agents to borrow at a lower rate. Second, it lowers the present-value tax rate from the perspective of constrained agents, thereby reducing the distortionary effects of income taxation. We study marginally delayed taxation in a dynamic optimal tax model, contrasting these gains with the fiscal costs caused by income effects and defaults. We then characterize optimal delayed tax reform in a simple calibrated model. This quantitative analysis reveals significant welfare gains from delayed taxation, even when the government has no lending advantage over private credit markets. These gains persist on top of other policies, such as age-dependent taxation, nonlinear tax schedules, and subsidized lending. The key source of these gains is delayed taxation's ability to undo distortions caused by income taxation. The final part of our paper empirically tests this mechanism using a natural experiment from Norway. There, a kinked income-contingent student debt conversion scheme replicates delayed taxation. Bunching analyses reveal elasticities that are considerably lower than those found for a regular income tax threshold. Consistent with our theory, proxies for financial constraints are associated with lower sensitivities to the de facto delayed tax, but not to the regular tax. Taken together, our theoretical, quantitative, and empirical findings underscore the potential for delayed taxation to be a powerful new component of optimal tax policy.

JEL: H21, G51, D15

Keywords: delayed taxation, credit constraints, optimal income taxation

*Spencer Bastani is at the Institute for Evaluation of Labour Market and Education Policy (IFAU) and the Department of Economics, Uppsala University; Research Institute of Industrial Economics (IFN), Stockholm. Marius Ring (corresponding author) is at the University of Texas at Austin. E-mail: mariuskallebergring[at]gmail.com. This paper previously circulated as "Financial frictions and the Non-Distortionary Effects of Delayed Taxation," available on SSRN as of February 2021. We thank seminar participants at MIT, UT Austin, IFN, Washington University in St. Louis, Nordic Public Policy Symposium, LMU Munich, Cornell University, Erasmus University, and Yale University for helpful comments and questions, and Anthony DeFusco, Victoria Marone, 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, Ragnhild Schreiner, and Nathaniel Hendren, Costas Meghir, Mayara Felix, Costas Arkolakis, Deborah Lucas, James Choi, and Veronica Guerrieri 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

Capital market imperfections render borrowing against future labor incomes expensive or even impossible. Both policymakers and economists recognize the welfare losses from these frictions. Governments therefore engage in a broad range of credit market interventions, particularly in markets where there may be positive externalities. Home purchasing, higher education, and entrepreneurship are activities that are believed to carry positive externalities and often benefit from government-subsidized financing. We argue that the motivations for government intervention in these markets carry over to the tax system.

Through the tax system, governments raise revenues by imposing costs on labor supply. A 30% marginal tax rate implies that an additional dollar of labor income costs 30 cents. Yet, neither policymakers nor the public finance literature devotes attention to whether income tax payments, a key cost of supplying labor, should be financed over a longer horizon. The idea that in-house financing may increase both revenues and welfare is central to other markets. Automakers routinely and profitably offer low-interest financing to reduce price elasticities (Einav, Jenkins, and Levin, 2012). This is optimal even without incorporating the additional welfare gains from increased consumption smoothing for credit-constrained customers. Governments that care about both revenues and taxpayers have even stronger incentives to offer in-house financing. For the tax system, this financing scheme would materialize as delayed taxation: taxes accrue on an ongoing basis, but payments may be delayed, subject to a subsidized interest rate.

The notion that credit market imperfections have implications for optimal tax design is not new. The age-dependent tax literature (Kremer 2002, Lozachmeur 2006, Weinzierl (2011), Bastani et al. 2013, Heathcote et al. 2020) document that letting tax rates vary by age may reduce the efficiency costs of taxation. While this literature typically abstracts from financial frictions, Weinzierl (2011) documents considerably larger welfare gains from age-dependent taxation when workers cannot save or borrow. Lower tax rates for young workers help smooth consumption over time. Nevertheless, even if young workers are more credit constrained, reducing tax rates for all young workers is a blunt instrument. The political feasibility of implementing age-discriminatory statutory tax rates is also uncertain. While the literature on optimal taxation under incomplete markets features borrowing constraints (Eaton and Rosen 1980, Varian 1980, Aiyagari 1995, Conesa and Krueger 2006, Heathcote, Storesletten, and Violante 2017), the key friction in this literature is the inability to insure against idiosyncratic shocks. Financial frictions simply exacerbate the welfare losses from incomplete insurance markets. Hence, this literature focuses on the natural policy to address income shocks, namely progressive taxation, rather than the optimal timing of tax payments.

This paper focuses on the optimal scheduling of tax payments in the presence of two key frictions. First, due to limited commitment, lenders must take costly action to disincentivize workers from defaulting. As a result, workers who wish to borrow against future labor earnings face high interest rates. Second, taxes may only be levied on labor earnings and thus distort labor supply. In this setting, we demonstrate that allowing workers to delay their income tax payments both reduces welfare losses from financial frictions and exploits these financial frictions to mitigate the distortionary effects of income taxation. Importantly, delayed taxation enables the government to shift the tax burden away from younger workers without imposing age-discriminatory tax rates.

Rather than explicit targeting, delayed taxation relies on self-selection. For borrowing-constrained workers who discount future cash flows at high interest rates, delaying taxes lowers the present-value tax rate and thus reduces tax distortions. Unconstrained workers, who would not delay their payments, remain unaffected.[1] This policy thus generates two sources of welfare gains for borrowing-constrained agents: enhanced intertemporal consumption smoothing and reduced income tax distortions. The first may be offset by defaults or high lending costs facing the government. The second, however, persists and distinguishes delayed taxation from other forms of government intervention in private credit markets.

To illustrate these effects, consider a simple model with two groups of workers subject to an income tax. One group has a flat or declining wage trajectory and wants to save. The other has an upward wage trajectory and wants to borrow against future earnings, but faces a higher interest rate on borrowing than on saving (and is therefore referred to as financially constrained). Suppose the government allows workers to delay their tax payments at the saving rate. Net savers are indifferent about delaying, while constrained workers strictly prefer it. For a worker, i, who faces a marginal interest rate of r^i and pays interest rate r_{dtax} on taxes delayed for T periods, the statutory marginal tax rate τ translates into a present value tax rate of:

$$PV^i(\tau) = \frac{(1 + r_{dtax})^T}{(1 + r^i)^T} \times \tau. \tag{1}$$

For net borrowers with $r^i > r_{dtax}$, the present-value tax rate falls below the statutory rate, reducing labor supply distortions. Importantly, this reduction would not materialize if the government simply offered unconditional subsidized loans. By shifting the timing of tax *payments*, delayed taxation ties present-value tax rates to each individual's marginal borrowing rate. This effectively allows the government to impose different tax rates on constrained and unconstrained workers without conditioning the tax schedule on observable taxpayer characteristics (i.e., "tagging"). This is an attractive feature of delayed taxation relative to explicitly discriminatory instruments such as age-dependent taxation.

Our study provides an in-depth examination of delayed taxation through a blend of theoretical analysis, numerical simulations, and empirical evidence. In the first part of our paper, we introduce delayed taxation into a dynamic optimal tax model with linear taxation. We consider heterogeneous workers who differ in their labor market productivity, which allows us to study the welfare gains from delayed taxation in the presence of distributional considerations. We 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

[1] Note that there is nothing conceptual that differentiates delayed taxation from offering subsidized loans where maximal loan amounts depend on income levels. We find the tax framing more natural because such an income-contingent lending scheme would generate externalities (higher labor supply) which the government internalizes *through the tax system*.

magnitude of financial frictions, measured by the difference between the interest rates faced by net borrowers and net savers. In other words, assessing the potential welfare gains from delayed taxation does not require any new types of behavioral elasticities.

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 five 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 the tax revenue of older workers, (iv) a negative income effect on the present value of tax revenue due to relaxed financial constraints,[2] and (v) a negative present-value fiscal effect due to costs of default. If the government simply offered unconditional low-interest loans to financially constrained agents, the positive effects on labor supply—that reduce the distortionary effects of income taxation—would be absent.

In the second part of our paper, we provide numerical solutions by calibrating the optimal tax model to the Norwegian economy.[3] The Norwegian register data, which cover the entire population, allow us to compute realized earnings trajectories for a large sample of 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 extent to which workers may wish to borrow against future earnings. We then consider the optimal tax schedule that the government would impose on these workers to maximize welfare in the presence of financial frictions.

The financial friction manifests as a wedge between the interest a wedge between the interest rate on borrowing and the interest rate on saving, which (as the wedge grows large) nests the standard "no-borrowing" constraint (see, e.g., Heathcote 2005). This wedge is due to workers being able to default on their liabilities. We focus on a no-default equilibrium in which lenders pay an upfront cost that allows them to seize workers' assets and thus disincentivize default. Importantly, while delayed taxes feature low interest rates equal to that on savings, there is no free lunch. By allowing workers to delay their tax payments, the government incurs the same upfront costs on the implied loan amount that private lenders would when issuing a loan. Hence, the government discounts delayed tax repayments at the same interest rate that individuals face in private credit markets. We show that deviations from this conservative assumption greatly increases the welfare gains of delayed taxation.[4]

Our baseline analyses consider the potential welfare gains from implementing delayed taxation on top of a benchmark economy in which taxes are linear and age-independent, and individuals may save at an interest of 3% but face a 20% interest rate on borrowing. In the benchmark tax system, the government imposes a high marginal tax rate in part to redistribute across workers

[2]With labor supply fixed, delaying a dollar of taxes increases the present-value wealth of constrained workers because of the high marginal interest rate they face.

[3]We show that our findings are robust to using instead data on U.S. households from the PSID.

[4]For example, one might otherwise assume that the government has superior collection technology through, e.g., their tax administration, that causes the government to be able to break even on loans with lower interest rates than private lenders.

and in part to provide lump-sum transfers that help smooth consumption over time. As a bridge between our theoretical and quantitative analyses, we use the quantitative framework to numerically value the theoretical welfare effect expressions. We demonstrate robust positive marginal welfare gains from delayed taxation. While the main source of welfare is enhanced intertemporal consumption smoothing, we show that welfare gains from reduced income-tax distortions is pivotal in making delayed taxation a desirable policy tool. The money-metric gains from reducing income tax distortions equal nearly two thirds of the direct welfare effect through intertemporal smoothing. Importantly, this reduced-distortion channel is not present for policies that only alleviate financial frictions through, e.g., unconditional subsidized lending.

We use our theoretical framework to compute the Marginal Value of Public Funds (Hendren and Sprung-Keyser 2022, Finkelstein and Hendren 2020). Under our main assumptions, the MVPF is 1.53, which says that that the money-metric welfare effects of marginally delaying taxation are 53% higher than its present-value cost. When the borrowing rate is lower (e.g., 7%) and the government has a cost advantage in lending to individuals, the MVPF is infinite, indicating that delayed taxation can be implemented as a Pareto-improving reform. We also show that the optimality of delayed taxation, ensured by $MVPF > 1$, is not sensitive to either the extent of borrowing frictions faced by individuals (measured as the interest spread on borrowing) or the potential costs of default. Even when the government faces a lending disadvantage and thus uses a higher interest rate than private markets to discount future debt repayments, the MVPF significantly exceeds 1.

Our main quantitative exercise studies the optimal implementation of delayed taxation reform. When the interest rate on borrowing is 20%, the government optimally chooses to delay 31% of tax payments and thereby achieves (money-metric) welfare gains of 2.2% of period-1 consumption. Welfare gains are material even when financial frictions are less severe. At a borrowing rate of 10%, welfare gains equal about 1.2% of period-1 consumption.

Policies such as age-dependent taxation, non-linear taxation, and subsidized lending may also address some of the welfare losses from financial frictions. A key question is therefore whether delayed taxation is appealing when the government already makes use of these other policy tools. We find this to consistently be the case. The 2.2% welfare gains are only reduced by 0.4 percentage points when the government can offer unconditional low-interest loans, and by 0.25 percentage points when the government can set age-dependent tax rates. Interestingly, when the government implements non-linear taxation as in Heathcote et al. (2020), the welfare gains nearly double.

In robustness tests, we further verify that delayed taxation yields material welfare gains when our model is calibrated to the U.S. PSID data used by Weinzierl (2011). Our qualitative results are also robust to varying both the severity of financial frictions faced by workers and the constraints on government borrowing.

In the final part of the paper, we empirically test the key distinguishing feature of delayed taxation as an optimal tax tool, namely that delayed taxation reduces income tax distortions when workers are financially constrained. Conducting such a test is challenging because it requires quasi-experimental variation in effective tax rates alongside variation in when taxes are paid. Few settings offer substantial variation in the timing of 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 behavioral responses to a student debt conversion scheme in Norway, which constitutes a *de-facto* delayed tax.

Norwegian students receive an annual loan of approximately \$13,000 from the government to pay for housing and other consumption. If students remain in good standing, half of the loan payment is usually forgiven at the end of the year. However, if the student earns above a threshold of about \$17,000, each additional dollar of earnings reduces the amount forgiven by 50 cents. In other words, additional earnings increases the student's debt. Since this debt is owed to the government, it is conceptually no different than a delayed tax liability. Thus, this program creates a large jump in the effective marginal tax rate, where marginal taxes can be financed with the same generous terms as subsidized student loans. 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 large increase in the effective tax rate at the earnings threshold is also more than significant enough for any student to be aware of: at the threshold, the marginal after-tax wage falls from 75 cents to 25 cents once we account for the increase in student debt. Despite this drastic reduction in the marginal 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 after-tax wage reduction that occurs. Our bunching analysis yields an implied elasticity of labor earnings to after-tax wages of only 0.016. This small elasticity may be due, in part, to labor supply optimization frictions that create a mismatch between bunching and structural elasticities. Accounting for these frictions, as measured by Kostøl and Myhre (2021) using Norwegian data, increases our implied elasticity estimate to 0.053. This elasticity is still considerably smaller than the frictionless elasticity estimate of Kostøl and Myhre (2021) of 0.30.

We contend that the reason our estimate is particularly low is that we are studying delayed taxation in a sample of likely highly constrained workers. For these students, the present value of the tax rate is much lower than the nominal value implied by the increase in debt. An objection to our claim is that optimization frictions are particularly prevalent among the students in our sample. Thus, the inelastic behavior we observe could be driven by optimization frictions as opposed to financial frictions. We address this concern in several ways. Our main approach is to also examine the bunching behavior of *students* at a regular tax threshold. This allows us to compare the bunching elasticity of the de facto delayed tax with that of a regular tax regime in a similar sample of individuals. The regular tax threshold occurs around \$6,000, where the marginal tax rate jumps from 0 to 25 percent. Using the bunching at this threshold, we estimate an implied labor supply elasticity of 0.130. This is about eight times larger than the elasticity implied by the delayed tax threshold. Assuming similar optimization frictions in the two samples, this large difference in elasticity (0.016 versus 0.130) can be explained by the fact that students face an average marginal borrowing rate of more than 20 percent and are thus much less responsive to the de facto delayed tax scheme created by the student loan program. We argue that it is unlikely that this difference is driven by the fact that the kinks occur at different income levels. Using a regression-based approach that controls for differences in observables such as occupation codes, age, and family background, we find a qualitatively similar difference in elasticity.

Our theoretical framework implies that the sensitivity to delayed but not regular taxation is moderated by the severity of borrowing constraints. This empirical prediction is strongly supported by the data. Those who do not respond to the delayed tax incentive (i.e., non-bunchers) have less liquid assets and are more likely to engage in unsecured borrowing. We do not find this pattern in the regular tax sample.

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 borrowing-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. This addresses the concern that less human capital as opposed to less liquidity is driving the down the bunching elasticities. 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 less liquid students facing a 10 percentage point higher marginal interest rate.

The main empirical finding that delaying the payment of a tax reduces distortions is not surprising. In the presence of borrowing-constrained agents, this is what we would expect from economic theory. Rather, the contribution of this empirical exercise is to test the applicability of life-cycle model reasoning in modeling the labor supply decisions of constrained workers, which is necessary to assess the potential of delayed taxation as a new policy tool.

Literature. Our paper is related to the optimal tax literature that allows tax rates to depend on taxpayer characteristics, i.e., tagging (Akerlof, 1978). Delayed taxation causes financially constrained taxpayers to experience a reduction in effective (present value) tax rates. As taxpayers age and borrowing constraints are no longer binding, effective tax rates equal the higher nominal rate. In this sense, delayed taxation has a strong element of age-dependent taxation (Kremer 2002, Lozachmeur 2006; , Blomquist and Micheletto 2008; Weinzierl 2011; Gervais 2012; Bastani, Blomquist, and Micheletto 2013; da Costa and Santos 2018; Heathcote, Storesletten, and Violante 2020). The key differences are that (i) delayed taxation does not require the government to condition tax rates on taxpayer characteristics (which is likely to be controversial), and (ii) it does not rely on using age as a proxy for liquidity constraints. Instead, delayed taxation allows constrained borrowers to self-select into the system, thereby introducing a voluntary type of history dependence. More broadly, this paper contributes to the literature on dynamic optimal taxation (see, for example, Ndiaye 2020, Yu 2021, and the surveys in Golosov and Tsyvinski 2015 and Stantcheva 2020). Most closely related are studies that consider changing the timing of tax payments (e.g., Lockwood 2020) or incorporating financial frictions (e.g., Andreoni 1992, Dávila and Hébert 2019, Coven et al. 2024). The central contribution of this paper is to propose and study the welfare implications of the simple idea that—in the presence of credit market imperfections—changing the timing of income tax payments can offer substantial welfare gains, in large part by reducing the distortionary effects of income taxation. We strengthen this contribution by providing quasi-experimental evidence that delayed taxation does indeed reduce income tax distortions. To our knowledge, this idea has not been explored before, either theoretically or empirically.

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; Asatryan and Peichl 2017; Tazhitdinova 2018; Søgaard 2019; Tazhitdinova 2020; Glogowsky 2021; and the review by Kleven 2016), loan term thresholds (see, e.g., Bachas, Kim, and Yannelis 2021; Bäckman, van Santen et al. 2020; DeFusco and Paciorek 2017; DeFusco, Johnson, and Mondragon 2020; de Silva 2023; Le Barbanchon 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 that is not present in studies that consider sensitivity to taxation. This paper also relates to the emerging literature on the effects of debt on labor supply (see, e.g., Zator 2019; Bruze, Hilsløv, and Maibom 2024; Bernstein 2021; Doornik et al. 2021; Brown and Matsa 2020; Donaldson et al. 2019). We also contribute to research that considers how different tax instruments may affect behavioral elasticities. For example, Kostøl and Myhre (2021) consider how labor supply elasticities are affected by providing more information about 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.

The paper is organized 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 system in Norway to test some of the behavioral implications of our theoretical framework. Section 6 briefly discusses implementation issues and unmodeled trade-offs associated with the introduction of delayed taxation.

2 The welfare gains of delayed taxation

2.1 The model

We consider an economy consisting of heterogeneous agents who live and work for two periods and differ in their exogenous and deterministic lifetime labor productivity profiles (w_1^i, w_2^i), where w_t^i denotes the market productivity of agent i in period $t = 1, 2$. In each period, an agent earns an income of $y_t^i = w_t^i \ell_t^i$, where ℓ_t^i is the labor supply.

For tractability reasons, we focus on a dynamic extension of the linear (progressive) taxation framework (Sheshinski, 1972). This setting allows us to study the welfare gains from delayed taxation in the presence of distributional considerations. Later, we will also consider delayed taxation in the context of a nonlinear tax schedule as in Heathcote et al. (2020). The structure of the tax and transfer schedules are given by

$$
\begin{aligned}
T_1(y_1) &= -G_1 + \delta\tau_1 y_1, \\
T_2(y_1, y_2) &= -G_2 + \tau_2 y_2 + (1+r)(1-\delta)\tau_1 y_1,
\end{aligned}
$$

where $\tau_t \geq 0$ denotes the nominal (statutory) marginal tax rate in period t and $G_t \geq 0$ denotes the lump-sum transfer.[5] The parameter $\delta \in [0, 1]$ determines the timing of tax payments: a fraction δ of period-2 taxes must be paid in period 1, while the remaining fraction $1 - \delta$ carries an interest rate of r and is paid in period 2.

Consumption c_t^i is given by the per-period budget constraints,

$$
c_1^i = w_1^i \ell_1 [1 - \delta\tau_1] + G_1 - s^i + x^i, \tag{2}
$$

$$
c_2^i = w_2^i \ell_2 [1 - \tau_2] + G_2 - (1+r)[1-\delta]\tau_1 w_1^i \ell_1^i + R(s^i) - (1+r)x^i, \tag{3}
$$

where $R(s)$ defines the saving technology. $R(s)$ is the amount by which disposable income in the second period increases if the individual saves the amount s. $R(s)$ plays a key role in our analysis as it captures the extent of financial frictions, materializing as a higher marginal interest rate, $R'(s)$ on borrowing (i.e., $s < 0$) relative to saving ($s \geq 0$).[6] We discuss $R(s)$ and the financial friction in more detail below. Finally, x^i is the amount of subsidized (low-interest) loans that individual i receives in period 1. This loan is subject to an interest rate of r.

Individual preferences are represented by the utility function:

$$
u_1(c_1^i) - v(\ell_1^i) + \beta[u_2(c_2^i) - v(\ell_2^i)], \tag{4}
$$

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

Savings technology and the financial friction. We assume that net savings, $s \geq 0$, is subject to a marginal interest rate of r. Borrowing ($s < 0$) is subject to a higher marginal interest rate: $R'(s) > 1 + r$. There are two possible micro-foundations for this credit penalty. One foundation is that individuals may save freely in international bonds but can only borrow from private lenders. These private lenders have market power and thus charge markups. The other foundation, which we focus on, is one in which borrowers can choose to default but lenders may take costly upfront action to deter defaults.

We build this microfoundation in Appendix H. Private lenders must pay an upfront fee

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

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

proportional to the loan amount to be able to seize borrowers' income in the event of default. Anticipating this, defaulting workers do not supply any labor. Since there is no income to seize, lenders only lend when they can ensure no default. Since we make the restrictive assumption that lenders may not seize the government transfer, G_2,[7] lenders must implement loan limits to ensure no defaults. That is, the utility of not defaulting must exceed utility from taking out the maximal loan size, defaulting, and then consuming only G_2 in period 2. We quantitatively verify that these loan limits do not bind in our calibrations and thus ignore them in our analyses that assume a no-default equilibrium. We further assume that lending markets are competitive, which generates a break-even condition for lenders that determine the interest rate on borrowing, r_b.

$$(1+r)(1+z) = 1 + r_b, \tag{5}$$

where r is the savings interest rate and z is the upfront lending cost. The left hand side is the (period-2) opportunity cost of lending \$1. The right hand side is the repayment. Since no borrowers default in equilibrium, the borrowing interest rate depends only on the cost of avoiding default (z) and the interest rate on savings (r).

For our main theoretical results, we assume that z is a constant. This implies a piece-wise linear savings technology, $R(s)$, given by:

$$R(s, r_b) = \begin{cases} (1+r)s & \text{if } s \geq 0, \\ (1+r_b)s & \text{if } s < 0, \end{cases} \tag{6}$$

where r_b is the interest rate borrowers face in private credit markets (determined by equation 5), and thus $r_b - r > 0$ reflects the credit penalty faced by borrowers. Note that (6) can be seen as a generalization of the standard "no-borrowing constraint" in macroeconomics (obtained when $r_b \to \infty$). Before introducing the government's optimization problem, we discuss how the government similarly faces upfront lending costs that affect the fiscal costs of delayed taxation or subsidized lending policies.

The individual's problem. The problem solved by an individual i with a lifetime wage profile of (w_1^i, w_2^i) is to choose ℓ_1^i, ℓ_2^i, s^i in order to maximize (4) subject to constraints (2) and (3). The first-order conditions are:

$$(\ell_1) : u_1'(c_1^i)w_1^i[1 - \delta\tau_1] - \beta(1+r)u_2'(c_2^i)[1-\delta]\tau_1 w_1^i = v'(\ell_1^i), \tag{7}$$

$$(\ell_2) : u_2'(c_2^i)w_2^i[1 - \tau_2] = v'(\ell_2), \tag{8}$$

$$(s) : u_1'(c_1^i) = \beta u_2'(c_2^i)R'(s^i). \tag{9}$$

The first order condition for ℓ_1 (equation 7) can be written:

$$u_1'(c_1^i)w_1^i\left(1 - \tau_1\Big[\delta + [1-\delta]\theta\Big]\right) = v'(\ell_1^i) \quad \text{where} \quad \theta^i = \frac{\beta(1+r)u'(c_2^i)}{u'(c_1^i)}. \tag{10}$$

[7] That is, the government does not allow lenders to cause "starvation" in the event of default.

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

As can be seen from the simplified FOC for ℓ_1 (10), the effects of delayed taxation depend on the degree of consumption smoothing. Assuming an interior solution for s^i such that the Euler equation (9) holds, we may substitute (9) into (10) to obtain

$$u_1'(c_1^i)w_1^i\left(1-\tilde{\tau}_1^i\right) = v'(\ell_1^i), \tag{11}$$

where

$$\tilde{\tau}_1^i = \tau_1\left[1-(1-\delta)\Delta_r^i\right] \tag{12}$$

is the effective period-1 tax rate and

$$\Delta_r^i = \frac{R'(s^i)-(1+r)}{R'(s^i)} = 1 - \frac{1+r}{R'(s^i)} \tag{13}$$

is the interest wedge faced by net borrowers, representing the proportional difference between their marginal interest rate $R'(s^i)$ and the interest rate on savings $1+r$.

The effective tax rate equals the nominal tax rate if there is no delayed taxation ($1-\delta = 0$) or no financial frictions ($\Delta_r^i = 0$). Equation (11) illustrates that delaying taxes (increasing $1-\delta$) reduces the distortion on period 1 labor supply at a rate determined by the difference in interest rates between government and private agents, Δ_r^i, which captures the "strength" of the financial frictions faced by agent i. This yields two testable implications that we examine in Section 5. First, when agents face financial frictions, their behavioral responses to delayed taxes are smaller compared to regular taxes. Second, the magnitude of these behavioral responses decreases with the severity of an individual's borrowing constraints.

Note that period-1 and period-2 labor supply can be related by substituting (9) into (11) and then substituting in (8):

$$\beta R'(s^i)\frac{\tilde{w}_1^i}{\tilde{w}_2^i} = \frac{v'(\ell_1^i)}{v'(\ell_2^i)}. \tag{14}$$

where $\tilde{w}_1^i = w_1^i\left(1-\tilde{\tau}_1^i\right)$ and $\tilde{w}_2^i = w_2^i(1-\tau_2)$ are the after-tax wages (taking into account the effective tax rate $\tilde{\tau}_1^i$ in period 1). This condition relates labor supply across periods through the discounted after-tax wage ratio, making clear how taxes and borrowing frictions (via $R'(s^i)$)

[8] For the second period, the first-order condition for ℓ_2 (equation 8) is the same as in a standard model without delayed taxation. Second-period labor supply is only affected indirectly through c_2.

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

To characterize the relationship between delayed taxation and regular taxation, it is useful to define a conversion factor that measures how much the effective tax rate responds to changes in the delay parameter δ relative to how much it responds to changes in the nominal tax rate τ_1.

Definition 1 *Define the δ-to-τ_1 conversion factor*

$$\Gamma^i(\delta) \equiv \tau_1 \frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}. \tag{15}$$

By construction, $\Gamma^i(\delta) = \frac{\partial \tilde{\tau}_1^i/\partial \delta}{\partial \tilde{\tau}_1^i/\partial \tau_1}$, representing the relative marginal influence of δ versus τ_1 on the effective tax rate $\tilde{\tau}_1^i$.

Government lending costs. Like private lenders, the government faces an upfront lending cost in order to avoid default, but we allow this cost to differ from that of private lenders. While private lenders always wish to avoid default, this is not necessarily the case for the government. Letting all borrowers default on their liabilities effectively boils down to age-dependent taxation in our model.[9] Hence, we make the assumption that the government always wishes to avoid defaults. Specifically, let z^g denote the government's upfront lending cost, which determines its break-even interest rate, r_b^g, through

$$(1+r)(1+z^g) = 1 + r_b^g. \tag{16}$$

While the government does not need to break even, the interest rate, r_b^g, determines the cost of capital associated with lending. Since the government offers a subsidized interest rate, $r < r_b^g$, it incurs a marginal present-value cost for each loan:

$$\Delta_{r,g}^i = \mathbb{1}[s^i < 0] \left(1 - \frac{1+r}{1+r_b^g}\right) \geq 0. \tag{17}$$

This cost is zero for net savers ($s^i \geq 0$) and equals $1 - \frac{1+r}{1+r_b^g}$ for net borrowers ($s^i < 0$). In our baseline analysis, we assume the government faces the same lending costs as private lenders ($z^g = z$), implying equal borrowing interest rates ($r_b^g = r_b$) and marginal costs ($\Delta_{r,g}^i = \Delta_r^i$). However, the government could have advantages in lending ($z^g < z$) due to superior enforcement capabilities or monitoring technology, resulting in a lower borrowing interest rate ($r_b^g < r_b$) and enhancing the appeal of delayed taxation. In the limiting case where the government can costlessly enforce no defaults and thus discounts debt repayments at $1+r$, then $\Delta_{r,g}^i = 0$, and neither delayed taxation nor government lending affect the budget constraint.

To keep the analysis tractable and isolate the core mechanism, we adopt the following assumption:

[9] When borrowers can default on delayed taxes without penalty, the government is effectively lowering τ_1. When borrowers can default on (uniform) loans, the government is effectively just increasing G_1.

Assumption 1 (Credit Markets) *The savings interest rate r and lending costs z and z^g are exogenous, and lending does not generate welfare-relevant profits.*

Thus, similar to Lozachmeur (2006) and Weinzierl (2011), we treat financial frictions and interest rates as exogenous and assume that financial intermediation is irrelevant for welfare.[10] Consequently, when policy reduces borrowing in credit markets, there are no welfare-relevant effects on lenders' profits. There are also no equilibrium effects due to market-clearing in the savings asset. One justification exogenous interest rates is that saving and borrowing ultimately occurs in international bonds, and the net effect of delayed taxation on the demand for bonds is limited as reduced private borrowing is offset by increased government borrowing.

The government's problem. The government's problem is to maximize social welfare, which we express as a utilitarian sum over individual utilities (though our results extend to more general social welfare functions with different welfare weights). Let π^i denote the population share of agents of type i. Then the government solves:

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

subject to:

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

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

where M is an exogenous revenue requirement that is not refunded to agents, x is a uniformly-sized loan provided to all agents,[11] and $\Delta_{r,g}^i x$ is the present-value cost (per agent) of providing a uniform loan of $x \geq 0$. The $(1-\delta)\Delta_{i,g}^r \tau_1 w_1^i \ell_1^i$ term is a present-value adjustment that accounts that a fraction, $1-\delta$ of period-1 taxes are paid in the next period. If the government has no lending advantage, $\Delta_{r,g}^i = \Delta_r^i$, and the left-hand-side of (19) can be written:

$$\sum_i \pi^i \left(\tilde{\tau}_1^i w_1^i \ell_1^i + \frac{\tau_2 w_2^i \ell_2^i}{1+r} - \Delta_r^i x \right). \tag{21}$$

A few remarks are in order. Dynamic models of optimal taxation often assume that the budget constraint is satisfied in expectation (as in equation 19). This allows for unrestricted borrowing if, for example, the government runs a deficit in period 1. In reality, governments are usually subject to borrowing limits. In our numerical simulations, we extend the standard life-cycle framework by explicitly incorporating borrowing constraints (see section 3.5). Although

[10] Weinzierl (2011) considers two main cases. In the first case, there is (exogenously) neither borrowing nor saving. In the second case, where welfare gains are considerably lower, individuals may save or borrow at an (exogenous) interest rate of 5%. Lozachmeur (2006) imposes a no-borrowing constraint but lets individuals save at an interest rate of zero.

[11] Whether all agents receive the loan or only the agents who face high marginal interest rates is inconsequential for most of our analyses. Net savers, with $R'(s^i) = 1 + r$ are indifferent regarding whether to receive the loan, and the loan, for these agents, imposes no present-value costs on the government.

we do not adopt an OLG model, these borrowing constraints mimic some of the fiscal discipline found in OLG frameworks, where borrowing and spending must be carefully managed over time due to concerns about intergenerational equity. Nevertheless, our focus on delayed taxation is primarily concerned with redistribution within cohorts rather than across generations.

2.2 Characterizing optimal tax policy

Now we want to characterize the optimal solution to the government's problem outlined in (18). For this purpose, let λ denote the Lagrange multiplier associated with the government budget constraint (19). The Lagrangian is

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

It is useful to define the following auxiliary variables:

$$g_1^i = \frac{u_1'(\cdot)}{\lambda}, \tag{22}$$

$$g_2^i = \beta(1+r)\frac{u_2'(\cdot)}{\lambda}, \tag{23}$$

$$\varepsilon_{ts}^i = \frac{1-\tau_s}{y_t^i}\frac{dy_t^i}{d(1-\tau_s)}, \tag{24}$$

$$\varepsilon_t^i = \frac{1-\tau}{y_t^i}\frac{dy_t^i}{d(1-\tau)}, \tag{25}$$

$$\rho^i = \frac{d}{dG_1}\left(\tau_1 y_1^i \left[1-(1-\delta)\Delta_{r,g}^i\right] + \frac{\tau_2 y_2^i}{1+r} \right) \leq 0, \tag{26}$$

where (22)–(23) defines the social value of giving an additional dollar to an agent of type i in period $s = 1, 2$ (in money metric terms) and (24) is the elasticity of period t income with respect to the period s net-of-tax rate. This elasticity captures labor supply adjustments to within-period tax changes as well as across-period tax changes (intertemporal labor substitution effects). Equation (25) is the elasticity of period t income with respect to a change in $1 - \tau$ (a change in the net-of-tax rate in both periods). Equation (26) defines an income effect parameter that represents the reduction in present value taxes caused by an increase in lump-sum transfers or unearned income in period 1, adjusted for the government's discount factor when delayed taxation is present ($\delta < 1$) and the government faces lending costs ($\Delta_{r,g}^i > 0$).

We start by characterizing the regime that we refer to as our benchmark, which is a linear tax scheme where taxes and transfers do not depend on age and the amount of delayed taxation $(1-\delta)$ is exogenous (and possibly zero).[12]

Proposition 1 (Benchmark Linear Tax Scheme) *Consider the government's optimization program (18) with $\tau_1 = \tau_2 = \tau$ and $G_1 = G_2 = G$, a fixed amount of delayed taxation $1-\delta$, and no subsidized lending ($x = 0$).*

[12] A corresponding characterization of the optimal age-dependent linear tax system is provided in appendix B.

(i) The optimal marginal tax rate τ satisfies

$$\sum_i \pi_i y_1^i \left[\delta g_1^i + (1-\delta) g_2^i\right] + \frac{1}{1+r} \sum_i \pi_i y_2^i \left[g_2^i\right] \tag{27}$$

$$= \sum_i \pi_i y_1^i \left[1 + \frac{\tau}{1-\tau}\varepsilon_1^i\right] + \frac{1}{1+r} \sum_i \pi_i y_2^i \left[1 + \frac{\tau}{1-\tau}\varepsilon_2^i\right]. \tag{28}$$

(ii) The optimal per-period transfer G satisfies

$$\sum_i \pi^i \Big(g_1^i + \frac{1}{1+r} g_2^i \Big) = \sum_i \pi^i \left(1 + \frac{1}{1+r} - \left[1 + \frac{1}{R'(s^i)}\right] \rho^i \right). \tag{29}$$

Proof. See Appendix A.1. ■

The formulations in equations (27) and (29) reflect a dynamic extension of the seminal work on optimal linear (affine) taxation by Atkinson and Stiglitz (1980), among others. The determination of the optimal linear tax rate involves balancing equity, as reflected on the left-hand side of the equation (27), with efficiency, as reflected on the right-hand side of the same equation. Similarly, the optimal level of the transfer depends on a balance between equity, shown on the LHS of equation (29), and costs, shown on the RHS of equation (29). These costs are the direct costs of providing the transfer, adjusted for the resulting negative impact on the tax base due to income effects. In Marginal Value of Public Funds (MVPF) language, these optimality conditions state that the marginal value of public funds equals one for both tax and transfer policies, i.e., $MVPF_\tau = MVPF_G = 1$.

Next, we 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 *Assume the individual's marginal borrowing rate is locally constant at the baseline, namely, for each i, $R'(s)$ is well-defined and constant for marginal changes in incentives around s^i. Then, for $t \in \{1, 2\}$,*

$$\frac{d\ell_t^i}{d(1-\delta)} = \Gamma^i(\delta) \frac{d\ell_t^i}{d(1-\tau_1)}, \tag{30}$$

where $\Gamma^i(\delta)$ is the conversion factor defined in (15).

Proof. See Appendix A.2. ■

The result follows from the fact that δ and τ_1 enter the period-1 intratemporal condition only through the same effective wedge, $\tilde{\tau}_1^i = \tau_1 \left[1 - (1-\delta)\Delta_r^i\right]$. Under a locally constant marginal borrowing rate, a marginal change in $1-\delta$ therefore moves labor supply exactly like a marginal change in $1-\tau_1$, scaled by the conversion factor $\Gamma^i(\delta)$. [13]

[13] The local-constancy assumption on $R'(s)$ is mild in the sense that it requires only that marginal policy changes do not move an individual across a kink or into a region where the marginal rate varies; it is satisfied by piecewise-linear schedules away from kinks and, more generally, by any schedule that is locally affine at s^i. If $R''(s^i) \neq 0$ at the margin, a small additional term arises because $d(1-\delta)$ perturbs s^i and hence $R'(s^i)$; in that case (30) remains a first-order approximation evaluated at the baseline.

The conversion factor $\Gamma^i(\delta)$ has a clear economic interpretation as a policy-equivalence measure: locally, a 1 percentage point increase in $1-\delta$ has the same effect on labor supply as a $\Gamma^i(\delta)$ percentage point increase in $1-\tau_1$. This factor is increasing in the severity of the borrowing friction, Δ_r^i, equals zero for savers ($\Delta_r^i = 0$), and simplifies to $\tau_1\Delta_r^i$ when there is no delay ($\delta = 1$). For $\delta < 1$, the denominator $1-(1-\delta)\Delta_r^i$ amplifies the effect for borrowers.

Proposition 2 characterizes the optimal amount of delayed taxation $1-\delta$. To better convey the effects of delayed taxation, we express this proposition in terms of compensated tax elasticities (indicated by the superscript c).

Proposition 2 (Optimal Delayed Taxation) *Consider the optimization program (18), given some fixed values of τ_1, τ_2, G_1, G_2 (not necessarily optimal). Assume the local-constancy condition used in Lemma 1 holds. Assuming an interior solution for δ, the optimal share of delayed taxation, $1-\delta$, satisfies:*

$$\tau_1 \cdot \sum_i \pi_i y_1^i \left(g_1^i - g_2^i\right) = \underbrace{-\sum_i \pi_i \Delta_{r,g}^i \tau_1 y_1^i}_{\textit{Direct fiscal cost}}$$
$$\underbrace{-\sum_i \pi_i y_1^i \, \Gamma^i(\delta) \left(\frac{\tau_1 \left\{1-(1-\delta)\Delta_{r,g}^i\right\}}{1-\tau_1} \varepsilon_{11}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{21}^{i,c} + \rho^i \right)}_{\textit{Behavioral effects scaled by conversion factor}}. \quad (31)$$

Proof. See Appendix A.3. ■

The LHS is the welfare effect of increased consumption smoothing and the RHS captures the fiscal externalities of marginally delayed taxation. The first term on the RHS, $-\sum_i \pi_i \Delta_{r,g}^i \tau_1 y_1^i$, represents the direct present-value cost to the government of delaying tax collection when the government faces lending costs. The second term accounts for the behavioral effects of delayed taxation, which contains three components: (i) the period-1 substitution component: the compensated elasticity $\varepsilon_{11}^{i,c}$ scaled by the revenue-relevant effective marginal tax rate $\tau_1[1-(1-\delta)\Delta_{r,g}^i]$ and normalized by the untaxed share $(1-\tau_1)$; (ii) the period-2 labor supply response through intertemporal substitution effects $\varepsilon_{21}^{i,c}$; and (iii) the overall income effect on present-value tax revenues ρ^i. If there were no financial frictions, that is, $\Delta_r^i = \Delta_{r,g}^i = 0$ for all i, then both the left and right sides of (31) would equal zero. In other words, if no one is financially constrained, it does not matter whether the government delays taxation. There are no consumption-smoothing benefits, since agents can already borrow freely at the government rate, and on the right-hand side there are no fiscal externalities, since the present value of the delayed tax from the perspective of both agents and government 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 (31) is positive and equal to the marginal welfare gains from improved consumption smoothing. The right-hand side is also nonzero due to fiscal externalities caused by constrained agents who now face a lower effective tax rate in period 1.

Note the first order condition (31) fails to emphasize a tension between the two welfare gains from delayed taxation that we emphasized in the introduction. As more and more taxes are delayed, private borrowing is crowded out. Once some agents seize to borrow, their marginal interest rates equal that on savings, and thus $\Delta_r^i = 0$. This implies that when delayed taxation achieves maximal welfare gains from improving welfare gains from improving intertemporal consumption smoothing, delayed taxation no longer reduces the distortions of the income tax. Note that this may require delaying more than 100% of taxes. However, under moderate implementations of delayed taxation (which are indeed optimal in our quantitative framework), individuals still face high interest rates on the margin, implying a that key channel through which delayed taxation generates welfare is by reducing income tax distortions.

2.3 Welfare effects of marginal reforms

An interesting question is under what conditions the introduction of delayed taxation increases welfare in an economy without delayed taxation. We start with our benchmark economy, which is characterized by a tax system with $\tau_1 = \tau_2 = \tau$ and no delayed taxation $1 - \delta = 0$. We then consider a marginal delay in taxation (i.e., a marginal increase in $1 - \delta$). Note that this reform has no mechanical cost to the government.

By evaluating Proposition 2 at $\tau_1 = \tau_2 = \tau$ and setting $\delta = 1$, and noting that welfare effects only arise from borrowers (since $\Delta_r^i = 0$ for savers with $s^i > 0$), we obtain the marginal welfare effects of delayed taxation. Making the additional simplifying assumption that the return function $R(s)$ is piecewise linear around zero savings ($s = 0$) and that no agent has exactly zero savings ($s^i \neq 0$ for all i), we can formalize these insights in the following corollary.

Corollary 1 (Marginal Welfare Effects of Delayed Taxation) *Assume that $R(s)$ is piecewise linear around $s = 0$, $s^i \neq 0$ for all i, and consider the benchmark economy without delayed taxation ($\delta = 1$). The money-metric welfare effect of a marginal introduction of delayed taxation can be decomposed as:*

$$
\begin{aligned}
\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\bigg|_{\delta=1} = \tau \sum_{i:s^i<0} \pi^i \Bigg[& \underbrace{\Delta_r^i y_1^i \cdot \frac{1}{\lambda} \cdot u'(c_1^i)}_{\substack{\textit{Welfare gains}\\ \textit{from consumption}\\ \textit{smoothing}}} + \underbrace{\Delta_r^i \frac{1}{1-\tau} y_1^i \varepsilon_{11}^{i,c}}_{\substack{\textit{Period-1 tax revenue}\\ \textit{gains (substitution)}}} \\
& + \underbrace{\frac{1}{1+r}\Delta_r^i \frac{1}{1-\tau} y_2^i \varepsilon_{21}^{i,c}}_{\substack{\textit{Period-2 tax revenue}\\ \textit{losses (substitution)}}} + \underbrace{\Delta_r^i y_1^i \rho^i}_{\substack{\textit{Present-value tax}\\ \textit{revenue effects (income)}}} - \underbrace{y_1^i\, \Delta_{r,g}^i}_{\substack{\textit{Direct fiscal cost of}\\ \textit{government lending}}} \Bigg],
\end{aligned} \tag{32}
$$

where $\Delta_r = 1 - \frac{1+r}{1+r_b} > 0$ represents the private borrowing friction, $\Delta_{r,g}^i \geq 0$ represents the government's lending costs, and all welfare effects originate from net borrowers since $\Delta_r^i = 0$ for agents with $s^i > 0$. Note that the substitution effects are scaled by $\tau \cdot \Delta_r^i \frac{1}{1-\tau} = \frac{\Gamma^i(\delta=1)}{1-\tau}$, while the income effect is scaled by $\tau \cdot \Delta_r^i = \Gamma^i(\delta = 1)$.

Proof. See Appendix A.4. ∎

Corollary 1 provides a decomposition revealing five distinct channels through which delayed taxation affects welfare. The first channel represents *direct welfare gains* from improved consumption smoothing, as delayed taxation allows financially constrained agents to effectively borrow from the government at favorable rates. The second channel captures *positive fiscal externalities* from increased period-1 tax revenues, as the lower effective tax rate encourages higher labor supply through substitution effects. The third channel shows *intertemporal substitution effects*, which partially offset the period-1 gains by reducing period-2 labor supply and associated tax revenues. The fourth channel reflects *income effects* on present-value tax revenues, which tends to be negative as higher effective income reduces labor supply. The fifth channel represents the *direct present-value cost to the government of delaying tax collection*, reflecting the government's lending costs ($\Delta^i_{r,g} \geq 0$). This cost is zero when the government can lend at the risk-free rate and positive when the government faces some lending costs, though these are typically lower than private borrowing rates. To better understand the economic mechanisms at work, it is instructive to compare delayed taxation with an alternative policy instrument that also provides liquidity to constrained agents: direct government lending.

Lemma 2 (Marginal Welfare Effects of Uniform Government Lending) *Under the same assumptions as Corollary 1, the money-metric welfare effect of the government offering a marginal loan, $dx > 0$, at an interest rate of r is:*

$$\frac{1}{\lambda}\frac{dW}{dx}\bigg|_{\delta=1} = \sum_i \pi^i \left(g^i_1 - g^i_2\right) + \sum_i \pi^i \Delta^i_r \rho^i - \sum_i \pi^i \Delta^i_{r,g} = \Delta_r \sum_{i:s^i<0} \pi^i (g^i_1 + \rho^i) - \Delta_{r,g} \sum_{s^i<0} \pi^i, \quad (33)$$

where the second equality follows from the fact that only borrowers ($s^i < 0$) benefit from subsidized loans, and for these agents, the intertemporal optimality condition implies $(g^i_1 - g^i_2) = \Delta^i_r g^i_1$.

Proof. See Appendix A.5. ■

With the marginal welfare effects of both policies now characterized in Corollary 1 and Lemma 2, we can establish Proposition 3 which shows how the welfare gains from delayed taxation can be decomposed into two components: (i) a uniform lending effect that provides the same liquidity benefits as direct government loans, plus (ii) additional welfare gains arising from delayed taxation's inherent targeting mechanism.

Proposition 3 (Decomposing Delayed Taxation into Lending and Tax Effects) *Assume that $R(s)$ is piecewise linear around $s = 0$, $s^i \neq 0$ for all i, and consider the benchmark economy with no delayed taxation. Then, the marginal welfare effect of introducing delayed taxation can be written as:*

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\bigg|_{\delta=1} = \underbrace{\tau\bar{y}_1 \cdot \frac{1}{\lambda}\frac{dW}{dx}\bigg|_{\delta=1}}_{\substack{\text{Lending}\\ \text{effect}}} + \tau\Delta_r \Bigg[\underbrace{\sum_{i:s^i<0} \pi^i \frac{1}{1-\tau}\left(y^i_1 \varepsilon^{i,c}_{11} + \frac{1}{1+r} y^i_2 \varepsilon^{i,c}_{21}\right)}_{\substack{\text{Reduced tax}\\ \text{distortions}}} \quad (34)$$

$$+ \underbrace{\sum_{i:s^i<0} \pi^i (y^i_1 - \bar{y}_1) g^i_1}_{\substack{\text{Loan}\\ \text{targeting}}} + \underbrace{\sum_{i:s^i<0} \pi^i (y^i_1 - \bar{y}_1) \rho^i}_{\substack{\text{Income effect}\\ \text{targeting}}} \Bigg] - \underbrace{\tau \sum_{i:s^i<0} \pi^i (y^i_1 - \bar{y}_1) \Delta^i_{r,g}}_{\substack{\text{Differential fiscal}\\ \text{cost targeting}}}$$

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

Proof. See Appendix A.6. ■

Proposition 3 provides a decomposition revealing how delayed taxation can be understood as uniform government lending plus four additional effects. The decomposition isolates the *lending effect*, which captures the welfare gains from providing uniform subsidized loans equal to $\tau\bar{y}_1$ (the total tax burden of borrowers). This represents the pure consumption smoothing benefit that any government lending program would provide. Beyond this baseline lending effect, delayed taxation generates four additional channels. The first additional channel captures *efficiency gains from reduced tax distortions*, as delayed taxation effectively lowers the period-1 tax rate for borrowers, generating positive substitution effects that increase labor supply and tax revenues. This channel highlights a key advantage of delayed taxation over pure lending: it simultaneously provides liquidity and reduces tax wedges. The second additional channel reflects *loan targeting effects*, representing the covariance between income and the marginal welfare value of period-1 consumption. For constrained borrowers, this covariance is typically negative, reflecting that higher-income agents have lower marginal utilities of consumption, making uniform lending somewhat better targeted than the income-proportional transfers implicit in delayed taxation. As we discuss, this may be offset by lower fiscal costs of lending to high-income agents. The third additional channel accounts for *income effect targeting*, capturing how income effects vary across the income distribution. Since delayed taxation provides transfers proportional to period-1 income while uniform lending does not, this term reflects the covariance between income and income effect parameters.

The fourth and last channel represents *differential fiscal costs* between delayed taxation and uniform lending when the government faces lending costs ($\Delta_{r,g} > 0$). This term captures how the fiscal costs of income-proportional transfers (delayed taxation) differ from uniform transfers (lending). Under our baseline assumption that $\Delta^i_{r,g} = \Delta^i_r$ (government has same technology as agents), and assuming these are constant across borrowers ($\Delta^i_r = \Delta_r$), this term is zero.[14] We include this term to highlight that a feature that may differentiate delayed taxation from lending in a more comprehensive model. If $\Delta^i_{r,g}$ and y^i_1 are negatively correlated, then delayed taxation will involve lower default costs than uniform lending. The intuition is that those with higher period-1 incomes may be easier to collect from for two reasons. The first reason is that high-income borrowers may borrow less and thus be less risky (materializing as high-income borrowers facing lower marginal interest rates). The second reason is that high-income borrowers may otherwise be easier to collect from, due to, e.g., lower risks of default.

Additional discussion. Before proceeding to the quantitative analysis, we clarify a key assumption in our theoretical framework regarding the interest rate on delayed taxes. We set $r_{dtax} = r$, where r is both the government's borrowing rate and the return on savings for unconstrained agents. This assumption has several important implications that merit discussion.

Behavioral neutrality for unconstrained agents. For workers whose marginal borrowing rate

[14]When $\Delta^i_{r,g} = \Delta^i_r = \Delta_r$ for $s_i < 0$, we can factor it out: $\Delta_r \sum_{i:s^i<0} \pi^i (y^i_1 - \bar{y}_1) = 0$ by definition of $\bar{y}_1$.

equals r, delayed taxation does not affect behavior since it leaves the present value of their tax burden unchanged. This ensures that delayed taxation selectively benefits only those facing financial constraints, making it a precisely targeted policy tool.

Fiscal costs depend on government lending advantage. When the government faces the same lending costs as private markets ($\Delta^i_{r,g} = \Delta^i_r$), delaying tax collection imposes fiscal costs equal to the interest wedge. However, if the government has a lending advantage ($\Delta^i_{r,g} < \Delta^i_r$) or can lend costlessly ($\Delta^i_{r,g} = 0$), the fiscal costs are reduced or eliminated entirely.

Robustness to frictionless environments. In economies without financial frictions (where $r_b = r$ for all agents), delayed taxation has no effects on either behavior or welfare. This confirms that the policy's benefits arise specifically from addressing financial market imperfections.

While one might consider more flexible designs where r_{dtax} deviates from r, such approaches would fundamentally alter the policy's character. As we discuss in Section 2.4, if the government could freely choose any $r_{dtax} \in \mathbb{R}$, it could replicate any age-dependent tax system by appropriate choice of parameters.[15] However, such flexibility would sacrifice delayed taxation's key advantage: its ability to automatically target transfers based on individuals' revealed financial constraints. Our approach preserves this targeting mechanism while maintaining fiscal discipline.

2.4 Model Extensions

Myopic agents. Appendix C analyzes the marginal welfare effect of introducing delayed taxation when a subset of agents are more impatient than the government. We assume these agents share a common discount factor $\beta < \beta^*$, where β^* is the government's discount factor used in welfare evaluation. Relative to the non-paternalistic case, equation (C5) differs in exactly two ways. First, borrowers' experienced consumption-smoothing gain is unchanged, but its social contribution is reweighted under paternalism and can turn negative when $\frac{\beta^*}{\beta} > \frac{R'(s^i)}{1+r}$. Second, when the government values future utility more than agents do, period-2 labor choices are no longer at the social optimum, creating an additional positive welfare term. Since $\varepsilon^i_{12} < 0$, a marginal delay reduces period-2 labor supply. Not only does the paternalistic government want myopic agents to consume more in period 2, the government would also like them to work less, which delayed taxation achieves. In short, paternalism lowers the social value of smoothing for impatient agents but introduces an offsetting leisure gain; other behavioral-revenue and fiscal terms are unchanged.

Making the interest rate on delayed taxation a policy tool. In our baseline analysis, we constrain the interest rate on delayed taxes to equal the risk-free rate, $r_{dtax} = r$. This constraint has important implications for the policy's targeting mechanism: delayed taxation has no effect on the behavior of workers whose marginal return on savings equals r (i.e., whose $R'(s^i) = 1 + r$), since it does not change the present value of their tax burden. This implies that delayed taxation does not affect the behavior of net savers in partial equilibrium, nor does it have any effect in the

[15] For instance, setting $\tau = \tau_2$, $r_{dtax} = -1$, and $\delta = \tau_1/\tau_2$ would emulate the system (τ_1, τ_2). If $\tau_2 = 0$, the parameter δ becomes undefined, but this is inconsequential since delayed taxation becomes irrelevant when $\tau = \tau_2 = 0$.

absence of financial frictions (when $r_b = r$).

A natural extension of our framework is to treat the interest rate on delayed taxes as a policy instrument, allowing r_{dtax} to differ from r. The key benefit of this extension is that it expands the set of feasible policies. In our baseline framework, the constraints $\delta \in [0, 1]$ and $r_{dtax} = r$ limit which age-dependent tax systems can be replicated. However, with full flexibility over $r_{dtax} \in \mathbb{R}$, the government can replicate any age-dependent marginal tax system characterized by $(\tau_1, \tau_2) \in \mathbb{R}^2_+$ by setting $\tau = \tau_2$, $r_{dtax} = -1$, and $\delta = \tau_1/\tau_2$.[16] This allows replication even when $\tau_1 > \tau_2$, which is impossible under the baseline constraints.

Beyond this replication result, delayed taxation with flexible interest rates can potentially Pareto dominate age-dependent taxation because it preserves the automatic targeting feature: benefits are allocated based on agents' revealed marginal borrowing costs rather than applying uniform rates regardless of individual financial constraints.

Delayed taxation in a three-period model. Note that the model can be extended to three periods, where agents work during periods 1 and 2, and are retired in period 3. Since retired workers do not supply labor, delayed taxation avoids the post-mortem default problem that would otherwise plague unconditional subsidized loans. Suppose taxes can only be delayed for one period. Since retired workers have no labor income and thus cannot borrow through the delayed tax system, they cannot default on their liabilities post-mortem. However, a new challenge arises: agents may optimally choose to delay period-2 taxes and then default once retired in period 3. This challenge also affects private lending markets, but may be resolved by conditioning credit supply on age. If we restrict delayed tax policy to be *age-independent*, preventing default in period 3 requires that workers willingly save for retirement, since these constitute the entirety of seizable disposable income (assuming the government-transfer, G, is not seizable). Workers will only choose to save if the benefits of repaying and accessing these savings in period 3 outweigh the benefits of defaulting in period 2. This trade-off imposes an upper bound on the amount of delayed taxes. if $1 - \delta$ is too large, default becomes optimal, and anticipating default, workers do not save in period 2.[17] Under the assumption that delayed taxation is age-independent, discouraging period-3 default thus limits the welfare gains (by limiting $1 - \delta$) that can be achieved to smooth consumption between periods 1 and 2. This may be partially offset by the government choosing a lower unconditional transfer, G, which would reduce the amount of consumption in default. We also note that while private lenders would need to discourage default (see Appendix H) to break even on loans, it may be optimal for the government to allow some workers to default. Strategic default is essentially a transfer from workers to the government, which limits welfare losses from default. We conjecture that there may be environments in which these transfers produce welfare gains by targeting those with low retirement-saving incentives (due to, e.g., low period-2 earnings).

[16] If $\tau_2 = 0$, δ is not well defined, but this is irrelevant since δ becomes immaterial when $\tau = \tau_2 = 0$.

[17] For intuition, we can consider the case when $1 - \delta \to \infty$. In this case, defaulting workers can borrow and consume an infinite amount in period 2 and are limited to consuming the government transfer, G, in period 3. Nevertheless, the life-time utility of default tends to infinity (barring, e.g., Leontief preferences over consumption). Hence, workers will always default.

Nonlinear income taxation. Most modern tax systems feature a progressive marginal tax schedule that cause average tax rates to be lower for lower-income younger workers. This leaves the question of whether delayed taxation still has material welfare gains on top of a progressive marginal tax schedule. We investigate this by extending our model such that the benchmark economy features a nonlinear marginal tax schedule as in Heathcote et al. (2020):

$$T_t(y_t) = y_t - a_t y_t^{1-k_t} + G_t, \tag{35}$$

where a_t, k_t, and G_t are policy parameters that determine the progressivity of the tax schedule. Note that unlike Heathcote et al. (2020), we include the lump sum transfer, $G_t \geq 0$ in the tax and transfer schedule to stay consistent with our previous analyses.

3 A quantitative investigation of delayed taxation

3.1 Calibration

We assume that the utility function has the form

$$u(c) - v(\ell) = \frac{c^{1-\sigma} - 1}{1-\sigma} - \xi \frac{\ell^{1+\frac{1}{k}}}{1+\frac{1}{k}}, \tag{36}$$

which implies that σ is the inverse of the elasticity of intertemporal substitution and k is the constant consumption elasticity of labor supply, while ξ is a scaling parameter reflecting the intensity of the disutility of labor. In the simulations below, we choose $\sigma = 5$ and $k = 0.5$.[18] We also set $\xi = 1$. Our baseline analyses consider constant social welfare weights.

We calibrate our model to Norwegian workers between the ages of 20 and 30 in 1990 who are employed and not in school. There are 100 agents. For each decile in the 1990 wage distribution, there are 10 agents corresponding to their decile in the 2011 wage distribution. We set π_i, $i = 1, \ldots, 100$ equal to the population fraction of each of these types. Following Weinzierl (2011), we assume perfect foresight to obtain heterogeneity in expected inflation-adjusted wage trajectories.[19] We set the exogenous revenue requirements, M_t, equal to 15% of GDP_t (sum of labor income in period t) measured in the benchmark economy when $r_b = 3\%$. The revenue requirements thus do not vary by r_b or policy choices.

Appendix Table A.1 provides summary statistics for the wage trajectories. We see that the median cumulative real wage growth is 0.88. Annualizing this over a 21-year period yields annual real wage growth of 5.16%. This masks considerable heterogeneity. The fifth percentile of cumulative real wage growth is -28% and the ninety-fifth percentile is 464%.

[18] $\sigma = 5$ implies an elasticity of intertemporal substitution of 0.2, which is in line with recent quasi-experimental evidence (Best, Cloyne, Ilzetzki, and Kleven 2020; Ring 2024; Ring and Thoresen 2025. We follow Heathcote, Storesletten, and Violante (2017) in assuming a Frisch elasticity of $k = 0.5$, which is broadly consistent with empirical evidence (Keane 2011; Chetty 2012)

[19] See, e.g., Herbst and Hendren (2023) who argue that young workers (students) possess substantial private knowledge about their future earnings, academic persistence, employment, and likelihood of loan repayment beyond what is captured by observable characteristics.

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

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

Policies. The *Benchmark* policy involves the government optimally choosing τ and G, where the restrictions are that tax rates and transfers are age-independent ($\tau_1 = \tau_2 = \tau$ and $G_1 = G_2 = G$). The benchmark policy does not feature delayed taxation or subsidized lending (i.e., $1 - \delta = 0$ and $x = 0$). For *Age-Dependent* taxation (AD), the government may choose different tax rates, i.e., we allow $\tau_1 \neq \tau_2$, but we require $\tau_t \geq 0$. Allowing $\tau_1 \neq \tau_2$ is not an option under either the benchmark policy or delayed taxation. We also require $G_1 = G_2 = G \geq 0$, except when we explicitly an age-dependent tax *and* transfer system (AD T&T). In the AD T&T case, we require $G_t \geq 0$ but allow $G_1 \neq G_2$.

Under *Delayed Taxation* (DT), we restrict the fraction of period 1 taxes payable in period 1 to be within $[0, 1]$. $1 - \delta \leq 1$ implies that agents cannot borrow from the government in excess of the amount of taxes they accrue, which is typically binding when financial frictions are severe. Imposing $1-\delta \geq 0$ 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 do not arise in our model.[20]

Individual-level delayed taxes. We further impose that agents dissave rather than delay taxes if they weakly prefer to. This occurs when $R'(s^i) \leq 1 + r_{dtax}$ and is only relevant for welfare calculations when we, in robustness checks, impose borrowing limits on the government.

Numerical implementation of R(s). To avoid the computational issues associated with agents facing kinked budget constraints, we employ an $R(s)$ function that approximates a piecewise-linear function with a kink at $s = 0$ but is continuously differentiable. The functional form, discussed in detail in Appendix I, ensures that marginal interest rates, $R'(s) \approx 1 + r$ when $s \geq 0$ and $R'(s) \approx 1 + r_b$ when $s \leq 0.15$.

3.2 Marginal effects of delayed taxation

The Marginal Value of Public Funds. This subsection quantitatively evaluates delayed taxation using the marginal value of public funds methodology (Hendren and Sprung-Keyser 2022, Finkelstein and Hendren 2020, Bastani 2025). We calculate the MVPF of marginally changing

[20] One way to achieve an optimal $1 - \delta < 0$ is to make the government "paternalistic". In Choukhmane and Palmer (2023), for example, the government uses a different time discount factor (i.e., a higher β) than consumers when calculating welfare.

some policy parameter, ω, as

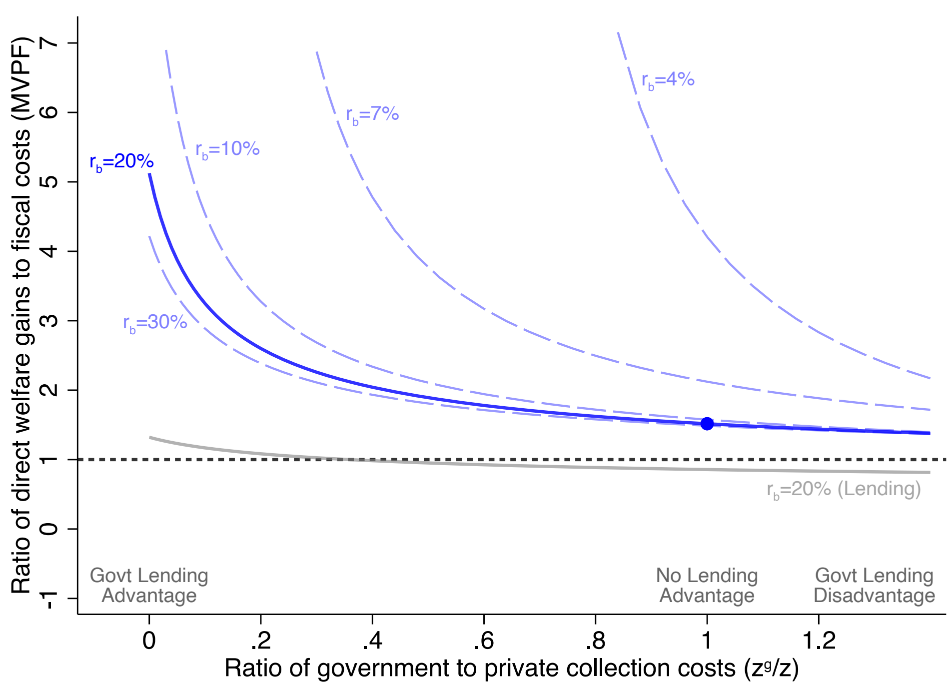
$$MVPF = \frac{\sum_i \pi^i \frac{dV^i}{d\omega}/\lambda}{\frac{dREV}{d\omega}}, \tag{37}$$

where V_i is an individual's life-time utility, λ is the Langrange multiplier on the government budget constraint in the benchmark economy, and REV is present-value government tax revenues net of any cost-of-capital effects (due to the possibility of default). The MVPF is thus the ratio of money-metric utility gains from changing ω to the marginal net cost faced by the government. If the numerator is positive but net cost is negative (e.g., if the tax reform increases revenues), the ratio is set to infinity (Hendren and Sprung-Keyser, 2022).

We value the MVPF using the analytical expressions derived in Section 2. These expressions are written in terms of elasticities and not allocations. Hence, we derive analytical expressions for the elasticities that depend only on allocations (i.e., consumption and labor supply of an agent) in Appendix E. We can thus plug in the numerical allocations for the simulated benchmark economy (or any economy for that matter) to obtain the exact MVPFs for different tax reforms.

Figure 1: Marginal Value of Public Funds

This figure plots the marginal value of public funds (MVPF) for different assumptions about the interest rate on borrowing, r_b. The x-axis varies the government collection advantage relative to private lenders. When $z^g/z = 1$, there is no advantage and the government discounts future delayed tax repayments at the same interest rate as agents face in private lending markets (i.e., $r_b^g = r$) When $z^g/z = 1$, $r_b^g = r$, and thus there is no discounting of future repayments in excess of the interest rate on saving (r). For $z^g/z > 1$, the government faces a lending disadvantage and thus discounts repayments at a higher rate than marginal interest rates in private credit markets. The bottom solid gray line provides the MVPF from offering subsidized loans when $r_b = 20\%$. Figures are restricted to $MVPF < 7$ except for when $r_b = 4\%$ (when we restrict $MVPF < 8$). For $r_b = 4\%$, the MVPF is infinite when $z^g/z \leq 0.11$. For $r_b = 7\%$, the MVPF is infinite when $z^g/z \leq 0.60$.



One of our central findings is that delayed taxation produces substantial welfare gains. Figure 1 emphasizes the robustness of this result by showing how the MVPF varies with both interest rates on borrowing (r_b) and relative lending costs. The solid blue line shows the MVPF for the case when $r_b = 20\%$. Most of our analysis are based on the case when the government has no lending advantage, which is when $z^g/z = 1$ and the government's cost of capital, r_g^b equals r_b. This baseline is marked by a circle in the figure and provides an $MPVF = 1.53$. This value implies that for each marginal dollar in reduced present-value revenues, the government generates \$1.53 in welfare gains through improved intertemporal consumption smoothing. Reduced distortions from income taxation play a pivotal role by reducing fiscal costs and thus the numerator. Introducing a lending *disadvantage*, wherein the government discounts marginal repayments at higher interest rates than the agents face in private credit markets, does not materially change the MVPF. This can be seen from the flat slope of the $MVPF$ curve for values of z^g/z around 1. However, if we introduce an advantage by imposing $z^g/z < 1$, the MVPF rapidly increases to about 5.

The dashed lines consider the MVPF under different assumptions on the interest rate on borrowing. Since we keep the interest rate on savings, r fixed at 3%, varying r_b can be thought of as varying the severity of financial frictions. Interestingly, we find that reducing the extent of financial frictions *increases* the MVPF. The reason is that income effects decrease relative to substitution effects when r_b is reduced. Hence, a lower r_b causes the numerator to shrink. If we also reduce lending costs for the government by decreasing z^g/z, the $MVPF$ becomes infinite. This occurs because income effects are small relative to substitution effects, and when we additionally reduce the present-value costs of lending (i.e., default costs), the numerator approaches zero or even becomes negative.

The bottom solid line shows that the MVPF from offering uniform (unconditional on income) subsidized loans carry an MVPF close to 1. Only when the government has a lending advantage does the lending policy provide an MVPF above one. This stands in stark contrast with delayed taxation, which consistently carries an MVPF well above unity. The takeaway is that whether a subsidized lending policy is optimal depends crucially on the cost of capital associated with lending. The optimality of delayed taxation is considerably more robust.

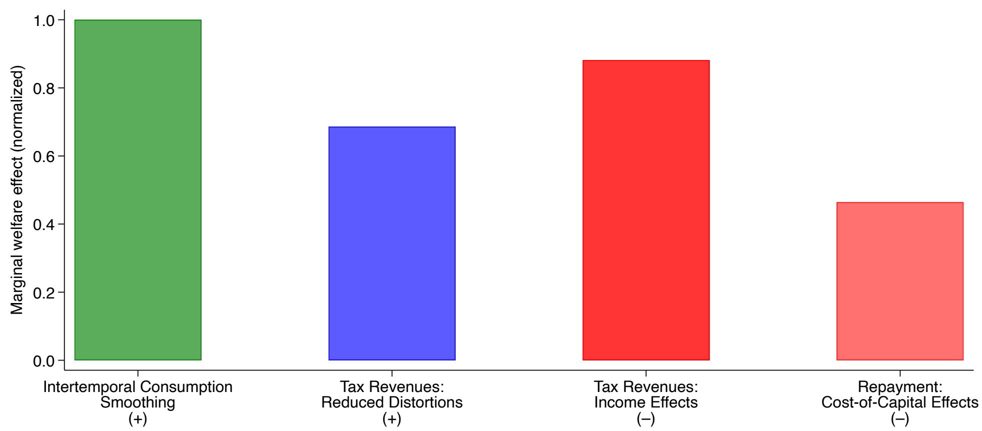
Decomposing the marginal welfare effects. To gain some insight into the source of welfare gains, we use Corollary 1 to decompose the gains into four terms. We present the decomposition in Panel A of Figure 2. We first normalize the welfare gains by dividing through by the gains achieved by enhanced intertemporal consumption smoothing among constrained agents. The second (blue) bar shows the marginal welfare gains from increasing labor supply through reducing the distortionary effects of income taxation, which increases government revenues. These indirect gains account for about two thirds of the direct gains from improved intertemporal smoothing. In terms of negative welfare effects, we find that income effects account for about 85% of the intertemporal gains. The costs of lending, due to discounting future repayments at $r_b > r$, equals about 45% of the intertemporal gains. While this decomposition shows that the main welfare sourcesis enhanced intertemporal consumption smoothing, it emphasizes the importance of reduced distortions. If the reduced-distortion channel were not present, the marginal welfare gains of delayed taxation would clearly be negative. This provides important motivation for investigating this channel in more detail using empirical evidence from a de-facto delayed tax scheme in Norway in Section 5.

Panel B of Figure 2 uses Proposition 3 to provide differential welfare gains from delayed taxation relative to subsidized lending. In line with our previous discussion, the marginal welfare effects of lending are negative (first bar). The positive welfare gains of delayed taxation are driven by the addition of more sizable positive effects from reduced distortions (second bar). The three additional terms address the fact that delayed taxes and subsidized loans are allocated differently. As anticipated, targeting higher-income individuals reduce the intertemporal consumption smoothing effect relative to a uniform lending scheme. There are also slightly weaker income effects from uniform lending. Finally, there is an even more slight advantage to delayed taxation in that the default costs associated with lending to higher-income individuals are lower. In our quantitative framework, this is caused by higher-income borrowers having lower demand for credit and thus being more likely to choose s^i such that $R'(s^i)$ is in the non-linear region where $r_b < R'(s^i) < r$.

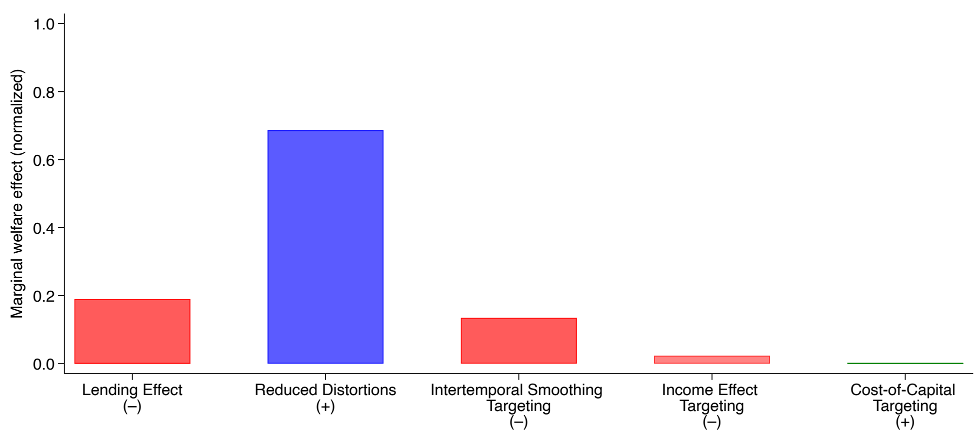
Figure 2: Decomposing Marginal Welfare Effects

Panel A plots the marginal money-metric welfare effects of marginally introducing delayed taxation, $\frac{dW}{d(1-\delta)}\frac{1}{\lambda}$, when $r_b = 20\%$. We assume no government lending advantage ($z^g = z$ and thus $r_b^g = r_b$). The decomposition relies on equation (32), where the period-1 and period-2 distortions are combined, valued using allocations from the benchmark economy. The terms are normalized by dividing through by the value of the intertemporal consumption smoothing term. The first two bars indicate positive welfare effects, and the second two bars provide the absolute value of negative welfare effects. Panel B decomposes the marginal welfare effect into a lending effect plus additional terms using the decomposition results from Proposition 3. In Panel B, we similarly normalize by dividing all terms by the value of the intertemporal consumption smoothing term from delayed taxation.

Panel A: Decomposed Welfare Effects of Delayed Taxation



Panel B: Decomposed into a Lending Effect and Additional Terms



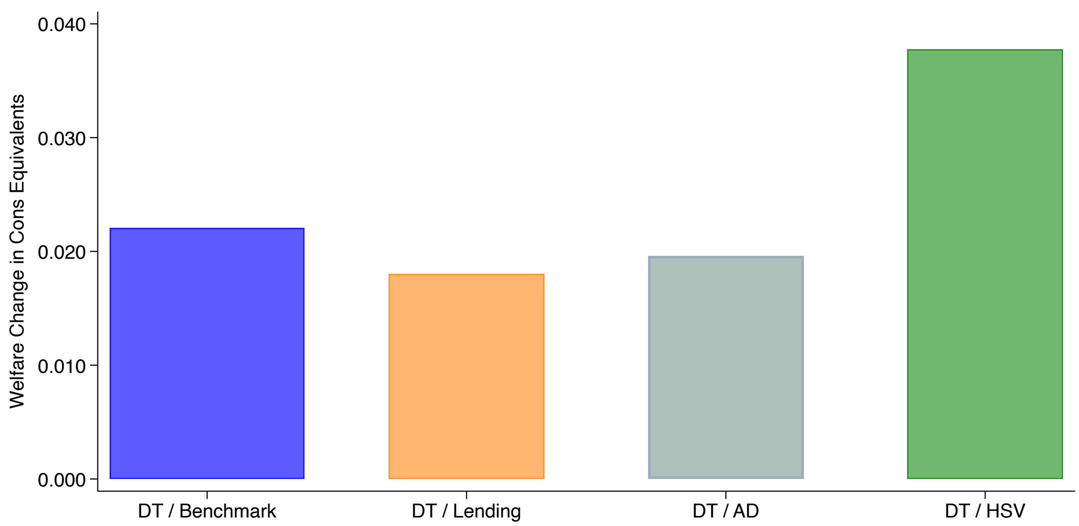
3.3 Welfare effects of Delayed Tax Reform

We now move beyond analyzing the effects of marginal delayed taxation to study the welfare impact of full delayed tax reform. These analyses continue to make the conservative assumption that the government has no lending advantage ($s^g = z$). That is, the government discounts debt or delayed tax repayments at an individual's marginal interest rate in private credit markets.

Figure 3 provides the money-metric welfare gains from implementing delayed taxation. These gains from a given policy constitute the amount by which the government would have to be compensated to obtain the same welfare levels without implementing the policy. The first bar provides welfare gains (in period-1 aggregate consumption units, C_1) relative to the benchmark policy with linear age-independent taxation and no subsidized lending. The second bar provides the differential welfare gains from implementing delayed taxation on top of a subsidized lending scheme. That is, we implement two sets of policies: (i) subsidized lending and (ii) subsidized lending combined with delayed taxation. The differential welfare gains is the welfare gains from the policy combination minus the welfare gains from only implementing subsidized lending. Interestingly, most of the welfare gains from delayed taxation stay intact—consistent with behavioral responses in terms of increased labor supply constituting a key welfare channel for delayed taxation relative to subsidized lending.

Figure 3: Welfare Gains from Introducing Delayed Taxation (On Top of Other Policies)

This figure plots the differential welfare gains from implementing delayed taxation on top of other policies when $r_b = 20\%$. The first bar provides the money-metric welfare gains from implementing delayed taxation relative to the benchmark economy. The second bar provides the difference in welfare gains between (i) implementing both delayed taxation and subsidized lending and (ii) only implementing delayed taxation. Money-metric welfare gains are measured the amount of money that would make the government indifferent between the benchmark policy and any (combined) reform under consideration; this monetary value is then divided by the benchmark economy's period-1 aggregate consumption. The third bar similarly provides the differential welfare gains of implementing delayed taxation on top of age-dependent taxation. The fourth bar provides the gains from implementing delayed taxation on top of a benchmark economy that features an HSV-style non-linear (progressive) tax schedule. Appendix Table A.3 provides additional statistics for the age-dependent and HSV policies.



The third bar similarly provides the welfare gains from implementing delayed taxation on top of an age-dependent tax policy. Age-dependent taxes may also address financial frictions by letting young borrowers pay lower taxes. They may also reduce period-1 income tax distortions by lowering τ_1. Nevertheless, age-dependent taxation is bluntly targeted to all workers. The fact that delayed taxation targets constrained workers lets it provide substantial welfare gains even on top of an age-dependent policy.

The fourth bar provides the gains relative to a non-linear HSV tax schedule. More progressive taxes may reduce financial frictions in that average tax rates are lower when agents have low incomes. Interestingly, we find that an HSV-style benchmark *increases* the gains from delayed taxation. Appendix Table A.3 shows that this is driven by complementarities between delayed and nonlinear taxation. With HSV, the government fully delays taxes ($1 - \delta = 100\%$) and increases the parameter, a, that governs the progressivity of the tax schedule.

Optimal tax schedule and allocations. Table 1 provides optimal policy parameters and average allocations for different policy reforms. The first two columns provide statistics for the benchmark economy with and without delayed taxation. The third column consider the case when the government can implement subsidized lending, and the fourth column further allows the government to implement delayed taxation. We see that the differential welfare gain from implementing delayed taxation on top of the benchmark economy is 2.2% of aggregate period-1 consumption. This differential is only reduced to 1.8% when the government engages in subsidized lending. Interestingly, we see that the government refrains from engaging in subsidized lending, by setting the loan limit very close to zero when it has access to delayed taxation.

Table 1: Optimal Tax Schedules and Allocations

This table provides summary statistics for different policy combination. The first column provides statistics for the benchmark economy. The second column allows for delayed taxation. The third column allows for subsidized lending. The fourth column allows for subsidized lending *and* delayed taxation. The present value function calculates present values according to $1+r$. The (money-metric) welfare gain is the exogenous shock to revenue that the government must experience in the benchmark economy to be equally well off as with the given policy reform(s). This number is measured as a fraction of the aggregate period-1 consumption, C_1. Govt loan limit is the uniform subsidized loan given under the subsidized lending reform. B is the amount the government borrows to satisfy its first-period budget constraint. Appendix Table A.3 considers the cases with Age-Dependent and non-linear HSV taxation as well.

	Tax schedule and allocations with $r_b = 20\%$			
	Benchmark	+ DT	Lending	+ DT
τ_1	0.646	0.678	0.649	0.678
τ_2	0.646	0.678	0.649	0.678
G_1	0.544	0.561	0.543	0.561
G_2	0.544	0.561	0.543	0.561
$1-\delta$	0.000	0.312	0.000	0.312
r_{dtax}	0.030	0.030	0.030	0.030
Govt loan limit, x	0.000	0.000	0.017	0.000
Govt borrowing, B	0.125	0.369	0.134	0.369
Welfare Gain, % C_1 (rel. to Benchmark)	0.000	2.206	0.403	2.206
Δ Welfare from DT	.	2.206	.	1.803
π^i-weighted means				
$l_1 w_1$	0.815	0.739	0.804	0.739
$l_2 w_2$	1.677	1.913	1.714	1.913
c_1	0.887	0.920	0.891	0.920
c_2	1.011	0.947	1.001	0.947
PV($l_1 w_1$,$l_2 w_2$)	1.717	1.767	1.725	1.767
PV($l_1 w_1 \tau_1$,$l_2 w_2 \tau_1$)	1.110	1.199	1.119	1.199
$(TaxPaid_1 - G_1)/y_1$	-0.077	-0.302	-0.086	-0.302
$(TaxPaid_2 - G_2)/y_2$	0.262	0.362	0.272	0.362
l_1	0.803	0.728	0.790	0.728
l_2	0.827	0.920	0.846	0.920
Net savings, s	-0.055	0.036	-0.066	0.036
Impl. govt loan amount, IL	0.000	0.045	0.005	0.045

We also see that delayed taxation reduces period-1 labor earnings, $l_1 w_1$, but increases period-2 labor earnings. This differs from our theoretical results on the marginal effects of delayed taxation due to the government also changing the tax rate, τ and G. When the interest rate on borrowing is high (e.g., $r_b = 20\%$), income effects dominate substitution effects in how taxes affect labor supply. Hence, lowering the effective period-1 tax rate through delayed taxation decreases labor supply, causing $l_1 w_1$ to decrease. At the same time, the period-2 tax rate increases (as τ is increased), causing period-2 labor supply to increase.[21]

Table 1 also shows that the benchmark economy already features elements of age-dependent taxation. The ratio of taxes paid minus transfers to gross income is -0.077 in period 1 and 0.262 in period 2. This is due to the restriction that $G_1 = G_2$, which implies that the lump-sum transfer is tied to inflation (which is zero) rather than real wage growth (whose median is 88%). In quantifying welfare effects of delayed taxation, requiring $G_1 = G_2$ is a conservative choice. The most alternative would be to tie G_t to real wage growth, in which case G_2 would be higher than G_1. Ceteris paribus, this would shift consumption towards period-1, increasing the wedge between marginal utilities and thus the welfare gains from delayed taxation.

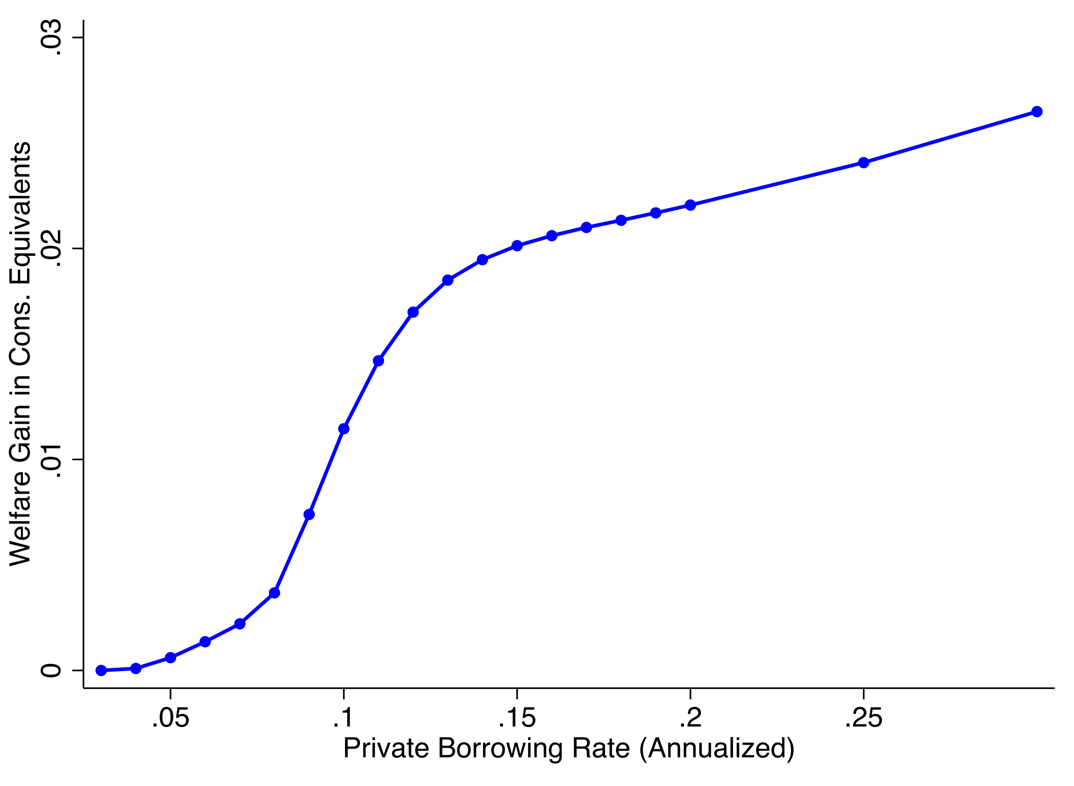
While we show negative marginal welfare effects from delayed taxation in Figure 2, we nevertheless find modest positive gains from a full lending reform. Studying Table 1 reveals that this is due to the complementarities between lending and optimally choosing τ and G (which were held fixed when considering marginal effects). When implementing subsidized lending, the government keeps G approximately fixed but increases τ. We can think of subsidized lending as being an alternative to a lump-sum transfer (G) that better targets those with low marginal utilities in period 1. When allowed this new transfer technology, the government chooses to engage in some lending, financed via higher tax rates. Nevertheless, the extent is modest. The third column shows that the government offers a loan (x) of about 0.017, which is about 3.3% of average period-1 taxes, $\tau_1 \ell_1 w_1$. This is modest relative to the extent of delayed taxation at 31.2%.

Welfare gains when varying the severity of financial frictions. Figure 4 shows the welfare gains from allowing the government to optimally delay taxes relative to the benchmark economy. We plot the money metric welfare gains for different assumptions on r_b. A higher r_b implies higher marginal interest rates for borrowers and is thus a measure of the extent of financial frictions affecting households. We see that welfare gains are low when r_b is low. This is both because (i) there are modest gains from improving intertemporal consumption smoothing and (ii) because the behavioral effect on labor supply is small. When marginal interest rates are already low, delayed taxation also has a small effect on effective present-value tax rates, which renders the welfare gains from reducing the distortions of taxation to be modest. The leveling off (i.e., the concavity for $r_b > 0.1$) is due to individuals effectively ceasing to borrow when r_b becomes large. Hence, additional increases in r_b has little effect on individuals' actual marginal interest rates (and thus on their intertemporal marginal utility wedges).We demonstrate this in Appendix Figure A.3.

[21]The combined effect of delaying taxes and increasing τ is actually to slightly increase the effective period-1 tax rate, but the fact that G_1 increases as well works to reduce $l_1 w_1$.

Figure 4: Welfare Gains from Introducing Delayed Taxation Under Different Assumptions on Borrowing Rates

This figure plots the money-metric welfare gains from implementing delayed taxation on top of the benchmark policy for different borrowing rates. Money-metric welfare gains equal the exogenous period-1 budget injection the government would need to obtain the same welfare levels in the benchmark policy as with delayed taxation. These money-metric gains are then divided by aggregate period-1 consumption in the benchmark economy.



3.4 U.S. calibration

A distinct feature of our calibration is that a large majority of workers see real wage increases. Accordingly, a majority wishes to borrow and thus more workers are affected by the financial frictions. This creates an important source of welfare gains for policies that enhance intertemporal consumption smoothing. However, it is useful to verify that the results apply to other settings as well. We therefore use data from the Panel Study of Income Dynamics (PSID) obtained from the replication package in Weinzierl (2011). To match our empirical setting, we consider workers aged 20–30 in 1980 for whom Weinzierl (2011) has calculated effective wages in both 1980 and 2001 (i.e., the same 21-year gap). We use then use these wage trajectories to provide numerical solutions.

The U.S. calibration features very different wage trajectories (w_1^i, w_2^i) (Appendix Table A.2) than the Norwegian calibration (Appendix Table A.1). In the PSID, the median 21-year *cumulative* wage growth is only 2% (0.1% annual). In the Norwegian data, median 21-year wage growth is 88%. Since heterogeneity in borrowing demand is solely driven by wage profiles in our framework, there are substantially fewer borrowing constrained individuals in the PSID calibration. Hence, welfare gains from policies that address financial frictions will be more modest. This is evident from Appendix Table 2, which shows optimal tax schedules and allocations for the PSID sample. For example, we see that the government is a net saver in the benchmark economy ($B < 0$), and the optimal tax schedule features average tax rates (taxes net of transfers divided by labor earnings) that decrease with age. This is contrasts the Norwegian calibration where the government borrows and ATRs increase with age.

Table 2: U.S. PSID Sample: Optimal Tax Schedules and Allocations

This table provides statistics for different policy combinations as in Table 1 when *the model is calibrated to the U.S. PSID data*. The first column provides statistics for the benchmark economy. The second column allows for delayed taxation. The third column allows for subsidized lending. The fourth column allows for subsidized lending *and* delayed taxation. The fifth and sixth columns consider an age-dependent baseline policy; and the seventh and eight consider an nonlinear-HSV policy. The present value function calculates present values according to $1 + r$. The (money-metric) welfare gain is the exogenous shock to revenue that the government must experience in the benchmark economy to be equally well off as with the given policy reform(s). This number is measured as a fraction of the aggregate period-1 consumption, C_1. Govt loan limit is the uniform subsidized loan given under the subsidized lending reform. B is the amount the government borrows to satisfy its first-period budget constraint. IL is the individual-level implied amount that the individual borrows from the government, including delayed taxes (see Appendix I)

	Tax schedule and allocations with $r_b = 20\%$							
	Benchmark	+ DT	Lending	+ DT	AD	+ DT	HSV	+ DT
τ_1	0.63	0.63	0.63	0.63	0.58	0.58	0.63	0.63
τ_2	0.63	0.63	0.63	0.63	0.67	0.68	0.63	0.63
G_1	0.48	0.48	0.48	0.48	0.47	0.47	0.54	0.54
G_2	0.48	0.48	0.48	0.48	0.47	0.47	0.54	0.54
a_t (HSV)	.	.	.	.	.	.	0.30	0.31
k_t (HSV)	.	.	.	.	.	.	-0.24	-0.20
$1-\delta$	0.00	0.07	0.00	0.07	0.00	1.00	0.00	0.04
r_{dtax}	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Govt loan limit	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
B	0.01	0.06	0.01	0.06	0.03	0.62	0.01	0.04
Welfare Gain, % C_1								
(rel. to Benchmark)	0.00	0.27	0.08	0.27	0.64	1.29	0.31	0.44
Δ Welfare	.	0.27	.	0.19	.	0.65	.	0.14
means								
$(TaxPaid_1 - G_1)/y_1$	0.06	0.03	0.06	0.03	0.04	-0.01	0.04	0.03
$(TaxPaid_2 - G_2)/y_2$	-0.06	-0.04	-0.06	-0.04	-0.06	-0.04	-0.15	-0.12

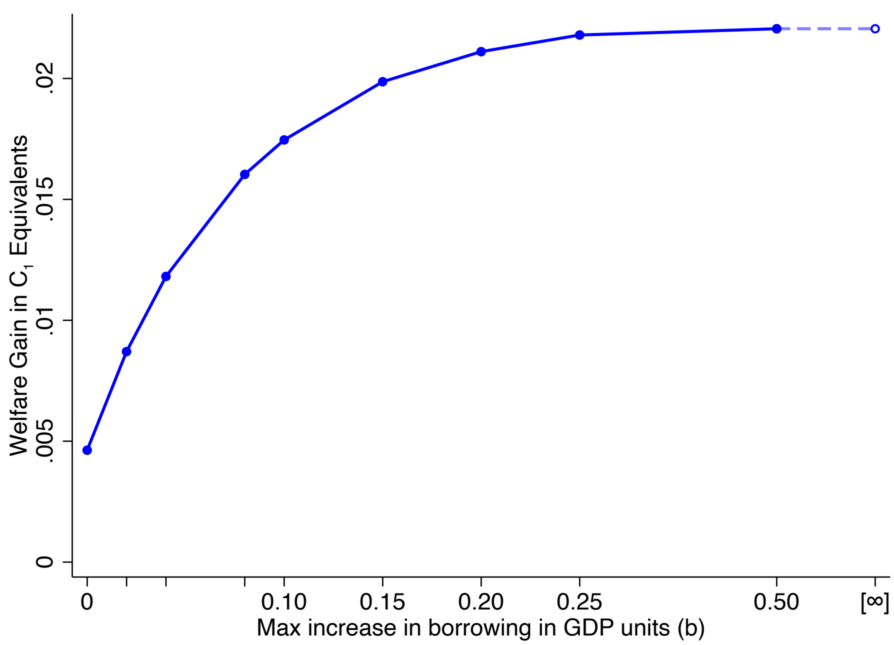
Nevertheless, we still find meaningful welfare gains from implementing delayed taxation, ranging from 0.14% to 0.65% of GDP. On top of the benchmark policy, the government optimally chooses a modest tax delay, $1 - \delta$, of 7%, which delivers welfare gains of 0.27% of benchmark GDP. This is three times larger than the 0.08% welfare gains from the subsidized lending scheme. We also see that when the government can engage in both subsidized lending and delayed taxation, it effectively only chooses delayed taxation by setting the loan limit to ≈ 0. Interestingly, implementing delayed taxation on top of an age-dependent tax scheme *doubles* the welfare gains relative to only implementing age-dependent taxation.

3.5 Restrictions on government borrowing

We examine the potential welfare gains from delayed taxation when the government is limited in how much it can borrow. We define the government borrowing amount as the period-1 budget deficit, and limit this to some fraction, b, of benchmark GDP. Figure 5 shows that delayed taxation delivers material welfare gains even when the government's borrowing ability is limited. When the government cannot increase borrowing at all, welfare gains are still 0.5%. Welfare gains increase rapidly up to the point when the government borrows about 10% of benchmark GDP, indicating that delayed tax reform can achieve substantial welfare gains even without substantial changes to government borrowing. The caveat, demonstrated by the bottom green line in Figure 5 is that if the government already engages in age-dependent taxation, larger borrowing amounts are needed to obtain substantial welfare gains.

FIGURE 5: WELFARE EFFECTS WHEN THE GOVERNMENT FACES BORROWING RESTRICTIONS

This figure plots the money-metric welfare effects (in units of aggregate period-1 consumption, C_1) of implementing either delayed taxation on top of the benchmark policy. The values of the x-axis denote the maximal increase in government borrowing (period-1 deficit) the government from the benchmark policy (in benchmark GDP units). $b = 0$ allows no additional borrowing. $b = 0.10$ allows the government to increase borrowing by 10% of benchmark GDP.



4 Does delayed taxation already exist?

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

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

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

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

The idea that social security contributions can reduce welfare in the presence of financial frictions is not new. Hubbard et al. (1986) argue that when there are liquidity constraints, social security contributions lead to reduced welfare and excessive saving. In discussing the paper, Larry Summers notes that "reversing the direction of transfers" under social security would seem a natural way to reduce welfare losses from financial frictions. However, the key point that negative Social Security taxes (i.e., delayed taxation) can also reduce welfare losses from the distortionary effects of income taxation is missing.[22] The possibility that social security distorts labor supply, however, is mentioned by Robert Barro (see Barro et al. 1979) when discussing the potential crowd-out effects of social security on saving: "[...] the absence of a downward effect on saving does not eliminate the harmful economic effects of the social security, system that involve distortions of labor-market decisions."

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

Student stipends. A case of delayed taxation is when student financial aid is based on

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

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

Retirement savings accounts. Tax-deductible contributions to retirement accounts where subsequent distributions are taxed as income (such as traditional IRAs in the U.S) effectively delays income taxes for savers.[23] However, these deductions are often capped at modest levels, which means that many workers (who contribute the maximal amount) see no effect on their effective marginal tax rate. Roth IRAs do not resemble delayed taxation since they do not involve lower effective tax rates on income while saving (Beshears et al., 2017). Instead, they shield eligible contributions from subsequent capital gains taxation. In addition, both traditional and Roth IRAs are irrelevant for the labor supply decisions of constrained workers whose marginal saving would go toward reduced borrowing as opposed to IRA contributions.

5 De facto Delayed Taxation in Norway

5.1 Empirical setting

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

Our empirical study focuses on the years 2004 to 2011. During these years, most Norwegian students faced an earnings threshold ranging from NOK 104,500 ($17,000) in 2004 to NOK 140,823 in 2011. Monthly transfers ranged from NOK 8000 ($1,300) in 2004 to NOK 9785 in 2011. These transfers are initially given as a loan, but 40% can be forgiven (converted to a stipend) as long as students pass their courses and stay below the earnings thresholds mentioned above. Students are notified of the amount of the transfer at the beginning of the academic year. These notification letters include a breakdown of the transfers, noting the amount (40% of the total) that will be given as a conversion loan, and stating that the conversion of the loan to a stipend is contingent upon income being below an income threshold. The following year, students are notified how much of their loan has been converted, based on grades reported by the educational institutions

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

and income reported to the tax authorities. Loans must generally be repaid within 20 years of graduation. No interest is charged while the student is receiving aid. Thereafter, interest rates are slightly above the risk-free rate and loan payments can be delayed at the (former) student's discretion for up to a total of 3 years.[24]

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

Summary statistics are given in Table 3. The average student is 23 years old. This is reasonable given that high school graduation occurs at age 18 and that we condition on students being enrolled in higher education for both semesters within a given year. The summary statistics reveal a significant spread in the amount of liquid assets available to students. While students at the 25th percentile have only NOK 8,000 ($1,300) in liquid assets, students at the 75th percentile have almost *ten* times more. A similar spread can be observed in the liquid assets of the students' parents. We also see that the average student earns around NOK 100,000 ($17,000), which is a direct consequence of our sample restrictions caused by focusing on students around the debt conversion threshold. Four years later, the average student has a much higher income of around NOK 360,000 ($60,000).

TABLE 3: SUMMARY STATISTICS

This table presents summary statistics. The main sample period is 2004-2011. The financial variables are denominated in NOK. The USD/NOK exchange rate was around 6 in 2010. The main sample is restricted to students who had earned income within 50,000 of the debt conversion threshold. Liquid assets consist of deposits, investment funds and ownership of public shares. Labor earnings are censored to be below NOK 1,000,000 in 2010 NOK. The Bottom Tax Threshold is only considered for the years 2005-2011.

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

Salience. In order to meaningfully compare the implied elasticity of the debt conversion threshold with that of the regular tax thresholds, the conversion threshold must be similarly salient. As one of the authors is a former participant in this program, we certainly believe this

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

to be the case. Beyond anecdotal evidence, however, it is useful to consider the magnitude of the effective tax increase. A 50 percentage point reduction in the ”net of debt” wage is unlikely to go unnoticed. Moreover, students are informed of the existence of such a cap in a loan agreement letter that they must sign, and they also receive letters informing them of any conversion that has occurred. Even if students do not expect to receive a reduction in their debt through conversion, they will want to read these letters to confirm that their institution has accurately recorded and reported their academic progress. Non-passing grades in courses also reduce debt conversion. Students are also informed of their annual student debt balances when they receive their annual prefilled tax returns, which also include information about their income tax liabilities.

5.2 Bunching methodology

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

$$e = \frac{\Delta y^*/y^*}{\Delta\tau/(1-\tau)}, \tag{38}$$

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

$$B = \int_{y^*}^{y^*+\Delta y^*} h_0(y)dy \approx h_0(y^*)\Delta y^*. \tag{39}$$

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

$$\frac{\Delta y^*}{y^*} = \frac{B}{h_0(y^*)y^*} = \frac{b}{y^*}. \tag{40}$$

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

We empirically estimate b, the relative excess mass at the threshold, using the methodology in Chetty et al. (2011), which we call the bunching estimate. The empirical analog of y^* is the (average) debt conversion threshold, expressed in the same units (thousands) as the empirical earnings bins.[25] We write our estimated compensated labor earnings elasticity as

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

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

In a standard model without adjustment frictions, the estimator $\hat{e}$ is considered to estimate the Frisch elasticity (Saez 2010 and Kleven 2016). When preferences are additively separable as in our calibration (equation 36), this implies that $\hat{e}$ identifies the structural Frisch elasticity, k, in a standard frictionless model. 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 (11) can be written as:

$$u'(c_1^i)\cdot w_1^i(1-\tau_1(1-\delta)\Delta_r^i)=v'(\ell_1^i). \tag{42}$$

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

$$\varepsilon^{i,F}_{\ell_1,1-\tau_1}=\left(1-(1-\delta)\Delta_r^i\right)\frac{v'(\ell_1^i)}{\ell_1^i v''(\ell_1^i)}=\left(1-(1-\delta)\Delta_r^i\right)\,k, \tag{43}$$

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

$$\varepsilon^{i,F}_{\ell_1,1-\tau_1}=\left(1-\Delta_r^i\right)\,k. \tag{44}$$

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

$$E[\hat{e}]=E[\zeta\varepsilon^{i,F}_{\ell_1,1-\tau_1}]=\underbrace{E\left[1-\Delta_r^i\right]}_{\text{Delayed taxation effect}}\quad\underbrace{\cdot\ \zeta}_{\text{net of bias factor}}\quad\underbrace{\cdot k.}_{\text{structural Frisch}} \tag{45}$$

5.3 Bunching at the debt-conversion threshold

Figure 6 summarizes the empirical analysis. Panel A verifies that earnings above the threshold lead to an increase in debt in the next period. Most students are on the expected kinked trajectory, where each additional NOK of earnings increases debt by NOK 0.50. The blue fitted line illustrates how we obtain our first-stage measure of the effect of excess earnings on debt accumulation. We find that the slope increases by 0.47. This is close to the nominal increase of 0.50 because there are very few non-compliers.[26] In terms of the previous notation, this means that $\widehat{\Delta\tau}=0.47$.

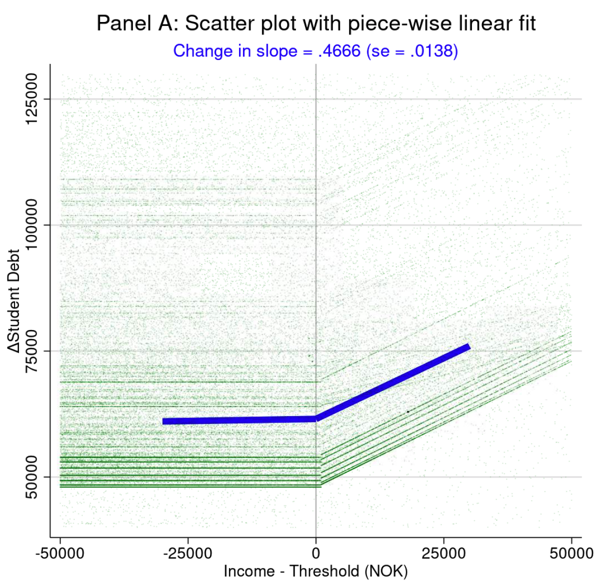
In Panel (B), the yellow dotted line shows the distribution of students around the earnings threshold. The green line is the counterfactual distribution, which is a 5th order polynomial

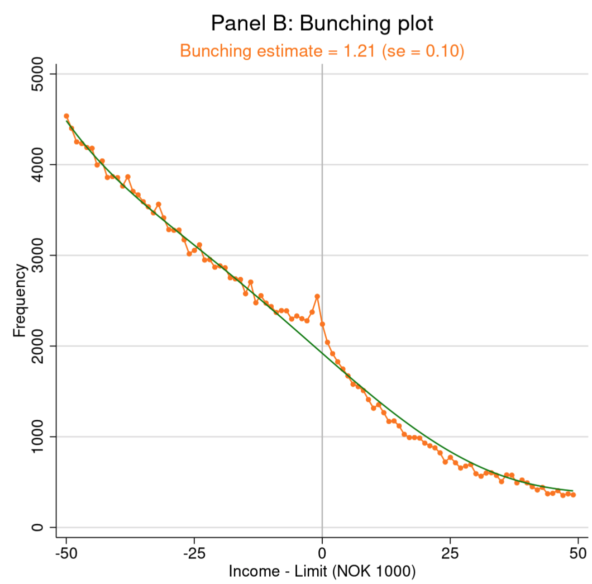
[26]Some non-compliers exist, for example, because they may have moved in with their parents during the fall semester, which would exclude them from receiving a conversion for fall semester loans. Such moves must be reported to the Educational Loan Fund, but not to the tax authorities from which we receive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

fitted to the non-bunching region. By comparing the actual and counterfactual distributions, we obtain a measure of the excess mass of individuals near the threshold. This provides a bunching estimate, b, of 1.21, which means that there are 121% more individuals around the threshold than the counterfactual distribution implies. Dividing 1.21 by the average threshold amount (120.162 in NOK 1,000s), per equation 41, we obtain a remarkably low elasticity of labor earnings to the net-of-tax (or net-of-debt-increase) wage of 0.0162.[27] The standard error is 0.0013.[28]

Figure 6: Verifying the Effect of Excess Earnings on Future Debt and Examining Bunching Responses

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





This analysis shows that students are remarkably unresponsive to de facto delayed taxation. However, if we account for the delayed taxation effect, the elasticity estimate squares well with recent evidence from Kostøl and Myhre (2021) who estimate a structural frisch elasticity of 0.3 and a downward bias of 70%, which translates into $k = 0.3$ and $\xi = 0.3$. If we assume a risk free rate of 2% and an average marginal borrowing rate among students of 20%, we can quantify our expectation of the the elasticity in equation (45).

$$E[\hat{e}] = \underbrace{\left(\frac{1+2\%}{1+20\%}\right)^{10}}_{\text{Delayed taxation effect}} \quad \underbrace{\cdot(1-0.7)}_{\text{net of bias factor}} \quad \underbrace{\cdot 0.3}_{\text{structural Frisch}} = 0.0178, \tag{46}$$

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

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

which is virtually identical to the elasticity that we estimate.

In robustness checks, show that the results are virtually identical when considering students employed in likely highly flexible hospitality and sales positions in Figure A.2. This suggests that the small elasticity is in fact due to a delayed taxation effect as opposed to higher adjustment or optimization frictions among students. We also find qualitatively similar results when we consider bunching around the conversion cap. Here, additional earnings no longer increase student debt because students are no longer eligible for *any* conversion from loans to stipends. We report these results in Figure A.1. We find that the bunching estimate is negative, in line with theory, but statistically close to zero (t-stat=-1.64).

5.4 Determinants of non-bunching

We now examine potential determinants of this (non-)bunching behavior. Our main approach is to plot students' characteristics against their position relative to the conversion threshold.[29] This is a visual exercise in which we try to draw conclusions from visual breaks in the relationship between a given characteristic and students' earnings that occur around the conversion threshold.

In Figure 7, Panel (A), we see that the amount of ex-ante liquid wealth drops sharply just above the threshold. This suggests that non-bunchers have less liquid assets, which is consistent with these students being financially constrained. Panel (B) of Figure 7 shows how future labor earnings vary with the student's position relative to the threshold. This shows no sharp increase or decrease in realized future earnings above the threshold, suggesting that non-bunchers do not differ significantly in terms of medium-term earnings prospects.

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

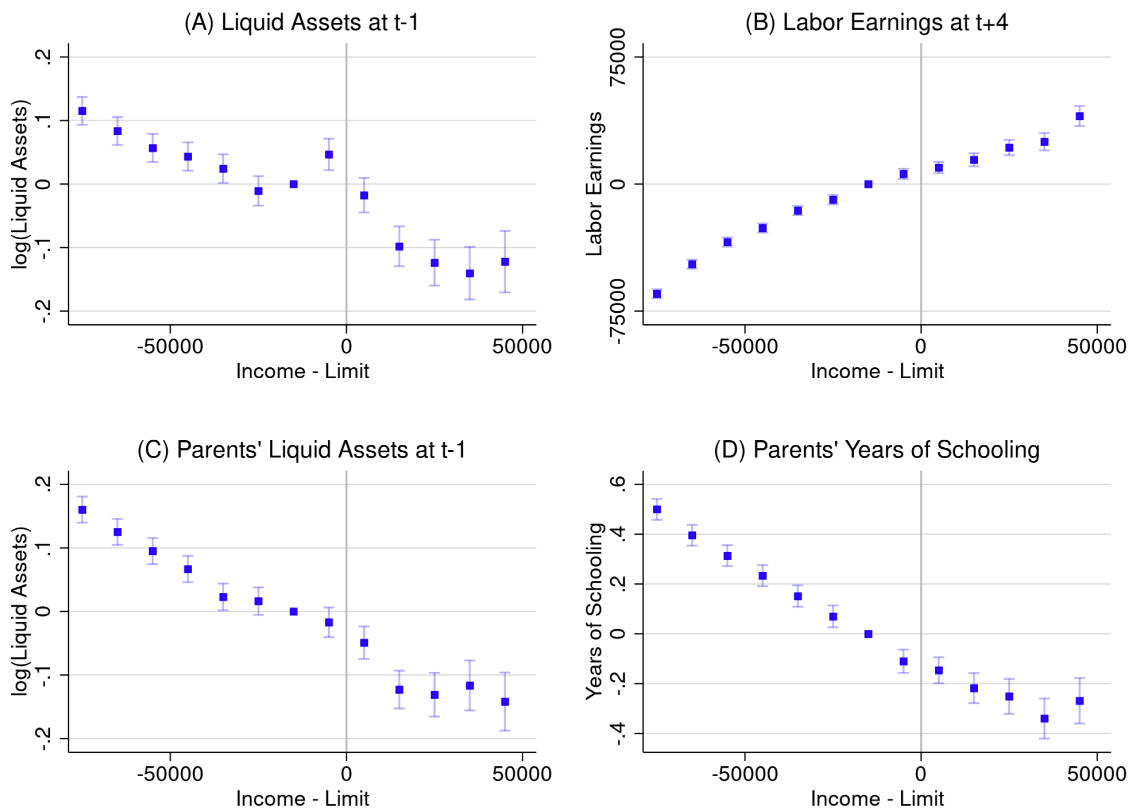
To investigate this liquidity channel further, Panel (C) of Figure 7 shows how *parents'* liquidity correlates with the student's earnings location. This documents a notable negative relationship between parents' financial resources and the child's in-school labor earnings. This suggests that parents play an important role in determining the amount of time students are able to devote to their studies. More relevant to the present study is the finding that parental wealth declines just above the earnings threshold. This suggests that non-bunchers have access to fewer financial resources, which is consistent with financial frictions playing a key role in the observed non-responsiveness to the conversion threshold. However, wealth may be a proxy for human capital, which influences tax responsiveness (Bastani and Waldenström, 2021). Therefore, we plot parental education on the y-axis in panel (D). This shows that there is no break in the relationship between educational attainment, measured by the maximum number of years of schooling of the

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

parents, and the position of the child relative to the conversion threshold. This argues against the hypothesis that fewer resources, in a human capital rather than a financial sense, can explain the irresponsiveness to the threshold. If anything, extrapolating from the relationship below the threshold, non-bunchers may have more highly educated parents. To the extent that this is correlated with students' lifetime wealth, it may explain some of the students' desire to front-load consumption by taking on higher student loans.

Figure 7: Characteristics of Students Below and Above the Income-contingent Debt-conversion Threshold

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

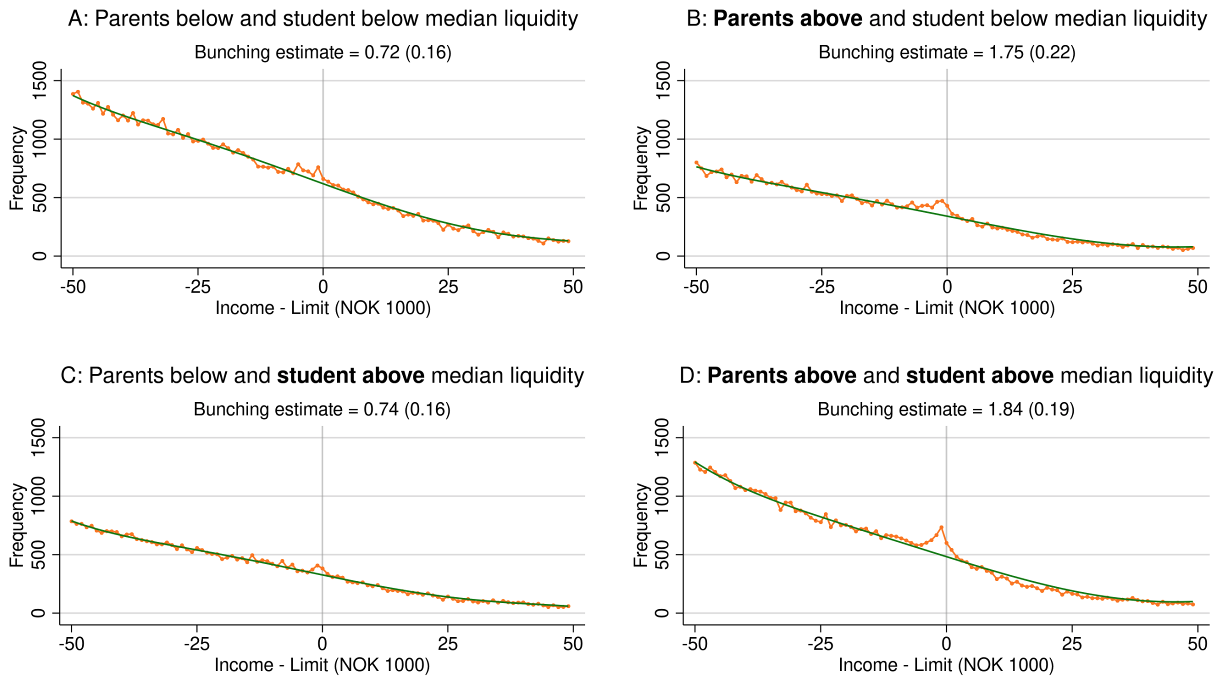


5.5 Bunching heterogeneity

We proceed with a complementary, more standard approach to examine the heterogeneity in the earnings sensitivity to the threshold in Figure 8. This approach splits the sample into subsets based on student and parental characteristics to compute heterogeneous bunching elasticities. We see that the largest contribution to the total excess mass in the previous figure 6 comes from students who themselves and their parents have above-average liquid assets. Figure 8 also suggests that the main driver of bunching responses is parents' rather than students' own liquid assets. Moving from the left to the right panel, which improves parental liquidity, more than doubles the bunching estimates.[30]

Figure 8: Heterogeneity in Bunching by Amount of Liquid Assets

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



What can this heterogeneity tell us about how the severity of financial constraints varies with liquidity? In Figure 8, we see that the elasticity increases from 0.72 to 1.84 as we move from below to above the median for both student and parent resources. Using our expression (45) for the expectation of the estimator $\hat{e}$, we can write

$$\frac{1.84}{0.72} = \frac{\mathbb{E}[\hat{e}_{\text{above}}]}{\mathbb{E}[\hat{e}_{\text{below}}]} = \frac{E_{\text{above}}\left[(1-\tau_1)(1-\Delta_r^i)\right] \quad \cdot \quad \zeta \quad \cdot k}{E_{\text{below}}\left[(1-\tau_1)(1-\Delta_r^i)\right] \quad \cdot \quad \zeta \quad \cdot k}. \tag{47}$$

From this expression, we get that

$$\frac{1.84}{0.72} = \frac{E_{\text{above}}[1-\Delta_r^i]}{E_{\text{below}}[1-\Delta_r^i]} = \frac{E_{\text{above}}[R'(s^i)]}{E_{\text{below}}[R'(s^i)]}. \tag{48}$$

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

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

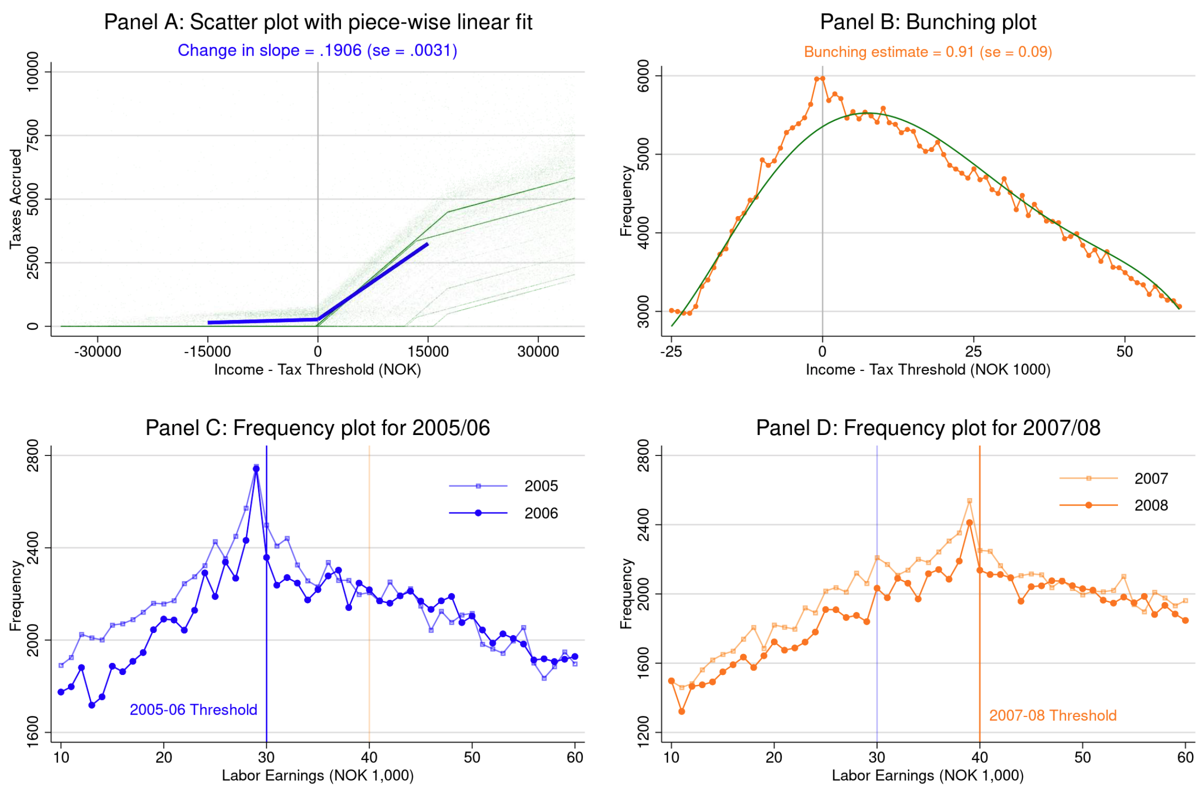
choice between not borrowing and accumulating credit card debt at interest rates close to 20%. If the marginal rate at which they would borrow is 10%, these students would borrow 0, and thus we would not observe any (realized) interest rates for them.

5.6 Analysis of bunching at a regular tax threshold

In this section, we repeat the introductory analyses performed in Figure 6 using a *tax* threshold instead of the debt conversion threshold. The purpose of this exercise is to obtain a reference estimate of the implied labor earnings elasticity at a tax threshold at which marginally accrued taxes are not delayed. We focus on the first tax threshold in the progressive income tax system. This threshold was NOK 30,000 in 2005-06 and NOK 40,000 in 2007-2011.[31] At this threshold, the marginal income tax increases from 0 to about 25 percent for most taxpayers.

Figure 9: Bunching at a Regular Tax Threshold

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



In Figure 9, we examine this complementary empirical setting. Panel (A) provides a scatterplot that verifies the presence of an increase in the marginal income tax rate by plotting total taxes accrued in the year against income. It also provides the fitted line, from which we infer an

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

average increase in the marginal tax rate of 19 percentage points at the threshold. The coefficient is lower than the nominal increase of 25 percentage points because some individuals may be eligible for higher standard deductions.

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

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

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

$$\frac{0.13}{0.0162} = \frac{E_{\text{delayed}}\left[(1-(1-0)\Delta_r^i)\right]\ \zeta\ k}{E_{\text{regular}}\left[(1-(1-1)\Delta_r^i)\right]\ \zeta\ k} = (1 - E_{\text{delayed}}[\Delta_r^i]). \tag{49}$$

which implies an average annual interest rate over 10 years of $\left(\frac{0.13}{0.162}\right)^{\frac{1}{10}} - 1 = 23\%$. This figure is comparable to average credit card rates, which are slightly above 20%.[32] This suggests that some students are willing to borrow from the Education Loan Fund at an interest rate higher than that offered by financial institutions. This may be partly due to credit rationing, but probably mainly due to the fact that the loan fund does not require payments while students are still in school, and generally has a long maturity, with the additional option of delaying payments for up to three years.

We can use this implied elasticity to get an idea of how much bunching would be caused by the debt conversion threshold in the absence of financial frictions. In other words, how much bunching would there be in Figure 6 if students responded to the debt conversion threshold as if it were a regular income tax threshold? To find out, we reverse the calculation used to derive labor supply elasticities from the bunching estimates. This yields a counterfactual bunching estimate

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

of 23.43.[33] This is considerably larger than the empirical bunching estimate of 1.21.

Figure 10: Contrasting Characteristics of Bunchers at the Delayed Tax and Regular Tax Thresholds

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

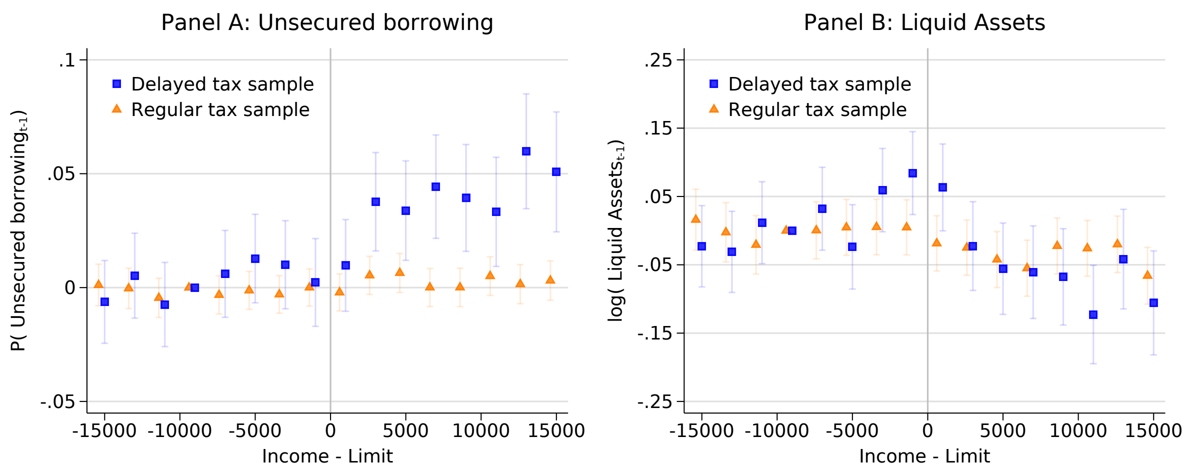


Figure 7 shows that those who bunch at the *delayed tax* threshold have more liquid wealth. This is intuitive because more constrained students have higher marginal borrowing rates on average and are thus less sensitive to a tax that is payable in the future. However, it is useful to show that proxies for financial constraints matter more for delayed tax bunching than for regular tax bunching to rule out the possibility that financial frictions are simply correlated with labor supply adjustment frictions. Accordingly, we empirically investigate whether regular-tax bunchers also appear to be more or less constrained, and we contrast this with the characteristics of bunchers at the delayed-tax threshold in Figure 10.

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

5.7 Accounting for differences in observables when comparing elasticities.

In this section, we pool the samples used to examine bunching at the debt conversion (delayed tax) and regular tax thresholds. We develop a regression-based approach that allows us to compare

[33]=0.13*(120162/1000)*(75/50)

the underlying elasticities while holding observables fixed.[34] This addresses the fact that higher earning students in the debt-conversion sample may have different characteristics than those in the lower earning regular tax sample. We want to address the fact that differences in observable characteristics, such as occupation, may partially explain differences in bunching behavior due to, for example, differences in labor supply adjustment frictions.

We first define the individual-level elasticity as

$$e_i = \left[\underbrace{\frac{1[y_i \in BR_s] - \hat{P}^{cf}(y_i \in BR_s)}{\hat{P}^{cf}(y_i \in BR_s)/N_s^{\text{BR bins}}}}_{b_i} \cdot 1 \Big/ y_s^* \right] \Big/ (\widehat{\Delta\tau_s}/(1-\tau_s)), \tag{50}$$

where $\hat{P}^{cf}$ denotes the estimated (counterfactual) probabilities of being in the bunching region in the absence of any tax or debt conversion kinks. This is estimated using the frequencies in the earnings bins around the bunching region as in Saez (2010). $N_s^{\text{BR bins}}$ is the number of bins in the bunching region. b_i as defined by the curly braces is the individual-level excess mass, the sample-specific mean of which provides an estimate of the relative excess mass at the given threshold (see Appendix D.2). By multiplying further by the ratio of bin widths to the mean threshold value and dividing by the relative change in the after-tax wage, we obtain e_i. The s sample mean, $\hat{E}$, of e_i provides an estimate of the implied earnings elasticity. For the delayed tax sample, this mean is about 0.0155, which is very close to our baseline estimate of 0.0162.[35]

We then estimate regression equations of the following form.

$$e_i = \alpha + \beta \mathbb{1}[\text{regular tax sample}]_i + \gamma' X_i + \varepsilon_i, \tag{51}$$

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

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

[35] 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 9 shows that this is likely to be a very conservative estimate. Differences arise because in the regression-based approach we take the simpler approach of estimating P^{cf}s using the observed number of observations in the two income bins just below and the two income bins just above the bunching region, BR_s (as in Saez 2010), rather than estimating a higher order polynomial (as in Chetty et al. 2011).

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

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

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

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

We note that the methodology developed here can be used in other settings where one wishes to compare responses to different kinks. It is important to note that the standard definition of relative excess mass, b, produces a measure that is not invariant with respect to bin width. The implied elasticity, however, is invariant because y^* is expressed in bin-width units.

6 Discussion

When workers are unable to borrow against future earnings, the timing of cash flows (and especially tax payments) becomes an important policy tool. This is recognized in the existing literature on age-dependent taxation, but not fully exploited. We propose a simple policy tool, delayed taxation, that both addresses the welfare losses from credit market imperfections and exploits them by reducing the distortionary effects of income taxation. Delayed taxation allows agents to delay the payment, but not the accrual, of income taxes. This decoupling is a novel feature of our study. From a horizontal equity perspective, it is attractive in that substantial welfare gains can be achieved without conditioning tax rates on taxpayer characteristics such as age.

Our numerical analyses show that the welfare gains can be substantial. Overall, the results highlight delayed taxation as a promising new tool in optimal taxation and a fertile ground for further theoretical and empirical research. Towards the latter, we make some progress by studying a *de-facto* delayed tax scheme affecting young workers in Norway. The empirical results confirm one of the key mechanisms in our model, namely that delayed taxation reduces the distortionary effects of income taxation when workers are financially constrained.

References

Aiyagari, S. R. (1995): "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting," *Journal of Political Economy*, 103, 1158–1175.

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.

Asatryan, Z. and A. Peichl (2017): "Responses of firms to tax, administrative and accounting rules: Evidence from Armenia," .

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.

Barro, R. J., M. Darby, M. Feldstein, and A. Munnell (1979): "Social security and private saving: another look," *Social Security Bulletin*, 42, 33.

Bastani, S. (2025): "The marginal value of public funds: a brief guide and application to tax policy," *International Tax and Public Finance*, 32, 919–956.

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.*

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2017): "Does front-loading taxation increase savings? Evidence from Roth 401 (k) introductions," *Journal of Public Economics*, 151, 84–95.

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.

Best, M. C., J. S. Cloyne, E. Ilzetzki, and H. J. Kleven (2020): "Estimating the elasticity of intertemporal substitution using mortgage notches," *The Review of Economic Studies*, 87, 656–690.

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.

Bruze, G., A. K. Hilsløv, and J. Maibom (2024): "The Long-Run Effects of Individual Debt Relief," Tech. rep., CESifo.

Chetty, R. (2012): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80, 969–1018.

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.

Choukhmane, T. and C. Palmer (2023): "How Do Consumers Finance Increased Retirement Savings?" *Unpublished Working Paper*.

Conesa, J. C. and D. Krueger (2006): "On the optimal progressivity of the income tax code," *Journal of Monetary Economics*, 53, 1425–1450.

Coven, J., S. Golder, A. Gupta, and A. Ndiaye (2024): "Property Taxes and Housing Allocation Under Financial Constraints," Tech. rep., CESifo Working Paper.

da Costa, C. E. and M. R. Santos (2018): "Age-Dependent Taxes with Endogenous Human Capital Formation," *International Economic Review*, 59, 785–823.

Dávila, E. and B. Hébert (2023): "Optimal corporate taxation under financial frictions," *The Review of Economic Studies*, 90, 1893–1933.

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

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

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

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

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

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

Eaton, J. and H. S. Rosen (1980): "Optimal redistributive taxation and uncertainty," *The Quarterly Journal of Economics*, 95, 357–364.

Einav, L., M. Jenkins, and J. Levin (2012): "Contract pricing in consumer credit markets," *Econometrica*, 80, 1387–1432.

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.

Finkelstein, A. and N. Hendren (2020): "Welfare analysis meets causal inference," *Journal of Economic Perspectives*, 34, 146–167.

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.

Glogowsky, U. (2021): "Behavioral responses to inheritance and gift taxation: Evidence from Germany," *Journal of Public Economics*, 193, 104309.

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. (2005): “Fiscal policy with heterogeneous agents and incomplete markets,” *The Review of Economic Studies*, 72, 161–188.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): “Optimal tax progressivity: An analytical framework,” *The Quarterly Journal of Economics*, 132, 1693–1754.

——— (2020): “Optimal progressivity with age-dependent taxation,” *Journal of Public Economics*, 189, 104074.

Hendren, N. and B. Sprung-Keyser (2022): “The case for using the MVPF in empirical welfare analysis,” Tech. rep., National Bureau of Economic Research.

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.*

Holmström, B. and J. Tirole (1998): “Private and public supply of liquidity,” *Journal of political Economy*, 106, 1–40.

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. R. and A. S. Myhre (2021): “Labor supply responses to learning the tax and benefit schedule,” *American Economic Review*, 111, 3733–3766.

Kremer, M. (2002): “Should Taxes be Independent of Age,” Working 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.

Lucas, D. and D. Moore (2010): “Guaranteed versus direct lending: the case of student loans,” in *Measuring and managing federal financial risk*, University of Chicago Press, 163–205.

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

Rampini, A. A. and S. Viswanathan (2010): “Collateral, risk management, and the distribution of debt capacity,” *The Journal of Finance*, 65, 2293–2322.

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

Ring, M. A. (2024): “Wealth taxation and household saving: Evidence from assessment discontinuities in Norway,” *Review of Economic Studies*, rdae100.

Ring, M. A. and T. O. Thoresen (2025): “Wealth Taxation and Charitable Giving,” *Review of Economics and Statistics*, 1–45.

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

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

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

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.

Tazhitdinova, A. (2018): “Reducing evasion through self-reporting: Evidence from charitable contributions,” *Journal of Public Economics*, 165, 31–47.

——— (2020): “Do only tax incentives matter? Labor supply and demand responses to an unusually large and salient tax break,” *Journal of Public Economics*, 184, 104162.

Townsend, R. M. (1979): “Optimal contracts and competitive markets with costly state verification,” *Journal of Economic theory*, 21, 265–293.

Varian, H. R. (1980): "Redistributive taxation as social insurance," *Journal of public Economics*, 14, 49–68.

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

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

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

A Proofs

A.1 Proof of Proposition 1

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

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

Substituting in for g_t^i and rearranging yields

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

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

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

Equation (27) follows by re-arrangement and using the operator $\mathbb{E}_t[x] = \sum_i \pi^i y_t^i x,\ t = 1,2$. Condition (29) follows by noting that the the FOC for G can be written as:

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

A.2 Proof of Lemma 1

We assume that $R'(s^i)$ is well-defined and constant for marginal changes in economic incentives. When using the piecewise-linear parametric formulation of $R(s)$ where the marginal interest rate is higher for $s < 0$, we require that $s^i \neq 0$. Under these assumptions, 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. Define $\bar{\delta}^i = \delta + (1-\delta)\frac{1+r}{R'(s^i)}$. Using the expression derived in (E11), we can write:

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{u'(c_1^i)w_1^i\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-\ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\tilde{\tau}_1^i\right)}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-u_1''(c_1^i)w_1^{i2}\left(1-\tilde{\tau}_1^i\right)^2}. \tag{A5}$$

From this, we see that marginal changes in δ and τ_1 only affect l_1 through the term $\tilde{\tau}_1^i = \tau_1[1-(1-\delta)\Delta_r^i]$. To obtain the desired result, note that $\bar{\delta}^i = \delta + (1-\delta)\frac{1+r}{R'(s^i)} = \delta + (1-\delta)(1-\Delta_r^i) = 1-(1-\delta)\Delta_r^i$, so that $\bar{\delta}^i\tau_1 = \tilde{\tau}_1^i$. We can then write:

$$\frac{d\ell_1^i}{d(1-\delta)} = \frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)}\frac{d(\bar{\delta}^i\tau_1)}{d(1-\delta)} = \frac{d\ell_1^i}{d(\tilde{\tau}_1^i)}\frac{d\tilde{\tau}_1^i}{d(1-\delta)}, \tag{A6}$$

$$\frac{d\ell_1^i}{d(1-\tau_1)} = \frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)}\frac{d(\bar{\delta}^i\tau_1)}{d(1-\tau_1)} = \frac{d\ell_1^i}{d(\tilde{\tau}_1^i)}\frac{d\tilde{\tau}_1^i}{d(1-\tau_1)}. \tag{A7}$$

Since $\tilde{\tau}_1^i = \tau_1[1-(1-\delta)\Delta_r^i]$, we have $\frac{d\tilde{\tau}_1^i}{d(1-\delta)} = \tau_1\Delta_r^i$ and $\frac{d\tilde{\tau}_1^i}{d(1-\tau_1)} = -(1-(1-\delta)\Delta_r^i)$. Therefore:

$$\frac{d\ell_1^i}{d(1-\delta)} = \frac{\tau_1\Delta_r^i}{1-(1-\delta)\Delta_r^i}\frac{d\ell_1^i}{d(1-\tau_1)} = \Gamma^i(\delta)\frac{d\ell_1^i}{d(1-\tau_1)}, \tag{A8}$$

which establishes (30) for $t=1$.

For $t=2$, substitute the Euler equation (9) into the period-1 intratemporal FOC (11) to relate ℓ_1 and ℓ_2, differentiate to relate $d\ell_1$ and $d\ell_2$, and substitute into (A5). Then the above logic applies, since marginal changes in δ and τ_1 only affect ℓ_2 through $\tilde{\tau}_1^i$, yielding the $t=2$ case of (30).

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:

$$u_1'(c_1)w_1^i\left(1-\tau_1\bar{\delta}^i\right) = v'(\ell_1) \tag{A9}$$

$$\frac{u_1'(c_1)}{\beta(1+r_b)}w_2^i[1-\tau_2] = v'(\ell_2) \tag{A10}$$

$$c_1 + \frac{c_2}{1+r_b} = w_1\ell_1(1-\tau_1\bar{\delta}^i) + \frac{w_2\ell_2(1-\tau_2)+G_2}{1+r_b}. \tag{A11}$$

The first condition is just (11), obtained by inserting the intertemporal FOC (9) into (7), the second condition is obtained by inserting (9) into (8), and the third constraint is just the life-time budget constraint.[36] 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

$$\frac{d\ell_1}{d\tau_1} = \frac{d\ell_1}{d(\tau_1\bar{\delta}^i)}\frac{d(\tau_1\bar{\delta}^i)}{d\tau_1} = \bar{\delta}^i\frac{d\ell_1}{d(\tau_1\bar{\delta}^i)} \tag{A12}$$

$$\frac{d\ell_1}{d\delta} = \frac{d\ell_1}{d(\tau_1\bar{\delta}^i)}\frac{d(\tau_1\bar{\delta}^i)}{d\delta} = \tau_1\left(1 - \frac{1+r}{R'(s)}\right)\frac{d\ell_1}{d(\tau_1\bar{\delta}^i)} \tag{A13}$$

Substituting from (A12) into (A13), we get:

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

Since $\left(1 - \frac{1+r}{R'(s)}\right) = \Delta_r^i$ and $\bar{\delta}^i = 1 - (1-\delta)\Delta_r^i$, we can rewrite this as:

$$\frac{d\ell_1}{d\delta} = \frac{\tau_1\Delta_r^i}{1-(1-\delta)\Delta_r^i}\frac{d\ell_1}{d\tau_1} = \Gamma^i(\delta)\frac{d\ell_1}{d\tau_1}. \tag{A14}$$

Converting to the $(1-\delta)$ and $(1-\tau_1)$ forms (noting that $\frac{d\ell_1}{d(1-\delta)} = -\frac{d\ell_1}{d\delta}$ and $\frac{d\ell_1}{d(1-\tau_1)} = -\frac{d\ell_1}{d\tau_1}$), we obtain:

$$\frac{d\ell_1}{d(1-\delta)} = \Gamma^i(\delta)\frac{d\ell_1}{d(1-\tau_1)}, \tag{A15}$$

which establishes (30) for $t = 1$. The proof for ℓ_2 is analogous.

A.3 Proof of Proposition 2

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. From the budget constraint $\tau_1 y_1^i[1-(1-$

[36] The latter is derived by noticing that

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

$\delta)\Delta^i_{r,g}]$, the derivative with respect to $(1-\delta)$ gives both behavioral and mechanical effects:

$$\sum_i \pi^i \left(u'(c_1)\tau_1 y_1 - \beta u'(c_2)(1+r)\tau_1 y_1\right) + \\ \lambda \sum_i \pi^i \left(\tau_1[1-(1-\delta)\Delta^i_{r,g}]\frac{dy^i_1}{d(1-\delta)} + \tau_2 \frac{1}{1+r}\frac{dy^i_2}{d(1-\delta)} - \Delta^i_{r,g}\tau_1 y^i_1\right) = 0. \quad \text{(A16)}$$

We then assume $s^i \neq 0$ and use Lemma 1 to substitute behavioral responses. The coefficient on period-1 responses is $\tau_1[1-(1-\delta)\Delta^i_{r,g}]$, which after applying Lemma 1 becomes:

$$\left(\tau_1[1-(1-\delta)\Delta^i_{r,g}]\Gamma^i(\delta)\ \frac{dy^i_1}{d(1-\tau_1)} + \tau_2 \frac{1}{1+r}\Gamma^i(\delta)\ \frac{dy^i_2}{d(1-\tau_1)} - \Delta^i_{r,g}\tau_1 y^i_1\right). \quad \text{(A17)}$$

We use the Slutsky equation to rewrite $\frac{dy^i_1}{d(1-\tau_1)} = \left(\frac{dy^i_1}{d(1-\tau_1)}\right)^c + y_1 \frac{dy_1}{dG_1}$ and the cross-price Slutsky equation to rewrite $\frac{dy^i_2}{d(1-\tau_1)} = \left(\frac{dy^i_2}{d(1-\tau_1)}\right)^c + y_1 \frac{dy_2}{dG_1}$. The expression above becomes

$$\left(\tau_1[1-(1-\delta)\Delta^i_{r,g}]\Gamma^i(\delta)\left\{\left(\frac{dy^i_1}{d(1-\tau_1)}\right)^c + y_1\frac{dy_1}{dG_1}\right\} + \\ \tau_2\frac{1}{1+r}\Gamma^i(\delta)\left\{\left(\frac{dy^i_2}{d(1-\tau_1)}\right)^c + y_1\frac{dy_2}{dG_1}\right\} - \Delta^i_{r,g}\tau_1 y^i_1\right). \quad \text{(A18)}$$

Further rearranging and using elasticity notation yields

$$y^i_1\frac{1}{1-\tau_1}[1-(1-\delta)\Delta^i_{r,g}]\Gamma^i(\delta)\ \left(\left\{\tau_1\varepsilon^{i,c}_{1,1-\tau_1} + (1-\tau_1)\tau_1\frac{dy_1}{dG_1}\right\}\right. \\ \left. + \frac{1}{1+r}\left\{\tau_2\frac{y^i_2}{y^i_1}\varepsilon^{i,c}_{2,1-\tau_1} + (1-\tau_1)\tau_2\frac{dy_2}{dG_1}\right\}\right) - \Delta^i_{r,g}\tau_1 y^i_1. \quad \text{(A19)}$$

Further using the definitions $\rho^i = \frac{d}{dG_1}\left(\tau_1 y^i_1\left[1-(1-\delta)\Delta^i_{r,g}\right] + \frac{\tau_2 y^i_2}{1+r}\right)$, we may now rewrite the government's FOC with respect to $1-\delta$ as

$$\tau_1\sum_i \pi^i y^i_1\left(\frac{1}{\lambda}\Delta^i_r \cdot u'(c^i_1) + \frac{\Gamma^i(\delta)}{\tau_1}\left(\frac{\tau_1}{1-\tau_1}[1-(1-\delta)\Delta^i_{r,g}]\varepsilon^{i,c}_{1,1-\tau_1}\right.\right. \\ \left.\left. + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y^i_2}{y^i_1}\varepsilon^{i,c}_{2,1-\tau_1} + \rho^i\right)\right) \\ - \lambda\sum_i \pi^i\Delta^i_{r,g}\tau_1 y^i_1 = 0. \quad \text{(A20)}$$

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

$$\tau_1 \sum_i \pi^i y_1^i \left(g_1^i - g_2^i\right) = -\sum_i \pi^i \Delta_{r,g}^i \tau_1 y_1^i - \sum_i \pi^i y_1^i \left[\Gamma^i(\delta)\left(\frac{\tau_1}{1-\tau_1}[1-(1-\delta)\Delta_{r,g}^i]\varepsilon_{1,1-\tau_1}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^{i,c} + \rho^i\right)\right]. \quad \text{(A21)}$$

Using Lemma 1 yields (31).

A.4 Proof of Corollary 1

Using the derivations from the proof of Proposition 2 (see Appendix A.3), and setting $\tau_1 = \tau_2 = \tau$ and $\delta = 1$ as assumed in the corollary, we have:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\bigg|_{\delta=1} = \tau \mathbb{E}_1\left(g_1^i - g_2^i\right) + \mathbb{E}_1\left[\tau\Delta_r^i\left(\frac{1}{1-\tau}\varepsilon_{11}^{i,c} + \frac{1}{1+r}\frac{1}{1-\tau}\frac{y_2^i}{y_1^i}\varepsilon_{21}^{i,c} + \rho^i\right)\right] - \mathbb{E}_1[\tau\Delta_{r,g}^i]. \quad \text{(A22)}$$

Using the intertemporal optimality condition $g_2^i = g_1^i \frac{1+r}{R'(s^i)}$, we get $(g_1^i - g_2^i) = g_1^i\left[1 - \frac{1+r}{R'(s^i)}\right] = g_1^i \Delta_r^i$ for borrowers. Given the piecewise linear return technology with $R'(s^i) = 1+r$ when $s^i > 0$ and $R'(s^i) = 1+r_b$ when $s^i < 0$, and noting that welfare effects only arise from borrowers ($\Delta_r^i = 0$ for $s^i > 0$), we obtain:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\bigg|_{\delta=1} = \tau \sum_{i:s^i<0} \pi^i \left[\Delta_r^i y_1^i g_1^i + \Delta_r^i \frac{1}{1-\tau} y_1^i \varepsilon_{11}^{i,c} + \Delta_r^i \frac{1}{1+r}\frac{1}{1-\tau} y_2^i \varepsilon_{21}^{i,c} + \Delta_r^i y_1^i \rho^i - y_1^i\, \Delta_{r,g}^i\right]. \quad \text{(A23)}$$

Substituting $g_1^i = \frac{u'(c_1^i)}{\lambda}$ and rearranging to match the corollary statement completes the proof.

A.5 Proof of Lemma 2

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

$$\frac{dW}{dx} = \sum_i \pi^i \left(u'(c_1^i) - (1+r)\beta u'(c_2^i)\right) + \lambda \sum_i \pi^i \left(\frac{dy_1^i}{dx} + \frac{1}{1+r}\frac{dy_2^i}{dx}\right). \quad \text{(A24)}$$

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

$$\frac{dW}{dx} = \sum_i \pi^i \left(u'(c_1^i) - (1+r)\beta u'(c_2^i)\right) + \lambda \sum_i \pi^i \left(1 - \frac{1+r}{R'(s^i)}\right)\left(\tau_1 \frac{dy_1^i}{dG_1} + \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_1}\right). \quad \text{(A25)}$$

Further rearranging and using the fact that for government lending (with $\delta = 1$), our definition of ρ^i reduces to $\rho^i = \tau_1 \frac{dy_1^i}{dG_1} + \frac{\tau_2}{1+r}\frac{dy_2^i}{dG_1}$,

$$\frac{1}{\lambda}\frac{dW}{dx} = \sum_i \pi^i \left(g_1^i - g_2^i\right) + \sum_i \pi^i \left(1 - \frac{1+r}{R'(s^i)}\right)\rho^i. \tag{A26}$$

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

A.6 Proof of Proposition 3

Multiplying (33) in Lemma 2 by $\bar{y}_1 = \sum_{i:s^i<0} \pi^i y_1^i$ yields:

$$\frac{\bar{y}_1}{\Delta_r}\frac{dW}{\lambda dx} = \bar{y}_1 \sum_{i:s^i<0} \pi^i (g_1^i + \rho^i). \tag{A27}$$

From Corollary 1, we have:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\bigg|_{\delta=1} = \tau \sum_{i:s^i<0} \pi^i \Big[\Delta_r y_1^i g_1^i + \Delta_r \frac{1}{1-\tau}\left(y_1^i \varepsilon_{11}^{i,c} + \frac{1}{1+r} y_2^i \varepsilon_{21}^{i,c}\right) + \Delta_r y_1^i \rho^i\Big] - \tau \sum_{i:s^i<0} \pi^i y_1^i \Delta_{r,g}^i. \tag{A28}$$

From Lemma 2, we know that:

$$\frac{1}{\lambda}\frac{dW}{dx}\bigg|_{\delta=1} = \Delta_r \sum_{i:s^i<0} \pi^i (g_1^i + \rho^i) - \Delta_r \sum_{i:s^i<0} \pi^i \Delta_{r,g}^i. \tag{A29}$$

Therefore:

$$\tau \bar{y}_1 \cdot \frac{1}{\lambda}\frac{dW}{dx}\bigg|_{\delta=1} = \tau \Delta_r \sum_{i:s^i<0} \pi^i \bar{y}_1 (g_1^i + \rho^i) - \tau \Delta_r \sum_{i:s^i<0} \pi^i \bar{y}_1 \Delta_{r,g}^i. \tag{A30}$$

Substituting this into the corollary expression and rearranging:

$$\begin{aligned}\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\bigg|_{\delta=1} &= \tau \bar{y}_1 \cdot \frac{1}{\lambda}\frac{dW}{dx}\bigg|_{\delta=1} + \tau \Delta_r \sum_{i:s^i<0} \pi^i \frac{1}{1-\tau}\left(y_1^i \varepsilon_{11}^{i,c} + \frac{1}{1+r} y_2^i \varepsilon_{21}^{i,c}\right) \\ &+ \tau \Delta_r \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1) g_1^i - \tau \Delta_r \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1)\rho^i + \tau \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1)\Delta_{r,g}^i.\end{aligned}$$

This establishes (35).

B Extension: Age-dependent taxation

Consider an age-dependent tax system, allowing $\tau_1 \neq \tau_2$ that may be combined with an age-dependent transfer system, allowing $G_1 \neq G_2$, while considering a fixed amount of delayed taxation $1 - \delta$.

Proposition 4 (Optimal Age-Dependent Taxation) *Consider the optimization program* (18)*, with some fixed value of* $1-\delta$.

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

$$\sum_i \pi_i y_1^i \left[\delta g_1^i + [1-\delta] g_2^i\right] = \sum_i \pi_i y_1^i \left[1 - \frac{\tau_1}{1-\tau_1}\left(\varepsilon_{11}^i + \frac{1}{1+r}\frac{\tau_2}{\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{21}^i\right)\right], \tag{B1}$$

$$\sum_i \pi_i y_2^i \left[g_2^i\right] = \sum_i \pi_i y_2^i \left[1 - \frac{\tau_2}{1-\tau_2}\left(\varepsilon_{22}^i + (1+r)\frac{\tau_1}{\tau_2}\frac{y_1^i}{y_2^i}\varepsilon_{12}^i\right)\right]. \tag{B2}$$

(i) The optimal transfers G_1 *and* G_2 *satisfy:*

$$\sum_i \pi^i g_1^i = \sum_i \pi^i \left[1-\rho^i\right], \tag{B3}$$

$$\sum_i \pi^i g_2^i = \sum_i \pi^i \left[1-\rho^i \frac{1+r}{R'(s^i)}\right]. \tag{B4}$$

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

$$-\sum_i \pi^i y_1^i \Big(\delta u_1'(\cdot) + [1-\delta]u_2'(\cdot)\beta(1+r)\Big) + \lambda \sum_i \pi^i \left[y_1^i + \tau_1 \frac{dy_1^i}{d\tau_1} + \frac{\tau_2}{1+r}\frac{dy_2^i}{d\tau_1}\right] = 0, \tag{B5}$$

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

$$-\sum_i \pi^i y_1^i \Big(\delta u_1'(\cdot) + [1-\delta]u_2'(\cdot)\beta(1+r)\Big) + \lambda \sum_i \pi^i y_1^i \Big[1 - \frac{\tau_1}{1-\tau_1}\varepsilon_1^i - \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^i\Big] = 0. \tag{B6}$$

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

$$-\mathbb{E}_1\left[\delta g_1^i + [1-\delta]g_2^i\right] + \mathbb{E}_1\left[1 - \frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^i - \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^i\right] = 0. \tag{B7}$$

Alternatively, we may write it as

$$\frac{\tau_1}{1-\tau_1} = \frac{\mathbb{E}_1\left[1-\delta g_1^i + [1-\delta]g_2^i\right]}{\mathbb{E}_1\left[\varepsilon_1^i\right]} - \frac{\tau_2}{1-\tau_1}\frac{\mathbb{E}_1\left[\frac{y_2^i}{y_1^i}\varepsilon_{2,1}^i\right]}{\mathbb{E}_1\left[\varepsilon_1^i\right]} = 0. \tag{B8}$$

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

$$-\sum_i \pi^i y_2^i \beta u_2'(\cdot) + \lambda \sum_i \pi^i \left[\tau_1 \frac{dy_1}{d\tau_2} + \frac{1}{1+r}\left(y_2^i + \tau_2 \frac{dy_2^i}{d\tau_2}\right)\right] = 0, \tag{B9}$$

which we may rewrite as

$$-\frac{1}{1+r}\mathbb{E}_1\left[\frac{y_2^i}{y_1^i} g_2^i\right] + \mathbb{E}_1\left[-\frac{\tau_1}{1-\tau_2}\varepsilon_{1,1-\tau_2}^i + \frac{1}{1+r}\left(\frac{y_2^i}{y_1^i} - \frac{\tau_2}{1-\tau_2}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_2}^i\right)\right] = 0. \tag{B10}$$

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

$$\sum_i \pi^i u_1'(\cdot) = \lambda \sum_i \pi^i \left[1 - \tau_1 \frac{dy_1^i}{dG_1} - \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_1}\right], \tag{B11}$$

$$\beta \sum_i \pi^i u_2'(\cdot) = \lambda \sum_i \pi^i \left[\frac{1}{1+r} - \tau_1 \frac{dy_1^i}{dG_2} - \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_2}\right]. \tag{B12}$$

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

$$\beta \sum_i \pi^i u_2'(\cdot) = \lambda \sum_i \pi^i \left[\frac{1}{1+r} - \left(\tau_1 \frac{dy_1^i}{dG_1} + \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_1}\right)\frac{1}{R'(s^i)}\right]. \tag{B13}$$

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. ■

The weighted average of the social weights on the LHS of (B1) reflects that the burden of a marginal increase in τ_1 is borne partly in period 1 and partly in period 2 when $1-\delta > 0$. If there is either no delayed taxation ($\delta = 1$) or no agents are financially constrained ($R'(s^i) = 1+r$ for all i), then the LHS simplifies to $\sum_i \pi_i y_1^i g_1^i$.[37] The RHS of (B1) and (B2) reflect that a marginal change in a tax rate in one period affects labor supply in *both* periods. The intertemporal substitution of labor supply in response to the age-dependent tax rates τ_1 and τ_2 is an important feature of our framework.

Equations (B3) and (B4) require that G_t, for $t = 1, 2$, are set so that the average social value of giving everyone an additional dollar in period t ($\sum_i \pi^i g_t^i$) is exactly equal to the resource cost of an additional dollar ($\sum_i \pi^i = 1$) minus the loss of tax revenue due to fiscal externalities (individuals reduce their labor supply when transfers are increased). Note again that $g_2^i = \frac{1+r}{R'(s^i)} g_1^i$. If $R'(s) = 1+r$ when $s > 0$ and $R'(s) > 1+r$ when $s < 0$, then if at least one agent borrows, we have $\sum_i \pi^i g_1^i > \sum_i \pi g_2^i$, working towards $G_1 > G_2$. Note that since $\frac{1+r}{R'(s^i)} \leq 1$, the negative income effects on tax revenue are generally less severe for period 2 labor supply than they are for period 1 labor supply. This is because financially constrained agents discount future cash flows

[37] Note that $g_2^i = \frac{1+r}{R'(s^i)} g_1^i$ by virtue of (9), thus the value of δ is irrelevant if $R'(s^i) = 1+r$ for all i.

at a higher rate.

C Extension: Delayed taxation with paternalism

Suppose that some subset of agents optimize according to $\beta^i = \beta < \beta^*$, where β represents the common discount factor of impatient agents and β^* is the government's discount factor used in welfare evaluation. Let V^i denote their decision-relevant utility, and V^{*i} the utility from the perspective of the social planner who uses β^* in welfare calculations (i.e., paternalism). V^i and V^{*i} differ whenever $\beta^i = \beta < \beta^*$. We describe agents with $\beta^i < \beta^*$ as impatient.

In these analyses, it matters whether agents who are indifferent towards receiving delayed taxes actually receive them. Hence, we assume that (marginal) indifferent agents do not (marginally) delay any taxes. This can be microfounded as follows. Assume that the interest rate on delayed taxes equals $r + \epsilon < r_b$, and that agents can choose whether to delay their taxes. With $\epsilon > 0$, those with $s^i > 0$ always strictly prefer not to delay taxes, while those with $s^i < 0$ strictly prefer to delay. If $\epsilon = 0$, those with $s^i > 0$ *weakly* prefer not to delay taxes, while net borrowers still strictly prefer to delay. We thus assume that ϵ is strictly positive (and deters indifferent agents from delaying taxes) but small enough not to otherwise affect welfare.

For brevity, we assume that there is no subsidized lending (i.e., $x = 0$). We can then rewrite the Lagrangian as

$$\begin{aligned} W^* &= \sum_i \pi^i V^i - \lambda^* \left(-\sum_i \pi^i \left(\tau_1 w_1^i \ell_1^i + \frac{\tau_2 w_2^i \ell_2^i}{1+r} - (1-\delta)\Delta_{r,g}^i \tau_1 w_1^i \ell_1^i \right) + G_1 + \frac{G_2}{1+r} + M \right) \\ &+ \sum_{i:\beta^i<\beta^*} \pi^i (V^{i*} - V^i), \end{aligned} \tag{C1}$$

where the last term is a quasi-error term equal to

$$(\beta^* - \beta) \sum_{i:\beta^i<\beta^*} \pi^i \left[u(c_2^i) - v(\ell_2^i) \right]. \tag{C2}$$

This term accounts for the fact that some agents optimize according to a rate-of-time preference, β, that is lower than that of the government, β^*.

The marginal welfare effect of increasing $(1-\delta)$ can be written as

$$\begin{aligned} \frac{dW^*}{d(1-\delta)}\Big|_{\delta=1} &= \tau_1 \cdot \sum_{i:R'(s^i)<1+r} \pi_i y_1^i \left(u'(c_1^i) - \beta^i (1+r) u'(c_2^i) \right) \\ &- \lambda^* \sum_i \pi_i y_1^i \left[\tau_1 \Delta_{r,g}^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{11}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{21}^{i,c} + \rho^i \right) \right] \\ &- \lambda^* \sum_i \pi^i \tau_1 w_1^i \ell_1^i \Delta_{r,g}^i \\ &+ (\beta^* - \beta) \sum_{i:\beta^i<\beta^*, R'(s^i)>1+r} \pi^i \left[u'(c_2^i) \frac{dc_2^i}{d(1-\delta)} - v'(\ell_2^i) \frac{d\ell_2^i}{d(1-\delta)} \right], \end{aligned} \tag{C3}$$

where we use FOC implied by Proposition 2 on all terms except the error term and value derivatives at $1-\delta=0$. In the error term, we have imposed that only those who strictly prefer to delay taxes ($R'(s^i)>1+r$) do so. This is relevant for the welfare of $\beta^i<\beta^*$ agents because the government strictly prefers that they do not delay taxes.

We rewrite the error term as

$$
\begin{aligned}
& (\beta^*-\beta) \sum_{i:\beta^i<\beta^*, R'(s^i)>1+r} \pi^i \left[u'(c_2^i)\frac{dc_2^i}{d(1-\delta)} - v'(\ell_2^i)\frac{d\ell_2^i}{d(1-\delta)} \right] \\
= & (\beta^*-\beta) \sum_{i:\beta^i<\beta^*, R'(s^i)>1+r} \pi^i \left[u'(c_2^i)\left\{ -(1+r)\tau_1 w_1^i \ell_1^i - (1+r)[1-\delta]\tau_1 w_1^i \frac{d\ell_1^i}{d(1-\delta)} \right\} - v'(\ell_2^i)\frac{d\ell_2^i}{d(1-\delta)} \right]. \\
= & (\beta^*-\beta) \sum_{i:\beta^i<\beta^*, R'(s^i)>1+r} \pi^i \left[u'(c_2^i)\left\{ -(1+r)\tau_1 w_1^i \ell_1^i \right\} - v'(\ell_2^i)\frac{d\ell_2^i}{d(1-\delta)} \right]. \\
= & (\beta^*-\beta) \sum_{i:\beta^i<\beta^*, R'(s^i)>1+r} \pi^i \left[u'(c_2^i)\left\{ -(1+r)\tau_1 y_1^i \right\} - v'(\ell_2^i)\frac{d\ell_2^i}{d(1-\delta)} \right]. \\
= & (\beta^*-\beta) \sum_{i:\beta^i<\beta^*, R'(s^i)>1+r} \pi^i u'(c_2^i) \left[\left\{ -(1+r)\tau_1 y_1^i \right\} - w_2^i(1-\tau_2)\frac{d\ell_2^i}{d(1-\delta)} \right]. \\
= & (\beta^*-\beta) \sum_{i:\beta^i<\beta^*, R'(s^i)>1+r} \pi^i u'(c_2^i) \left[\left\{ -(1+r)\tau_1 y_1^i \right\} - \Delta_r^i \tau_1 w_2^i(1-\tau_2)\frac{d\ell_2^i}{d(1-\tau_1)} \right]. \\
= & (\beta^*-\beta) \sum_{i:\beta^i<\beta^*, R'(s^i)>1+r} \pi^i u'(c_2^i) \left[\left\{ -(1+r)\tau_1 y_1^i \right\} - \Delta_r^i \tau_1 (1-\tau_2)\frac{y_2^i}{1-\tau_1}\varepsilon_{12}^i \right].
\end{aligned}
$$

We can now move the curly-bracket term above out the error term into the first welfare effect term, and thus rewrite the marginal welfare effect as

$$
\begin{aligned}
\frac{dW^*}{d(1-\delta)} = & \tau_1 \cdot \sum_{R'(s^i)>1+r} \pi_i y_1^i \Big(u'(c_1^i) - (\beta^i + [\beta^*-\beta^i])(1+r)u'(c_2^i) \Big) \\
& - \lambda^* \sum_i \pi_i y_1^i \left[\tau_1 \Delta_{r,g}^i \left(\frac{\tau_1}{1-\tau_1}\varepsilon_{11}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{21}^{i,c} + \rho^i \right) \right] \\
& - \lambda^* \sum_i \pi^i \tau_1 w_1^i \ell_1^i \Delta_{r,g}^i \\
& + (\beta^*-\beta) \sum_{i:\beta^i<\beta} \pi^i u'(c_2^i) \cdot \Delta_r^i \frac{\tau_1}{1-\tau_1}(1-\tau_2) y_2^i \big(-\varepsilon_{12}^i \big).
\end{aligned} \tag{C4}
$$

In the last term, the subsetting over $i: R'(s^i)>1+r$ is no longer needed due to $\Delta_r^i=0$ whenever $R'(s^i)=1+r$. Substituting in the expressions for $g_1^{i*}=u'(c_1^i)/\lambda^*$ and $g_2^{i*}=\beta^*(1+r)u'(c_2^i)$ and then substituting in for $u'(c_2^i)$ using the Euler equations, we may rewrite the welfare effects as

$$
\begin{aligned}
\frac{dW^*}{d(1-\delta)}\frac{1}{\lambda^*} &= \tau_1 \cdot \sum_{i:\beta^i<\beta^*, s^i<0} \pi_i y_1^i g_1^{i*}\left(1-\frac{1+r}{R'(s^i)}\frac{\beta^*}{\beta}\right) \\
&+ \tau_1 \cdot \sum_{i:\beta^i=\beta^*, R'(s^i)>1+r} \pi_i y_1^i g_1^{i*}\left(1-\frac{1+r}{R'(s^i)}\right) \\
&- \sum_i \pi_i y_1^i \left[\tau_1 \Delta_{r,g}^i \left(\frac{\tau_1}{1-\tau_1}\varepsilon_{11}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{21}^{i,c} + \rho^i\right)\right] \\
&+ \sum_i \pi^i \tau_1 y_1^i \Delta_{r,g}^i \\
&+ \frac{\beta^*-\beta}{\beta}\sum_{i:\beta^i<\beta} \pi^i \frac{g_2^{i*}}{R'(s^i)} \cdot \Delta_r^i \frac{\tau_1}{1-\tau_1}(1-\tau_2) y_2^i \left(-\varepsilon_{12}^i\right).
\end{aligned} \tag{C5}
$$

In the first term, welfare gains from intertemporal consumption smoothing among impatient borrowers decrease with paternalism. If the relative "paternalistic wedge", β^*/β, exceeds the interest rate wedge, $R'(s^i)/(1+r)$, then welfare gains from intertemporal smoothing are negative. This effect, however, is partially offset by being away from the envelope, and thus obtaining welfare gains from reduced labor supply in the second period (last term).[38] The government not only wants impatient workers to consume more in period 2—they also want workers to enjoy more leisure. Nevertheless, we recognize that there may be cases, with very strong paternalism (β^*/β is large and $\beta^i < \beta^*$ for many workers), where the total marginal welfare gains are negative. Interestingly, this then implies that *negatively* delayed taxation is optimal. Negatively delayed taxation is equivalent to forced saving (as in social security).

The preceding analysis allows us to formalize the key differences between paternalistic and non-paternalistic welfare evaluation of delayed taxation. The following proposition summarizes how equation (C5) differs from the baseline case.

Proposition 5 (Marginal Welfare Effects of DT under Paternalism) *Assume: (i) a subset of agents share a common discount factor $\beta < \beta^*$ (all others have $\beta^i = \beta^*$); (ii) no subsidized lending ($x = 0$); (iii) the interest rate on delayed taxes is $r + \epsilon$ with $\epsilon > 0$ small so indifferent savers do not delay; (iv) the marginal reform is evaluated at the benchmark with no delay ($\delta = 1$); and (v) $R'(s^i)$ is locally constant. Then the marginal change in social welfare from introducing delayed taxation is given by (C5) and differs from the non-paternalistic case in exactly two channels:*

1. ***Consumption smoothing.*** *Borrowers ($s^i < 0$) enjoy the same* experienced *gain proportional to $\left(1-\frac{1+r}{R'(s^i)}\right)$, but its* social *contribution is reweighted under paternalism:*

$$
g_1^{i*}\left(1-\frac{1+r}{R'(s^i)}\frac{\beta^*}{\beta}\right),
$$

which turns negative when $\frac{\beta^*}{\beta} > \frac{R'(s^i)}{1+r}$.

[38] This hinges on $\varepsilon_{12}^i < 0$, which must be the case. Both the substitution and income effects of lower period-1 taxes reduce period-2 labor supply with standard preferences.

2. ***Leisure gains.*** *Because the planner values future utility more ($\beta^* > \beta$), the planner prefers* less *period-2 labor than the agent. A marginal delay lowers period-1 distortions and, via $\varepsilon^i_{12} < 0$, reduces period-2 labor, creating an additional* positive *welfare term:*

$$\frac{\beta^* - \beta}{\beta} \sum_{i:s^i<0} \pi^i \frac{g_2^{i*}}{R'(s^i)} \Delta_r^i \frac{\tau_1}{1-\tau_1} (1-\tau_2)\, y_2^i\, (-\varepsilon^i_{12}),$$

which is positive because $\varepsilon^i_{12} < 0$ (a lower period-1 tax raises period-1 consumption and reduces period-2 labor).

All other behavioral-revenue and fiscal terms are identical to the non-paternalistic baseline and are omitted here for brevity; see (C5).

D Bunching Analysis Appendix

D.1 Additional figures

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 6 for further info on the methodology.

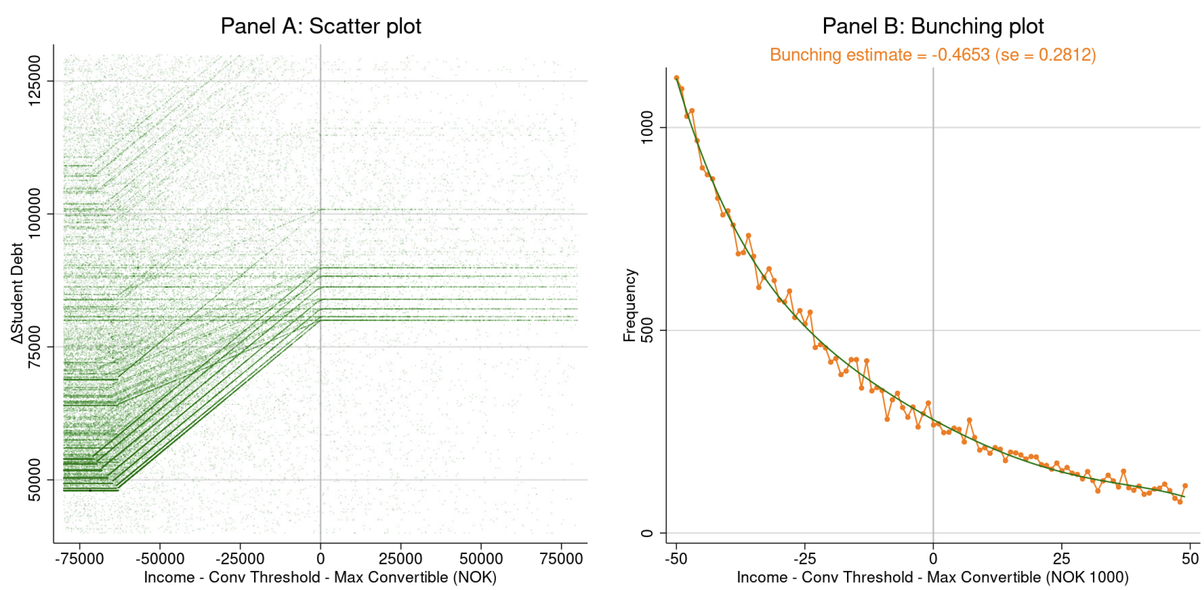
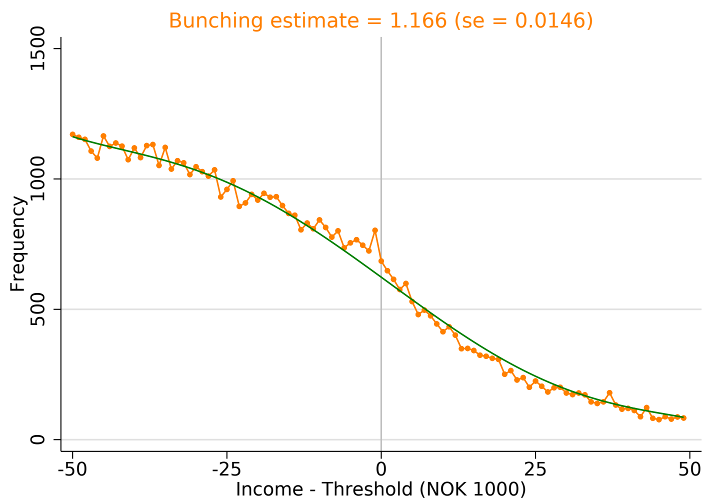


Figure A.2: Bunching at Debt-Conversion Threshold for Workers with Sales and Hospitality Occupations

We repeat the exercise in Panel (B) of Figure 6 on a subset of workers with hospitality (4-digit "STYRK-98"occupation code = 5123, waiters and bartenders) and store sales/clerk jobs (4-digit occupation code = 5221.



D.2 Individual-level bunching elasticities

We expand here on how individual level elasticities are constructed (see main text section 5.7).

In the standard bunching framework, b is the the relative excess mass in the bin right below the kink, where all bunchers in the (wider) bunching region are assumed to be in the one bin below the kink. We discuss how to map this more visual notion of excess mass into relative excess probabilities.

First, the estimated excess mass can be written

$$\hat{B} = \Big[P[y_i \in BR] \cdot N^{sample} - P^{cf}[y_i \in BR] \cdot N^{sample}\Big] \times \text{Bin width}. \tag{D1}$$

Then to obtain $\hat{b}$, we need to divide by the baseline mass. That is, the counterfactual mass of individuals at the single bin right below the kink.

$$\hat{b} = \frac{\Big[P[y_i \in BR] \cdot N^{sample} - P^{cf}[y_i \in BR] \cdot N^{sample}\Big] \times \text{Bin width}}{\frac{P^{cf}[y_i \in BR]}{N^{\text{BR Bins}}} \times N^{sample} \times \text{Bin width}}. \tag{D2}$$

Note that, in the denominator, we divide $P^{cf}[y_i \in BR]$ by the number of bins in the bunching region, $N^{\text{BR Bins}}$, to get an estimate of the probability that someone is in the single bin right below the kink. We simplify to get

$$\hat{b} = \frac{\Big[P[y_i \in BR] - P^{cf}[y_i \in BR]\Big]}{\frac{P^{cf}[y_i \in BR]}{N^{\text{BR Bins}}}}. \tag{D3}$$

From this we can define an individual level b_i as

$$b_i = \frac{\mathbb{1}[y_i \in BR] - P^{cf}[y_i \in BR]}{\frac{P^{cf}[y_i \in BR]}{N^{\text{BR Bins}}}}, \tag{D4}$$

which satisfies the property that

$$\widehat{\mathbb{E}}[b_i] = \hat{b}, \tag{D5}$$

namely that the sample average of b_i equals $\hat{b}$.

Now, in order to obtain an individual-level elasticity, we divide b_i by y^* (as in equation 41) and then further by the relative change in the after-tax wage.

E 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.

E.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.

$$dc_1 = d\ell_1 w_1(1 - \delta\tau_1) - \ell_1 w_1 d(\delta\tau_1) - ds. \tag{E1}$$

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

$$ds = \frac{1}{R'(s)}\left[dc_2 + \ell_1 w_1 d\left\{(1-\delta)(1+r)\tau_1\right\} + (1-\delta)(1+r)w_1\tau_1 d\ell_1 - (1-\tau_2)w_2 d\ell_2\right]. \tag{E2}$$

Substituting (E2) into (E1), and using the expression for $\bar{\delta}^i$, yields

$$dc_1 = d\ell_1 w_1(1 - \bar{\delta}^i\tau_1) - \ell_1 w_1 d(\bar{\delta}^i\tau_1) - \frac{1}{R'(s^i)}dc_2 + \frac{1}{R'(s^i)}(1-\tau_2)w_2 d\ell_2. \tag{E3}$$

Now we can differentiate the second period intratemporal FOC (8) to get $d\ell_2 = \frac{u''(c_2)}{v''(l_2)}w_2^i(1-\tau_2)dc_2$ and substitute this into (E3) and collect multiplicative terms on dc_2 to get

$$dc_1 = d\ell_1 w_1(1 - \bar{\delta}^i\tau_1) - \ell_1 w_1 d(\bar{\delta}^i\tau_1) - \frac{1}{R'(s^i)}\left(1 - [(1-\tau_2)w_2]^2\frac{u''(c_2)}{v''(l_2)}\right)dc_2. \tag{E4}$$

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

$$dc_1\left[1+\left(1-[(1-\tau_2)w_2]^2\frac{u''(c_2)}{v''(l_2)}\right)\frac{u''(c_1)}{\beta u''(c_2)}\frac{1}{R'(s)^2}\right] = d\ell_1 w_1(1-\bar{\delta}^i\tau_1) - \ell_1 w_1 d(\bar{\delta}^i\tau_1). \tag{E5}$$

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 11, which relies on the Euler equation 9):

$$u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)dc_1 - u'(c_1^i)w_1^i d\left(\bar{\delta}^i\tau_1\right) \quad = \quad v''(\ell_1^i)d\ell_1^i. \tag{E6}$$

We then substitute (E6) into (E5). We denote the term in the brackets in (E5) as ι^i.

$$\left[\frac{v''(\ell_1^i)d\ell_1^i}{u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)} + \frac{u'(c_1^i)w_1^i d\left(\bar{\delta}^i\tau_1\right)}{u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}\right]\iota^i = d\ell_1 w_1(1-\bar{\delta}^i\tau_1) - \ell_1 w_1 d(\bar{\delta}^i\tau_1). \tag{E7}$$

Then we re-arrange to obtain

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{-u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\iota^i - w_1^i(1-\bar{\delta}^i\tau_1)u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}. \tag{E8}$$

Then we re-arrange to obtain

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{-u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\iota^i - w_1^i(1-\bar{\delta}^i\tau_1)u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}. \tag{E9}$$

This may also be written as

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{-u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u_1''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i u_1''(c_1^i)\tilde{w}_1^i}. \tag{E10}$$

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

$$\frac{d\ell_1^i}{d(\bar{\delta}^i\tau_1)} = \frac{-u'(c_1^i)w_1^i\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right] - \ell_1^i w_1^i u_1''(c_1^i)w_1^i\left(1-\bar{\delta}^i\tau_1\right)}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right] - u_1''(c_1^i)w_1^{i2}\left(1-\bar{\delta}^i\tau_1\right)^2}, \tag{E11}$$

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

E.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.

$$dc_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)}((1-\tau_2)w_2 d\ell_2 - dc_2) \tag{E12}$$

(b) Now find an expression for $d\ell_2$. We use the implied intratemporal FOC for labor (14), and differentiate it to get

$$d\ell_2 = \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} d\ell_1. \tag{E13}$$

(c) Now find an expression for dc_2. We differentiate the Euler equation (9).

$$dc_2 = \frac{u''(c_1)}{\beta R'(s)u''(c_2)} dc_1 \tag{E14}$$

(d) Now find an expression for dc_1. We differentiate the intratemporal FOC for ℓ_1, (11), which relies on the intertemporal FOC, (9).

$$dc_1 = \frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1. \tag{E15}$$

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

$$dc_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)}\left((1-\tau_2)w_2 \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} d\ell_1 - \frac{u''(c_1)}{\beta R'(s)u''(c_2)} dc_1\right) \tag{E16}$$

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

$$\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)}\left((1-\tau_2)w_2 \frac{w_2(1-\tau_2)v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} d\ell_1 - \frac{u''(c_1)}{\beta R'(s)u''(c_2)} \frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} d\ell_1\right) \tag{E17}$$

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

$$\left(\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} - \tilde{w}_1 - \frac{1}{R'(s)}\left[\frac{[w_2(1-\tau_2)]^2 v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} - \frac{1}{\beta R'(s)u''(c_2)}\frac{v''(\ell_1)}{\tilde{w}_1}\right]\right) d\ell_1 = dG_1 \tag{E18}$$

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

$$\left(\frac{v''(\ell_1)}{u''(c_1)\tilde{w}_1} - \frac{1}{R'(s)}\left[\frac{[w_2(1-\tau_2)]^2 v''(\ell_1)}{\beta R'(s)\tilde{w}_1 v''(\ell_2)} - \frac{1}{\beta R'(s)u''(c_2)}\frac{v''(\ell_1)}{\tilde{w}_1}\right] - \tilde{w}_1\right) d\ell_1 = dG_1 \tag{E19}$$

$$\left(v''(l_1)\left[1 - u''(c_1)\frac{1}{\beta R'(s)^2}\left[\frac{[w_2(1-\tau_2)]^2}{v''(\ell_2)} - \frac{1}{u''(c_2)}\right]\right] - \tilde{w}_1 u''(c_1)\tilde{w}_1\right) d\ell_1 = u''(c_1)\tilde{w}_1 dG_1 \tag{E20}$$

$$\left(v''(l_1)\left[1+\left(1-\frac{[w_2(1-\tau_2)]^2u''(c_2)}{v''(\ell_2)}\right)\frac{u''(c_1)}{u''(c_2)}\frac{1}{\beta R'(s)^2}\right]-\tilde{w}_1u''(c_1)\tilde{w}_1\right)d\ell_1=u''(c_1)\tilde{w}_1dG_1 \tag{E21}$$

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

$$\frac{d\ell_1^i}{dG_1}=\frac{u''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-\tilde{w}_1^iu''(c_1^i)\tilde{w}_1^i}, \tag{E22}$$

or using the definition of ι^i,

$$\frac{d\ell_1^i}{dG_1}=\frac{u''(c_1^i)\tilde{w}_1}{v''(\ell_1^i)\iota^i-\tilde{w}_1u''(c_1^i)\tilde{w}_1^i}, \tag{E23}$$

E.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,

$$\frac{d\ell_1^i}{d\tilde{w}_1^i}=\left(\frac{d\ell_1^i}{d\tilde{w}_1^i)}\right)^c+\frac{d\ell_1^i}{dG_1}\ell_1^i. \tag{E24}$$

If we keep $\bar{\delta}^i$ fixed, $d\tilde{w}_1^i=-\bar{\delta}^iw_1d\tau_1=\bar{\delta}^iw_1d(1-\tau_1)$. Hence, the relevant Slutsky equation becomes

$$\frac{d\ell_1^i}{d(1-\tau_1)}=\left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c+\frac{d\ell_1^i}{dG_1}\ell_1^i\bar{\delta}w_1^i. \tag{E25}$$

Substituting in for the LHS using equation (E11) and the second-term on the RHS using (E23), we obtain

$$-\bar{\delta}^i\frac{-u'(c_1^i)w_1^i\iota^i-\ell_1^iw_1^iu_1''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i)\iota^i-\tilde{w}_1^iu_1''(c_1^i)\tilde{w}_1^i}=\left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c+\frac{u''(c_1^i)\tilde{w}_1}{v''(\ell_1^i)\iota^i-\tilde{w}_1u''(c_1^i)\tilde{w}_1^i}\ell_1^iw_1^i\bar{\delta}^i. \tag{E26}$$

Rearrange and cancel out to get

$$\left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c=\bar{\delta}^i\frac{u'(c_1^i)w_1^i\iota^i}{v''(\ell_1^i)\iota^i-\tilde{w}_1^iu_1''(c_1^i)\tilde{w}_1^i}. \tag{E27}$$

In terms of an elasticity, we can write it as

$$\varepsilon^c_{1,1-\tau_1}=\bar{\delta}^i\frac{1-\tau_1}{\ell_1^i}\frac{u'(c_1^i)w_1^i\iota^i}{v''(\ell_1^i)\iota^i-\tilde{w}_1^iu_1''(c_1^i)\tilde{w}_1^i}. \tag{E28}$$

We may rewrite this using the intratemporal FOC (11), which says that $u'(c_1^i)\tilde{w}_1^i=v'(\ell_1)$ and thus $u'(c_1^i)w_1^i = \frac{1}{1-\bar{\delta}^i\tau 1}v'(\ell_1)$ to get

$$\varepsilon^c_{1,1-\tau_1} = \frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\frac{1-\tau_1}{\ell_1^i}\frac{v'(\ell_1)\iota^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i u_1''(c_1^i)\tilde{w}_1^i}. \tag{E29}$$

Writing out the ι^i terms, we get

$$\varepsilon^c_{1,1-\tau_1} = \frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\frac{1-\tau_1}{\ell_1^i}\frac{v'(\ell_1)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-\tilde{w}_1^i u_1''(c_1^i)\tilde{w}_1^i}. \tag{E30}$$

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

$$\varepsilon^c_{1,1-\tau_1} = \frac{\frac{\delta+(1-\delta)\frac{1+r}{R'(s^i)}}{1-\left[\delta+(1-\delta)\frac{1+r}{R'(s^i)}\right]\tau_1}\frac{1-\tau_1}{\ell_1^i}v'(\ell_1)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]}{v''(\ell_1^i)\left[1+\left(1-\frac{[w_2^i(1-\tau_2)]^2u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{u''(c_2^i)}\frac{1}{\beta R'(s^i)^2}\right]-\left[w_1^i\right]^2\left(1-\tau_1\left[\delta+(1-\delta)\frac{1+r}{R'(s^i)}\right]\right)^2u_1''(c_1^i)}. \tag{E31}$$

F Data for calibration

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

Wages in 1990

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

winsorize at the 99.9th percentile.

Wages in 2011

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

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

Wage trajectories for calibration

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

Table A.1: Wage Trajectories for Calibration

This table provides summary statistics for the wage trajectories used in our calibration

	Panel A: Exogenous wage heterogeneity					
	p5	p25	p50	p75	p95	mean
w_1	0.53	0.80	0.95	1.20	1.75	1.03
w_2	0.76	1.38	1.73	2.55	5.13	2.16
w_2/w_1	0.80	1.45	1.94	2.69	4.15	2.16

Table A.2: Wage Trajectories for Calibration Using PSID Data

This table provides summary statistics for the wage trajectories used in our calibration to the U.S. PSID Data.

	Panel A: Exogenous wage heterogeneity				
	p5	p25	p50	p75	p95
w_1	0.42	0.70	0.97	1.35	2.06
w_2	0.33	0.68	0.99	1.45	2.49
w_2/w_1	0.30	0.59	1.02	1.68	3.30

G Additional quantitative results

Figure A.3: Marginal Borrowing Rates

This figure provides a scatter plot of $R'(s^i)$ against r_b, where each r_b corresponds to a different set of allocations optimal under the benchmark (age-independent, no delayed taxation, no subsidized lending) policy. The figure shows that, as r_b increases, more and more individuals respond by reducing their borrowing (and thus increase s^i) into the region where $R'(s^i)$ lies between r and r_b. Essentially, more and more individuals bunch at zero saving/borrowing, and the spread between nominal interest rate, r_b, and average marginal interest rates, $R'(s)$ becomes larger.

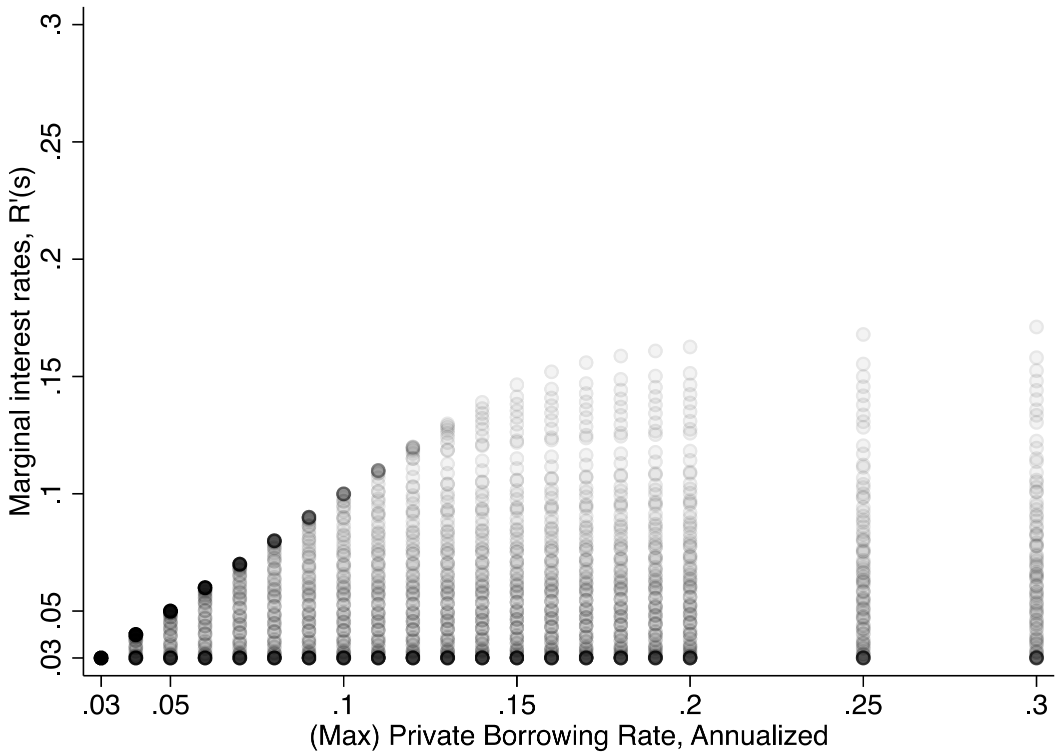


Table A.3: Extended Table: Optimal Tax Schedules and Allocations

This table provides summary statistics for different policy combinations as in Table 1. The first column provides statistics for the benchmark economy. The second column allows for delayed taxation. The third column allows for subsidized lending. The fourth column allows for subsidized lending *and* delayed taxation. The fifth and sixth columns consider an age-dependent baseline policy; and the seventh and eight consider an nonlinear-HSV policy. The present value function calculates present values according to $1 + r$. The (money-metric) welfare gain is the exogenous shock to revenue that the government must experience in the benchmark economy to be equally well off as with the given policy reform(s). This number is measured as a fraction of the aggregate period-1 consumption, C_1. Govt loan limit is the uniform subsidized loan given under the subsidized lending reform. B is the amount the government borrows to satisfy its first-period budget constraint. IL is the implied amount that the individual borrows from the government, including delayed taxes (see Appendix I).

	Tax schedule and allocations with $r_b = 20\%$								
	Benchmark	+ DT	Lending	+ DT	AD	+ DT	HSV	+ DT	
τ_1	0.65	0.68	0.65	0.68	0.36	0.40		.	.
τ_2	0.65	0.68	0.65	0.68	0.69	0.70		.	.
G_1	0.54	0.56	0.54	0.56	0.43	0.46	0.48	0.39	
G_2	0.54	0.56	0.54	0.56	0.43	0.46	0.48	0.39	
a_1 (HSV)	.	.	.	.	.	.	0.42	0.51	
a_2 (HSV)	.	.	.	.	.	.	0.42	0.51	
k_1 (HSV)	.	.	.	.	.	.	0.14	0.33	
k_1 (HSV)	.	.	.	.	.	.	0.14	0.33	
$1 - \delta$	0.00	0.31	0.00	0.31	0.00	1.00	0.00	1.00	
r_{dtax}	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	
Govt loan limit	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	
B	0.12	0.37	0.13	0.37	0.23	0.59	0.12	0.54	
Welfare Gain, % C_1									
(rel. to Benchmark)	0.00	2.21	0.40	2.21	7.66	9.62	0.10	3.87	
Δ Welfare	.	2.21	.	1.80	.	1.96	.	3.77	
means									
$l_1 w_1$	0.82	0.74	0.80	0.74	0.88	0.84	0.82	0.76	
$l_2 w_2$	1.68	1.91	1.71	1.91	1.68	1.79	1.66	1.88	
PV($l_1 w_1$,$l_2 w_2$)	1.72	1.77	1.73	1.77	1.78	1.80	1.71	1.77	
PV($l_1 w_1 \tau_1$,$l_2 w_2 \tau_1$)	1.11	1.20	1.12	1.20	0.94	1.01	1.11	1.14	
$(TaxPaid_1 - G_1)/y_1$	-0.08	-0.30	-0.09	-0.30	-0.17	-0.26	-0.07	-0.37	
$(TaxPaid_2 - G_2)/y_2$	0.26	0.36	0.27	0.36	0.36	0.36	0.26	0.41	
l_1	0.80	0.73	0.79	0.73	0.88	0.84	0.81	0.75	
l_2	0.83	0.92	0.85	0.92	0.80	0.83	0.82	0.90	
Net savings, s	-0.05	0.04	-0.07	0.04	0.03	0.32	-0.05	0.21	
Impl. govt loan amount, IL	0.00	0.04	0.00	0.04	0.00	0.02	0.00	0.17	
GDP	0.85	0.78	0.84	0.78	0.91	0.88	0.86	0.80	

H Micro-foundation for higher interest rates on borrowing

Below, we discuss a possible micro-foundation for a wedge between the interest rate on borrowing, r_b, and the interest rate on saving r. The environment we lay out allows workers to default on liabilities but supports a no-default equilibrium as in, e.g., Holmström and Tirole 1998, Rampini and Viswanathan 2010, Dávila and Hébert 2023. However,

where the possibility of (off-equilibrium-path) defaults imposes costs on lenders. The environment requires lenders to pre-pay for a collection technology that eliminates lender losses from default (not default itself), and this upfront cost determines the interest rate spread.

Environment. Agents can choose to borrow up to a limit of $\bar{a}$ in period 1. They then owe a payment of $a(1+r_b)$ in period 2. Agents can choose to default on any payments (on, e.g., private loans, delayed taxes, loans from the government).

Private lenders can pay an up-front period-1 fee to seize the agents' labor incomes in the event that the agent does not repay (with interest) in period 2. This fee is equal to $z \times a$, where $z \geq 0$ is a fee parameter. This technology cost is proportional to the loan amount and independent of the rate, which yields a simple break-even condition. In Holmström and Tirole (1998) terminology, the lender must pre-pay to make the workers' income pledgeable. However, we acknowledge that the government cannot commit to letting any defaulting agent endure zero consumption. Instead, we assume that the government will provide some non-seizable monetary transfer, $\underline{Y}$, to any agent that defaults.[39] A natural candidate for $\underline{Y}$ is the second-period transfer, G_2. This transfer amount does not require the government to condition transfers on whether borrowers default.

Private lending break-even condition. We assume that private lending is competitive and thus subject to a break even condition for each dollar of lending that accounts for both the time-value of money (r) and the upfront cost (z). This equation determines r_b. Assuming that the lender pays the upfront fee and there are no (ex-post) defaults, we have that

$$(1+r)(1+z) = (1+r_b) \Leftrightarrow z = \frac{1+r_b}{1+r} - 1 = \Delta_r, \tag{H1}$$

where $1+r$ is the time value of money. This term is multiplied by $1+z$ which is the money not saved in period 1 due to lending and paying the upfront cost.

The setup so far also resembles costly-state-verification models (Townsend, 1979). Lenders must pay a cost, which is reflected in the interest rate, to seize assets. The fact that lenders can seize assets disincentivizes default. The key difference in our setup is that this cost is due up front (before and regardless of whether default occurs), which is how we obtain a credit penalty without default occurring on the equilibrium path.

The government's cost of capital. We assume that the government faces a fee parameter, z^g, potentially different from z. The government can choose to not break even on a loan (or delayed taxes), but their cost of capital—the interest rate with which they discount future repayment in a no-default equilibrium—is determined by

$$(1+r)(1+z^g) = (1+r_b^g). \tag{H2}$$

The left-hand side is the opportunity cost (in period-2 money) of lending \$1 today, which involves giving up $1+z^g$ today. Had $1+z^g$ been saved, the government would receive $(1+r)(1+z^g)$ in period 2.

No-default equilibrium with private borrowing. Suppose that there is no delayed

[39] We could alternatively assume that they may work but be subject to a high marginal tax rate, but the subsistence transfer assumption is analytically simpler.

taxation ($1-\delta=0$.) If the lender(s) pay the upfront collection fee, the agent does not supply labor when they default, because all earnings would be seized. Hence, unlike Holmström and Tirole (1998) and related work by Rampini and Viswanathan (2010), we only have "legal pledgeability". Workers are still free to effectively divert future labor income by choosing not to supply any labor. Thus, in the event of default, in period 2, workers only consume the non-seizable transfer, $\bar{Y}$ and have no disutility from supplying labor. Accordingly, their second-period utility is $u(\underline{Y})$. Anticipating this, defaulting agents borrow the maximal amount, $\bar{a}$ in period 1. Ensuring no default thus requires that

$$V(\tau_1, \tau_2, G_1, G_2, \bar{a} \mid \delta = 1) \geq u(G_1 + \ell_1^d w_1(1-\tau_1) + \bar{a}) - v(\ell_1^d) + \beta u(\underline{Y}), \tag{H3}$$

where V is their life-time utility when not defaulting and ℓ_1^d is their period-1 labor supply if they choose to default in period 2. Note that V potentially depends on $\bar{a}$.

The existence of an equilibrium without default and where $\bar{a}$ never binds intuitively depends on $\underline{Y}$. If $\underline{Y}$ were zero, agents would never default to avoid negative-infinity utility. We verify below (see 73), using the allocations from the simulations and solving for optimal choices given default, that there exists economically meaningful (i.e., "generous") values for $\underline{Y}$ that allow for an equilibrium where there are no defaults and the borrowing limit, $\bar{a}$, does not bind.

For example, when $r_b = 20\%$, setting $\underline{Y}$ equal to the government transfer amount, $G = 0.5134$, implies that the smallest $\bar{a}$ that ensures no defaults is equal to 12.5755, whereas the maximal borrowing amount is 0.0895 ($\min_i s^i = -0.0895$).

Note that we can obtain a no-default-lending equilibria even when de-facto pledgeability is zero due to the fact that diversion is costly (unlike, e.g., Holmström and Tirole 1998, Rampini and Viswanathan 2010). This utility cost arises because diverting (not working) reduces consumption, and the disutility of lower consumption exceeds the utility of not working.

No-default equilibrium with private borrowing plus delayed taxation. Now suppose there is delayed taxation ($1-\delta > 0$). We are interested in the conditions under which there is neither default on regular debt nor default on delayed taxes. Since all disposable income is assumed to be seized regardless of the type of default, borrowers either do not default or default on both debt and delayed taxes. Hence, the utility given default for an individual agent is similar to before. The only difference is that the effective tax rate under default equals $\delta\tau_1$ instead of τ_1.

$$V(\tau_1, \tau_2, G_1, G_2, \bar{a}, \delta) \geq u(G_1 + \ell_1^d w_1(1-\delta\tau_1) + \bar{a}) - v(\ell_1^d) + \beta u(\underline{Y}), \tag{H4}$$

This setup causes the right-hand-side value to increase (less taxes paid in period 1) which favors default, but delayed taxation also increases utility for borrowers which increases the left-hand-side (no-default) utility. Since delayed taxes are subject to a low interest rate of $3\% = r < r_b = 10\%$, the incentives to default on delayed taxes are much weaker than on regular debt. In addition, since delayed taxation crowds out private borrowing, there are also weaker incentives to default (since there is less high-interest debt to default on). Quantitatively, we find that the LHS increases by more. With delayed taxation, the lowest $\bar{a}$ that ensures no default increases. For example, in the same scenario as above ($r_b = 20\%$ and government borrowing is unrestricted), setting $\underline{Y} = G = 0.5259$ implies that the smallest $\bar{a}$ that ensures no defaults is equal to 15.9684, which is higher than the benchmark economy. We verify that this borrowing cap does not bind: the largest observed borrowing amount (private borrowing plus delayed taxes) is only 0.1306.[40]

No-default equilibrium with government lending. We now consider an environment without delayed taxation but with subsidized lending from the government, in which workers can borrow up to a uniform limit of $\bar{x}$. The no default condition says that

$$V(\tau_1, \tau_2, G_1, G_2, \bar{x}, \bar{a} \mid \delta = 1) \geq u(G_1 + \ell_1^d w_1(1 - \tau_1) + \bar{a} + \bar{x}) - v(\ell_1^d) + \beta u(\underline{Y}). \tag{H5}$$

The left-hand-side is now different because agents may borrow at a subsidized rate up to a cap of $\bar{x}$. The right hand side is different because defaulting borrowers will exhaust both the private borrowing cap, $\bar{a}$, and the government cap, $\bar{x}$. Ceteris paribus, a higher $\bar{x}$ increases the utility from defaulting (on the right-hand side), thus necessitating a lower private borrowing cap, $\bar{a}$.

When $r_b = 20\%$, and the government's own borrowing is unrestricted, $\underline{Y} = G = 0.5430$, and $\bar{x} = 0.1692$, we have that the necessary $\bar{a} + \bar{x} = 12.4969$. The largest amount of borrowing (including from both government and private lenders) is 0.1321. Hence, the $\bar{a}$ is not close to binding.

Agent's optimization when defaulting. As mentioned above, we verify that there exist economically meaningful values for $\underline{Y}$ that allow for no-default equilibria by solving for the optimal choices of an agent who chooses to default. Assuming additively separable preferences, an agent's utility when defaulting is given by

$$V^{\text{default}} = \max_{\ell_1} \frac{c_1^{1-\sigma} - 1}{1 - \sigma} - \xi \frac{\ell_1^{1+\frac{1}{k}}}{1 + \frac{1}{k}} + \beta u(\underline{Y}), \tag{H6}$$

$$\text{s.t. } c_1 = G_1 + (1 - \delta\tau_1)\ell_1 w_1 + \bar{a}. \tag{H7}$$

The agent optimally chooses not to work in the second period, since any income would be seized by creditors. Since the agent will default, they will always borrow up to the cap $\bar{a}$, hence $\bar{a}$ appears as unearned income in period 1. We assume that all (seizable) disposable income is seized regardless of whether the agent defaults on regular debt or delayed taxes. Hence, if an agent defaults, the agent defaults on both types of liabilities.

The FOC is

$$c_1^{-\sigma}(1 - \delta\tau_1)w_1 = \xi \ell_1^{1/k}, \tag{H8}$$

[40] Total borrowing amounts, for this comparison, are computed as $-min(s^i, 0) + IL^I$. Where IL^i is the implied loan amount from the government as defined in Appendix I.

which may be written

$$c_1 = \left[\frac{(1-\delta\tau_1)w_1}{\xi}\right]^{1/\sigma} \ell_1^{-\frac{1}{k\sigma}}. \tag{H9}$$

Hence, the optimal ℓ_1 is given by

$$\left[\frac{(1-\delta\tau_1)w_1}{\xi}\right]^{1/\sigma} \ell_1^{-\frac{1}{k\sigma}} = G_1 + (1-\delta\tau_1)\ell_1 w_1 + \bar{a}. \tag{H10}$$

This equation implicitly defines the optimal labor supply ℓ_1^d for a defaulting agent as a function of the model parameters. Using this solution, we can compute V^{default} and verify that for economically meaningful values of $\underline{Y}$, the no-default conditions in the preceding analysis are satisfied, ensuring that equilibria without default exist.

Government versus private lending costs. Our analyses focus on the case where $z^g = z$ and thus the government faces the same cost of capital in lending to workers as private lenders do. However, we find this to be a conservative baseline. A possible justification for $z^g < z$ is that public agencies already have tax collection systems in place that are either complementary to loan (and delayed tax) collection technology, or these systems have slack and are partially substitutable.

Lucas and Moore (2010) perform a detailed estimation of the costs associated with government involvement in the student loan market. In doing so, they highlight several potential sources for a $z^g - z$ wedge: no securitization fees, lower contracting costs (legal expenses), and higher recovery rates due to government collection technology (e.g., Treasury Offset Program and administrative wage garnishment). In the student loan context, they argue that larger loan amounts have lower default rates due to these borrowers attending more elite institutions. The analogy in our context is that delayed tax amounts are mechanically linked to pre-tax earnings levels, implying that delayed taxes are less risky (in a per dollar sense) than regular borrowing.

I Notes on numerical solutions and $R(s)$

The simplicity of some of our theoretical results rely on $R(s^i)$ being locally linear at s^i for all i. One way to satisfy this is by assuming that $R(s)$ is piecewise linear with a kink at $s = 0$ and no agents are located with $s^i = 0$. In our quantitative framework, we instead employ a continuously-differentiable $R(s)$ function that approximates such a piecewise linear function. This allows us to avoid computational issues with a kinked budget constraint and ensures that the first-order conditions (i.e., the Euler equation) always holds. Specifically, we define

$$R(s, r_b) = \int_0^s R'(\mathsf{s})d\mathsf{s} = (1+r)s - \frac{r_b - r}{m_1}\ln\left(\frac{1+e^{-m_1(s+m_2)}}{1+e^{-m_1 m_2}}\right), \tag{I1}$$

where the derivative with respect to the first argument, s, equals

$$R'(s, r_b) = 1 + \frac{r + \exp(-m_1\,(s + m_2))r_b}{1 + \exp(-m_1(s + m_2))}. \tag{I2}$$

Here, r_b is the borrowing rate parameter, m_1 is a smoothing parameter, and m_2 is a shifting parameter, and $\bar{x}$ is another shifting parameter that we will discuss shortly.

When $m_2 = 0$ and $m_1 \to \infty$, then these functions converge to their theoretical counterparts (for $s \neq 0$). However, given numerical constraints, we need to choose the smoothing parameter m_1 large but finite. We therefore set $m_1 = 100$. To ensure that virtually no one with $s > 0$ has $R'(s) > 1 + r$ when $\bar{x} = 0$, we set the shifting parameter $m_2 = 0.1$. The orange line in figure A.4 shows $R(s)$ when $r = 0.03$, $r_b = 0.10$, $m_1 = 100$, $\bar{x} = 0$ and $m_2 = 0.1$. With these assumptions, the nonlinearity primarily takes place in the interval $[-0.15, -0.05]$.

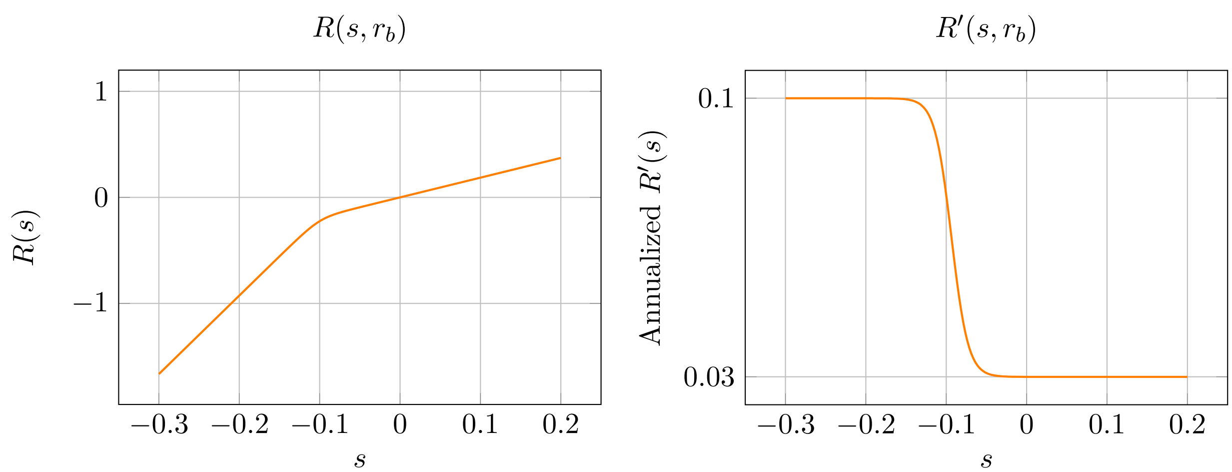


Figure A.4

Government's cost of capital. Appendix H provides the theoretical foundation for the interest rate on private loans (r_b) and the the government's cost-of-capital for providing loans or delayed taxation (r_b^g). For the government, we use eqs. (I1)–(I2) using r_b^g instead of r_b and letting $\bar{x} = 0$:

$$R(s, r_b^g, 0) \quad \text{and} \quad R'(s, r_b^g, 0). \tag{I3}$$

Fiscal costs of lending. We assume that the cost of lending L, either via a subsidized loan program or by delaying taxes, to a worker is given by

$$LC(L, s) = \int_{s-\bar{x}}^{s} \left(1 - \frac{1 + r}{R'(z, r_b^g)}\right). \tag{I4}$$

The first term inside the brackets equals the initial outlay. The second term is the present value of the marginal repayment. Repayment occurs at a gross interest rate of $1 + r$ and is discounted by the marginal interest rate $R'(s, r_b^g)$. This specification makes implicit assumptions about the (partial-equilibrium) timing. That is, agents first borrow in private markets, then s is determined, and then the government's lending costs are determined by s. Since $R'(s, r_b^g)$ is weakly decreasing in s, the more an agent borrows, the higher the government's lending costs. We write out the analytical expression for LC below.

$$
\begin{aligned}
LC(L, s) &= \int_{s-L}^{s} 1 - \frac{1+r}{R'(\mathtt{s}, r_b^g, 0)} d\mathtt{s} && \text{(I5)} \\
&= \int_{s-L}^{s} \left(1 - \frac{1+r}{1 + \frac{r + \exp(-m_1(\mathtt{s}+m_2)) r_b^g}{1+\exp(-m_1(\mathtt{s}+m_2))}} \right) d\mathtt{s} && \text{(I6)} \\
&= \frac{r_b^g - r}{m_1(1+r_b^g)} \ln \left(\frac{1 + r + (1+r_b^g)\exp(-m_1(s - L + m_2))}{1 + r + (1+r_b^g)\exp(-m_1(s+m_2))} \right). && \text{(I7)}
\end{aligned}
$$

Note that $LC(s, 0) = 0$. We plot $LC(s, d)$ below, fixing $L = 0.25$ and $r_b^g = r_b = 10\%$ (annualized). We see that $LC(s, 0.25) > 0$ for s equal to, e.g., 0.1. This occurs because the loan amount exceeds net savings (by about 0.15). Hence $s - L = -0.15$, and thus the (final dollar of) subsidized loans are discounted at $R'(-0.15, r_b^g) > r$. Also note that $LC(s, 0.25) \approx 0$ for large s. This is a desired property. For large s, individuals are indifferent about whether to take the loan. Hence, everyone would be equally well off if they did not delay, hence for large s, delayed taxation or subsidized loans should not impose any costs onto the government.

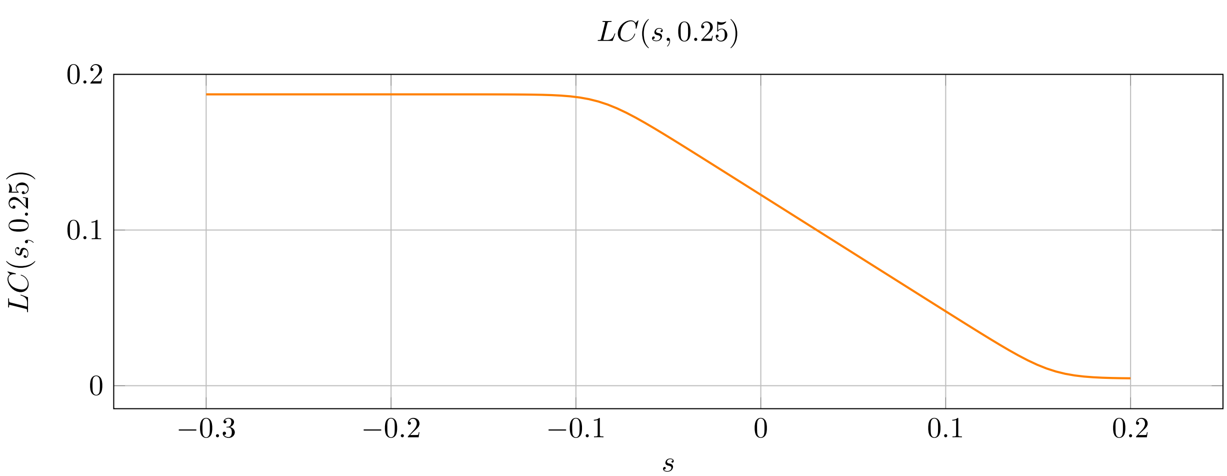


FIGURE A.5: Present-value cost of lending L=0.25 via either delayed taxation or subsidized lending

We also note that $LC(s, L)$ as defined above is consistent with a different implementation of delayed taxation. Suppose that the government instead guarantees the period-2 repayment of a loan of at most L. This removes default risk, and the lender only requires an interest rate of r. The marginal period-2 cash flows would be $1 + r$ (payment to private lender) and $1 + r$ (payment from agent). The payment to the lender would be discounted by $1 + r$ while the repayment from the agent would be discounted at $R'(s-L, r_b^g, 0)$. Simplifying the marginal costs, and taking the integral from $s-L$ to s produces the same formula, $LC(s, L)$.

Implied borrowing needs. When implementing restrictions on government borrowing, we require a measure of the implied loan amount relevant for calculating the government's borrowing needs. We define the implied loan amount, x, by equating values of interest subsidies for a liability $L = x+d$. In the hypothetical linear interest scheme case, the value is $x(r_b - r)$. In the simulations, the value equals the right hand side in the following expression.

$$IL(s, L)(r_b - r) = \int_{s-L}^{s} R'(z, r_b)dz - \int_{s-L}^{s} R'(z, r)dz, \tag{I8}$$

where the first term, on the left hand side, is the interest payment for someone with savings equal to s when nominal borrowing rates equal r_b. The second term is the interest payment when nominal borrowing rates equal r. This second term can be rewritten $(1+r)L$. We can thus write that

$$\begin{aligned} IL(s, L) &= \frac{\int_{s-L}^{s} R'(z, r_b)dz - (1+r)L}{r_b - r}, &\text{(I9)} \\ &= -\frac{1}{m_1}\ln\left(\frac{1+e^{-m_1(s+m_2)}}{1+e^{-m_1(s-L+m_2)}}\right). &\text{(I10)} \end{aligned}$$

This simplified expressions follow from letting the $(1+r)L$ and then the $r_b - r$ terms cancel out. This expression has some intuitively appealing properties. (i) For a fixed s, as L grows large, $IL(s, L)$ approaches L. That is, when someone borrows a large amount from the government, this amount equals the implied borrowing amount. (ii) For a fixed L, as s grows large, $IL(s, L)$ approaches zero. That is, when someone is a net saver, far from the interest kink, the implied loan amount is zero. These are the individuals who are indifferent to receiving subsidized loans or delayed taxes. (iii) As $s \to -\infty$, $IL(s, L) \to L$. That is, for deep borrowers, the implied borrowing amount equals L.

For the purpose of calculating government borrowing needs, we thus ignore tax-payment reductions and subsidized loan cash-flows in period 1 to calculate implied government borrowing,

$$B = -\sum_i \pi_i \left[\tau_1 y_1^i - G_1 - IL\Big(s^i, (1-\delta)\tau_1 y_1^i + x\Big) - M_1\right], \tag{I11}$$

where $x \geq 0$ is the uniform subsidized loan amount and M_1 is the exogenous period-1 revenue requirement.