Optimal Delayed Taxation in the Presence of Financial Frictions

Spencer Bastani and Marius A. K. Ring*

March 24, 2026 (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. 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 labor income taxation. We study delayed taxation in a dynamic optimal tax model, contrasting the welfare gains with the fiscal costs caused by income effects and defaults. We then characterize optimal delayed taxation in a simple calibrated model. This quantitative analysis reveals significant welfare gains, even when agents can default on their tax liabilities. These gains are robust to alternative bench mark tax systems such as age-dependent and progressive taxation. The key source of these gains is delayed taxation's ability to undo distortions caused by income taxation. Alternative credit market interventions aimed at reducing borrowing costs cannot achieve this. The final part of our paper provides quasi-experimental bunching evidence from a de facto delayed tax scheme in Norway that supports the key mechanism, namely that delaying taxation reduces income-tax distortions.

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, Deborah Lucas, James Choi, Håkan Selin, Guillermo Ordoñez, 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 regularly intervene in credit markets to help individuals spread the cost of home purchasing, entrepreneurship, and education over time. Yet, at the same time, governments impose distortionary labor income taxes that do not account for borrowing constraints. A natural question is whether these tax payments should be payable over a longer horizon. Would delaying tax payments reduce the welfare losses from financial frictions, and at the same time decrease the distortions from income taxation? To date, the public finance literature has largely overlooked this question.

That spreading payments over time reduces price sensitivities is well appreciated in other markets. Automakers, for example, routinely and profitably offer low-interest financing (Einav, Jenkins, and Levin, 2012). The key mechanism is that liquidity-constrained buyers become less price elastic when payments are spread over time. This mechanism also applies to the tax system. By allowing workers to delay their tax payments at a subsidized interest rate, the government may reduce the responsiveness of labor supply to taxation, and thus decrease the welfare losses from distortionary taxation. Governments have even stronger incentives to offer in-house financing than automakers because they internalize both the revenue effects and the welfare gains from increased consumption smoothing.

Credit market imperfections play a limited role in the optimal tax literature. While Weinzierl (2011) documents considerably larger welfare gains from age-dependent taxation when workers cannot save or borrow, the broader age-dependent tax literature (Kremer 2002, Lozachmeur 2006, Weinzierl 2011, Bastani et al. 2013, Heathcote et al. 2020) generally abstracts from financial frictions. Instead, this literature focuses on reducing tax distortions by tailoring tax rates to age-dependent labor supply elasticities. Financial frictions also play a limited role in other branches of the public finance literature. While models of optimal taxation under incomplete markets feature 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 primarily serve to exacerbate the welfare losses from incomplete insurance markets by limiting workers' ability to self-insure. 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, the government can only tax labor earnings rather than ability, and thus taxes distort labor supply. We show that allowing workers to delay their income tax payments both reduces welfare losses from financial frictions and leverages financial frictions to reduce tax distortions. The first effect arises because delayed taxation effectively allows individuals

to borrow at lower interest rates.[1] The second effect arises because constrained borrowers discount future repayments at a high marginal interest rate. This renders the present-value tax rate lower, which reduces income tax distortions.

To illustrate the effects of delayed taxation, consider a simple model with two groups of workers subject to a labor 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. This is exactly the friction Friedman (1957) identifies as the main objection to the permanent-income hypothesis. Suppose the government allows workers to delay their tax payments at the savings interest rate, r_{save}. Net savers are indifferent about delaying, while borrowers strictly prefer to. For a worker, i, who faces a marginal interest rate of r^i and an interest rate of r_{save} on taxes delayed for T periods, the statutory marginal tax rate τ translates into a present value tax rate of:

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

For net borrowers with $r^i > r_{save}$, setting $T > 0$ pushes the present-value tax rate below the statutory rate, thereby reducing labor supply distortions. By shifting the timing of tax *payments*, delayed taxation ties present-value tax rates to each individual's marginal borrowing rate.

Equation (1) relates delayed taxation to earlier work on "optimal elasticities." Slemrod and Kopczuk (2002) stress that behavioral responses to taxes depend "on the environment that individuals operate in, and may be manipulated by instruments controlled by the government." While Slemrod and Kopczuk focus on tax enforcement, delayed taxation follows their general insight. Many individuals operate in an environment characterized by high interest rates on borrowing, which equation (1) distills into a high r^i. The instrument controlled by the government is the extent of tax delay, T. Given high interest rates, r^i, governments may lower the sensitivity to the statutory tax rate, τ, by increasing T. Importantly, this allows the government to impose different effective tax rates on constrained and unconstrained workers without conditioning the tax schedule on observable taxpayer characteristics (i.e., "tagging"). This self-selection mechanism is an attractive feature of delayed taxation relative to explicitly discriminatory instruments such as age-dependent taxation.

Our study provides a comprehensive examination of delayed taxation through a blend of theoretical analysis, numerical simulations, and empirical evidence. We first introduce this policy into a dynamic optimal tax model with linear taxation, financial frictions, and heterogeneous workers who differ in labor market productivity. This extension of the Sheshinski (1972) framework allows for an intuitive decomposition of welfare effects into substitution and income effect terms (Harberger, 1971). We show that the optimal policy (i.e., the fraction of taxes that can be delayed) depends on standard behavioral elasticities of labor supply with respect to marginal tax rates and on the magnitude of financial frictions. Assessing the potential welfare gains from

[1] Delayed taxation is conceptually equivalent to 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.

delayed taxation does not require estimating new types of behavioral elasticities.

In our model, financial frictions manifest as a wedge between the interest rate on borrowing and the interest rate on saving. As the wedge grows large, the model nests the standard "no-borrowing" constraint (see, e.g., Heathcote 2005). This wedge arises because workers can strategically 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 thereby disincentivize default. Importantly, while workers may delay their taxes at the low interest rate on savings, there is no free lunch. By allowing delayed tax payments, the government incurs the same upfront costs on the delayed tax amount that private lenders incur when issuing loans. 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 increase the welfare gains of delayed taxation.[2]

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. We consider the case when individuals live and work for two periods and potentially face high interest rates on borrowing. 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 labor supply, which is (iii) partially offset by a negative intertemporal substitution effect on period-2 labor supply, (iv) a negative income effect on the present value of tax revenue due to relaxed financial constraints,[3] and (v) a negative present-value fiscal effect due to lending costs associated with default risk. We show that the distinguishing feature of delayed taxation is the second—the positive fiscal externality through reduced tax distortions. These welfare gains are not achievable from providing unconditional low-interest rate loans. They are further not attenuated if the government is paternalistic and places higher welfare weights on future consumption than workers. In fact, we show that paternalistic government preferences implies larger welfare gains from delayed taxation due to delayed taxation crowding out high-interest borrowing.

In the second part of the paper, we calibrate the optimal tax model to the Norwegian economy and solve for optimal tax schedules numerically. The Norwegian register data provide earnings trajectories for the entire population of workers. We consider all workers aged 20 to 30 in 1990 for whom we can compute effective wages in both 1990 and 2011.The resulting sample exhibits considerable variation in initial earnings and wage trajectories, and thus the extent to which workers wish to borrow against future earnings (as in, e.g., Heckman 1976).

Our baseline analyses consider the welfare gains from implementing delayed taxation on top of a benchmark economy in which taxes are linear and age-independent. Individuals can save at

[2]There are two reasons why this assumption is conservative. First, workers may face high interest rates in partly due to lenders charging high markups. Markups that are not compensation for default risk would not carry over as expected default costs to the government when offering credit-related products (e.g., delayed taxes). Second, to the extent that interest spreads are compensation for default risk, this default risk may be lower for the government than private lenders. For example, the government may have superior collection technology through, e.g., their tax administration, which may allow the government to be able to break even on loans with lower interest rates than private lenders.

[3]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.

an interest rate of 3% or borrow at an interest rate of 20%. We begin by evaluating the theoretical welfare decomposition using the calibrated model. We find robust marginal welfare gains from delayed taxation. While the main source of gains is enhanced intertemporal consumption smoothing, welfare gains from reduced income-tax distortions play a central role in making delayed taxation a desirable policy tool. The compensated effects from reducing income tax distortions equal about a quarter 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 baseline assumptions, the MVPF is 5.30, meaning that the money-metric welfare effects of marginally delaying taxation are 430% higher than its present-value cost. When the government has a cost advantage in lending, the MVPF is infinite, indicating that delayed taxation can be implemented as a Pareto-improving reform. The optimality of delayed taxation, ensured by an MVPF exceeding 1, is robust to varying both the severity of borrowing frictions and the potential costs of default.

Moving beyond marginal effects, our main quantitative exercise studies the optimal implementation of delayed taxation reform. At the baseline borrowing rate of 20%, the government optimally delays up to 39% of tax payments. With net savers opting out of delayed taxation, the average fraction of delayed taxes is modest at about 11%. However, each \$1 of delayed taxes delivers sizable money-metric welfare gains of \$0.38 on average. These welfare gains equal 2.1% of aggregate period-1 consumption. Welfare gains remain material when financial frictions are less severe: at a borrowing rate of 15%, welfare gains equal 1.9% of period-1 consumption.

Policies such as age-dependent taxation, nonlinear taxation, and subsidized lending may also address welfare losses from financial frictions. A key question is therefore whether delayed taxation produces material welfare gains when the government already employs these instruments. We consistently find this to be the case. When the government can implement subsidized lending alongside delayed taxation, it optimally refrains from lending, indicating that delayed taxation is a superior alternative. Incremental welfare gains from delayed taxation also remain high when the government can implement age-dependent taxation. A notable result arises with nonlinear marginal tax rates as in Heathcote, Storesletten, and Violante (2017, 2020). HSV-style tax progressivity adds minor welfare gains by itself but doubles the gains from delayed taxation.[4]

Our quantitative findings are robust across several dimensions. We show that calibrating our model to U.S. PSID data delivers comparable welfare gains. Since our baseline focuses on working life (as, e.g., Weinzierl 2011, Heathcote et al. 2020), we extend our model to include a retirement period. Including a retirement lowers borrowing demand and introduces complexities related to strategic default during retirement. Nevertheless, delayed taxation remains feasible and produces comparable welfare gains as our baseline model. We further show that our results are robust to varying the degree of financial frictions, and imposing constraints on government borrowing.

[4]The benchmark policy always includes unconditional transfers, so the pre-HSV benchmark already features progressivity in *average* tax rates, which reduces the potential welfare gains of implementing HSV.

In the final part of the paper, we empirically test the distinguishing feature of delayed taxation, namely that it reduces income tax distortions when workers are financially constrained. Conducting such a test requires quasi-experimental variation in both effective tax rates and the timing of tax payments. Few settings offer substantial variation in timing due to income tax withholding. We overcome this challenge by studying behavioral responses to a student debt conversion scheme in Norway, which constitutes a *de facto* delayed tax.

Norwegian students receive annual subsidized loans from the government to cover consumption expenditures. These loans are typically forgiven at the end of each year if the student remains in good standing. However, if the student has concurrent labor earnings in excess of about $17,000, each additional $1 of income increases their debt balance by 50 cents. Incomes above $17,000 are thus subject to a 50% marginal tax rate, where the payment can be delayed for up to 20 years at the subsidized student loan interest rate. This scheme effectively replicates delayed taxation and is thus well-suited for examining how financial frictions make delayed taxes less distortionary. Students are almost by definition financially constrained, with steep subsequent wage trajectories that generate strong demand for borrowing against future earnings. The effective tax rate increase at the threshold is also large and salient, with a nominal drop in the after-tax (and after-debt) wage from 75 to 25 cents. This sharp kink in the tax schedule thus lends itself naturally to a bunching framework for estimating behavioral responses to delayed taxes.

We find clear visual evidence that students bunch around the debt-conversion threshold. However, bunching responses are modest relative to the large after-tax wage reduction. Under the assumption that students value debt increases one-for-one with take-home wage reductions, our bunching analysis yields an implied elasticity of labor earnings of only 0.016. We consider two standard explanations for why the elasticity is low. First, income effects create a wedge between bunching estimators and structural elasticities. We provide a formal proof that the bunching estimator identifies a behavioral elasticity closely related to the Frisch elasticity. Applying our identification result, we find that income effects attenuate the bunching elasticity by 7.84% relative to the Frisch. Second, optimization frictions in labor supply may attenuate bunching estimators. Existing estimates of adjustment-friction attenuation imply a downward bias of 70% (Kostøl and Myhre, 2021). Accounting for both income effects and adjustment frictions increases our implied elasticity to 0.059, still well below the frictionless estimate of 0.30 found by Kostøl and Myhre (2021) for Norway.

We contend that the implied elasticity is low because students are borrowing-constrained and responding to a de facto delayed tax. For constrained students, the present value of the tax rate is much lower than the nominal value implied by the debt increase, as illustrated by equation (1). A natural concern is that optimization frictions are particularly prevalent among students. We consider this unlikely, since students generally have more flexible work arrangements than the general population. Nevertheless, we address the concern directly. Our main approach is to examine the bunching behavior of *students* at a regular tax threshold, allowing us to compare elasticities for delayed and regular taxation within the same population. Using the same bunching framework, we estimate an implied elasticity of 0.130, which is eight times larger than at the delayed tax threshold. Assuming similar optimization frictions across the two samples but allowing

for heterogenous income effects, the large difference in elasticities (0.016 versus 0.130) can be rationalized by students facing an average marginal interest rate of about 25%. A marginal interest rate of 25% implies a present-value tax rate far below the nominal rate, making students much less responsive to the de facto delayed tax. A remaining concern is that the kinks occur at different income levels. Using a regression-based approach that controls for observables such as occupation, age, and family background, we find a qualitatively similar difference in elasticities.

Our framework implies that sensitivity to delayed but not regular taxation is materially affected 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 high-interest unsecured borrowing. We do not find this pattern in the regular tax sample. We further examine how students' characteristics vary with their position relative to the debt conversion threshold. Non-bunchers (and their parents) have significantly lower liquid assets, but not lower future earnings. This is precisely what our framework predicts if non-bunching reflects binding borrowing constraints. We also find no evidence that parental education varies across the threshold in a way that could explain differences in bunching behavior, addressing the concern that less human capital as opposed to less liquidity is driving down the bunching elasticities.

Literature. This paper relates to the optimal tax literature on characteristic-dependent taxation, i.e., tagging (Akerlof, 1978). Delayed taxation reduces effective tax rates for financially constrained taxpayers. As taxpayers age and borrowing constraints no longer bind, 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, e.g., 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, analyze, and empirically test the 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. 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; Chetty, Friedman, Olsen, and Pistaferri 2011; Bastani and Selin 2014; Seim 2017; Asatryan and Peichl 2017; Tazhitdinova 2018; Søgaard 2019; Tazhitdinova 2020; Glogowsky 2021; Mavrokonstantis and Seibold 2022) and loan term thresholds (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; 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 absent from the existing literature. In taking intertemporal aspects seriously, we make an important contribution to this literature. We formalize the conjectures by Saez (2010) and Chetty (2012) that the bunching estimator approximately identifies the Frisch elasticity in a dynamic model. Our constructive proof shows how the approximation depends on income effects through the length of the planning horizon and structural preference parameters. Finally, this paper 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) and to research on how different tax instruments affect behavioral elasticities (Slemrod and Kopczuk 2002; Fack and Landais 2016; Kostøl and Myhre 2021; Ring and Thoresen 2021).

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 applies a bunching framework to a de facto delayed tax system in Norway to test the behavioral implications of our theoretical framework. Section 6 concludes.

2 The welfare gains of delayed taxation

2.1 The model environment

To focus on the key welfare trade-offs associated with delayed taxation, we incorporate financial frictions into a dynamic extension of the linear taxation framework (Sheshinski 1972, Dixit and Sandmo 1977). This approach allows us to highlight the roles of tax distortions and capital market imperfections, and to express welfare effects in terms of compensated elasticities and income effects (Harberger, 1971). We later consider delayed taxation in the context of a nonlinear tax schedule as in Heathcote, Storesletten, and Violante (2017, 2020).

We consider an economy with a single cohort 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. We thus focus on years in which individuals are in the labor market (as, e.g., Weinzierl 2011 and Heathcote et al. 2020). Appendix F extends this framework to include a retirement period.

The general structure of the tax and transfer schedules is given by

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

where $\tau_t \geq 0$ denotes the 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-1 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. Similarly, $x \geq$ represents a uniform loan from the government that is repaid at an interest rate of $1+r$. To ease notation, we assume that δ and x do not vary across agents. We discuss how letting agents opt out of delayed taxation or government loans matters when we introduce fiscal costs related to defaults.

2.2 Workers' problem

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

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

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

where $R(s)$ is the amount by which disposable income in the second period increases if the individual saves the amount $s^i \in \mathbb{R}$. We explain below how $R(s)$ embodies the financial frictions in our model.[6]

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. Financial frictions play a central role in our model by providing heterogeneity in marginal interest rates. We model financial frictions similarly to Hirshleifer (1958) by imposing higher interest rates on borrowing than on saving. Specifically, the return function, $R(s)$, satisfies $R'(s) \geq 1+r$ and $R''(s) \leq 0$. That is, marginal interest rates weakly exceed a savings interest rate, r, and the marginal interest rate is higher for borrowers. In our quantitative analysis, we construct $R(s)$ to be an approximately piecewise-linear function:

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

where $r_b > r$ is the interest rate deep borrowers face in private credit markets. Hence, for the purpose of describing our theoretical results, we describe borrowers ($s < 0$) as facing a marginal gross interest rate of $1+r_b > 1+r$. We describe savers as facing a marginal gross interest rate of $1+r$. However, our theoretical results apply to any $R(s)$ that has $R'(s) \geq 1+r$ and $R''(s) \leq 0$.

[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 attempt to introduce a poll tax in the United Kingdom between 1989 and 1990.

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

Note that $R(s)$ can be seen as a generalization of the standard "no-borrowing constraint" in macroeconomics (obtained when $r_b \to \infty$).

The kinked interest schedule from Hirshleifer (1958) has intuitive appeal but lacks theoretical underpinnings in a deterministic model. We therefore construct a microfoundation in Appendix A. Lenders must pay an upfront fee proportional to the loan amount to be able to seize income or assets from borrowers in the event of default. Imposing an exogenous fee to obtain costly borrowing mirrors costly-state-verification models (Townsend, 1979) in which lenders pay an ex-post fee to verify potential defaults. Our innovation is to require the fee to be paid ex-ante to address borrowers' inability to commit to repay. To impose realistic constraints on the ability to discourage default, we assume that lenders cannot seize the government transfer component of disposable income (G_2). Hence, some workers, particularly those with low period-2 wages, may wish to strategically default and consume $c_2 = G_2$. Conditional on defaulting, workers do not supply any labor. This implies that there is nothing for lenders to seize from defaulting workers. Accordingly, lenders only lend when they can ensure no default. In our framework, this is ensured by a uniform maximal borrowing amount that ensures that the life-time utility of not defaulting exceeds 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 assume that lending markets are competitive, which generates a break-even condition that determines the interest rate on borrowing, r_b:

$$(1+r)(1+z) = 1 + r_b, \tag{6}$$

where r is the savings interest rate and z is the upfront lending cost. These lending costs constitute a pure resource cost. The left hand side of (6) is the period-2 opportunity cost of lending \$1 in period 1. The right hand side is the repayment. Since no borrowers default in equilibrium, r_b depends only on the cost of avoiding default (z) and the interest rate on savings (r).

Worker's optimization program. 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 life-time utility (4) subject the budget constraints, (2) and (3). The first-order conditions are:

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

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

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

The FOC condition for ℓ_1 is central to our analyses. Individuals work up to the point where the marginal utility of more consumption equals the marginal disutility of working. Taxes affect this trade-off by reducing how much consumption that each unit of labor supply affords. The novel ingredient in our setting is that tax payments may be delayed at a gross interest rate of $1+r$. Hence, both period-1 and period-2 marginal utilities of consumption play a role in determining

this trade-off. Rearranging the FOC for ℓ_1 (7) makes the intertemporal aspect more salient:

$$u'(c_1^i)w_1^i\left(1-\tau_1\left[\delta+\left(1-\delta\right)\frac{\beta(1+r)u'(c_2^i)}{u'(c_1^i)}\right]\right)=v'(\ell_1^i), \tag{10}$$

which emphasizes that that delaying taxes (by reducing δ) scales down the effect of taxes, τ_1, when the period-1 marginal utility of consumption exceeds period-2 marginal utility.

Substituting the Euler equation (9) into (10), we obtain

$$u'(c_1^i)w_1^i\left(1-\underbrace{\tau_1\left[\delta+\left(1-\delta\right)\frac{1+r}{R'(s^i)}\right]}_{\text{Effective tax rate, }\tilde{\tau}_1}\right)=v'(\ell_1^i), \tag{11}$$

which implies that the nominal tax rate, τ_1, enter the optimization problem through the effective tax rate, defined as

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

where

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

is the (proportional) interest wedge faced by borrowers.[7]

The effective tax rate, $\tilde{\tau}_1$, 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 interest rate wedge, 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.

2.3 Government's optimization problem

Government lending costs. Like private lenders, the government faces an upfront lending cost in order to avoid default. 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. However, letting all borrowers default on their tax liabilities resembles age-dependent taxation in our model.[8] Hence, to clearly distinguish government credit market intervention from age-dependent taxation, 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

[7] Writing $\Delta_r^i=1-\frac{1+r}{R'(s^i)}=\frac{R'(s^i)-(1+r)}{R'(s^i)}$ makes it more clear that Δ_r^i is a *proportional* interest wedge.

[8] Allowing a worker to default on delayed taxes without penalty is equivalent to lowering their ex-ante pay-as-you-go tax rate.

interest rate, r_b^g, through

$$(1+r)(1+z^g) = 1 + r_b^g. \tag{14}$$

We next assume that marginal lending costs are zero for net savers ($s^i \geq 0$). This assumption is made for notational convenience and could be replaced with an assumption in which only borrowers (who face $R'(s^i) > 1+r$) select into delayed taxation.[9] Hence, an individual's marginal interest rate from a government cost-of-capital perspective is given by

$$R'^g(s) \approx \begin{cases} 1+r & \text{if } s \geq 0, \\ 1+r_b^g & \text{if } s < 0. \end{cases} \tag{15}$$

Since the government offers a subsidized interest rate, $r \leq r_b^g$, it incurs a marginal present-value cost for each loan. This marginal cost is equal to the interest rate wedge faced by borrowers from the perspective of the government,

$$\Delta_{r,g}^i = 1 - \frac{1+r}{R'^g(s^i)} \geq 0. \tag{16}$$

In our quantitative analyses, we assume the government faces the same up front lending costs as private lenders ($z^g = z$), implying that the government has no cost advantage in lending to households (i.e., $\Delta_{r,g}^i = \Delta_r^i$).

To keep the analyses tractable and isolate the core trade-offs, we adopt the following assumption:

Assumption 1 (Credit Markets) *The savings interest rate r and lending costs, z and z^g, are exogenous.*

Thus, similar to Lozachmeur (2006) and Weinzierl (2011), we treat financial frictions and interest rates as exogenous.[10] Consequently, when policies such as subsidized lending, age-dependent, or delayed taxation change net saving in the economy, there are no welfare-relevant effects through changes in the interest rate, r. One justification for exogenous interest rates is that saving and borrowing ultimately occurs in large international bond markets. Another justification is that

[9] The more correct assumption would be that agents can choose whether to delay their taxes (i.e., can choose 1-$\delta^i \in \{0, 1-\delta\}$) but only choose to delay when they strictly prefer to. This alternative assumption is strictly speaking the correct one since it is consistent with the microfoundations for the financial frictions (Appendix A). To deter defaults on delayed taxes, the government needs to incur upfront costs on *all* delayed taxes. When lending to net savers is costly, welfare is (discretely) higher when the government discourages net savers from delaying taxes (or taking subsidized government loans). Net savers can be discouraged by some vanishingly small fixed cost or interest premium on repayments. Since this cost or premium can be vanishingly small, we would consider the limit as it approaches zero, and all our results remain the same. However, this alternative approach would come at notational cost that we wish to avoid, and the same results are obtained by the imposing zero marginal lending costs for net savers.

[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.

the net effect of policies such as 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:

$$\max_{G_1, G_2, \tau_1, \tau_2, \delta, x} \sum_i \pi^i V^i, \tag{17}$$

subject to:

$$\sum_i \pi^i \left(\tau_1 w_1^i \ell_1^i + \frac{\tau_2 w_2^i \ell_2^i}{1+r} - (1-\delta)\Delta_{r,g}^i \tau_1 w_1^i \ell_1^i - \Delta_{r,g}^i x \right) \geq G_1 + \frac{G_2}{1+r} + K, \tag{18}$$

$$0 \leq \delta \leq 1, \tag{19}$$

where K 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 $(1-\delta)\Delta_{r,g}^i \tau_1 w_1^i \ell_1^i$ term is a present-value adjustment that accounts that a fraction, $1-\delta$ of period-1 taxes are paid in the next period.

A few remarks are in order. Dynamic models of optimal taxation often assume that the budget constraint is satisfied in expectation (as in equation 18). 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 3.6). Although 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.

2.4 Characterizing optimal tax policy

We now characterize the optimal solution to the government's problem outlined in (17). For this purpose, let λ denote the Lagrange multiplier associated with the government budget constraint (18). The Lagrangian is

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

[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.

It is useful to define the following auxiliary variables:

$$g_1^i = \frac{u_1'(\cdot)}{\lambda}, \tag{20}$$

$$g_2^i = \beta(1+r)\frac{u_2'(\cdot)}{\lambda}, \tag{21}$$

$$\varepsilon_{ts}^i = \frac{1-\tau_s}{y_t^i}\frac{dy_t^i}{d(1-\tau_s)}, \tag{22}$$

$$\varepsilon_t^i = \frac{1-\tau}{y_t^i}\frac{dy_t^i}{d(1-\tau)}, \tag{23}$$

$$\rho_1^i = \frac{d}{dG_1}\left(\tau_1 y_1^i\left[1-(1-\delta)\Delta_{r,g}^i\right] + \frac{\tau_2 y_2^i}{1+r}\right) \leq 0, \tag{24}$$

$$\rho_2^i = \frac{d}{dG_2}\left(\tau_1 y_1^i\left[1-(1-\delta)\Delta_{r,g}^i\right] + \frac{\tau_2 y_2^i}{1+r}\right) \leq 0, \tag{25}$$

where (20)–(21) 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 (22) is the uncompensated elasticity of period t income with respect to the period s net-of-tax rate. Equation (23) is the elasticity of period t income with respect to a change in $1-\tau$ (a change in the net-of-tax rate in both periods). Equations (24)–(25) define income effect parameters that represent the reduction in present-value taxes caused by an increase in lump-sum transfers in period 1 and period 2, respectively. When $R''(s) = 0$, $R'(s^i)\rho_2^i = \rho_1^i$.

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 (17) with $\tau_1 = \tau_2 = \tau$ and $G_1 = G_2 = G$, a fixed amount of delayed taxation $1-\delta$, and no subsidized lending ($x = 0$).*

(i) The optimal marginal tax rate τ satisfies

$$\sum_i \pi_i y_1^i\left[\delta g_1^i + (1-\delta)g_2^i\right] + \frac{1}{1+r}\sum_i \pi_i y_2^i\left[g_2^i\right] \tag{26}$$

$$= \sum_i \pi_i y_1^i\left[1-(1-\delta)\Delta_{r,g}^i\right]\left[1-\frac{\tau}{1-\tau}\varepsilon_1^i\right] + \frac{1}{1+r}\sum_i \pi_i y_2^i\left[1-\frac{\tau}{1-\tau}\varepsilon_2^i\right]. \tag{27}$$

(ii) The optimal per-period transfer G satisfies

$$\sum_i \pi^i\left(g_1^i + \frac{1}{1+r}g_2^i\right) = \sum_i \pi^i\left(1 + \frac{1}{1+r} - \rho_1^i - \rho_2^i\right), \tag{28}$$

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

Proof. See Appendix C.1. ■

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

The formulations in equations (26) and (28) 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 (26), with efficiency, as reflected on the right-hand side of the same equation. On the left hand side, delayed taxes ($1-\delta > 0$) shifts the dependency towards period-2 social values. On the right-hand side, the presence of delayed taxation reduces the importance of mechanical or behavioral responses to period-1 taxes. This is due to period-1 taxes having a lower present value when taxes are delayed and thus discounted more heavily due to potential default costs. We also see that the optimal level of the transfer, G, depends on a balance between equity, shown on the LHS of equation (28), and costs, shown on the RHS of equation (28). These costs are the direct mechanical costs of providing the transfer plus the negative revenue effects due to present-value income effects.

Proceeding to optimal delayed taxation, it is useful to characterize the relationship between delayed taxation and regular taxation. We define a conversion factor that measures how much the present-value 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*

$$\Gamma^i(\delta) \equiv \tau_1 \frac{\Delta_r^i}{1-(1-\delta)\Delta_r^i}. \tag{29}$$

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

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 $1-\tau_1$, up to a timing correction that arises when the marginal borrowing rate is not locally constant. This result allows us to characterize optimal delayed taxation in terms of standard labor supply elasticities.

Lemma 1 *For any $\delta \in [0,1]$ and tax rates (τ_1, τ_2), for $t \in \{1,2\}$,*

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

where $\Gamma^i(\delta)$ is the conversion factor defined in (29), $d\ell_t^i/d(1-\tau_1)$ is the response to a period-1 tax cut holding τ_2 fixed, and $\Psi_t^i(\delta)$ is the timing correction *defined by*

$$\mathbf{J}^i(\delta) \begin{pmatrix} \Psi_1^i \\ \Psi_2^i \\ \Psi_s^i \end{pmatrix} = \Delta \mathbf{b}_\Gamma^i(\delta) \equiv \mathbf{b}_{1-\delta}^i(\delta) - \Gamma^i(\delta)\, \mathbf{b}_{1-\tau}^i(\delta), \tag{31}$$

with $\mathbf{J}^i(\delta)$ the 3×3 Jacobian of the individual's FOC system (7)–(9) and $\mathbf{b}_{1-\delta}^i$, $\mathbf{b}_{1-\tau}^i$ the RHS vectors from differentiating with respect to $(1-\delta)$ and $(1-\tau_1)$, respectively (see Appendix C.2).

When $R''(s^i) = 0$, $\Psi_t^i(\delta) = 0$ for all δ, and the proportionality $d\ell_t^i/d(1-\delta) = \Gamma^i(\delta) \cdot d\ell_t^i/d(1-$

τ_1) *holds exactly.*

Proof. See Appendix C.2. ■

Lemma 1 follows from the fact that δ and τ_1 enter the period-1 intratemporal condition only through the same effective wedge, $\tilde{\tau}_1^i = \tau_1 \left[1 - (1-\delta)\Delta_r^i\right]$. When $R''(s^i) = 0$, $R'(s^i)$ is locally constant, and a marginal change in $1-\delta$ moves labor supply exactly like a marginal change in $1-\tau_1$, scaled by the conversion factor $\Gamma^i(\delta)$. When $R''(s^i) \neq 0$, marginal changes in δ and τ_1 create different savings responses: a marginal increase in $(1-\delta)$ shifts income from period 1 to period 2 via the delayed-payment obligation, while a marginal increase in $(1-\tau_1)$ of equivalent effective magnitude operates entirely through the effective tax rate $\tilde{\tau}_1^i$. The resulting difference in savings adjustments changes $R'(s^i)$, creating the timing correction $\Psi_t^i(\delta)$.

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 (17), given some fixed values of τ_1, τ_2, G_1, G_2 (not necessarily optimal). Assuming an interior solution for δ, the optimal share of delayed taxation, $1-\delta$, satisfies:*

$$\begin{aligned}\tau_1 \cdot \sum_i \pi^i y_1^i \left(g_1^i - g_2^i\right) = \underbrace{\sum_i \pi^i \Delta_{r,g}^i \tau_1 y_1^i}_{\text{Direct fiscal cost}} \\ \underbrace{- \sum_i \pi^i y_1^i \, \Gamma^i(\delta) \left(\frac{\tau_1 \left\{1-(1-\delta)\Delta_{r,g}^i\right\}}{1-\tau_1} \varepsilon_{11}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{21}^{i,c} \right)}_{\text{Tax revenue effects (compensated)}} \\ \underbrace{- \sum_i \pi^i \left(y_1^i \, \Gamma^i(\delta) \, \bar{\delta}^i \rho_1^i + \tau_1 \left\{1-(1-\delta)\Delta_{r,g}^i\right\} w_1^i \Psi_1^i + \frac{\tau_2}{1+r} w_2^i \Psi_2^i \right)}_{\text{Tax revenue effects (income)}}, \end{aligned} \tag{32}$$

where $\Psi_t^i(\delta)$ is the timing correction from Lemma 1. When $R''(s) = 0$, $\Psi_t^i(\delta) = 0$ and the income term reduces to $-\sum_i \pi^i y_1^i \Gamma^i(\delta) \bar{\delta}^i \rho_1^i$.

Proof. See Appendix C.3. ■

The LHS of (32) 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 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_1^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 (32) 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 (32) 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.

The marginal trade-offs emphasized by (32) fail to emphasize a key tension between the two welfare gains from delayed taxation. As more taxes are delayed, private borrowing is crowded out. Once some agents cease to borrow, their marginal interest rates equal that on savings, and thus $\Delta_r^i = 0$. Once $\Delta_r^i = 0$, delayed taxation no longer reduces the distortions from income taxation. This implies that the optimal share of delayed taxes, $1-\delta$, is not necessarily very large, and that the largest welfare gains from delayed taxation can be achieved from modest implementations.[13]

2.5 Welfare effects of marginal reforms

We now decompose the marginal welfare gains from introducing 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 $1-\delta$) from a starting point with no delayed taxation $(1-\delta = 0)$, no subsidized lending $(x = 0)$, and age-independent tax rates.

Corollary 1 (Marginal Welfare Effects of Delayed Taxation) *Consider the benchmark economy without delayed taxation ($\delta = 1$, $\tau_1 = \tau_2 = \tau$, $G_1 = G_2 = G$). The money-metric welfare effect of a marginal introduction of delayed taxation can be decomposed as:*

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \tau \sum_i \pi^i \Bigg[\underbrace{\Delta_r^i y_1^i \cdot \frac{1}{\lambda} \cdot u'(c_1^i)}_{\substack{\text{Welfare gains} \\ \text{from consumption} \\ \text{smoothing}}} + \underbrace{\Delta_r^i \frac{\tau}{1-\tau} y_1^i \varepsilon_{11}^{i,c} + \frac{1}{1+r}\Delta_r^i \frac{\tau}{1-\tau} y_2^i \varepsilon_{21}^{i,c}}_{\substack{\text{Tax revenue effects} \\ \text{(compensated)}}}$$

$$+ \underbrace{y_1^i \left(\rho_1^i - (1+r)\rho_2^i\right)}_{\substack{\text{Tax revenue effects} \\ \text{(income)}}} - \underbrace{y_1^i \, \Delta_{r,g}^i}_{\substack{\text{Direct fiscal cost of} \\ \text{government lending}}} \Bigg], \tag{33}$$

[13] In addition, even if all taxes are delayed $(1-\delta = 1)$, this may not be enough to completely crowd out borrowing. Hence even an economy with $1-\delta = 1$ may feature agents that face $\Delta_r^i > 0$.

where ρ_1^i and ρ_2^i are the income effects on tax revenues from period-1 and period-2 transfers, defined in (24)–(25). All welfare effects originate from borrowers since $\Delta_r^i = 0$ and $\rho_1^i - (1+r)\rho_2^i = 0$ for agents with $\Delta_r^i = 0$. When $R''(s) = 0$, $\rho_1^i = R'(s^i)\rho_2^i$ and the income term reduces to $\Delta_r^i y_1^i \rho_1^i$.

Proof. See Appendix C.4. ■

Corollary 1 decomposes the welfare effect into four channels. The first 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 captures *compensated tax revenue effects*: the lower effective tax rate encourages higher period-1 labor supply through substitution effects, partially offset by reduced period-2 labor supply. This is the channel through which delayed taxation may reduce the distortionary effects of income taxation (Harberger, 1971). The third reflects *income effects* on tax revenues: by delaying taxation, the government provides an implicit loan of τy_1^i in period 1 that must be repaid with interest $(1+r)\tau y_1^i$ in period 2. The period-1 transfer reduces labor supply (captured by $\rho_1^i < 0$), while the period-2 repayment obligation increases it (captured by $-(1+r)\rho_2^i > 0$). The fourth channel represents the *direct present-value cost to the government of delaying tax collection* ($\Delta_{r,g}^i \geq 0$). 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:*

$$\begin{aligned} \frac{1}{\lambda}\frac{dW}{dx}\Big|_{\delta=1} &= \sum_i \pi^i \left(g_1^i - g_2^i\right) + \sum_i \pi^i \left[\rho_1^i - (1+r)\rho_2^i\right] - \sum_i \pi^i \Delta_{r,g}^i \\ &= \sum_{i:s^i<0} \pi^i \left[\Delta_r^i g_1^i + \rho_1^i - (1+r)\rho_2^i - \Delta_{r,g}^i\right], \end{aligned} \tag{34}$$

where the second equality follows from the fact that only borrowers ($s^i < 0$) benefit from subsidized loans: for non-borrowers ($s^i \geq 0$), $g_1^i = g_2^i$, the behavioral effects of a loan vanish ($\rho_1^i - (1+r)\rho_2^i = 0$), and $\Delta_{r,g}^i = 0$. For borrowers, the intertemporal optimality condition implies $(g_1^i - g_2^i) = \Delta_r^i g_1^i$.

Proof. See Appendix C.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 three components: (i) a uniform lending effect that provides the same liquidity benefits as direct government loans, (ii) welfare gains from reducing income-tax distortions, plus (iii) three "error" terms that capture how delayed taxation and unconditional lending have different targeting properties.

Proposition 3 (Decomposing Delayed Taxation into Lending and Tax Effects) *Assume $s^i \neq 0$ 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:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\bigg|_{\delta=1} = \underbrace{\tau\bar{y}_1 \cdot \frac{1}{\lambda}\frac{dW}{dx}\bigg|_{\delta=1}}_{\text{(i) Lending effect}} + \underbrace{\tau \sum_{i:s^i<0} \pi^i \Delta_r^i \frac{\tau}{1-\tau}\left(y_1^i \varepsilon_{11}^{i,c} + \frac{1}{1+r} y_2^i \varepsilon_{21}^{i,c}\right)}_{\text{(ii) Reduced tax distortions}} \tag{35}$$
$$+ \underbrace{\tau \sum_{i:s^i<0} \pi^i \Delta_r^i (y_1^i - \bar{y}_1) g_1^i}_{\text{(iii) Loan targeting}} + \underbrace{\tau \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1)\left(\rho_1^i - (1+r)\rho_2^i\right)}_{\text{(iv) Income effect targeting}} - \underbrace{\tau \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1)\Delta_{r,g}^i}_{\text{(v) Differential fiscal cost targeting}}$$

where $\bar{y}_1 = \frac{\sum_{i:s^i<0} \pi^i y_1^i}{\sum_{i:s^i<0} \pi^i}$. *When* $R''(s) = 0$, Δ_r^i *is constant across borrowers, so a common factor* $\tau\Delta_r$ *can be pulled out of terms (ii)–(iv), recovering the formula with piecewise-linear* $R(s)$.

Proof. See Appendix C.6. ■

The decomposition in equation (36) isolates the *lending effect*. This effect captures the welfare gains from providing uniform subsidized loans equal to $\tau\bar{y}_1$ (the average tax burden of borrowers). These gains represent the pure consumption smoothing benefit that any subsidized lending program would provide. Beyond this baseline lending effect, delayed taxation has a novel first-order component due to the compensated labor supply responses to a lower present-value tax rate. This *reduced tax-distortion* channel highlights a key advantage of delayed taxation over pure lending: it simultaneously provides liquidity and reduces tax wedges.

The additional three welfare terms reflect *differential targeting*. The loan targeting term (iii) represents 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. The income-effect targeting term (iv) accounts for 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 last term (v) 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 $\Delta_{r,g}^i = \Delta_r^i$ (government has same technology as agents), and assuming these are constant across borrowers ($\Delta_r^i = \Delta_r$), this term is zero.[14] We include this term to highlight a feature that may differentiate delayed taxation from lending in a more comprehensive model. 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).[15] The second reason is that high-income borrowers may otherwise be easier to collect from, due to,

[14] When $\Delta_{r,g}^i = \Delta_r^i = \Delta_r$ for $s_i < 0$, we can factor it out: $\Delta_r \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1) = 0$ by definition of $\bar{y}_1$.

[15] This is the case in our numerical exercise where $R(s)$ *approximates* a kinked return function. For s near 0, the numerical $R'(s)$ is decreasing in s (i.e., marginal interest rates increase with the borrowing amount).

e.g., lower risks of default. These two factors would cause $\Delta^i_{r,g}$ and y^i_1 to be negatively correlated, which implies that delayed taxation features lower default costs than uniform lending.

The key methodological ingredient in Proposition 3 is the Slutsky equation. By decomposing the sensitivity of labor supply into compensated responses and income effects, we can highlight that it is only the compensated response (which reduces tax distortions) that is inherently unique to delayed taxation.

2.6 Model extensions and limitations

Delayed taxation in a three-period model. Appendix F extends our two-period model to include a third retirement period. This extended model formalizes the concern that workers may choose to delay taxes in their second (high-earning) period and then default on the payments in retirement. We show quantitatively that a no-default equilibrium, with comparable welfare gains, is obtainable in this three-period model even when workers may delay the same fraction of taxes in all periods.

Myopic agents. Appendix E analyzes the marginal welfare effect of introducing delayed taxation when a non-empty 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 baseline non-paternalistic case studied in Corollary 1, there is now an additional strictly positive welfare effect. This new term arises because delayed taxation crowds out high-interest private borrowing. Even if there is only partial crowd-out, the effect on period-2 consumption is strictly positive due to private borrowing carrying a higher interest rate. When the government is paternalistic, it places a higher value on this additional period-2 consumption than the individuals themselves. This paternalistic welfare term thus works to increase the welfare gains from delayed taxation.

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 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 $\delta = \tau_1/\tau_2$. 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.

Deterministic wages. Following Weinzierl (2011), we assume perfect foresight to obtain heterogeneity in expected wage trajectories.[16] Since this renders the model completely deterministic, we obtain lending costs (and default-related costs of delayed taxation) from strategic defaults. This framework allows us to focus on the key trade-offs associated with delayed taxation. The welfare gains from intertemporal smoothing and reduced distortions are weighed against fiscal costs caused by income effects and default-related costs. Introducing ex-ante income uncertainty would modify the welfare gains from delayed taxation in two potentially offsetting ways. Adding uncertainty to period-1 wages increases welfare gains, as delayed taxation helps workers self-insure through intertemporal smoothing. Adding uncertainty to period-2 wages decreases welfare gains, as the precautionary savings motive would reduce borrowing demand and thus the gains from enabling intertemporal smoothing.

Competitive lending markets. We impose that financial intermediation is irrelevant for welfare by assuming competitive lending markets. An alternative assumption would be that lenders have market power and that the proportional interest wedge, $1 - \frac{1+r}{1+r_b} > 0$, is partly due to markups (see, e.g., Herkenhoff and Raveendranathan 2025). If this were the case, we conjecture that the welfare gains from delayed taxation would be higher due reduced monopolistic rents and redistribution of surplus from lenders to borrowers.[17]

3 A quantitative investigation of delayed taxation

3.1 Calibration

We assume that the per-period 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

[16] 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.

[17] Consider the baseline case where the government faces the same proportional resource cost of lending (z) as private lenders. The more $\Delta_r \equiv 1 - \frac{1+r}{1+r_b}$ is driven by markups, the lower is the implied $z = z^g$. This implies the government's marginal cost of lending, Δ_r^g, is lower than the interest rate wedge faced by borrowers, Δ_r. Hence, by engaging in delayed taxation, which crowds out private borrowing, the government essentially transfers private-market lending profits to borrowers. These profits would constitute monopolistic rents, and the welfare effects of reallocating them via the government budget would presumably be positive.

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. Appendix Table A.2 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 3.05%. This masks considerable heterogeneity. The fifth percentile of cumulative real wage growth is -28% and the ninety-fifth percentile is 464%.

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 exogenous revenue requirement $K = K_1 + \frac{K_2}{1+r}$, where K_t is 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 which policy is implemented. We set the annual base interest rate, r, equal 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}. Similarly, we set $\beta = 0.97^{21}$. In most of our analyses, we consider an interest rate on borrow of $r_b = 20\%$. This is comparable to interest rates on credit cards and in line with the life-cycle interest rate estimates by Heckman (1976).

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), transfers remain age-independent ($G_1 = G_2 = G \geq 0$) but the government may choose different marginal tax rates, i.e., we allow $\tau_1 \neq \tau_2$ as long as $\tau_1, \tau_2 \geq 0$.

When implementing *HSV*-style nonlinear marginal tax rates, the tax and transfer schedule is summarized by disposable income,

$$y_t - T_t(y_t) = G_t + a_t y_t^{1-u'}, \tag{37}$$

where a_t, u', 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. To isolate the effect of progressive taxation, all tax parameters (a_t, u', G_t), are time invariant.

[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)

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. $1 - \delta \geq 0$, which never binds, implies that the government cannot force workers to save an amount proportional to their accrued taxes.[19]

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 H, ensures that marginal interest rates, $R'(s) \approx 1 + r$ when $s \geq 0$ and $R'(s) \approx 1 + r_b$ when $s \leq 0.15$.

Individual-level delayed taxes. We further impose that agents do not delay taxes or take on government-subsidized loans if they are indifferent (i.e., $R'(s^i) \leq 1 + r_{dtax}$). This assumption matters for computing government lending costs, as we describe in Appendix H. It also matters when we impose restrictions on the amount that the government can borrow.

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 some policy parameter, ω, as

$$MVPF = \frac{\sum_i \pi^i \frac{dV^i}{d\omega}/\lambda}{-\frac{dREV}{d\omega}}, \tag{38}$$

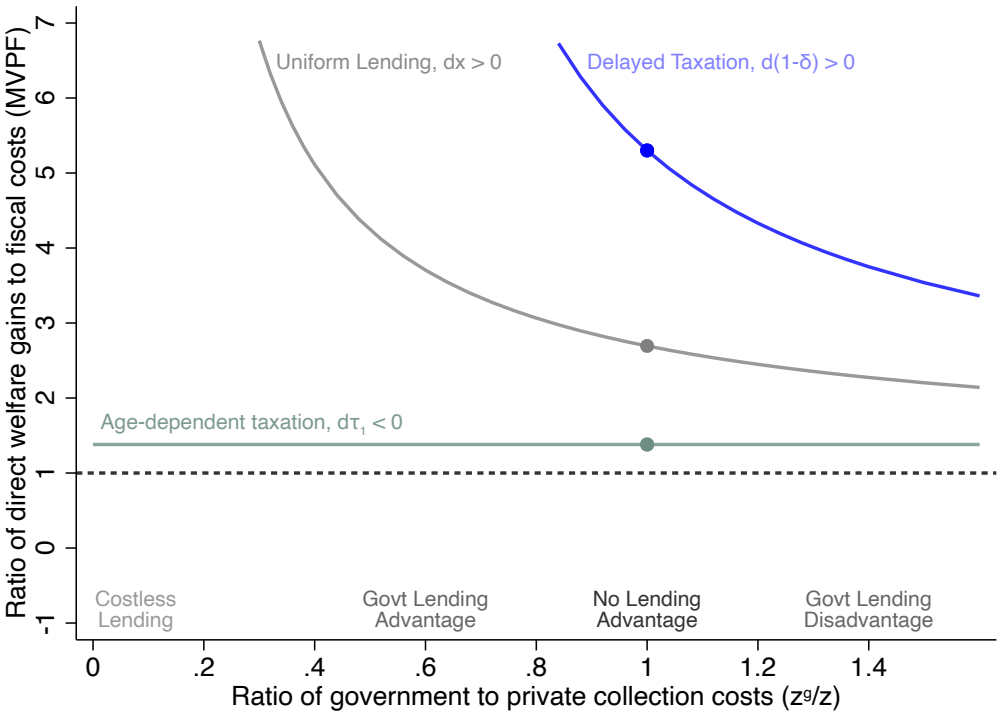
where V_i is an individual's life-time utility, λ is the Langrange multiplier on the government budget constraint in the benchmark economy, and REV is present-value government tax revenues net of any cost-of-capital effects (due to the possibility of default). The MVPF is thus the ratio of money-metric 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 B. We can thus use the numerical allocations for the simulated benchmark economy to quantify the MVPFs for different tax reforms.

[19] $1 - \delta < 0$ resembles a social security tax. One way to achieve an optimal $1 - \delta < 0$ would be to make the government highly "paternalistic" (see Appendix E).

Figure 1: Marginal Value of Public Funds

This figure plots the marginal value of public funds (MVPF) for delayed taxation ($d(1-\delta) > 0$), subsidized lending ($dx > 0$), and a reduction in period-1 tax rates ($d\tau_1 < 0$) in the benchmark economy ($1 - \delta = 0$, $x = 0$, $\tau_1 = \tau_2$). The x-axis varies the government lending cost 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 graph is restricted to $MVPF < 7$. The $MVPF$ of delayed taxation is infinite when $z^g/z \leq 0.4$. The points marked by a circle indicate MVPFs of 5.3013 (DT), 2.6944 (Lending), and 1.3808 (Age-dependent reduction τ_1 when $z^g/z = 1$.



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 private lenders' cost of capital (r_b). This baseline is marked by a circle in the figure and provides an $MPVF = 5.30$. This value implies that for each marginal dollar in reduced present-value revenues (e.g., due to income effects and default costs), the government generates \$5.30 in welfare gains through improved intertemporal consumption smoothing. Reduced distortions from income taxation play a pivotal role by reducing fiscal costs and thus the denominator. 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 may go as high as 5.

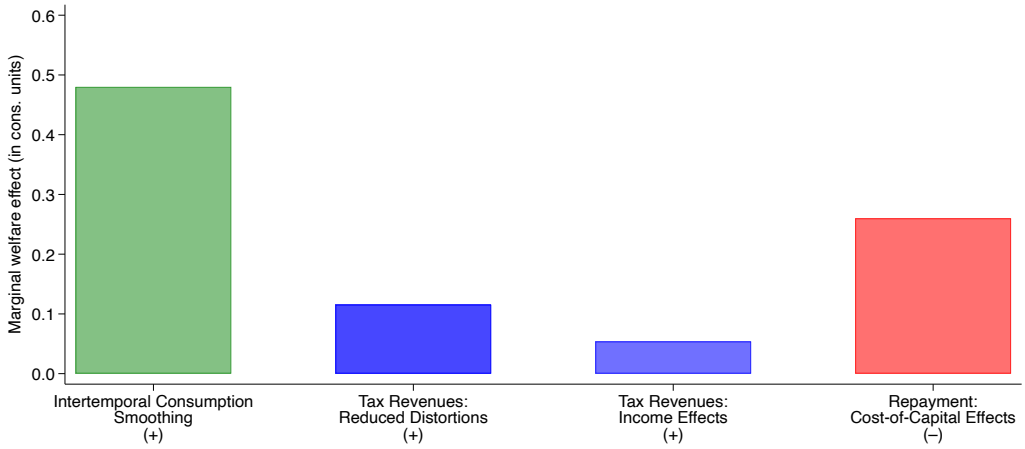
The middle solid line shows that the MVPF from offering uniform subsidized loans carry an MVPF less than half of delayed taxation. When there is no lending advantage, the $MVPF$

of lending is about 2.69. This is substantial, but less than half that of delayed taxation. We also contrast these gains with those achievable from an age-dependent reduction in the period-1 tax rate. At the benchmark economy, the government is restricted to set $\tau_1 = \tau_2$. Since many individuals are constrained ($\Delta_r^i > 0$), there may be substantial welfare gains from allowing the government to reduce τ_1. While we find this to be the case with an $MVPF$ of 1.38, these gains are modest relative to those of delayed taxation and lending.

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 Figure 2. We normalize the marginal welfare gains by dividing through by aggregate period-1 consumption. The first bar (green) shows that the gains from intertemporal consumption smoothing equal 0.41 consumption units. The welfare gains from reduced tax distortions equal about 0.1 consumption units. We further find that the income effects *increase* tax revenues, producing welfare gains of about 0.05. The implicit loan in delayed taxation creates a period-2 repayment obligation that induces enough additional labor supply to more than offset the period-1 income effect. The costs of lending, due to discounting future repayments at $r_b > r$, equals about -0.22.

FIGURE 2: DECOMPOSING MARGINAL WELFARE EFFECTS

This figure shows marginal money-metric welfare effects of marginally introducing delayed taxation, $(1/\lambda)dW/d(1-\delta)$, 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 (33), 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 aggregate period-1 consumption, $\sum_i \pi_i c_1^i$. The first two bars indicate positive welfare effects, and the second two bars provide the absolute value of negative welfare effects.



We can use these quantities to decompose the MVPF. Absent the reduced tax distortions, we would divide the smoothing gains (0.4055) by the lending costs net of income effects (0.1740). This would produce an MVPF of 2.33. However, accounting for the significantly lower fiscal costs when tax distortions are reduced, the MVPF more than doubles to 5.30. These findings highlight the role of reduced tax distortions as significantly reducing the fiscal costs of delayed taxation. We find this to be particularly interesting from a policy perspective in which fiscal costs of are

more easily measured than the welfare gains.

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. We keep the assumption that the government has no lending advantage ($z^g = z$). This section graphically illustrates the main welfare results. Afterwards, we examine the optimal tax schedules and allocations, before considering several robustness checks.

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.5 provides additional statistics for the age-dependent and HSV policies.

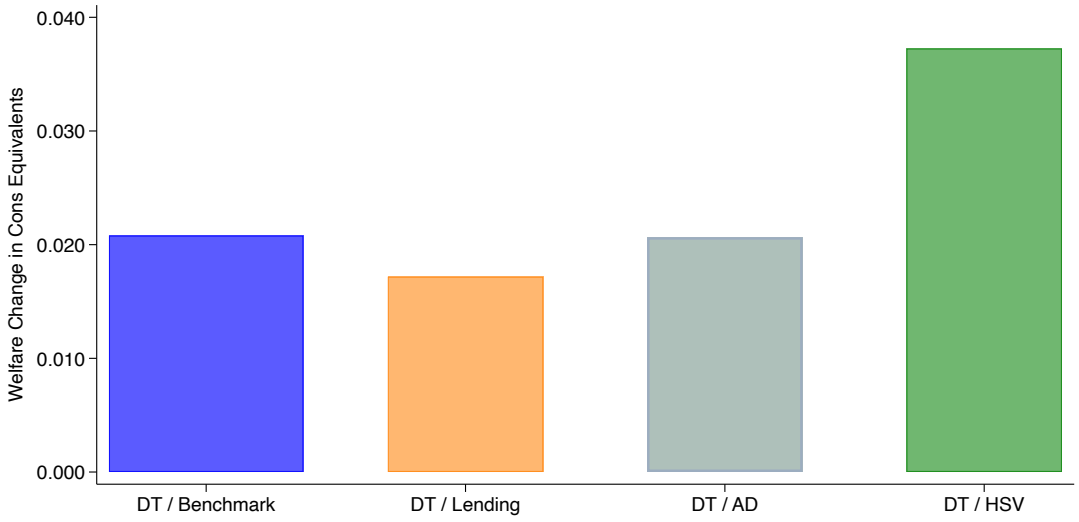


Figure 3 provides the money-metric welfare gains from implementing delayed taxation. These gains from a given policy constitute the amount by which the government would have to be compensated to obtain the same welfare levels without implementing the policy. The first bar provides welfare gains (in period-1 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. We find that most of the welfare gains from delayed taxation stay intact. Later, we show that the government in fact chooses to stop subsidized lending when it can delay taxes.

The third bar similarly provides the welfare gains from implementing delayed taxation on top of an age-dependent tax policy. Age-dependent and delayed taxation can both reduce period-1 tax rates, but differ in their targeting. It is therefore interesting that the welfare gains from delayed taxation are virtually unaffected. The fact that delayed taxation targets constrained workers lets it provide material 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.5 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, p, that directly governs the progressivity of the tax schedule.

3.4 Optimal delayed 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. We see that the benchmark economy features high marginal tax rates with $\tau = 0.644$. At the same time, transfers are also large at $G = 0.552$, causing the average tax rates to be considerably lower. In period-1, the ATR is negative at -0.10, and in period-2 it is quite modest at 0.251. This age-dependency in the ATRs is driven by the increase in average earnings from period-1 to period-2.

Column (2) shows that optimal delayed taxation implies a statutory delayed fraction of 38.8%. However, since net savers do not delay, the average fraction of delayed taxes is lower at 10.9% (see Appendix H.1). This overall modest extent of delayed taxation nevertheless produces material welfare gains equal to 2.08% of aggregate period-1 consumption in the benchmark economy. Another way to quantify the welfare gains is as a fraction of the amount of delayed taxes. Doing this, we find that each \$1 of delayed taxes produces money-metric welfare gains of \$0.38. Given that the money-metric welfare gains account for offsetting income effects on tax revenues and lending costs, this quantity is substantial. To underline this point, we can compare delayed taxation with a hypothetical investment opportunity. This investment has an annualized return in excess of the savings rate of $\tilde{r}-r$ and the present-value of investing \$1 must equal 1.3817:

$$\left(\frac{1+\tilde{r}}{1+r}\right)^{T} = 1.3817, \tag{39}$$

which implies that $\tilde{r} - r = 1.60$ percentage points—a substantial excess return for a risk free investment.[20]

We see that the government slightly increases marginal tax rates under delayed taxation from 0.644 to 0.680. We also find that delayed taxation reduces period-1 labor earnings, $l_1 w_1$, but increases period-2 labor earnings. We note that these are the sum of both substitution and

[20] This return is risk free in the sense that the underlying welfare calculations account for the potential of default by discounting delayed tax repayments at the private borrowing rate.

income effects. While delayed taxation leads to higher period-1 labor earnings through reducing marginal present-value tax rates, the income effects of increased consumption smoothing are large when $r_b = 20\%$. These income effects also work to increase period-2 labor supply. In addition, on the path to the optimal $1 - \delta$, the government also changes other policy parameters, which complicates the link between the theoretical results on marginal reforms and the quantitative results on a full implementation.

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. $1-\delta^i$ is the ratio of delayed taxes that imposes no delayed taxation for indifferent workers (see Appendix H.1). Welfare Gain in % of D is the money metric welfare gain divided by the amount of delayed taxes ($\sum_i \pi^i(1-\delta^i)\tau_1 \ell_1^i w_1^i$). Appendix Table A.5 considers the cases with Age-Dependent and non-linear HSV taxation as well.

	Tax schedule and allocations with $r_b = 20\%$			
	(1)	(2)	(3)	(4)
	Benchmark	+ DT	Lending	+ DT
τ_1	0.644	0.680	0.646	0.680
τ_2	0.644	0.680	0.646	0.680
G_1	0.552	0.571	0.551	0.571
G_2	0.552	0.571	0.551	0.571
$1-\delta$	0.000	0.388	0.000	0.388
r_{dtax}	0.030	0.030	0.030	0.030
Govt loan limit	0.000	0.000	0.015	0.000
B	0.131	0.416	0.138	0.416
Welfare Gain, % C_1 (rel. to Benchmark)	0.000	2.080	0.362	2.080
Δ Welfare, % C_1	.	.	.	1.718
Welfare Gain, % D	.	38.168	.	38.168
Δ Welfare, % D	.	.	.	31.526
π^i-weighted means				
$1-\delta^i$	0.000	0.109	0.000	0.109
$l_1 w_1$	0.804	0.721	0.793	0.721
$l_2 w_2$	1.662	1.912	1.695	1.912
c_1	0.893	0.927	0.897	0.927
c_2	1.016	0.947	1.006	0.947
PV($l_1 w_1$,$l_2 w_2$)	1.697	1.749	1.704	1.749
PV($l_1 w_1 \tau_1$,$l_2 w_2 \tau_1$)	1.093	1.189	1.101	1.189
$(TaxPaid_1 - G_1)/y_1$	-0.100	-0.260	-0.109	-0.260
$(TaxPaid_2 - G_2)/y_2$	0.251	0.323	0.261	0.323
l_1	0.792	0.711	0.781	0.711
l_2	0.820	0.914	0.836	0.914
s	-0.055	-0.015	-0.066	-0.015
Impl. govt loan amount, IL	0.000	0.049	0.004	0.049

Column (3) considers the case when the government can implement subsidized lending. While we show negative marginal welfare effects from subsidized lending in Figure 2, column (3) nevertheless demonstrates 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. Column (3) shows that the government offers a maximal loan (x) of about 0.015. This produces an average implied loan amount, IL, of 0.004, which is about 1.4% of average period-1 taxes, $\tau_1 \ell_1 w_1$. This is modest relative to the extent of delayed taxation at 10.9%.

Column (4) allows the government to implement delayed taxation on top of a subsidized lending scheme. This leads to a welfare gain of 2.08%, which is 1.7 percentage point higher than with only subsidized lending. Interestingly, we see that delayed taxation does not complement the subsidized lending scheme. Instead, when the government engages in delayed taxation it refrains from offering subsidized loans (setting $x \approx 0$). In other words, delayed taxation is a superior *alternative* to subsidized lending in this calibration.

Appendix Table A.5 also considers age-dependent and progressive (HSV-style) tax schemes. The results indicate that delayed taxation and age-dependent taxation do not act as substitutes. The additional welfare gains from implementing delayed taxation on top of age-dependent taxation equal 2.07% of period-1 consumption. Which is only slightly lower than implementing delayed taxation on top of the linear benchmark policy (which delivers 2.08%). We also see that the degree of age-dependency in marginal tax rates barely changes when the government can implement delayed taxation. The age-spread, $\tau_2 - \tau_1$ decreases modestly from 0.33 to 0.29. Interestingly, the incremental welfare gains from delayed taxation on top of an HSV benchmark are considerably higher, now equal to 3.73%. While non-linear taxes themselves only produce modest gains of 0.1%, they double the welfare gains from delayed taxation. On top of the HSV benchmark, the government optimally sets the statutory delayed fraction,$1 - \delta$, equal to the upper bound of 1, which produces an average delayed tax fraction of 37%. At the same time, the government imposes a more progressive tax system (increasing k from 0.15 to 0.31).

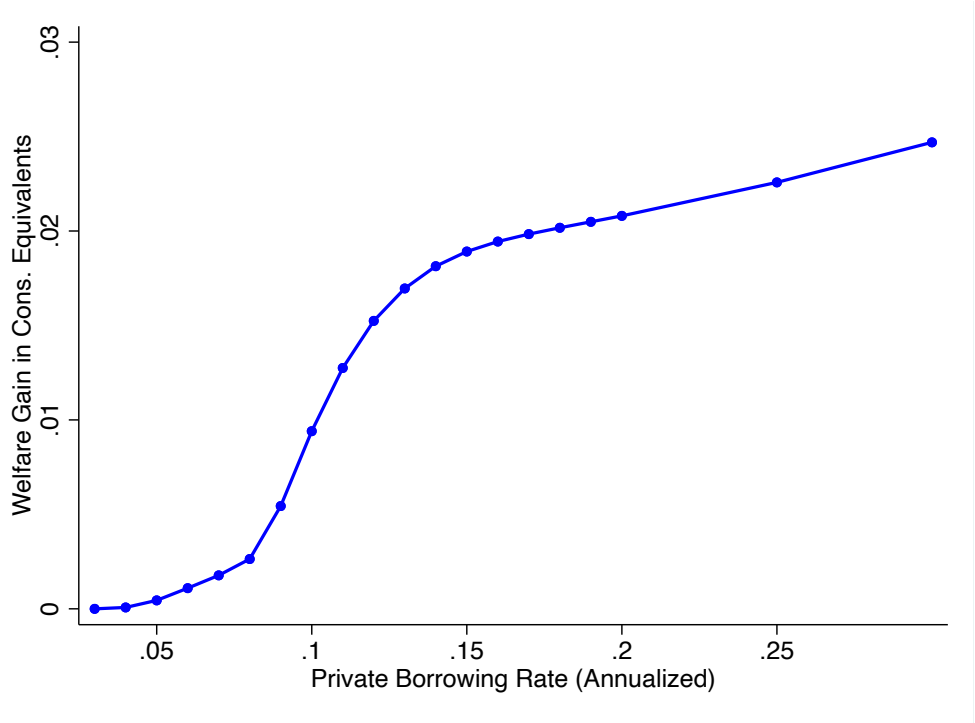
3.5 Welfare gains when varying the severity of financial frictions.

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 remain high for a wide range of plausible borrowing rates. Naturally, welfare gains approach zero when r_b approaches the interest rate on savings. When marginal interest rates are already low, delayed taxation 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 interest rates above 10%) is due to individuals borrowing less 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.

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.

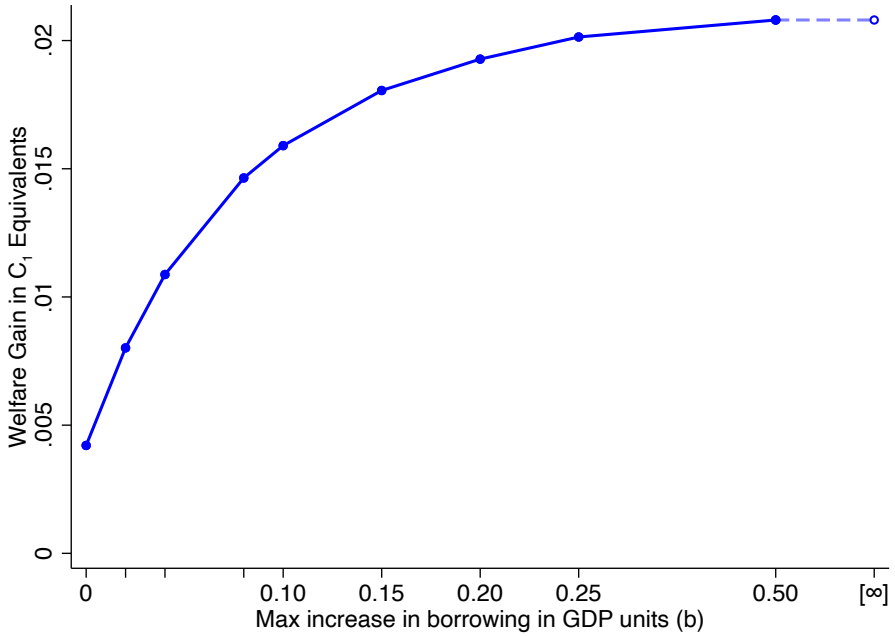


3.6 Restrictions on government borrowing

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 period-1 GDP (total labor earnings in period 1). 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 about 0.4%. Welfare gains are effectively maximized when the government borrows 25% of period-1 GDP. However, since gains are concave, large welfare gains can be achieved with modest government borrowing (e.g., two thirds can be achieved by borrowing 10% of period-1 GDP).

FIGURE 5: WELFARE EFFECTS WHEN 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 period-1 GDP units, defined as total labor earnings). $b = 0$ allows no additional borrowing. $b = 0.10$ allows the government to increase borrowing by 10% of aggregate benchmark labor earnings in period-1 under the benchmark policy.



3.7 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) for the years 1990 and 2011, where we replicate as closely the PSID sample criteria used by Weinzierl (2011).

The U.S. calibration (Appendix Table A.3) features different wage trajectories (w_1^i, w_2^i) than the Norwegian calibration (Appendix Table A.2). In the PSID, the median 21-year *cumulative* wage growth is only 51% (1.97% annual). In the Norwegian data, median 21-year wage growth is 87%. Since heterogeneity in borrowing demand is solely driven by wage profiles in our framework, there are naturally fewer borrowing constrained individuals in the PSID calibration. Hence, welfare gains from policies that address financial frictions may be more modest.

Nevertheless, we still find meaningful welfare gains from implementing delayed taxation. On top of the benchmark policy, the government optimally chooses to fully delay taxes, choosing $1 - \delta = 1$, which delivers welfare gains of 1.79% of period-1 consumption in the benchmark economy. These gains are seven times larger than the 0.24% welfare gains from a pure subsidized lending scheme. Note that setting $1 - \delta = 1$ does not mean that all taxes are delayed. When we impose that those who are indifferent choose not to delay, the average fraction of delayed taxes

is merely 11%, which resembles the fraction in the Norwegian calibration of 11%. If we allow unrestricted delayed taxation (i.e., allow $1-\delta > 1$), the government chooses $1-\delta = 1.07$. This increases the average fraction of delayed taxes to 15%, and modestly adds 0.01 percentage points in welfare gains.

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 zero. Interestingly, implementing delayed taxation on top of either an age-dependent or progressive (HSV) tax scheme increases the additional welfare gains from delayed taxation. We further note that allowing for HSV-style nonlinear taxation by itself adds no value. The government effectively chooses a linear tax (due to $k \approx 0$) and obtains welfare gains (relative to the linear benchmark) of approximately 0%. It is only when combined with delayed taxation, that the government chooses a progressive tax schedule ($k = 0.25$).

4 Does delayed taxation already exist?

We define delayed taxation as a scheme in which tax payment, but not tax accrual, is delayed and subject to a subsidized interest rate. This creates a wedge between effective present-value tax rates and statutory tax rates for borrowing-constrained tapayers. There are several tax schemes in place today that have related 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 typically not withheld, meaning that high-income Swedish workers face a roughly one year delay on a portion of their income tax. The limited duration of the delay, coupled with the implied targeting of only high-income workers, emphasizes that there is only a limited similarity with our notion of delayed taxation.

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 negatively delayed taxation. 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 concern 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 cause 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.[21] 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.

Retirement savings accounts. Tax-deductibility of 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.[22] The key distinction is that this tax deductibility targets net savers as opposed to net borrowers, hence welfare gains from intertemporal consumption smoothing are likely limited. Nevertheless, we view the possibility that this tax deductibility reduces distortions as a highly complementary avenue for research. The caveat is that 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

[21] 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.

[22] 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.

from subsequent capital gains taxation.

Student stipends. Income-contingency in student financial aid bears a close resemblance to delayed taxation. 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.

5 De facto Delayed Taxation in Norway

5.1 Empirical setting

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

Our empirical study focuses on the years 2004 to 2011. During these years, most Norwegian students faced an earnings threshold ranging from NOK 104,500 ($17,000) in 2004 to NOK 140,823 in 2011. Monthly transfers ranged from NOK 8000 ($1,300) in 2004 to NOK 9785 in 2011. These transfers are initially given as a loan, but 40% can be forgiven (converted to a stipend) as long as students pass their courses and stay below the earnings thresholds mentioned above. Students are notified of the amount of the transfer at the beginning of the academic year. These notification letters include a breakdown of the transfers, noting the amount (40% of the total) that will be given as a conversion loan, and stating that the conversion of the loan to a stipend is contingent upon income being below an income threshold. The following year, students are notified how much of their loan has been converted, based on grades reported by the educational institutions and income reported to the tax authorities. Loans must generally be repaid within 20 years of graduation. No interest is charged while the student is receiving aid. Thereafter, interest rates are slightly above the risk-free rate and loan payments can be delayed at the (former) student's discretion for up to a total of 3 years.[23]

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

[23]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.

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 2. 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 2: Summary Statistics

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

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

Salience. In order to meaningfully compare the implied elasticity of the debt conversion threshold with that of the regular tax thresholds, the conversion threshold must be similarly salient. As one of the authors is a former participant in this program, we certainly believe this to be the case. Beyond anecdotal evidence, however, it is useful to consider the magnitude of the effective tax increase. A 50 percentage point reduction in the "net of debt" wage is unlikely to go unnoticed. Moreover, students are informed of the existence of such a cap in a loan agreement letter that they must sign, and they also receive letters informing them of any conversion that has occurred. Even if students do not expect to receive a reduction in their debt through conversion, they will want to read these letters to confirm that their institution has accurately recorded and reported their academic progress. Non-passing grades in courses also reduce debt conversion. Students are also informed of their annual student debt balances when they receive their annual prefilled tax returns, which also include information about their income tax liabilities.

5.2 Bunching methodology

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

$$e = \frac{\Delta y^*/y^*}{\Delta\tau/(1-\tau)}, \tag{40}$$

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

$$B = \int_{y^*}^{y^*+\Delta y^*} h_0(y)dy \approx h_0(y^*)\Delta y^*. \tag{41}$$

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

$$\frac{\Delta y^*}{y^*} = \frac{B}{h_0(y^*)y^*} = \frac{b}{y^*}. \tag{42}$$

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.[24] We write our estimated compensated labor earnings elasticity as

$$\hat{e} = \frac{\hat{b}/y^*}{\widehat{\Delta\tau}/(1-\tau)}, \tag{43}$$

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

5.3 What the bunching estimator identifies

In a standard dynamic model without adjustment frictions, the estimator $\hat{e}$ is *conjectured* to estimate the Frisch elasticity (Saez 2010, Chetty 2012, Kleven 2016). An additional contribution of our paper is to move beyond a conjecture and formalize the identification arguments in Appendix J.1. We state the full theorem (applicable to settings with or without delayed tax payments) and provide the proof in Appendix J.1. When there are no labor-supply adjustment frictions and preferences are additively separable as in our calibration (equation 36), the bunching estimator

[24] 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.

identifies:

$$\varepsilon^{\text{bunch}} = \underbrace{Q_D}_{\substack{\text{Delayed Taxation} \\ \text{PV Adjustment}}} \cdot \underbrace{\frac{1}{1+\phi}}_{\text{Income-Effect Adjustment}} \cdot \underbrace{k,}_{\text{Structural Frisch}} \tag{44}$$

where $Q_D \in (0, 1]$ is a present-value adjustment for (delayed) tax payments, ϕ is an income-effect adjustment factor, and k is the structural Frisch elasticity. The standard assumption that income effects are effectively zero in dynamic models (Saez 2010, Chetty 2012, Kleven 2016) amounts to setting $\phi = 0$ in our framework. The argument is that income effects have a small impact when they are smoothed out over a long life-cycle. According to our Theorem 1,

$$\phi = k \cdot \sigma \cdot \frac{(1-\tau)y^*}{W}, \tag{45}$$

where σ is the inverse of the elasticity of intertemporal substitution and $\frac{(1-\tau)y^*}{W}$ is the ratio of the buncher's after-tax current income to wealth (*present-value* life-time incomes). If we disallow saving (and the intertemporal Euler equation does not hold), W should instead equal the concurrent period's disposable income. In which case, $\varepsilon^{\text{bunch}}$, as defined by equation (44), exactly equals the *Hicksian* elasticity.

If we specify quasi-linear preferences (i.e., $u(c) = c$), then income effects cease, the Hicksian and Frisch equate, and thus $\varepsilon^{\text{bunch}}$ exactly equals both the Frisch and the Hicksian. However, with standard concave preferences ($\sigma > 0$) and finite life-cycle horizons, income effects may play an important role. Financial frictions, in particular, may amplify these income effects since higher marginal interest rates mechanically lower the present-value of future earnings (thus decreasing W and increasing ϕ). Hence, in our setting, it seems ex-ante important to account for income effects.

Once we discuss our specific estimates, we also account for the role of adjustment frictions in creating a multiplicative wedge between structural and estimated elasticities (as in Kostøl and Myhre 2021 and Mavrokonstantis and Seibold 2022). We then consider our estimator to consist of four terms.

$$\varepsilon^{\text{bunch}} = \underbrace{Q_D}_{\substack{\text{Delayed taxation} \\ \text{PV adjustment}}} \cdot \underbrace{\frac{1}{1+\phi}}_{\text{Income-effect adjustment}} \cdot \underbrace{\zeta}_{\text{Adjustment-friction bias}} \cdot \underbrace{k,}_{\text{Structural Frisch}} \tag{46}$$

We provide additional discussion of our identification results and some numerical examples in Appendix J.1. One group faces interest rates of 20% while the other faces interest rates of 2%. This wedge causes the bunching estimator to identify a 16% lower elasticity for the high-interest rate group, despite the two groups having identical structural Frisch elasticities.

5.4 Bunching at the debt-conversion (de facto delayed tax) 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.[25] 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 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) and again by the relative reduction in the take-home rate, per equation 43, we obtain a remarkably low elasticity of labor earnings to the net-of-tax (or net-of-debt-increase) wage of 0.0162.[26] The standard error is 0.0013.[27]

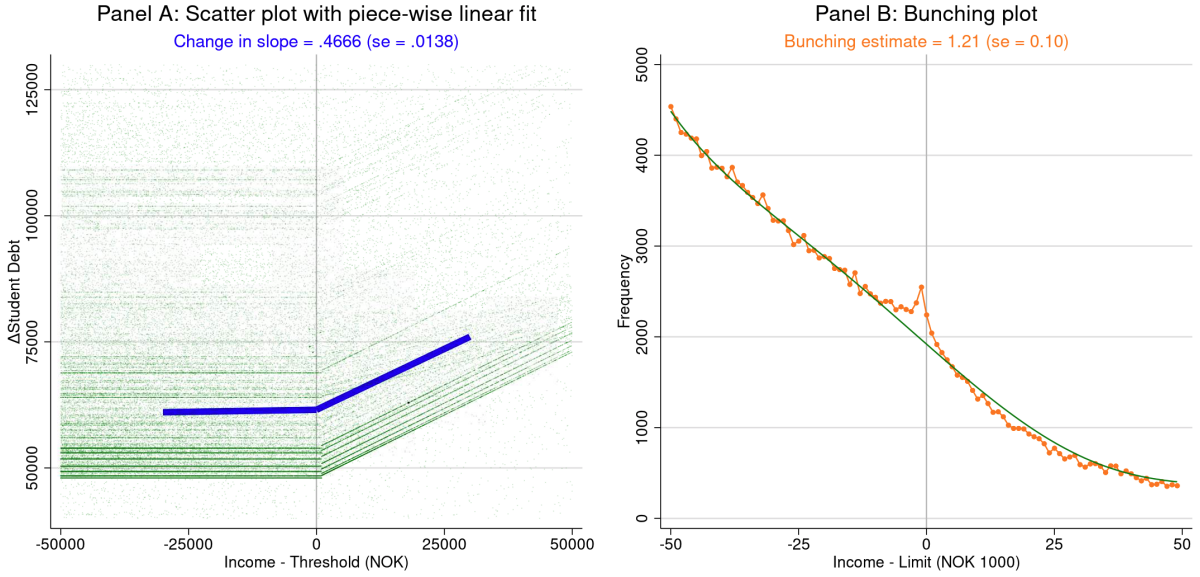
[25] Some non-compliers exist, for example, because they may have moved in with their parents during the fall semester, which would exclude them from receiving a conversion for fall semester loans. Such moves must be reported to the Educational Loan Fund, but not to the tax authorities from which we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

[26] 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%.

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

Figure 6: Bunching at the De Fact Delayed Tax Threshold

Panel (A) shows a scatterplot in green of the relationship between debt accumulation and student earnings around the debt conversion threshold. This verifies the kinked and delayed (debt) tax incentive. 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 (due to adjustment frictions) of 70%, which translates into $k = 0.3$ and $\zeta = 0.3$. 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 elasticity in equation (46).

$$E[\hat{e}] = \underbrace{\left(\frac{1+2\%}{1+20\%}\right)^{10}}_{\text{Delayed taxation effect}} \cdot \underbrace{\frac{1}{1+7.84\%}}_{\text{Income-effect adjustment}} \cdot \underbrace{(1-0.7)}_{\text{Net of adj. frictions}} \cdot \underbrace{0.3}_{\text{Structural Frisch}} = 0.01636, \quad (47)$$

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.5. 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.4. We find that the bunching estimate is negative, in line with theory, but statistically close to zero (t-stat=-1.64).

5.5 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.[28] 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.

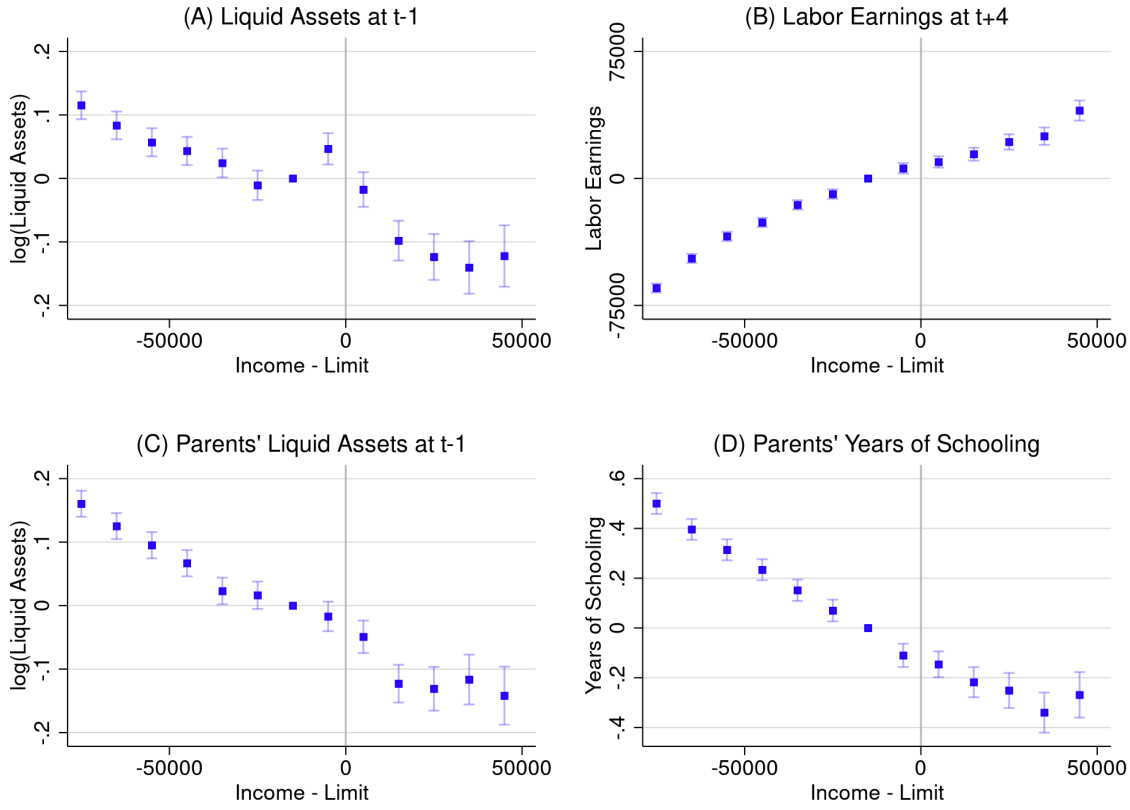
Taken together, these results highlight financial frictions as a key channel driving the insensitivity to the conversion threshold. Those who earn above the threshold have similar future earnings prospects, but hold significantly less liquid assets. Holding fewer assets may both causally affect the extent 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 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.

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

FIGURE 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.



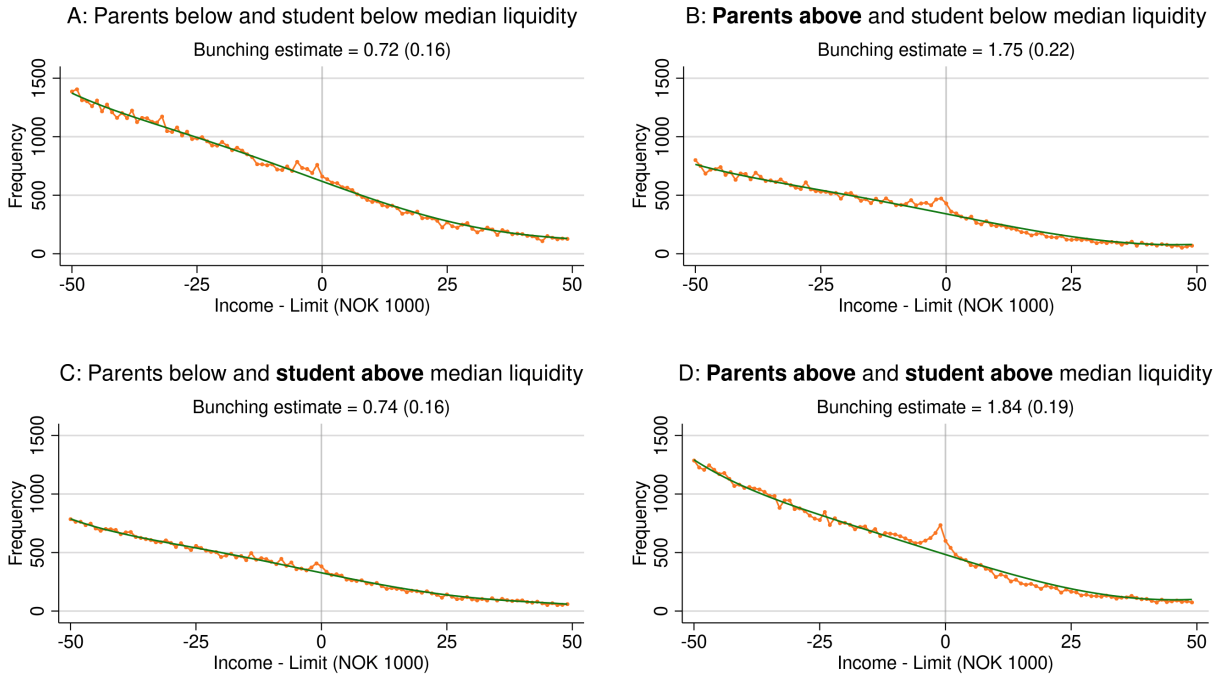
5.6 Bunching heterogeneity

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.[29]

[29] 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.

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 excess mass increases from 0.72 to 1.84 as we move from below to above the median for both student and parent resources. The corresponding labor supply elasticities change from 0.0096 to 0.0246. We first assume that the structural parameters (k and σ) are constant, and that the delayed tax is repaid after $D = 10$ years (equal to the average maturity on student loans). We can then using our expression (46) for the expectation of the estimator, $\hat{e}$, to write the ratio of estimands as

$$\frac{0.0246}{0.0096} = \frac{Q_D^{above}\ \xi\ \frac{1}{1+\phi_{above}}\ k}{Q_D^{below}\ \xi\ \frac{1}{1+\phi_{below}}\ k} = \left(\frac{1+r_{b,below}}{1+r_{b,above}}\right)^{10} \frac{1+\phi_{above}}{1+\phi_{below}}. \tag{48}$$

We note that the income-effect adjustment, ϕ, also depends on Q^D through life-time wealth (see Theorem 1 in Appendix J.1.2). Ignoring this dependence (i.e., assuming same income-effects adjustments in the two samples), we get that

$$\left(\frac{1+r_b^{below}}{1+r_b^{above}}\right) \approx \left(\frac{0.0246}{0.0096}\right)^{1/10} = 1.0987, \tag{49}$$

which says that the marginal interest rate is 1.0987 times greater for the below-median liquidity group. This corresponds to an 11.3 percentage point difference (assuming a baseline $r_b = 15\%$). Accounting for differences in income-effects adjustments produces a slightly moderated 10.59 percentage point difference (see Appendix J.2.1). This is a substantial difference in marginal

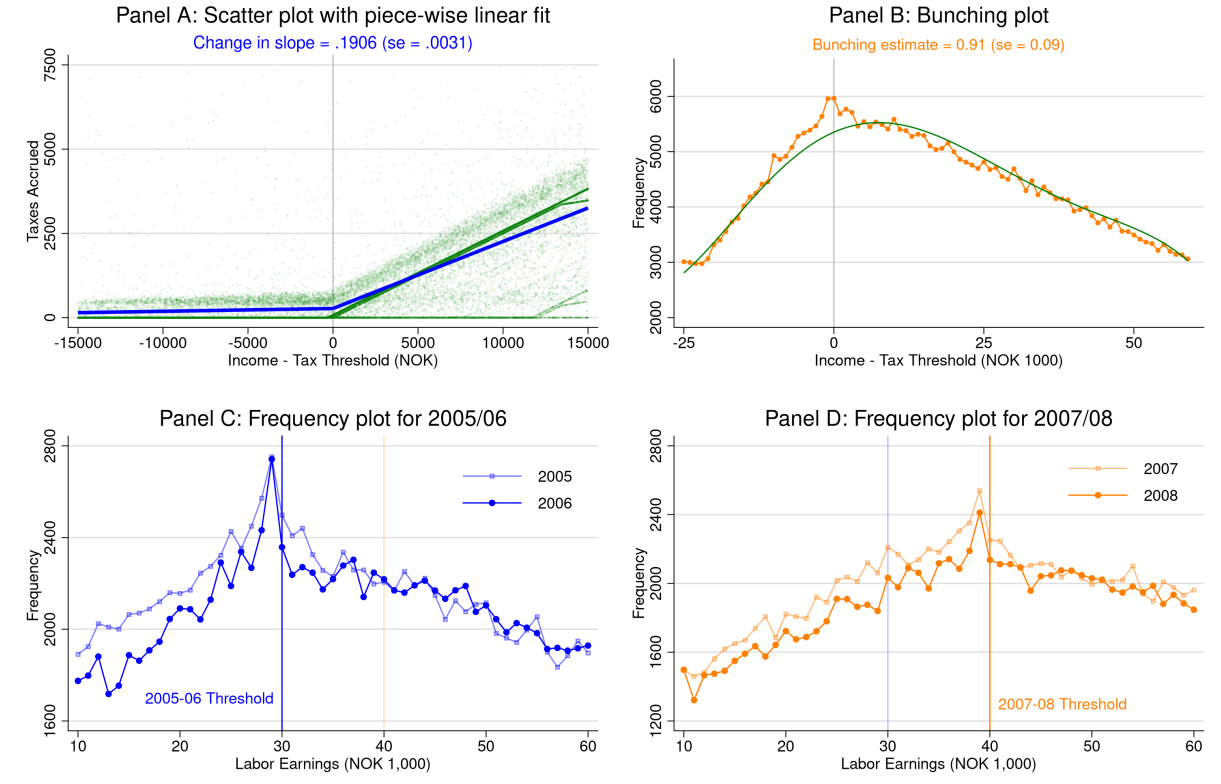
borrowing rates. While our theoretical framework allows us to calculate implied differences in marginal interest rates, we cannot derive them directly from the data. First, while we can calculate average interest rates on debt, we do not observe true marginal interest rates due to credit rationing through rigid credit card limits. These limits are likely small (or even zero) for students with relatively low incomes and short credit histories.

5.7 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.[30] 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.



[30] 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.

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 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 19.06%/100%, we get a more substantial elasticity of 0.1301.

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 quite high interest rates. If we assume that structural preference parameters, adjustment frictions, and the income effect adjustments are the same across samples, we can use expression (46) to infer that

$$\frac{0.1301}{0.0162} = \frac{Q_D^{\text{regular tax}}}{Q_D^{\text{delayed tax}}} = \frac{1}{Q_D^{\text{delayed tax}}} = \left(\frac{1+r_b}{1+r}\right)^D. \tag{50}$$

Assuming $D = 10$ years and an interest rate on delayed taxes of $r = 0.02$, the expression above implies an average marginal interest rate $r_b = 1.02 \cdot \left(\frac{0.1301}{0.0162}\right)^{\frac{1}{10}} - 1 = 25.63\%$. This figure is somewhat higher than average credit card rates of around 20%.[31] The fact that the implied marginal interest rate is above interest rates on credit cards is likely due to credit rationing. Some students would be willing to borrow at very high interest rates but are unable to obtain credit due to not having full-time employment or having exhausted available credit card limits.

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

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

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 excess-mass estimate, b, of 9.79.[32] 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.

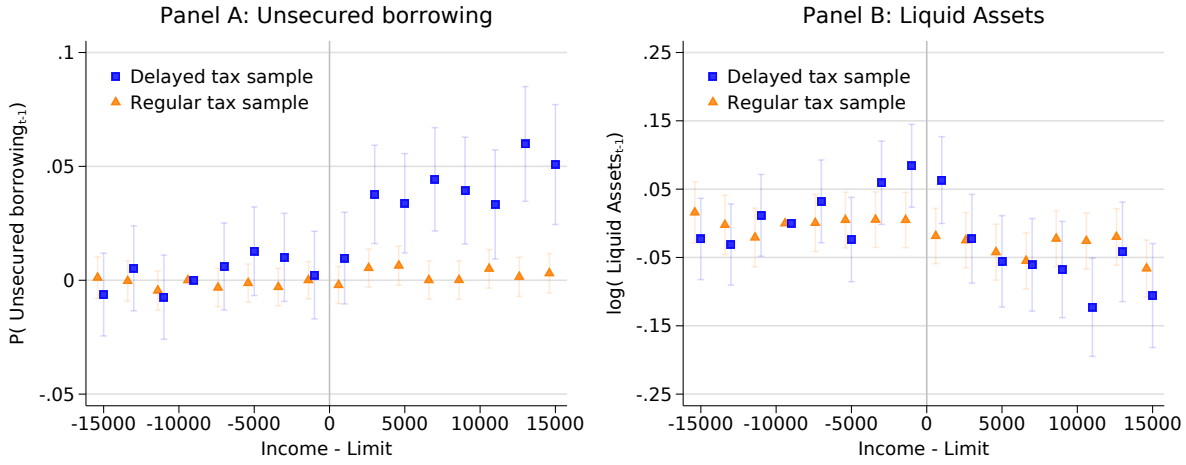


Figure 7 shows that those who bunch at the *delayed tax* threshold have more liquid wealth. This is intuitive because more constrained students have higher marginal borrowing rates on average and are thus less sensitive to a tax that is payable in the future. However, it is useful to show that proxies 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.

[32] =0.13*(120162/1000)*(0.47/0.75).

5.8 Accounting for differences in observables when comparing elasticities.

In this section, we pool the samples used to examine bunching at the debt conversion (delayed tax) and regular tax thresholds. We develop a regression-based approach that allows us to compare the underlying elasticities while holding observables fixed.[33] 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

$$e_i = \left[\underbrace{\frac{1[y_i \in BR_s] - \hat{P}^{cf}(y_i \in BR_s)}{\hat{P}^{cf}(y_i \in BR_s)/N_s^{\text{BR bins}}}}_{b_i} \cdot 1 \Big/ y_s^* \right] \Big/ (\widehat{\Delta\tau_s}/(1-\tau_s)), \tag{51}$$

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 J.5). 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.[34]

We then estimate regression equations of the following form.

$$e_i = \alpha + \beta \mathbb{1}[\text{regular tax sample}]_i + \gamma' X_i + \varepsilon_i, \tag{52}$$

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 3. 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.

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

[34] 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 0.13 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 3: 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

$$6.1 = \left(\frac{1 + r_b}{1 + r}\right)^{10}. \tag{53}$$

Setting $r = 0.02$, we obtain $r_b = 22.22\%$, which is qualitatively similar to the implied interest rate of 25.63% that we found earlier.

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 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. This empirical finding 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.

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.

Dixit, A. and A. Sandmo (1977): "Some simplified formulae for optimal income taxation," *The Scandinavian Journal of Economics*, 417–423.

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.

Fisher, I. (1930): *The theory of interest as determined by impatience to spend income and opportunity to invest it*, Macmillan.

Friedman, M. (1957): *A Theory of the Consumption Function*, no. 63 in National Bureau of Economic Research General Series, Princeton, NJ: Princeton University Press.

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)*.

Harberger, A. C. (1971): "Three basic postulates for applied welfare economics: an interpretive essay," *Journal of Economic literature*, 9, 785–797.

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.

Heckman, J. J. (1976): "A Life-Cycle Model of Earnings, Learning, and Consumption," *Journal of Political Economy*, 84, S9–S44.

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.*

Herkenhoff, K. F. and G. Raveendranathan (2025): "Who bears the welfare costs of monopoly? the case of the credit card industry," *Review of Economic Studies*, 92, 3067–3111.

Hirshleifer, J. (1958): "On the theory of optimal investment decision," *Journal of political economy*, 66, 329–352.

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.

Mavrokonstantis, P. and A. Seibold (2022): "Bunching and adjustment costs: Evidence from cypriot tax reforms," *Journal of Public Economics*, 214, 104727.

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

OECD (2025): "Net Pension Replacement Rates (Indicator)," Accessed: 2026-02-24.

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.

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

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

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

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

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.

Yaari, M. E. (1965): "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer1," *The Review of Economic Studies*, 32, 137–150.

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*.

Online Appendix

A Micro-foundation for higher interest rates on borrowing

Below, we introduce a 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). The distinction of our microfoundation is that the possibility of (off-equilibrium-path) defaults imposes upfront 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.

We note that this microfoundation is conservative in terms of the potential welfare gains from delayed taxation. An alternative micro-foundation for costly borrowing in private markets would be that lenders have market power in setting interest rates. If this were the case, welfare gains from delayed taxation would likely be considerably higher (see discussion in Appendix A.1.)

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' disposable incomes (after-tax labor income plus any savings) 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.[35] 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

$$(1+r)(1+z) = (1+r_b) \Leftrightarrow z = \frac{1+r_b}{1+r} - 1, \tag{A1}$$

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

[35] 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.

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

$$(1+r)(1+z^g) = (1+r_b^g). \tag{A2}$$

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 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, $\underline{Y}$ 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

$$V(\tau_1, \tau_2, G_1, G_2, \bar{a} \mid \delta = 1) \geq u(G_1 + \ell_1^d w_1(1-\tau_1) + \bar{a}) - v(\ell_1^d) + \beta u(\underline{Y}), \tag{A3}$$

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 4), 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 $\underline{Y} = G$, implies that the smallest $\bar{a}$ that ensures no defaults is equal to 11.73, whereas the maximal borrowing amount is 0.10.

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.

$$V(\tau_1, \tau_2, G_1, G_2, \bar{a}, \delta) \geq u(G_1 + \ell_1^d w_1(1 - \delta\tau_1) + \bar{a}) - v(\ell_1^d) + \beta u(\underline{Y}), \tag{A4}$$

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 $\underline{Y} = G$ implies that the smallest $\bar{a}$ that ensures no defaults is equal to 15.69, 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 0.26.[36]

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

$$V(\tau_1, \tau_2, G_1, G_2, \bar{x}, \bar{a} \mid \delta = 1) \geq u(G_1 + \ell_1^d w_1(1 - \tau_1) + \bar{a} + \bar{x}) - v(\ell_1^d) + \beta u(\underline{Y}). \tag{A5}$$

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, $\underline{Y} = G$, we have that the necessary $\bar{a} + \bar{x} = 11.70$. The largest amount of borrowing (including from both government and private lenders) is 0.1268. 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

$$V^{\text{default}} = \max_{\ell_1} \frac{c_1^{1-\sigma} - 1}{1 - \sigma} - \xi \frac{\ell_1^{1+\frac{1}{k}}}{1 + \frac{1}{k}} + \beta u(\underline{Y}), \tag{A6}$$

$$\text{s.t. } c_1 = G_1 + (1 - \delta\tau_1)\ell_1 w_1 + \bar{a}. \tag{A7}$$

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

[36] 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 H.

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

$$c_1^{-\sigma}(1-\delta\tau_1)w_1 = \xi \ell_1^{1/k}, \tag{A8}$$

which may be written

$$c_1 = \left[\frac{(1-\delta\tau_1)w_1}{\xi}\right]^{1/\sigma} \ell_1^{-\frac{1}{k\sigma}}. \tag{A9}$$

Hence, the optimal ℓ_1 is given by

$$\left[\frac{(1-\delta\tau_1)w_1}{\xi}\right]^{1/\sigma} \ell_1^{-\frac{1}{k\sigma}} = G_1 + (1-\delta\tau_1)\ell_1 w_1 + \bar{a}. \tag{A10}$$

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.

A.1 Alternative micro-foundation: market power

An alternative micro-foundation for an interest spread $r_b - r > 0$ is that lenders have market power. In the extreme case where the interest spread is purely a profit-generating markup, the welfare gains from delayed taxation would (likely) be considerably higher. The main reason is that there would be no fiscal costs associated with defaults on delayed taxes. Instead, there would be a welfare loss due to reducing lender profits. If we assume that these are welfare relevant due

to private lenders being owned by some agents (i.e., not welfare-irrelevant foreign entities) , we need to account for these agents' losses. However, these agents would (by definition) be net savers who would receive these profits in period 2. We conjecture that the net welfare change from redistributing period-2 incomes from net savers to either (i) all agents via increasing G or (ii) to net borrowers by increasing the ratio of delayed taxes, $1-\delta$, is strictly positive.

In reality, the fact that interest rates on debt exceed the risk-free rate is likely due to a combination of market power and default costs. Nevertheless, for our analyses, focusing on the pure-default cost scenario provides a conservative and simple framework. We also note that, even if default costs are the source of the interest spread, these may not be fully transferrable to the government. For example, the government may possess superior collection technology (through tax withholding) or better monitoring abilities. Hence, there is, in our view, considerable scope for increasing the quantitative welfare gains from delayed taxation. Figure 1 demonstrates this by considering how the MVPF exponentially increases when we reduce the ratio of government lending costs (z^g) relative to private lending costs (z).

B Dynamic Uncompensated and Compensated Elasticities

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

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

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

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

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

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

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

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

Now we can differentiate the second period intratemporal FOC (8) to get $d\ell_2 = \frac{u''(c_2)}{v''(\ell_2)} w_2^i (1-\tau_2) dc_2$ and substitute this into (B3) and collect multiplicative terms on dc_2 to get

$$dc_1 = d\ell_1 w_1 (1 - \bar{\delta}^i \tau_1) - \ell_1 w_1 d(\bar{\delta}^i \tau_1) - \frac{1}{R'(s^i)} \left(1 - [(1-\tau_2)w_2]^2 \frac{u''(c_2)}{v''(\ell_2)}\right) dc_2. \tag{B4}$$

We now eliminate dc_2 using the differentiated Euler equation for general $R(s)$. Differentiating $u'(c_1) = \beta R'(s) u'(c_2)$ yields

$$u''(c_1)\, dc_1 = \beta R''(s)\, u'(c_2)\, ds + \beta R'(s)\, u''(c_2)\, dc_2. \tag{B5}$$

We eliminate ds using the period-1 budget identity (B1): $ds = w_1(1-\delta\tau_1)\, d\ell_1 - \ell_1 w_1\, d(\delta\tau_1) - dc_1$. Substituting (B5) into (B4) and using this expression for ds, we obtain after rearranging:

$$dc_1 \cdot \iota^i = d\ell_1 w_1 (1 - \bar{\delta}^i \tau_1) - \ell_1 w_1 d(\bar{\delta}^i \tau_1), \tag{B6}$$

where the auxiliary variable ι^i is defined as

$$\iota^i \equiv 1 + \frac{\left(1 - \frac{[w_2^i(1-\tau_2)]^2\, u''(c_2^i)}{v''(\ell_2^i)}\right) u''(c_1^i)}{\beta \left[R'(s^i)^2\, u''(c_2^i) + \left(1 - \frac{[w_2^i(1-\tau_2)]^2\, u''(c_2^i)}{v''(\ell_2^i)}\right) R''(s^i)\, u'(c_2^i)\right]}. \tag{B7}$$

When $R''(s^i) = 0$, the $R''(s^i)\, u'(c_2^i)$ term vanishes and ι^i simplifies to $\iota_0^i \equiv 1 + (1 - B_2^i)\, u''(c_1^i)/(\beta R'(s^i)^2\, u''(c_2^i))$, the standard expression for a locally constant marginal rate. 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):

$$u''(c_1^i) w_1^i \left(1 - \bar{\delta}^i \tau_1\right) dc_1 - u'(c_1^i) w_1^i d\left(\bar{\delta}^i \tau_1\right) \quad = \quad v''(\ell_1^i) d\ell_1^i. \tag{B8}$$

We then substitute (B8) into (B6).

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

Then we re-arrange to obtain

$$\frac{d\ell_1^i}{d(\bar{\delta}^i \tau_1)} = \frac{-u'(c_1^i) w_1^i \iota^i - \ell_1^i w_1^i u''(c_1^i) w_1^i \left(1 - \bar{\delta}^i \tau_1\right)}{v''(\ell_1^i) \iota^i - w_1^i (1 - \bar{\delta}^i \tau_1) u''(c_1^i) w_1^i \left(1 - \bar{\delta}^i \tau_1\right)}. \tag{B10}$$

This may also be written as

$$\frac{d\ell_1^i}{d(\bar{\delta}^i \tau_1)} = \frac{-u'(c_1^i) w_1^i \iota^i - \ell_1^i w_1^i u''(c_1^i) \tilde{w}_1^i}{v''(\ell_1^i) \iota^i - \tilde{w}_1^i u''(c_1^i) \tilde{w}_1^i}, \tag{B11}$$

where ι^i is given by (B7). Writing out ι^i, the dependence on $R'(s^i)$ and $R''(s^i)$ is explicit.

B.2 Period-1 income effects

We want to derive an expression for $\frac{d\ell_1^i}{dG_1}$, which we may use together with (B11) in the Slutsky decomposition. All derivations assume $s^i \neq 0$.

We first note that period-1 and period-2 labor supply can be related by substituting the Euler equation (9) into (11) and then substituting in the FOC for ℓ_2 (8):

$$\beta R'(s^i)\frac{\tilde{w}_1^i}{\tilde{w}_2^i} = \frac{v'(\ell_1^i)}{v'(\ell_2^i)}, \tag{B12}$$

where $\tilde{w}_1^i = w_1^i\,(1-\tilde{\tau}_1^i)$ 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 $R'(s^i)$) reallocate labor intertemporally.

Deriving an expression for income effects. Under a dG_1 perturbation (holding all tax parameters fixed), differentiating the two budget constraints and eliminating ds via the period-1 budget identity yields

$$dc_1 = dG_1 + \tilde{w}_1 d\ell_1 + \frac{1}{R'(s)}\big((1-\tau_2)w_2\,d\ell_2 - dc_2\big). \tag{B13}$$

The remaining steps — eliminating $d\ell_2$ via the period-2 intratemporal FOC (8), dc_2 via the Euler equation (including the $R''(s)$ term as in (B5)), ds via the period-1 budget identity, and dc_1 via the differentiated period-1 intratemporal FOC (11) — follow the same procedure as in Section B.1. The result is

$$\frac{d\ell_1^i}{dG_1} = \frac{u''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i\,u''(c_1^i)\tilde{w}_1^i}, \tag{B14}$$

with the same ι^i from (B7).

B.3 Period-1 compensated elasticity via Slutsky

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,

$$\frac{d\ell_1^i}{d\tilde{w}_1^i} = \left(\frac{d\ell_1^i}{d\tilde{w}_1^i}\right)^c + \frac{d\ell_1^i}{dG_1}\ell_1^i. \tag{B15}$$

If we keep $\bar{\delta}^i$ fixed, $d\tilde{w}_1^i = -\bar{\delta}^i w_1^i d\tau_1 = \bar{\delta}^i w_1^i d(1-\tau_1)$. Hence, the relevant Slutsky equation becomes

$$\frac{d\ell_1^i}{d(1-\tau_1)} = \left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c + \frac{d\ell_1^i}{dG_1}\ell_1^i\bar{\delta}^i w_1^i. \tag{B16}$$

Substituting in for the LHS using equation (B11) and the second-term on the RHS using (B14), we obtain

$$-\bar{\delta}^i \frac{-u'(c_1^i)w_1^i\iota^i - \ell_1^i w_1^i u''(c_1^i)\tilde{w}_1^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i u''(c_1^i)\tilde{w}_1^i} = \left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c + \frac{u''(c_1^i)\tilde{w}_1}{v''(\ell_1^i)\iota^i - \tilde{w}_1 u''(c_1^i)\tilde{w}_1^i}\ell_1^i w_1^i \bar{\delta}^i. \tag{B17}$$

Rearrange and cancel out to get

$$\left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c = \bar{\delta}^i \frac{u'(c_1^i)w_1^i\iota^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i u''(c_1^i)\tilde{w}_1^i}. \tag{B18}$$

In terms of an elasticity, we can write it as

$$\varepsilon^c_{1,1-\tau_1} = \bar{\delta}^i \frac{1-\tau_1}{\ell_1^i} \frac{u'(c_1^i)w_1^i\iota^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i u''(c_1^i)\tilde{w}_1^i}. \tag{B19}$$

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 $u'(c_1^i)w_1^i = \frac{1}{1-\bar{\delta}^i\tau_1}v'(\ell_1^i)$ to get

$$\varepsilon^c_{1,1-\tau_1} = \frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1} \frac{1-\tau_1}{\ell_1^i} \frac{v'(\ell_1)\iota^i}{v''(\ell_1^i)\iota^i - \tilde{w}_1^i u''(c_1^i)\tilde{w}_1^i}. \tag{B20}$$

Writing out ι^i from (B7) and using the definitions of $\tilde{w}_1^i$ and $\bar{\delta}^i$, the dependence on $R'(s^i)$ and $R''(s^i)$ is again explicit.

B.4 Period-2 compensated cross elasticity via Slutsky

We derive the compensated cross-period elasticity $\varepsilon^{i,c}_{2,1-\tau_1}$, which measures the response of period-2 labor supply to a compensated change in the period-1 tax rate.

By Slutsky,

$$\frac{d\ell_2^i}{d(1-\tau_1)} = \left(\frac{d\ell_2^i}{d(1-\tau_1)}\right)^c + \frac{d\ell_2^i}{dG_1}\ell_1^i\bar{\delta}^i w_1^i. \tag{B21}$$

Differentiating the intertemporal labor condition (B12) with respect to G_1, holding $\tilde{w}_1^i$ and $\tilde{w}_2^i$ constant, eliminating $ds = (\iota^i - 1)\,dc_1$ via the period-1 budget identity, and substituting into the Euler equation (B5) to eliminate dc_2, yields

$$\frac{d\ell_2^i}{dG_1} = \frac{\tilde{w}_2^i\, v''(\ell_1^i)}{\beta R'(s^i)\,\tilde{w}_1^i\, v''(\ell_2^i)} \left(1 - \frac{R''(s^i)\,u'(c_1^i)}{R'(s^i)\,u''(c_1^i)}(\iota^i - 1)\right) \frac{d\ell_1^i}{dG_1}. \tag{B22}$$

The factor $\left(1 - \frac{R''(s^i)\,u'(c_1^i)}{R'(s^i)\,u''(c_1^i)}(\iota^i - 1)\right)$ reflects the fact that changing s changes $R'(s)$, which modifies dc_2/dc_1 through the $\beta R''(s)\,u'(c_2)\,ds$ term in the Euler differentiation (B5). Since $ds = (\iota^i - 1)\,dc_1$ is perturbation-independent (it holds for dG_1, $d(1-\tau_1)$, and $d(1-\delta)$ reforms), the same factor multiplies all cross-period expressions.

For the uncompensated response $d\ell_2^i/d(1-\tau_1)$, $\tilde{w}_1^i = w_1^i(1-\bar{\delta}^i\tau_1)$ varies with τ_1. Differentiating (B12) with respect to $(1-\tau_1)$ and eliminating ds and dc_2 as above yields

$$\frac{d\ell_2^i}{d(1-\tau_1)} = \left(1 - \frac{R''(s^i)\,u'(c_1^i)}{R'(s^i)\,u''(c_1^i)}(\iota^i - 1)\right)\left[\frac{\tilde{w}_2^i\,v''(\ell_1^i)}{\beta R'(s^i)\,\tilde{w}_1^i\,v''(\ell_2^i)}\,\frac{d\ell_1^i}{d(1-\tau_1)} - \frac{v'(\ell_2^i)}{v''(\ell_2^i)}\frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\right]. \quad \text{(B23)}$$

The first bracketed term applies the cross-period ratio to $d\ell_1^i/d(1-\tau_1)$, and the second captures the direct reallocation from the change in $\tilde{w}_1^i/\tilde{w}_2^i$ (absent from (B22) because $\tilde{w}_1^i$ is constant under dG_1). The leading factor multiplies both terms because both operate through dc_2/dc_1.

Substituting (B23) and (B22) into the Slutsky equation:

$$\begin{aligned}
\left(\frac{d\ell_2^i}{d(1-\tau_1)}\right)^c &= \frac{d\ell_2^i}{d(1-\tau_1)} - \frac{d\ell_2^i}{dG_1}\,\ell_1^i\bar{\delta}^i w_1^i \\
&= \left(1 - \frac{R''(s^i)\,u'(c_1^i)}{R'(s^i)\,u''(c_1^i)}(\iota^i - 1)\right)\left\{\frac{\tilde{w}_2^i\,v''(\ell_1^i)}{\beta R'(s^i)\,\tilde{w}_1^i\,v''(\ell_2^i)}\left[\frac{d\ell_1^i}{d(1-\tau_1)} - \frac{d\ell_1^i}{dG_1}\,\ell_1^i\bar{\delta}^i w_1^i\right] - \frac{v'(\ell_2^i)}{v''(\ell_2^i)}\frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\right\} \\
&= \left(1 - \frac{R''(s^i)\,u'(c_1^i)}{R'(s^i)\,u''(c_1^i)}(\iota^i - 1)\right)\left[\frac{\tilde{w}_2^i\,v''(\ell_1^i)}{\beta R'(s^i)\,\tilde{w}_1^i\,v''(\ell_2^i)}\left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c - \frac{v'(\ell_2^i)}{v''(\ell_2^i)}\frac{\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\right],
\end{aligned} \quad \text{(B24)}$$

where in the second line we factor the R'' correction and use the factored form of (B23), and in the third line we recognize the period-1 Slutsky decomposition $\frac{d\ell_1^i}{d(1-\tau_1)} - \frac{d\ell_1^i}{dG_1}\,\ell_1^i\bar{\delta}^i w_1^i = \left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c$. In terms of elasticities, $\varepsilon^{i,c}_{2,1-\tau_1} = \frac{1-\tau_1}{\ell_2^i}\left(\frac{d\ell_2^i}{d(1-\tau_1)}\right)^c$, so

$$\varepsilon^{i,c}_{2,1-\tau_1} = \left(1 - \frac{R''(s^i)\,u'(c_1^i)}{R'(s^i)\,u''(c_1^i)}(\iota^i - 1)\right)\left[\frac{\tilde{w}_2^i\,v''(\ell_1^i)}{\beta R'(s^i)\,\tilde{w}_1^i\,v''(\ell_2^i)}\,\frac{\ell_1^i}{\ell_2^i}\,\varepsilon^{i,c}_{1,1-\tau_1} - \frac{v'(\ell_2^i)}{\ell_2^i\,v''(\ell_2^i)}\,\frac{(1-\tau_1)\,\bar{\delta}^i}{1-\bar{\delta}^i\tau_1}\right]. \quad \text{(B25)}$$

B.5 Explicit timing correction Ψ^i at $\delta = 1$

The timing correction $\Psi^i = (\Psi_1^i, \Psi_2^i, \Psi_s^i)$ is defined by $\mathbf{J}^i\Psi^i = \Delta\mathbf{b}_\Gamma^i$ (eq. 31). At $\delta = 1$, $J_{12} = J_{21} = 0$ (block-sparse), so the system can be solved by block elimination.

At $\delta = 1$ with $\tau_1 = \tau_2 = \tau$ and $\Gamma^i(1) = \tau\Delta_r^i$:

$$\Delta\mathbf{b}_\Gamma^i\big|_{\delta=1} = \begin{pmatrix} -(1-\Delta_r^i)\,u''(c_1)\tau y_1\tilde{w}_1 \\ u''(c_2)(1+r)\tau y_1\tilde{w}_2 \\ -\tau y_1(1-\Delta_r^i)\left[u''(c_1) + \beta R'(s)^2 u''(c_2)\right] \end{pmatrix}, \quad \text{(B26)}$$

where we used the Euler equation $\beta(1+r)u'(c_2) = (1-\Delta_r^i)\,u'(c_1)$ to simplify $\Delta b_{\Gamma,1}$.

From the first two rows of the block-sparse system:

$$\Psi_1^i = \frac{\Delta b_{\Gamma,1}^i + u''(c_1)\tilde{w}_1\,\Psi_s^i}{u''(c_1)\tilde{w}_1^2 - v''(\ell_1)}, \tag{B27}$$

$$\Psi_2^i = \frac{\Delta b_{\Gamma,2}^i - u''(c_2)\tilde{w}_2 R'(s)\,\Psi_s^i}{u''(c_2)\tilde{w}_2^2 - v''(\ell_2)}. \tag{B28}$$

Substituting into the third row and solving for Ψ_s^i:

$$\Psi_s^i = \frac{\Delta b_{\Gamma,3}^i - \dfrac{u''(c_1)\tilde{w}_1\,\Delta b_{\Gamma,1}^i}{u''(c_1)\tilde{w}_1^2 - v''(\ell_1)} + \dfrac{\beta R'(s)\,u''(c_2)\tilde{w}_2\,\Delta b_{\Gamma,2}^i}{u''(c_2)\tilde{w}_2^2 - v''(\ell_2)}}{J_{33} + \dfrac{u''(c_1)^2\tilde{w}_1^2}{u''(c_1)\tilde{w}_1^2 - v''(\ell_1)} + \dfrac{\beta R'(s)^2\,u''(c_2)^2\tilde{w}_2^2}{u''(c_2)\tilde{w}_2^2 - v''(\ell_2)}}, \tag{B29}$$

where $J_{33} = -u''(c_1) - \beta R''(s)u'(c_2) - \beta R'(s)^2 u''(c_2)$.

Once Ψ_s^i is determined, Ψ_1^i and Ψ_2^i follow from (B27)–(B28). The fiscal contribution of the timing correction is $X_\Psi^i = \tau w_1^i \Psi_1^i + \tau w_2^i \Psi_2^i/(1+r)^T$.

When $R''(s^i) = 0$, the proportionality in Lemma 1 holds exactly, giving $\Psi_1^i = \Psi_2^i = 0$ and $\Psi_s^i = (1{+}r)\tau y_1^i/R'(s^i)$: the agent adjusts savings to cover the delayed payment, with no effect on labor supply. For non-borrowers ($R'(s^i) = 1{+}r$), $\Psi_s^i = \tau y_1^i$.

Cancellation of Ψ^i in the welfare formula at $\delta = 1$. At $\delta = 1$, the RHS residual $\Delta\mathbf{b}_\Gamma^i$ that defines Ψ^i (eq. 31) satisfies $\Delta\mathbf{b}_\Gamma^i = (1{+}r)\tau y_1^i\left[\mathbf{b}_{G_1}^i/R'(s^i) - \mathbf{b}_{G_2}^i\right]$, where $\mathbf{b}_{G_t}^i$ are the RHS vectors for lump-sum transfer perturbations. By linearity, $\Psi_t^i = (1{+}r)\tau y_1^i\left[d\ell_t^i/dG_1/R'(s^i) - d\ell_t^i/dG_2\right]$, and the fiscal contribution becomes $X_\Psi^i = -(1{+}r)\tau y_1^i\,\eta^i$ where $\eta^i \equiv \rho_2^i - \rho_1^i/R'(s^i)$. Combined with the identity $\Delta_r^i\,\rho_1^i = [\rho_1^i - (1{+}r)\rho_2^i] + (1{+}r)\eta^i$, the timing correction and income effect correction cancel exactly in the welfare formula, yielding the clean income term $y_1^i[\rho_1^i - (1{+}r)\rho_2^i]$ in Corollary 1.

C Proofs

C.1 Proof of Proposition 1

For this proof, we allow for a general $1-\delta$ but restrict the loan limit for the subsidized loan to be zero ($x = 0$). For convenience, we rewrite the Lagrangian as

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

where the bracketed term accounts for the net-of-default-cost period-1 tax revenue effect.

Assuming $\tau_1 = \tau_2 = \tau$, differentiating the Lagrangian of the government's optimization

problem with respect to τ, invoking the envelope theorem on V_i terms,

$$V^i = \max_{\ell_1^i \ell_2^i, c_1^i, c_2^i, s^i} \Big(u(c_1^i) - v(\ell_1^i) + \beta[u(c_2^i) - v(\ell_2^i)] \Big), \quad \text{subject to (2) and (3)}, \tag{C1}$$

and setting the differentiated Lagrangian equal to zero, yields

$$- \sum_i \pi^i \Big(\delta u_1'(\cdot) y_1^i + \Big([1-\delta](1+r) y_1^i + y_2^i \Big) u_2'(\cdot) \beta \Big) \tag{C2}$$

$$+ \lambda \sum_i \pi^i \left(\Big[1 - (1-\delta)\Delta_{r,g}^i \Big] \left[y_1^i + \tau \frac{dy_1^i}{d\tau} \right] + \frac{1}{1+r} \left[y_2^i + \tau \frac{dy_2^i}{d\tau} \right] \right) = 0. \tag{C3}$$

Substituting in for g_t^i and $\ell_t^i w_t^i = y_t^i$ and rearranging yields

$$- \sum_i \pi^i \Big(\delta g_1^i y_1^i + \Big([1-\delta] y_1^i + \frac{1}{1+r} y_2^i \Big) g_2^i \Big) \tag{C4}$$

$$+ \sum_i \pi^i \left(\Big[1 - (1-\delta)\Delta_{r,g}^i \Big] \left[y_1^i + \tau \frac{dy_1^i}{d\tau} \right] + \frac{1}{1+r} \left[y_2^i + \tau \frac{dy_2^i}{d\tau} \right] \right) = 0. \tag{C5}$$

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

$$\sum_i \pi^i \Big(\delta g_1^i y_1^i + \Big([1-\delta] y_1^i + \frac{1}{1+r} y_2^i \Big) g_2^i \Big) = \tag{C6}$$

$$\sum_i \pi^i \left(\Big[1 - (1-\delta)\Delta_{r,g}^i \Big] \left(y_1^i - \frac{\tau}{1-\tau} y_1^i \varepsilon_{1,1-\tau}^i \right) + \frac{1}{1+r} \left(y_2^i - \frac{\tau}{1-\tau} y_2^i \varepsilon_{2,1-\tau}^i \right) \right). \tag{C7}$$

Equation (26) follows by re-arrangement.

The FOC for G can be written as:

$$\sum_i \pi^i \Big(u'(c_1) + \beta u'(c_2) \Big) = \lambda \sum_i \pi^i \left(1 + \frac{1}{1+r} - \tau \frac{d}{dG} \left[\Big[1 - (1-\delta)\Delta_{r,g}^i \Big] dy_1^i - \tau \frac{1}{1+r} dy_2^i \right] \right). \tag{C8}$$

$$\sum_i \pi^i (u'(c_1^i) + \beta u'(c_2^i)) = \lambda \sum_i \pi^i \left(1 + \frac{1}{1+r} - \left[\tau [1 - (1-\delta)\Delta_{r,g}^i] \frac{dy_1^i}{dG} + \frac{\tau}{1+r} \frac{dy_2^i}{dG} \right] \right) \tag{C9}$$

Condition (28) follows by noting that increasing G amounts to an equal increase in per-period transfers, $dG_1 = dG_2 = dG$. Using ρ_1^i and ρ_2^i to denote the revenue effects of period-1 and period-2 transfers (defined in (24)–(25)), the behavioral revenue response to dG is $(\rho_1^i + \rho_2^i)\, dG$. Substituting into the FOC for G produces

$$\sum_i \pi^i (u'(c_1^i) + \beta u'(c_2^i)) = \lambda \sum_i \pi^i \left(1 + \frac{1}{1+r} - \rho_1^i - \rho_2^i \right). \tag{C10}$$

Dividing by λ and using the expressions for g_1^i and g_2^i produces (28).

C.2 Proof of Lemma 1

The endogenous variables (allocations) are $\mathbf{a}^i = (\ell_1^i, \ell_2^i, s^i)$ satisfying the FOC system $\mathbf{F}(\mathbf{a}; \delta, \tau_1, \tau_2) = 0$. Total differentiation gives $\mathbf{J}^i\, d\mathbf{a} = \mathbf{b}_p$ where $\mathbf{b}_p = -\partial \mathbf{F}/\partial p\big|_{\mathbf{a}}$ for parameter p.

General-δ Jacobian. Let $\bar{\delta}^i \equiv 1 - (1-\delta)\Delta_r^i$, $\tilde{w}_1 = w_1(1-\delta\tau_1)$ and $\tilde{w}_2 = w_2(1-\tau_2)$. The Jacobian $\mathbf{J}^i(\delta) = \partial \mathbf{F}/\partial \mathbf{a}$ at general δ is:

$$
\begin{aligned}
J_{11} &= u''(c_1)\tilde{w}_1^2 + \beta(1{+}r)^2(1{-}\delta)^2\tau_1^2 w_1^2 u''(c_2) - v''(\ell_1),\\
J_{12} &= -\beta(1{+}r)(1{-}\delta)\tau_1 w_1\, u''(c_2)\, \tilde{w}_2,\\
J_{13} &= -u''(c_1)\tilde{w}_1 - \beta(1{+}r)(1{-}\delta)\tau_1 w_1\, u''(c_2)\, R'(s),\\
J_{21} &= -u''(c_2)(1{+}r)(1{-}\delta)\tau_1 w_1\, \tilde{w}_2,\\
J_{22} &= u''(c_2)\tilde{w}_2^2 - v''(\ell_2),\\
J_{23} &= u''(c_2)\tilde{w}_2\, R'(s),\\
J_{31} &= u''(c_1)\tilde{w}_1 + \beta R'(s)\, u''(c_2)(1{+}r)(1{-}\delta)\tau_1 w_1,\\
J_{32} &= -\beta R'(s)\, u''(c_2)\, \tilde{w}_2,\\
J_{33} &= -u''(c_1) - \beta R''(s)\, u'(c_2) - \beta R'(s)^2\, u''(c_2).
\end{aligned}
\tag{C11}
$$

At $\delta = 1$: $J_{12} = J_{21} = 0$ (block-sparse structure), and J_{11} simplifies to $u''(c_1)\tilde{w}_1^2 - v''(\ell_1)$, recovering the $\delta = 1$ Jacobian used in the elasticity derivations (Appendix B).

Differentiation with respect to $(1-\delta)$. Setting $p = 1-\delta$, at general δ:

$$
\begin{aligned}
b_{1-\delta,1} &= -u''(c_1)\tau_1 y_1 \tilde{w}_1 - u'(c_1) w_1 \tau_1 + \beta(1{+}r)\tau_1 w_1 u'(c_2) - \beta(1{-}\delta)(1{+}r)^2\tau_1^2 w_1 y_1 u''(c_2),\\
b_{1-\delta,2} &= u''(c_2)(1{+}r)\tau_1 y_1 \tilde{w}_2,\\
b_{1-\delta,3} &= -u''(c_1)\tau_1 y_1 - \beta R'(s)\, u''(c_2)(1{+}r)\tau_1 y_1.
\end{aligned}
\tag{C12}
$$

Differentiation with respect to $(1-\tau_1)$. Setting $p = 1-\tau_1$ and holding τ_2 fixed, at general δ:

$$
\begin{aligned}
b_{1-\tau,1} &= -u''(c_1)\delta y_1 \tilde{w}_1 - u'(c_1) w_1 \delta - \beta(1{+}r)(1{-}\delta) w_1 u'(c_2) + \beta(1{+}r)^2(1{-}\delta)^2\tau_1 w_1 y_1 u''(c_2),\\
b_{1-\tau,2} &= -u''(c_2)(1{+}r)(1{-}\delta) y_1 \tilde{w}_2,\\
b_{1-\tau,3} &= -u''(c_1)\delta y_1 + \beta R'(s)\, u''(c_2)(1{+}r)(1{-}\delta) y_1.
\end{aligned}
\tag{C13}
$$

Combining the two marginal changes. Since $\mathbf{J}\, d\mathbf{a}_{1-\delta} = \mathbf{b}_{1-\delta}$ and $\mathbf{J}\, d\mathbf{a}_{1-\tau} = \mathbf{b}_{1-\tau}$:

$$
\frac{d\mathbf{a}^i}{d(1-\delta)} = \Gamma^i(\delta)\frac{d\mathbf{a}^i}{d(1-\tau_1)} + (\mathbf{J}^i)^{-1}\Delta\mathbf{b}_\Gamma^i,
\tag{C14}
$$

which gives (30) with Ψ_t^i being the t-th component of $(\mathbf{J}^i)^{-1}\Delta\mathbf{b}_\Gamma^i$, where $\Delta\mathbf{b}_\Gamma^i \equiv \mathbf{b}_{1-\delta}^i - \Gamma^i(\delta)\,\mathbf{b}_{1-\tau}^i$.

Why $\Psi_t^i = 0$ when $R''(s) = 0$ ($t \in \{1,2\}$). When $R''(s^i) = 0$, $R'(s^i)$ is locally constant. The individual's labor supply ℓ_t depends on (δ, τ_1) only through the effective tax rate $\tilde{\tau}_1^i = \bar{\delta}^i \tau_1$, because the present value of period-1 taxes (discounted at rate $R'(s)$) equals $\tilde{\tau}_1 y_1$, which depends on (δ, τ_1) through $\bar{\delta}\tau_1$, and with $R'(s)$ constant, the Euler equation does not create a differential response to δ vs. τ_1. Therefore the chain-rule decomposition is valid:

$$\frac{d\ell_t^i}{d(1-\delta)} = \frac{d\ell_t^i}{d(1-\tilde{\tau}_1)} \cdot \tau_1 \Delta_r^i, \qquad \frac{d\ell_t^i}{d(1-\tau_1)} = \frac{d\ell_t^i}{d(1-\tilde{\tau}_1)} \cdot \bar{\delta}^i, \tag{C15}$$

giving $d\ell_t^i/d(1-\delta) = \Gamma^i(\delta) \cdot d\ell_t^i/d(1-\tau_1)$ and hence $\Psi_t^i(\delta) = 0$ for all δ.

C.3 Proof of Proposition 2

We 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 $\tau_1 y_1^i[1-(1-\delta)\Delta_{r,g}^i]$, the derivative with respect to $(1-\delta)$ gives both behavioral and mechanical effects:

$$\sum_i \pi^i \left(u'(c_1^i)\tau_1 y_1^i - \beta u'(c_2^i)(1+r)\tau_1 y_1^i\right) + \lambda \sum_i \pi^i \left(\tau_1[1-(1-\delta)\Delta_{r,g}^i]\frac{dy_1^i}{d(1-\delta)} + \tau_2 \frac{1}{1+r}\frac{dy_2^i}{d(1-\delta)} - \Delta_{r,g}^i \tau_1 y_1^i\right) = 0. \tag{C16}$$

We then use Lemma 1 to substitute $\frac{dy_t^i}{d(1-\delta)} = \Gamma^i(\delta)\frac{dy_t^i}{d(1-\tau_1)} + w_t^i \Psi_t^i(\delta)$. The behavioral revenue for agent i becomes:

$$\Gamma^i(\delta)\left(\tau_1\left[1-(1-\delta)\Delta_{r,g}^i\right]\frac{dy_1^i}{d(1-\tau_1)} + \tau_2\frac{1}{1+r}\frac{dy_2^i}{d(1-\tau_1)}\right) + \tau_1\left[1-(1-\delta)\Delta_{r,g}^i\right]w_1^i\Psi_1^i + \frac{\tau_2}{1+r}w_2^i\Psi_2^i - \Delta_{r,g}^i\tau_1 y_1^i. \tag{C17}$$

We use the Slutsky equation to rewrite $\frac{dy_t^i}{d(1-\tau_1)} = \left(\frac{dy_t^i}{d(1-\tau_1)}\right)^c + y_1^i\bar{\delta}^i\frac{dy_t^i}{dG_1}$ (the compensating variation of a unit increase in $(1-\tau_1)$ is $y_1^i\bar{\delta}^i$ in G_1-equivalent terms, since the tax cut increases period-1 income by δy_1^i and period-2 income by $(1-\delta)(1+r)y_1^i$; see Appendix B). Further rearranging and using elasticity notation yields

$$y_1^i\frac{1}{1-\tau_1}\Gamma^i(\delta)\left(\left[1-(1-\delta)\Delta_{r,g}^i\right]\left\{\tau_1\varepsilon_{11}^{i,c} + (1-\tau_1)\bar{\delta}^i\tau_1\frac{dy_1^i}{dG_1}\right\} + \frac{1}{1+r}\left\{\tau_2\frac{y_2^i}{y_1^i}\varepsilon_{21}^{i,c} + (1-\tau_1)\bar{\delta}^i\tau_2\frac{dy_2^i}{dG_1}\right\}\right) - \Delta_{r,g}^i\tau_1 y_1^i. \tag{C18}$$

Further using the definition $\rho_1^i = \frac{d}{dG_1}\left(\tau_1 y_1^i\left[1-(1-\delta)\Delta_{r,g}^i\right] + \frac{\tau_2 y_2^i}{1+r}\right)$, the Γ-part of the behavioral revenue can be written as:

$$y_1^i\,\Gamma^i(\delta)\left(\frac{\tau_1\{1-(1-\delta)\Delta_{r,g}^i\}}{1-\tau_1}\varepsilon_{11}^{i,c} + \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{21}^{i,c} + \bar{\delta}^i\rho_1^i\right). \tag{C19}$$

The Ψ correction contributes the fiscal term $\tau_1[1-(1-\delta)\Delta_{r,g}^i]\,w_1^i\Psi_1^i + \frac{\tau_2}{1+r}\,w_2^i\Psi_2^i$, which we

group with the income effect $\bar{\delta}^i\rho_1^i$ in (C19). Using $g_1^i = u'(c_1^i)/\lambda$ and $g_2^i = \beta(1+r)u'(c_2^i)/\lambda$ to rewrite the envelope term, and re-arranging yields (32).

C.4 Proof of Corollary 1

Using the derivations from the proof of Proposition 2 (Section C.3), and setting $\tau_1 = \tau_2 = \tau$ and $\delta = 1$ as assumed in the corollary, we have $\Gamma^i(1) = \tau\Delta_r^i$, $\bar{\delta}^i = 1$, and $1-(1-\delta)\Delta_{r,g}^i = 1$. The compensated revenue terms give $\tau\Delta_r^i[\frac{\tau}{1-\tau}y_1^i\varepsilon_{11}^{i,c} + \frac{1}{1+r}\frac{\tau}{1-\tau}y_2^i\varepsilon_{21}^{i,c}]$. The income terms $y_1^i\Gamma^i\bar{\delta}^i\rho_1^i + \tau w_1^i\Psi_1^i + \frac{\tau}{1+r}w_2^i\Psi_2^i$ combine as follows: the $\Gamma\bar{\delta}\rho_1 = \tau\Delta_r^i\rho_1^i$ part contributes $\tau y_1^i\Delta_r^i\rho_1^i$, and the Ψ fiscal contribution is $X_\Psi^i = -(1+r)\tau y_1^i\eta^i$ (see Section B.5). Using $\Delta_r^i\rho_1^i = [\rho_1^i - (1+r)\rho_2^i] + (1+r)\eta^i$, the η terms cancel:

$$\tau y_1^i\Delta_r^i\rho_1^i + X_\Psi^i = \tau y_1^i[\rho_1^i - (1+r)\rho_2^i] + (1+r)\tau y_1^i\eta^i - (1+r)\tau y_1^i\eta^i = \tau y_1^i[\rho_1^i - (1+r)\rho_2^i].$$

Using the intertemporal optimality condition $(g_1^i - g_2^i) = g_1^i\Delta_r^i$, substituting $g_1^i = u'(c_1^i)/\lambda$, factoring τ, and rearranging yields (33).

C.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.

$$\frac{dW}{dx} = \sum_i \pi^i\left(u'(c_1^i) - (1+r)\beta u'(c_2^i)\right) + \lambda\sum_i \pi^i\left(\tau_1\frac{dy_1^i}{dx} + \frac{\tau_2}{1+r}\frac{dy_2^i}{dx} - \Delta_{r,g}^i\right). \tag{C20}$$

A loan of dx gives $+dx$ in period 1 and $-(1+r)dx$ in period 2, which is equivalent to $dG_1 = dx$ and $dG_2 = -(1+r)dx$ simultaneously. By the implicit function theorem, $\frac{dy_t^i}{dx} = \frac{dy_t^i}{dG_1} - (1+r)\frac{dy_t^i}{dG_2}$. Using the definitions of ρ_1^i and ρ_2^i (which at $\delta = 1$ reduce to $\rho_s^i = \tau_1\frac{dy_1^i}{dG_s} + \frac{\tau_2}{1+r}\frac{dy_2^i}{dG_s}$ for $s = 1, 2$), the behavioral revenue effect of the loan is $\rho_1^i - (1+r)\rho_2^i$. Hence,

$$\frac{1}{\lambda}\frac{dW}{dx} = \sum_i \pi^i\left(g_1^i - g_2^i\right) + \sum_i \pi^i\left[\rho_1^i - (1+r)\rho_2^i - \Delta_{r,g}^i\right]. \tag{C21}$$

The result follows by noting that for non-borrowers ($s^i \geq 0$), $g_1^i = g_2^i$ and a loan has no behavioral effect ($\rho_1^i - (1+r)\rho_2^i = 0$ since the loan has zero present value at $R'(s^i) = 1+r$). For borrowers, $g_2^i = g_1^i \cdot (1+r)/R'(s^i)$, so $(g_1^i - g_2^i) = \Delta_r^i g_1^i$.

C.6 Proof of Proposition 3

From Corollary 1, we have:

$$\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\bigg|_{\delta=1} = \tau\sum_{i:s^i<0}\pi^i\Big[\Delta_r^i y_1^i g_1^i + \Delta_r^i\frac{\tau}{1-\tau}\left(y_1^i\varepsilon_{11}^{i,c} + \frac{1}{1+r}y_2^i\varepsilon_{21}^{i,c}\right) + y_1^i\left(\rho_1^i - (1+r)\rho_2^i\right)\Big] - \tau\sum_{i:s^i<0}\pi^i y_1^i\Delta_{r,g}^i. \tag{C22}$$

From Lemma 2:

$$\frac{1}{\lambda}\frac{dW}{dx}\Big|_{\delta=1} = \sum_{i:s^i<0} \pi^i \left[\Delta_r^i g_1^i + \rho_1^i - (1+r)\rho_2^i - \Delta_{r,g}^i\right]. \tag{C23}$$

Multiplying by $\tau\bar{y}_1$:

$$\tau\bar{y}_1 \cdot \frac{1}{\lambda}\frac{dW}{dx}\Big|_{\delta=1} = \tau \sum_{i:s^i<0} \pi^i \bar{y}_1 \left[\Delta_r^i g_1^i + \rho_1^i - (1+r)\rho_2^i - \Delta_{r,g}^i\right]. \tag{C24}$$

Subtracting this from the Corollary expression and rearranging:

$$\begin{aligned}\frac{1}{\lambda}\frac{dW}{d(1-\delta)}\Big|_{\delta=1} = \tau\bar{y}_1 \cdot \frac{1}{\lambda}\frac{dW}{dx}\Big|_{\delta=1} &+ \tau \sum_{i:s^i<0} \pi^i \Delta_r^i \frac{\tau}{1-\tau}\left(y_1^i \varepsilon_{11}^{i,c} + \frac{1}{1+r} y_2^i \varepsilon_{21}^{i,c}\right) \\ &+ \tau \sum_{i:s^i<0} \pi^i \Delta_r^i (y_1^i - \bar{y}_1) g_1^i + \tau \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1)\left(\rho_1^i - (1+r)\rho_2^i\right) - \tau \sum_{i:s^i<0} \pi^i (y_1^i - \bar{y}_1)\Delta_{r,g}^i.\end{aligned}$$

This establishes (36).

D 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 (17), with some fixed value of $1-\delta$.*

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

$$\sum_i \pi_i y_1^i \left[\delta g_1^i + [1-\delta] g_2^i\right] = \sum_i \pi_i y_1^i \left[\left[1 - (1-\delta)\Delta_{r,g}^i\right] - \frac{\tau_1}{1-\tau_1}\left(\left[1-(1-\delta)\Delta_{r,g}^i\right]\varepsilon_{11}^i + \frac{1}{1+r}\frac{\tau_2}{\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{21}^i\right)\right], \tag{D1}$$

$$\sum_i \pi_i y_2^i \left[g_2^i\right] = \sum_i \pi_i y_2^i \left[1 - \frac{\tau_2}{1-\tau_2}\left(\varepsilon_{22}^i + (1+r)\frac{\tau_1}{\tau_2}\left[1-(1-\delta)\Delta_{r,g}^i\right]\frac{y_1^i}{y_2^i}\varepsilon_{12}^i\right)\right]. \tag{D2}$$

(ii) The optimal transfers G_1 and G_2 satisfy:

$$\sum_i \pi^i g_1^i = \sum_i \pi^i \left[1 - \rho_1^i\right], \tag{D3}$$

$$\sum_i \pi^i g_2^i = \sum_i \pi^i \left[1 - (1+r)\rho_2^i\right]. \tag{D4}$$

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

$$-\sum_i \pi^i y_1^i \Big(\delta u_1'(\cdot) + [1-\delta] u_2'(\cdot)\beta(1+r)\Big) + \lambda \sum_i \pi^i \left[\Big[1-(1-\delta)\Delta_{r,g}^i\Big]\left(y_1^i + \tau_1 \frac{dy_1^i}{d\tau_1}\right) + \frac{\tau_2}{1+r}\frac{dy_2^i}{d\tau_1}\right] = 0, \tag{D5}$$

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:

$$-\sum_i \pi^i y_1^i \Big(\delta u_1'(\cdot) + [1-\delta] u_2'(\cdot)\beta(1+r)\Big) + \lambda \sum_i \pi^i y_1^i \left[\Big[1-(1-\delta)\Delta_{r,g}^i\Big]\left(1 - \frac{\tau_1}{1-\tau_1}\varepsilon_{1,1-\tau_1}^i\right) - \frac{1}{1+r}\frac{\tau_2}{1-\tau_1}\frac{y_2^i}{y_1^i}\varepsilon_{2,1-\tau_1}^i\right] = 0. \tag{D6}$$

We then substitute in for g_1^i and g_2^i and obtain

$$-\sum_i \pi^i y_2^i \beta u_2'(\cdot) + \lambda \sum_i \pi^i \left[\Big[1-(1-\delta)\Delta_{r,g}^i\Big]\tau_1 \frac{dy_1}{d\tau_2} + \frac{1}{1+r}\left(y_2^i + \tau_2 \frac{dy_2^i}{d\tau_2}\right)\right] = 0. \tag{D7}$$

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

$$\sum_i \pi^i u_1'(\cdot) = \lambda \sum_i \pi^i \left[1 - \Big[1-(1-\delta)\Delta_{r,g}^i\Big]\tau_1 \frac{dy_1^i}{dG_1} - \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_1}\right], \tag{D8}$$

$$\beta \sum_i \pi^i u_2'(\cdot) = \lambda \sum_i \pi^i \left[\frac{1}{1+r} - \Big[1-(1-\delta)\Delta_{r,g}^i\Big]\tau_1 \frac{dy_1^i}{dG_2} - \frac{1}{1+r}\tau_2 \frac{dy_2^i}{dG_2}\right]. \tag{D9}$$

We substitute the definition of ρ_2^i from (25) into the G_2 FOC to obtain (D4). Note that unlike (D3), the coefficient on ρ_2^i is $(1+r)$ rather than 1, reflecting the intertemporal price of a period-2 transfer. Similarly, we substitute ρ_1^i from (24) into the G_1 FOC to obtain (D3). ■

The weighted average of the social weights on the LHS of (D1) 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 (D1) and (D2) 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 (D3) and (D4) 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 income effects. 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

[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.

they are for period 1 labor supply. This is because financially constrained agents discount future cash flows at a higher rate.

E 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 (see, e.g., Choukhmane and Palmer (2023)). 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.

We note that $\Delta_r^i \equiv 1 - (1+r)/R'(s^i) \geq 0$ for each agent, with $\Delta_r^i > 0$ for borrowers ($s^i < 0$) and $\Delta_r^i = 0$ for non-borrowers ($s^i \geq 0$). Since all smoothing and error terms below are proportional to Δ_r^i, the contributions of non-borrowing agents vanish from these sums. At the optimum, the government would not extend delayed taxation to non-borrowers, as this would incur fiscal costs ($\Delta_{r,g}^i > 0$) with no smoothing benefit.

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

$$\begin{aligned} W^* &= \sum_i \pi^i V^i - \lambda^* \left(-\sum_i \pi^i \left(\tau_1 w_1^i \ell_1^i + \frac{\tau_2 w_2^i \ell_2^i}{1+r} - (1-\delta)\Delta_{r,g}^i \tau_1 w_1^i \ell_1^i \right) + G_1 + \frac{G_2}{1+r} + M \right) \\ &+ \sum_{i:\beta^i<\beta^*} \pi^i (V^{i*} - V^i), \end{aligned} \tag{E1}$$

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

$$(\beta^* - \beta) \sum_{i:\beta^i<\beta^*} \pi^i \left[u(c_2^i) - v(\ell_2^i) \right]. \tag{E2}$$

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

$$\begin{aligned} \frac{dW^*}{d(1-\delta)}\Big|_{\delta=1} &= \tau_1 \cdot \sum_i \pi_i y_1^i \left(u'(c_1^i) - \beta^i (1+r) u'(c_2^i) \right) \\ &- \lambda^* \sum_i \pi_i y_1^i \left[\tau_1 \Delta_{r,g}^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{11}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{21}^{i,c} \right) + \rho_1^i - (1+r)\rho_2^i \right] \\ &- \lambda^* \sum_i \pi^i \tau_1 w_1^i \ell_1^i \Delta_{r,g}^i \\ &+ (\beta^* - \beta) \sum_{i:\beta^i<\beta^*} \pi^i \left[u'(c_2^i) \frac{dc_2^i}{d(1-\delta)} - v'(\ell_2^i) \frac{d\ell_2^i}{d(1-\delta)} \right]. \end{aligned} \tag{E3}$$

This decomposition follows from the fact that W^* equals W plus the error term (E2). We can

then apply Corollary 1 to obtain the first three terms in the marginal welfare effects expression (E3).

Writing out the error term. The error term equals

$$(\beta^* - \beta) \sum_{i:\beta^i<\beta^*} \pi^i \left[u'(c_2^i) \frac{dc_2^i}{d(1-\delta)} - v'(\ell_2^i) \frac{d\ell_2^i}{d(1-\delta)} \right], \tag{E4}$$

which says that paternalism increases the welfare effects of delayed taxation if delayed taxation increases period-2 utility. Note that

$$\frac{dc_2}{d(1-\delta)} = -(1+r)\tau_1 w_1 \ell_1 - (1+r)[1-\delta]\tau_1 w_1 \frac{d\ell_1}{d(1-\delta)} + w_2(1-\tau_2)\frac{d\ell_2}{d(1-\delta)} + R'(s)\frac{ds}{d(1-\delta)}. \tag{E5}$$

Hence, the bracketed terms in (E4) that contain $d\ell_2$ can be written

$$u'(c_2)w_2(1-\tau_2)\frac{d\ell_2}{d(1-\delta)} - v'(\ell_2)\frac{d\ell_2}{d(1-\delta)}. \tag{E6}$$

This expression equals zero due to the FOC that $u'(c_2)w_2(1-\tau_2) = v'(\ell_2)$. Hence, period-2 labor supply adjustments do not matter. We can thus rewrite the error term as

$$(\beta^* - \beta) \sum_{i:\beta^i<\beta^*} \pi^i u'(c_2^i) \Bigg[\underbrace{-(1+r)\tau_1 w_1^i \ell_1^i}_{\text{DT Repayment}} + \underbrace{R'(s)\frac{ds}{d(1-\delta)}}_{\text{Reduced loan repayment}} \Bigg]. \tag{E7}$$

This expression says that paternalism amplifies welfare gains if the delayed tax repayment is lower than the decrease in debt repayment. Debt repayments decrease as delayed taxation crowds out borrowing. Since $R'(s)$ is higher than $1+r$, total repayments (delayed tax plus private borrowing) may decrease even if there's only partial crowd out of private borrowing. In other words, if delayed taxation increases period-2 liquidity, welfare gains are higher under delayed taxation. This is due to the government valuing period-2 utility higher than low-β^i individuals.

We proceed by rewriting the error term such that it may be expressed in terms of allocations and preference parameters. This allows us to show that it is weakly positive, and strictly positive when ever there exists impatient borrowers ($\beta^i < \beta^*$ and $s^i < 0$).

Rewriting the error term. We now simplify the bracket in the error term. From the period-1 budget constraint differentiated at $\delta = 1$,

$$\frac{ds}{d(1-\delta)} = \tau_1 y_1 + w_1(1-\tau_1)\frac{d\ell_1}{d(1-\delta)} - \frac{dc_1}{d(1-\delta)}. \tag{E8}$$

Differentiating the Euler equation $u'(c_1) = \beta R'(s)u'(c_2)$ gives

$$u''(c_1)\frac{dc_1}{d(1-\delta)} = \beta R''(s)\, u'(c_2)\frac{ds}{d(1-\delta)} + \beta R'(s)\, u''(c_2)\frac{dc_2}{d(1-\delta)}. \tag{E9}$$

Differentiating the period-1 labor FOC at $\delta = 1$ and substituting in the Euler equation gives

$$u''(c_1)w_1(1-\tau_1)\frac{dc_1}{d(1-\delta)} + u'(c_1)w_1\tau_1\Delta_r^i = v''(\ell_1)\frac{d\ell_1}{d(1-\delta)}, \tag{E10}$$

where $\Delta_r^i = 1 - \frac{1+r}{R'(s^i)} \geq 0$, with strict inequality for borrowers ($s^i < 0$). Using (E10) to eliminate $d\ell_1/d(1-\delta)$ from (E8), and analogously eliminating $d\ell_2/d(1-\delta)$ from the period-2 budget constraint, the system reduces to

$$(1-A_1^i)\frac{dc_1^i}{d(1-\delta)} = \tau_1 y_1^i\left(1 + \frac{v'(\ell_1^i)}{\ell_1^i\, v''(\ell_1^i)}\Delta_r^i\right) - \frac{ds^i}{d(1-\delta)}, \tag{E11}$$

$$(1-B_2^i)\frac{dc_2^i}{d(1-\delta)} = -(1+r)\tau_1 y_1^i + R'(s^i)\frac{ds^i}{d(1-\delta)}, \tag{E12}$$

where

$$A_1^i \equiv \frac{u''(c_1^i)\left[w_1^i(1-\tau_1)\right]^2}{v''(\ell_1^i)} < 0, \qquad B_2^i \equiv \frac{u''(c_2^i)\left[w_2^i(1-\tau_2)\right]^2}{v''(\ell_2^i)} < 0. \tag{E13}$$

Solving equations (E9)–(E12) for the bracket yields

$$-(1+r)\tau_1 y_1^i + R'(s^i)\frac{ds^i}{d(1-\delta)} = (1-B_2^i)\,\frac{dc_2^i}{d(1-\delta)} = \tau_1 y_1^i\, R'(s^i)\,\Delta_r^i\,\Theta^i, \tag{E14}$$

where the first equality follows from (E12) and $\Theta^i > 0$ is defined below.

Some intuition is in order. The bracketed error term can be decomposed as

$$-(1+r)\tau_1 y_1^i + R'(s^i)\frac{ds^i}{d(1-\delta)} = \underbrace{\left(R'(s^i) - (1+r)\right)\tau_1 y_1^i}_{\text{rate wedge on crowd-out}} - \underbrace{R'(s^i)\left(\tau_1 y_1^i - \frac{ds^i}{d(1-\delta)}\right)}_{\text{behavioral savings adjustment}}. \tag{E15}$$

The first term is the mechanical effect: keeping the amount of liabilities to be repaid fixed, period-2 consumption increases as liabilities switch to lower-interest delayed taxes. This term is always positive. The second term works to reduce welfare if agents increase period-1 borrowing by less than the increase in period-1 liquidity (i.e., $ds/d(1-\delta) < \tau_1 y_1$). When $R''(s^i) > 0$, the savings adjustment is amplified because the marginal rate also improves as borrowing falls.

Simplified error term. Specifically, Θ^i is given by

$$\Theta^i \equiv \frac{\frac{u''(c_1^i)}{\beta R'\,u''(c_2^i)}\left(1 - \frac{u''(c_2^i)[w_2^i(1-\tau_2)]^2}{v''(\ell_2^i)}\right)\left(1 + \frac{v'(\ell_1^i)}{\ell_1^i\,v''(\ell_1^i)}\right) + \frac{(1+r)\left(\frac{-R''\,u'(c_2^i)}{R'\,u''(c_2^i)}\right)\left(1 - \frac{u''(c_1^i)[w_1^i(1-\tau_1)]^2}{v''(\ell_1^i)}\right)}{R'\,\Delta_r^i}}{\left(1 - \frac{u''(c_1^i)[w_1^i(1-\tau_1)]^2}{v''(\ell_1^i)}\right)\left[R' + \frac{R''\,u'(c_2^i)}{u''(c_2^i)}\left(1 - \frac{u''(c_2^i)[w_2^i(1-\tau_2)]^2}{v''(\ell_2^i)}\right)\right] + \frac{u''(c_1^i)}{\beta R'\,u''(c_2^i)}\left(1 - \frac{u''(c_2^i)[w_2^i(1-\tau_2)]^2}{v''(\ell_2^i)}\right)} > 0, \tag{E16}$$

which is positive because the terms $1 - u''(\cdot)[\cdot]^2/v''(\cdot)$ exceed unity (since $u'' < 0$ and $v'' > 0$), u''/u'' ratios are positive, and $-R''u'/u'' \geq 0$ (since $R'' \geq 0$, $u' > 0$, $u'' < 0$). When $R''(s^i) = 0$, the second numerator term and the R'' term in the denominator vanish.

The error term therefore equals

$$(\beta^* - \beta) \sum_{i:\, \beta^i < \beta^*} \pi^i \, u'(c_2^i) \, \tau_1 y_1^i \, R'(s^i) \, \Delta_r^i \, \Theta^i \geq 0, \tag{E17}$$

which is strictly positive whenever there exists at least one impatient borrower ($\beta^i < \beta^*$ and $\Delta_r^i > 0$).

Marginal welfare effects of delayed taxation. We now rewrite the marginal welfare effects in (E3). We substitute 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 then substitute (E17) in for the error term.

$$\begin{aligned}
\frac{1}{\lambda^*} \frac{dW^*}{d(1-\delta)}\bigg|_{\delta=1} &= \tau_1 \sum_{i:\, \beta^i < \beta^*} \pi^i \, y_1^i \, g_1^{i*} \, \Delta_r^i \left[1 + \left(\frac{\beta^*}{\beta} - 1\right) \Theta^i\right] \\
&+ \tau_1 \sum_{i:\, \beta^i = \beta^*} \pi^i \, y_1^i \, g_1^{i*} \, \Delta_r^i \\
&- \sum_i \pi^i \, y_1^i \left[\tau_1 \Delta_{r,g}^i \left(\frac{\tau_1}{1-\tau_1} \varepsilon_{11}^{i,c} + \frac{1}{1+r} \frac{\tau_2}{1-\tau_1} \frac{y_2^i}{y_1^i} \varepsilon_{21}^{i,c}\right) + \rho_1^i - (1+r)\rho_2^i\right] \\
&- \sum_i \pi^i \, \tau_1 y_1^i \, \Delta_{r,g}^i.
\end{aligned} \tag{E18}$$

This expression shows that paternalism amplifies the intertemporal consumption smoothing gains.

The following proposition summarizes how equation (E18) differs from the baseline case.

Proposition 5 (Marginal Welfare Effects of DT under Paternalism) *Assume: (i) a nonempty subset of agents share a common discount factor $\beta^i = \beta < \beta^*$ (all others have $\beta^i = \beta^*$); (ii) at least one of these agents (with $\beta^i = \beta$) are borrowers ($s^i < 0$); (iii) no subsidized lending ($x = 0$); (iv) the marginal reform is evaluated at the benchmark with no delay ($\delta = 1$). Then the marginal change in social welfare from introducing delayed taxation is given by (E18). These gains strictly exceed those from the non-paternalistic case due to a reduced period-2 repayment channel.*

More specifically, borrowers ($s^i < 0$) enjoy the same experienced *gain proportional to $\Delta_r^i = \left(1 - \frac{1+r}{R'(s^i)}\right)$, but its* social *contribution increases under paternalism:*

$$\tau_1 \sum_{i:\, \beta^i < \beta^*} \pi^i \, y_1^i \, g_1^{i*} \, \Delta_r^i \left(\frac{\beta^*}{\beta} - 1\right) \Theta^i,$$

where $\Theta^i > 0$ is defined by (E16). Net savers ($s^i \geq 0$ and thus $\Delta_r^i = 0$) contribute zero to the smoothing and error terms. All other behavioral-revenue and fiscal terms are identical to the non-paternalistic baseline and are omitted here for brevity; see (E18).

This proposition shows that delayed taxation has additional welfare effects when the government has paternalistic preferences. The correct thought experiment that justifies a statement such as "paternalism increases welfare gains" is the following. Consider a baseline where individ-

uals differ in their β^i as in the proposition but where the government applies a non-paternalistic time discount factor of β (instead of $\beta^* > \beta$) when calculating the welfare contributions of these individuals. Now consider modifying this baseline, keeping allocations and tax policies constant, by increasing the time discount factor for low-β^i individuals from β to β^*. This would increase the marginal welfare gains from delayed taxation: the last three terms in equation (E18) stay fixed, but the first term is larger (because $\beta^*/\beta - 1 > 0$.

F Three-period model with retirement

This section extends our two-period modeling framework to include a retirement period. We set up the model, discuss new aspects related to the microfoundation of the financial frictions, and provide numerical solutions.

F.1 Three-period model

Our main two-period model implicitly assumes that period-2 workers can neither delay taxes nor borrow. Given our microfoundation in Appendix A, allowing period-2 workers to delay taxes or borrow would cause them to delay or borrow the maximal amounts and then default post-mortem. The assumption that individuals cannot accrue debts that would be repaid post-mortem is ubiquitous and dates back to early work dealing with finite life cycles (e.g., Fisher 1930, Yaari 1965).

We now consider how one can implement delayed taxation without making the ability to delay taxes explicitly age dependent. More specifically, we focus on implementing a delayed tax policy where individuals do not default in the second period, which is succeeded by a third "retirement" period. Specifically, we assume that agents work during periods 1 and 2, but may not work in period 3 (i.e., retirement). Imposing binding retirement ages is common when working with life cycle models, and may be justified by low old-age productivity coupled with (potentially increasing) labor market participation costs. That agents cannot work in period 3 mechanically avoids the issue of post-mortem default on period-3 delayed taxes. There are no taxes to default on. Instead, we focus on ensuring that agents do not default on their period-2 delayed taxes that must be repaid in period 3. This is not trivial in our setting because the threat of seizing disposable incomes is less severe when there are no labor earnings.

Private lenders. We assume that private lenders always pay the upfront cost, z, proportional to the loan amount, which allows them to seize disposable income except the basic government transfer, G, in the period of default. Appendix A formalizes how the upfront cost, z, gives determines the interest rate spread, $r_b - r$.

Private lenders may impose age-dependent borrowing limits. They allow period-1 loans up to a limit of $\bar{a}_1 \geq 0$ and period-2 loans up to a limit of $\bar{a}_2 \geq 0$. As we discuss below, these limits must support a no default equilibrium. We further make the standard assumption that private lenders bar individuals from borrowing in period 3.

Government lending costs. The government similarly pays the upfront cost z to be able to seize any disposable income in excess of G. The government, however, may not explicitly discriminate based on age. Hence, the when implementing delayed taxation, the government is restricted to choosing $\delta_1 = \delta_2$.

Workers' problem if no default. Workers solve the following optimization program

$$V(\tau_1, \tau_2, G_1, G_2, G_3, \delta_1, \delta_2) = \max_{\ell_1, \ell_2, s_1, s_2} u(c_1) - v(\ell_1) + \beta\big[u(c_2) - v(\ell_2)\big] + \beta^2 u(c_3), \tag{F1}$$

subject to the budget constraints,

$$\begin{aligned} c_1 &= y_1[1 - \delta_1\tau_1] + G_1 - s_1, && \text{(F2)} \\ c_2 &= y_2[1 - \delta_2\tau_2] + G_2 + R(s_1) - (1+r)[1 - \delta_1]\tau_1 y_1 - s_2, && \text{(F3)} \\ c_3 &= G_3 - (1+r)[1 - \delta_2]\tau_2 y_2 + R(s_2), && \text{(F4)} \end{aligned}$$

where $y_t = w_t \ell_t$. We further require that

$$\begin{aligned} s_1 &\geq -\bar{a}_1, && \text{(F5)} \\ s_2 &\geq -\bar{a}_2. && \text{(F6)} \end{aligned}$$

Note that we omit the superscript i for notational convenience: c_t for $t = 1, 2, 3$, ℓ_t for $t = 1, 2$, and s_t for $t = 1, 2$ should all be indexed by i.

Transfers during retirement. Before setting up the workers optimization programs in the event of default it is useful to outline the structure of the transfers, G_t. We will further assume that $G_1 = G_2 = G$, but we will set $G_3 > G$ to reflect pensions and other transfers to retirees. We assume that lenders may seize the excess transfer in period 3 ($G_3 - G > 0$) in the event of period-3 default. This ensures age-independent consumption levels for those who default.

Workers' problem if defaulting in period 2. Someone who defaults in period 2 will not supply any labor in period 2 since any labor earnings would be seized. They further borrow the maximal in period-1 since the debt will not be repaid in period 2. Similarly, they delay the maximal amount of period-1 taxes.

$$V_{D2}(\tau_1, \tau_2, G_1, G_2, G_3, \delta_1, \delta_2) = \max_{\ell_1, s_2} u(c_1) - v(\ell_1) + \beta u(c_2) + \beta^2 u(c_3), \tag{F7}$$

subject to

$$
\begin{aligned}
c_1 &= y_1[1-\delta_1\tau_1] + G_1 + \bar{a}_1, && \text{(F8)}\\
c_2 &= G_2 - s_2, && \text{(F9)}\\
c_3 &= G_3 + R(s_2), && \text{(F10)}\\
s_2 &\geq 0. && \text{(F11)}
\end{aligned}
$$

Note we have imposed $s_2 \geq 0$. Individuals do not borrow due the assumption that lenders can seize all period-2 disposable income, including cash flows from borrowing. We conservatively do not let lenders seize period-3 disposable incomes. Hence, agents may choose to save in period 2. While allowing lender to seize parts of the period-3 transfer, G_3, would help enforce a no-default equilibrium, this additional default penalty is not needed in our calibration.

Workers' problem if defaulting in period 3. If agents default in period 3, they can only consume $c_3 = G < G_3$. They will not wish to save in period 2 because $R(s_2)$ will be seized in period 3. However, they will wish to borrow the maximal amount, setting $s_2 = -\bar{a}_2$.

$$
V_{D3}(\tau_1,\tau_2,G_1,G_2,G_3,\delta_1,\delta_2) = \max_{\ell_1,\ell_2,s_1} u(c_1) - v(\ell_1) + \beta\big[u(c_2) - v(\ell_2)\big] + \beta^2 u(c_3), \qquad \text{(F12)}
$$

subject to the budget constraints,

$$
\begin{aligned}
c_1 &= y_1[1-\delta_1\tau_1] + G_1 - s_1, && \text{(F13)}\\
c_2 &= y_2[1-\delta_2\tau_2] + G_2 + R(s_1) - (1+r)[1-\delta_1]\tau_1 y_1 + \bar{a}_2, && \text{(F14)}\\
c_3 &= G, && \text{(F15)}
\end{aligned}
$$

where $c_3 = G$ follows from the assumption that lenders (incl. the government) can seize excess transfers ($G_3 - G > 0$) in the event of period-3 default.

Government's three-period problem. The government chooses time-invariant parameters (τ, G, δ) to maximize utility in the no-default equilibrium,

$$
\sum_i \pi^i V^i, \qquad \text{(F16)}
$$

where $\sum_i \pi^i = 1$, subject to the present-value budget constraint,

$$
\begin{aligned}
&\sum_i \pi^i \left(\tau_1 w_1^i \ell_1^i + \frac{\tau_2 w_2^i \ell_2^i}{1+r} - (1-\delta)\Delta_{r,g}^i \tau_1 w_1^i \ell_1^i - \frac{(1-\delta)\Delta_{r,g}^i \tau y_2^i}{1+r} \right) && \text{(F17)}\\
\geq\ & G + \frac{G}{1+r} + \frac{G_3}{(1+r)^2} + K_1 + \frac{K_2}{1+r} + \frac{K_3}{(1+r)^2}, && \text{(F18)}
\end{aligned}
$$

and the no-default conditions,

$$V^i \geq V^i_{D3} \text{ for all } i, \tag{F19}$$

$$V^i \geq V^i_{D2} \text{ for all } i, \tag{F20}$$

and the restriction that $G_3 - G = \gamma G$, where $\gamma > 0$ is an exogenous parameter determining effective pension generosity.

F.2 Three-period model calibration

All assumptions closely follow those in the main two-period model calibration. We assume that $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$ and $v(\ell) = \frac{\ell^{1+\frac{1}{k}}}{1+\frac{1}{k}}$, where $\sigma = 5$ and $k = 0.5$. We set $1 + r = 1.03^{21}$, $1 + r_b = 1.20^{21}$, and $\beta = 0.97^{21}$. We further impose $\gamma = 2$ which ensures that the extra transfer during retirmement, $G_3 - G$ generates a realistic pension replacement rate of 57% in our calibration.

We further impose that government lending costs, $\sum_i \pi^i (1-\delta)\Delta^i_{r,g}\tau_1 w^i_1 \ell^i_1$, equal

$$\sum_i \pi^i LC(s^i, (1-\delta)\tau_1 w^i_1 \ell^i_1)$$

for the purpose of the numerical implementation. $LC(\cdot,\cdot)$ is defined as in equation (H5). $R(s)$ and $R'(s)$ take the functional forms given in equations (H1) and (H2).

To parameterize the per-period revenue requirements, K_t, we assume $K_t = 0$ and solve the government's problem with $1 + r_b = 1.03^{21}$ (i.e., zero interest spread, $r_b - r$). K_t is then set to 15% of aggregate labor earnings in period t in this frictionless baseline. $K_3 = 0$ since there are no period-3 labor earnings.

F.3 Three-period model numerical approach

Ensuring no default. We do not directly impose the borrowing limits $(\bar{a}_1, \bar{a}_2)$ on the no-default individual optimization problems. Once we find an optimal no-default allocation, we set $\bar{a}_1$ equal to the largest borrowing amount in period 1, and $\bar{a}_2$ equal to the largest borrowing amount in period 2. We then verify that it is not optimal for agents to default given these borrowing limits. A more general (but computationally more involved) approach is to impose the no-default conditions as constraints in the numerical optimization problem. This would address the potential a trade-off between higher transfers and more delayed taxation. Increasing either $1 - \delta$ and G may make more individuals strategically default.

General approach. We then solve for the benchmark allocations with $1 + r_b = 1.20^{21}$ and the parameterized values for K_1 and K_2. This involves finding the optimal tax policy (τ, G) with $1 - \delta = 0$. We then solve for the delayed tax allocations, which allow $1 - \delta \geq 0$. To obtain the money metric welfare gain from delayed taxation, we numerically find the period-1 amount, M, that must be injected into the government's budget constraint in the benchmark economy (no delayed tax) to obtain the same welfare ($\sum_i \pi^i V^i$) as with delayed taxation.

F.4 Three-period model quantitative results

The key result is that delayed taxation delivers material welfare gains and is compatible with a no-default equilibrium. We find that zero agents prefer to default in either period 2 or period 3 under our calibration. This is no-default equilibrium can be sustained by imposing non-binding borrowing limits. That is, setting $\bar{a}_t = \max(-\min_i(s_i), 0)$ for $t = 1, 2$. This causes the life-time utility of default, $V^i_{D,t}$, to be lower than life-time utility when not defaulting for all agents.

Table A.1 proves the optimal tax policies and summary statistics for the allocations. The second column shows that the government optimally sets $1 - \delta = 0.1836$, which is half what we found for the two-period model (0.39). The average effective delayed tax fraction, $1 - \delta^i_1$, equals 8.18%, which is about three-quarters the fraction in the two-period model. Virtually no one delays taxes in period 2, with the average $1 - \delta^i_2 = 0.01\%$.

We see that the three-period model features lower taxes rates ($\tau = 0.43$ to 0.50) than the two-period model. At the same time, benchmark transfers are also smaller, partly due to the requirement of additional pension transfers of $2 \times G$. We see that this additional transfer amounts to a 57% replacement rate on average period-2 after-tax labor income.[38] This replacement rate is quite close to the average pension replacement rate in Norway of 54.8% (OECD, 2025).

In this three-period calibration, welfare gains are still material. In terms of period-1 consumption units, delayed taxation increases welfare by around 1%. This is slightly less than half of what we found in the two-period model. The reason welfare gains are lower is the presence of a third period in which individuals cannot work (and thus effectively face wages of zero). This reduces the demand to borrow from period-2 earnings, since period-2 earnings are now also used to increase consumption in period 3.

Table A.1 further shows that the money-metric welfare gains per \$1 of delayed taxes equal \$0.2414. To find the implied risk-free return on delayed taxes,[39] we rearrange

$$\left(\frac{1+\tilde{r}}{1.03}\right)^{21} = 1.2414 \tag{F21}$$

and obtain a return $\tilde{r} = 4.07\%$, which is substantially higher than the risk free rate, $r = 3\%$.

[38] Dividing $2G$ under delayed taxation by the average $(1-\tau) * \ell^i_2 w^i_2$ yields (2*0.2928)/(2.0434*(1-0.5028))=57.64%.

[39] The money-metric welfare gain is risk-free in the sense that the underlying welfare calculations account for the cost of of capital associated with lending (i.e., the upfront cost of avoiding default). The money-metric gain of 0.2414 is a present-value figure. Hence, we to contrast it to a hypothetical investment opportunity with return $\tilde{r}$, we cumulate the return into the next period ($T = 21$ years) and then discount it back at the interest rate on savings ($r = 3\%$).

TABLE A.1: OPTIMAL TAX SCHEDULE AND ALLOCATIONS IN THREE-PERIOD MODEL

Notes: $1-\delta_t^i$ is the average fraction of delayed taxes in period t (defined analogously to the two-period-model case in Appendix H.1). $TaxesPaid_t^i = \delta_t^i \tau_t y_t$. Welfare gain (% C_1) is the money metric welfare gain (equivalent injection into government's benchmark budget constraint) divided by aggregate benchmark period-1 consumption. Welfare gain (% D) instead divides the money-metric gains by the present-value amount of delayed tax, $D = \sum_{t=1}^{2} \left(\frac{1}{1+r}\right)^{t-1} \sum_i \pi^i (1-\delta_t^i)\tau_t \ell_t^i w_t^i$.

	Tax schedule and allocations with $r_b = 20\%$	
	Benchmark	+ Delayed Taxation
τ	0.4258	0.5028
G	0.2490	0.2928
$1-\delta$	.	0.1836
Welfare gain, % C_1 (rel. to Benchmark)	.	1.0461
Welfare gain, % D (rel. to Benchmark)	.	24.14
$\min(s_1^i)$	-0.1169	-0.1070
$\min(s_2^i)$	-0.0038	0.0102
π^i-weighted means		
$1-\delta_1^i$	0.0000	0.0818
$1-\delta_2^i$	0.0000	0.0001
$1[V_{D2}^i > V^i]$	0.0000	0.0000
$1[V_{D3}^i > V^i]$	0.0000	0.0000
c_1^i	0.8893	0.8945
$\ell_1^i w_1^i$	1.0203	0.9657
$\ell_2^i w_2^i$	1.9249	2.0434
$(TaxesPaid_1^i - G_1)/y_1^i$	0.1716	0.1438
$(TaxesPaid_2^i - G_2)/y_2^i$	0.2782	0.3361
$IL(s_1^i, L_1^i)$	0.0000	0.0385
$IL(s_2^i, L_2^i)$	0.0000	0.0001

G 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 (37.5 hours). We assign numeric values of 5, 15, 25, 32.5, and 37.5 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 hours worked is then max(*typical hours worked* $\times$ 47 weeks $\times$ *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 $\times$ 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 $\times$ 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 they are old enough to have finished the length of education that are reported to have in the 2011 census. That is, their age must exceed 7 (first year of schooling) plus their reported years of schooling by at least one year. We also require that we can compute 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.2: 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	0.86	1.45	1.87	2.69	4.28	2.17

TABLE A.3: 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. w_1s come from the 1990 wave (reflecting 1989 data), and w_2s come from the 2011 wave (reflecting 2010 data). Wages (USD) are in 2010 prices.

	Panel A: Exogenous wage heterogeneity				
	p5	p25	p50	p75	p95
w_1	0.69	1.10	1.69	2.23	3.70
w_2	0.86	1.66	2.30	4.17	9.13
w_2/w_1	0.58	1.03	1.51	2.37	4.15

H 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

$$R(s) = \int_0^s R'(\mathtt{s})d\mathtt{s} = (1+r)s - \frac{r_b - r}{m_1}\ln\left(\frac{1+e^{-m_1(s+m_2)}}{1+e^{-m_1 m_2}}\right), \tag{H1}$$

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

$$R'(s) = 1 + \frac{r + \exp(-m_1\,(s+m_2))r_b}{1+\exp(-m_1(s+m_2))}. \tag{H2}$$

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.1 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 $[-0.15, -0.05]$.

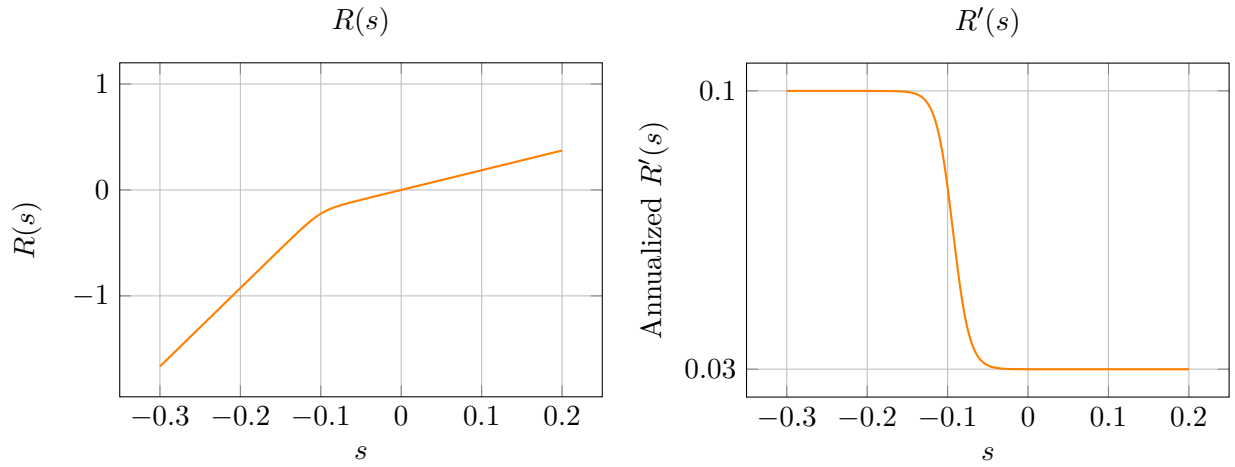


FIGURE A.1

Government's cost of capital. Appendix A provides the theoretical foundation for the interest rate on private loans (r_b) and the government's cost-of-capital for providing loans or delayed taxation (r_b^g). For the government, we use eqs. (H1)–(H2) using r_b^g instead of r_b and letting $\bar{x} = 0$:

$$R(s, r_b^g) \equiv R(s \big| r_b = r_b^g) \quad \text{and} \quad R'(s, r_b^g) \equiv R'(s \big| r_b = r_b^g). \tag{H3}$$

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

$$LC(s, L) = \int_{s-L}^{s} \left(1 - \frac{1+r}{R'(\mathbf{s}, r_b^g)}\right) d\mathbf{s}. \tag{H4}$$

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.

$$LC(s,L) = \int_{s-L}^{s} \left(1 - \frac{1+r}{R'(\mathtt{s}, r_b^g)}\right) d\mathtt{s} \tag{H5}$$

$$= \int_{s-L}^{s} \left(1 - \frac{1+r}{1 + \frac{r+\exp(-m_1(\mathtt{s}+m_2))r_b^g}{1+\exp(-m_1(\mathtt{s}+m_2))}}\right) d\mathtt{s} \tag{H6}$$

$$= \frac{r_b^g - r}{m_1(1+r_b^g)} \ln\left(\frac{1+r+(1+r_b^g)\exp(-m_1(s-L+m_2))}{1+r+(1+r_b^g)\exp(-m_1(s+m_2))}\right). \tag{H7}$$

Note that $LC(s,0) = 0$. We plot $LC(s,L)$ 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 $R'(-0.15, r_b^g) > 1+r$. 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.

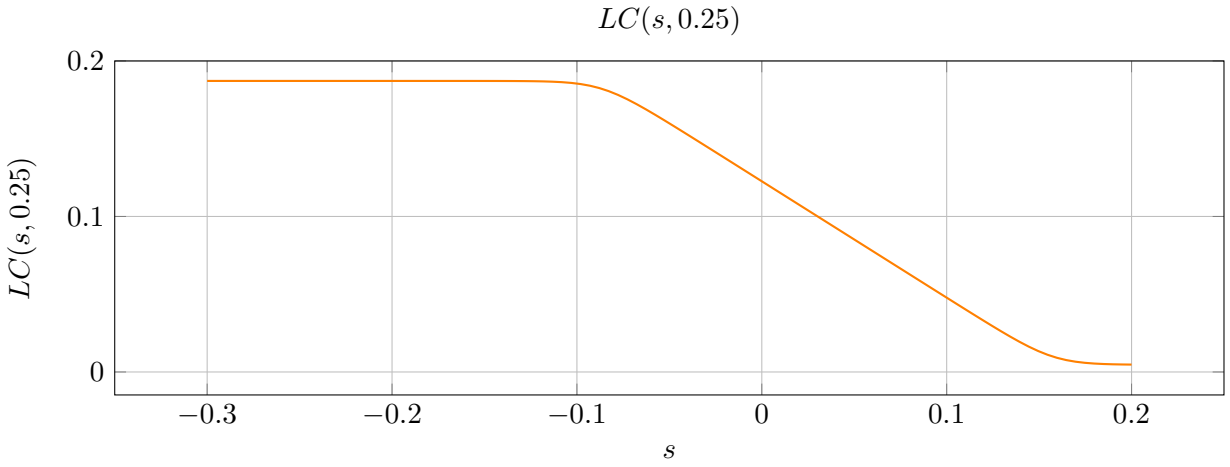


FIGURE A.2: Present-value cost of lending L=0.25 via either delayed taxation or subsidized lending

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 $R'(s-L, r_b^g)$. Simplifying the marginal costs, and taking the integral from $s-L$ to s produces the same formula, $LC(s,L)$.

Implied loan amounts. When we implement restrictions on government borrowing, we need to quantify how delayed taxes and subsidized loans affect the government's borrowing needs.

Suppose someone is allowed to delay taxes up to an amount of $d \geq 0$ and can obtain a subsidized loan upto a cap of $x \geq 0$. The maximal liability is $L = x + d$. However, many individuals (i.e., net savers) are indifferent and would opt out. Hence the implied loan amount, IL, is weakly smaller than L. We define the implied loan amount IL by equating period-2 values of the interest subsidy:

$$IL(s, L)(r_b - r) = \int_{s-L}^{s} R'(z)dz - \int_{s-L}^{s} R'(z|r_b = r)dz, \tag{H8}$$

where the LHS is the value of the interest subsidy of an loan of IL. The first term on the RHS is the interest payment for someone who borrows L and faces (unsubsidized) private credit markets. The second term on the RHS is the hypothetical interest payment for someone who faces a subsidized interest payment (where the nominal borrowing rate r_b is reduced to r). This second term can be rewritten $(1 + r)L$. Hence we can rearrange to obtain

$$\begin{aligned} IL(s, L) &= \frac{\int_{s-L}^{s} R'(z, r_b)dz - (1 + r)L}{r_b - r}, & \text{(H9)} \\ &= -\frac{1}{m_1} \ln \left(\frac{1 + e^{-m_1(s+m_2)}}{1 + e^{-m_1(s-L+m_2)}} \right), & \text{(H10)} \end{aligned}$$

where the integral is determined by imposing $\lim_{s\to\infty} IL(s, L) = 0$. This expression simplifies by letting the $(1 + r)L$ and then the $r_b - r$ terms cancel out. The definition of $IL(s, L)$ 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.

Government borrowing. With the function $IL(s, L)$ defined, we can calculate government borrowing as

$$B = -\sum_i \pi_i \left[\tau_1 y_1^i - G_1 - IL\Big(s^i, (1 - \delta)\tau_1 y_1^i + x\Big) - K_1 \right], \tag{H11}$$

where $x \geq 0$ is the uniform subsidized loan amount and K_1 is the exogenous period-1 revenue requirement.

H.1 Effective delayed tax fraction

$IL(s, L)$ is also well suited to calculate effective delayed tax amounts. Figure A.1 shows that when $r_b = 20\%$, marginal interest rates, $R'(s^i)$, are essentially equal to r for $s^i \geq 0$. These agents benefit little from delayed taxation (and introduce no material costs per the definition of $LC(\cdot, \cdot)$)) and would opt out if delayed taxation became slightly less attractive (e.g., by introducing vanishingly small participation costs or an interest rate slightly above r). To avoid that

statistics related to the average fraction of delayed taxes are potentially highly sensitive to small perturbations in participation costs, we define the effective delayed tax fraction of $1-\delta^i$, in a way that is conceptually aligned with our definition of lending costs. More specifically, we define

$$1-\delta^i \equiv \frac{d^i}{\tau_1 w_1^i \ell_1^i}, \tag{H12}$$

where

$$d^i \equiv IL(s^i, (1-\delta)\tau_1 w_1^i \ell_1^i). \tag{H13}$$

I Additional quantitative results

Table A.4: 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 unrestricted delayed taxation (i.e., does not require $1-\delta \leq 1$). The fourth column allows for subsidized lending. The fifth column allows for subsidized lending *and* delayed taxation. The sixth and seventh columns consider an age-dependent baseline policy; and the eight and ninth consider an nonlinear-HSV policy. 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.

	Tax schedule and allocations with $r_b = 20\%$								
	Benchmark	+ DT	+ DT UNR	Lending	+ DT	AD	+ DT	HSV	+ DT
τ_1	0.705	0.735	0.73	0.71	0.735	0.59	0.61	0.70	0.70
τ_2	0.705	0.735	0.73	0.71	0.735	0.77	0.77	0.70	0.70
G_1	0.727	0.749	0.75	0.73	0.749	0.70	0.71	0.73	0.63
G_2	0.727	0.749	0.75	0.73	0.749	0.70	0.71	0.73	0.63
a_t (HSV)	.	.	.	.	.	.	.	0.29	0.40
p_t (HSV)	.	.	.	.	.	.	.	-0.01	0.25
$1-\delta$	0.000	1.000	1.07	0.00	1.000	0.00	1.00	0.00	1.00
r_{dtax}	0.030	0.030	0.03	0.03	0.030	0.03	0.03	0.03	0.03
Govt loan limit, x	0.000	0.000	0.00	0.01	0.000	0.00	0.00	0.00	0.00
B	0.123	0.929	0.98	0.13	0.929	0.18	0.86	0.12	0.80
Welfare Gain, % C_1 (rel. to Benchmark)	0.000	1.794	1.80	0.24	1.794	3.32	5.59	0.00	2.70
Δ Welfare	.	.	.	.	1.556	.	2.27	.	2.70
Welfare Gain, % D	.	37.482	30.66	.	37.482	.	194.89	.	26.46
Δ Welfare, % D	.	.	.	.	31.526	.	79.06	.	26.46
means									
$1-\delta^i$	0.000	0.105	0.15	0.00	0.105	0.00	0.08	0.00	0.23
PV(l_1w_1,l_2w_2)	1.983	2.033	2.03	1.97	2.033	2.00	2.04	1.96	2.03
PV($l_1w_1\tau_1$,$l_2w_2\tau_1$)	1.398	1.494	1.49	1.39	1.494	1.35	1.41	1.38	1.43
$(TaxPaid_1 - G_1)/y_1$	-0.155	-0.352	-0.46	-0.20	-0.352	-0.20	-0.32	-0.19	-0.37
$(TaxPaid_2 - G_2)/y_2$	0.119	0.163	0.17	0.13	0.163	0.12	0.13	0.12	0.21

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.

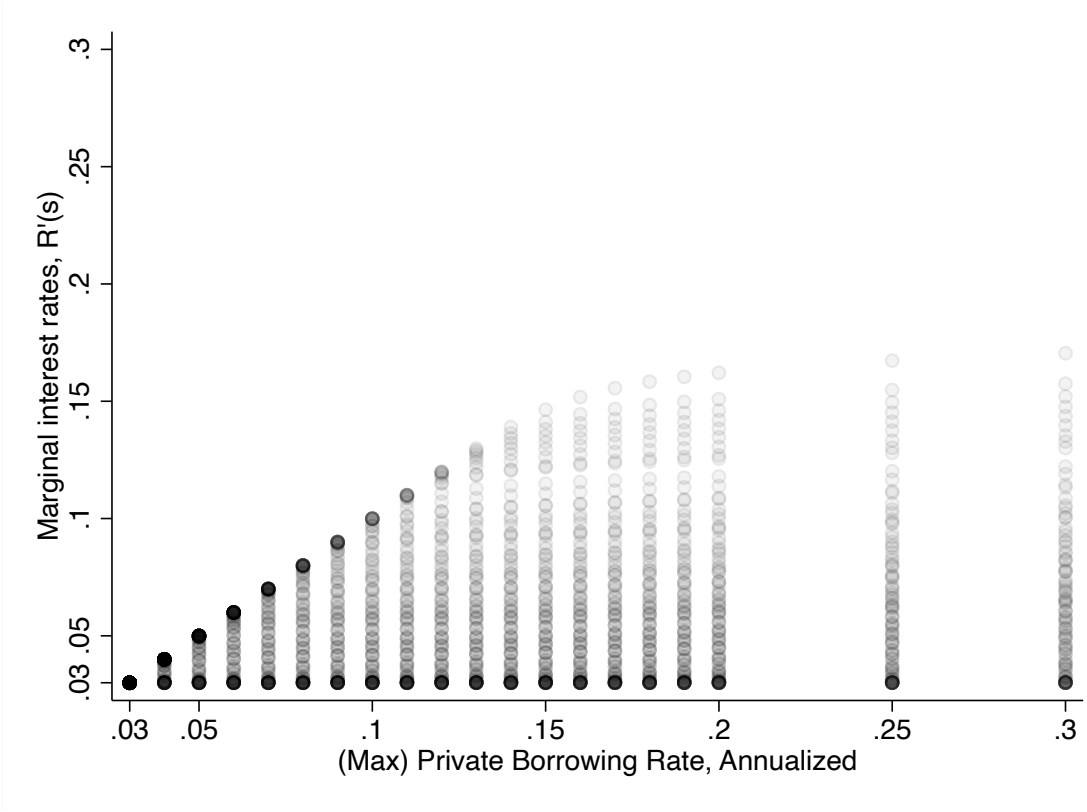


Table A.5: Extended Table: Optimal Tax Schedules and Allocations

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. Welfare Gain in % of D is the money metric welfare gain divided by the amount of delayed taxes ($\sum_i \pi^i(1-\delta^i)\tau_1 \ell_1^i w_1^i$). 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 H).

	Tax schedule and allocations with $r_b = 20\%$							
	Benchmark	+ DT	Lending	+ DT	AD	+ DT	HSV	+ DT
τ_1	0.64	0.68	0.65	0.68	0.36	0.41	0.64	0.64
τ_2	0.64	0.68	0.65	0.68	0.69	0.70	0.64	0.64
G_1	0.55	0.57	0.55	0.57	0.44	0.47	0.49	0.41
G_2	0.55	0.57	0.55	0.57	0.44	0.47	0.49	0.41
a_t (HSV)	.	.	.	.	.	.	0.42	0.50
p_t (HSV)	.	.	.	.	.	.	0.15	0.31
$1-\delta$	0.00	0.39	0.00	0.39	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	0.13	0.42	0.14	0.42	0.23	0.59	0.13	0.55
Welfare Gain, % C_1 (rel. to benchmark)	0.00	2.08	0.36	2.08	7.35	9.42	0.10	3.83
Δ Welfare, % C_1	.	2.08	.	1.72	.	2.07	.	3.73
Welfare Gain, % D	.	38.17	.	38.17	.	416.16	.	21.70
Δ Welfare, % D	.	38.17	.	31.53	.	91.28	.	21.11
means								
$1-\delta^i$	0.00	0.11	0.00	0.11	0.00	0.08	0.00	0.37
$l_1 w_1$	0.80	0.72	0.79	0.72	0.87	0.83	0.81	0.75
$l_2 w_2$	1.66	1.91	1.69	1.91	1.65	1.76	1.64	1.86
PV($l_1 w_1$,$l_2 w_2$)	1.70	1.75	1.70	1.75	1.76	1.78	1.69	1.75
PV($l_1 w_1 \tau_1$,$l_2 w_2 \tau_1$)	1.09	1.19	1.10	1.19	0.93	1.00	1.09	1.12
$(TaxPaid_1 - G_1)/y_1$	-0.10	-0.26	-0.11	-0.26	-0.19	-0.24	-0.09	-0.30
$(TaxPaid_2 - G_2)/y_2$	0.25	0.32	0.26	0.32	0.34	0.34	0.25	0.35
l_1	0.79	0.71	0.78	0.71	0.87	0.83	0.80	0.74
l_2	0.82	0.91	0.84	0.91	0.79	0.82	0.81	0.89
c_1	0.89	0.93	0.90	0.93	0.97	0.98	0.90	0.94
c_2	1.02	0.95	1.01	0.95	1.00	0.98	1.01	0.94
s	-0.06	-0.01	-0.07	-0.01	0.03	0.04	-0.05	-0.01
Impl. govt loan amount, IL	0.00	0.05	0.00	0.05	0.00	0.02	0.00	0.16

J Bunching Analysis Appendix

J.1 Identification of Labor Supply Elasticities via Bunching

Overview

The identification of elasticities via bunching at tax kinks was established by Saez (2010) for static models. In these static models, bunching identifies the compensated (Hicksian) elasticity because income effects are approximately the right size to move the estimand from the Frisch elasticity to the Hicksian Elasticity. The residual (i.e., income effects in excess of what the Hicksian should contain) is assumed to be small (i.e., "second order"). The typical assumption needed for the income effect residual to be small is that the change in the marginal tax rate at the kink is small. However, Bastani and Selin (2014) demonstrated subsequently via simulations that the zero residual income effect approximation holds well even for large kinks. Chetty (2012), section 4.4, conjectured that in life-cycle models, bunching instead identifies the Frisch elasticity because "the income effects of locating at the kink versus slightly above it are spread across the lifecycle." Kleven (2016) reiterates this point. The idea is that when income effects are spread over the life cycle, they become small enough for the bunching estimator to identify the Frisch elasticity instead of the Hicksian.

This appendix formalizes Chetty's conjecture and extends it to settings with financial frictions. The key insight is that income effects determine whether bunching identifies the Frisch or Hicksian elasticity. In standard life-cycle models, income effects are spread over the planning horizon; as the horizon grows, they vanish and bunching identifies the Frisch elasticity. With financial frictions, the present value of lifetime wealth is reduced (due to discounting at higher borrowing rates), which amplifies income effects and widens the gap between the identified elasticity and the Frisch. Financial frictions thus effectively shorten the planning horizon, pushing the identified elasticity toward the Hicksian.

J.1.1 Model Environment

Consider a T-period life-cycle model. The agent chooses consumption c_t and labor supply ℓ_t to maximize lifetime utility:

$$U = \sum_{t=1}^{T} \beta^{t-1} \left[\frac{c_t^{1-\sigma}}{1-\sigma} - \xi \frac{\ell_t^{1+1/k}}{1+1/k} \right] \tag{J1}$$

where $\sigma > 0$ is the coefficient of relative risk aversion, $k > 0$ governs the curvature of labor disutility (and equals the Frisch elasticity), $\beta \in (0, 1]$ is the discount factor, and $\xi > 0$ is a scaling parameter.

Delayed Taxation. We now define how delayed taxation enters our model environment. The general identification results in Section J.1.2 require only that Q_D be interpreted as a discount factor scaling the effective tax rate; the specific application to delayed taxation developed here is included to match our empirical setting. The agent faces a dynamic budget constraint where

earnings are received in period t, but the associated tax liability may be paid with a delay $D \geq 0$ periods ($D = 0$ for contemporaneous taxes, $D > 0$ for delayed taxation). We define the *discounting factor* Q_D as the present value of one unit of currency paid in period $t + D$ relative to period t. If the tax liability accrues interest at the policy rate r but the agent discounts at the marginal return on savings $R'(s)$ (which may exceed the market rate when borrowing):

$$Q_D \equiv \frac{(1+r)^D}{\prod_{u=1}^{D} R'(s_{t+u})} \tag{J2}$$

For contemporaneous taxes, $Q_0 = 1$. For delayed taxes with constant borrowing rate $r_b > r$, we have $Q_D = [(1+r)/(1+r_b)]^D$. The *effective* marginal tax rate faced by the agent is the present value of the statutory rate:

$$\tilde{\tau}_t \equiv \tau_t \cdot Q_D \tag{J3}$$

We allow the gross interest rate R_t to vary over time, capturing, e.g., the life-cycle transition from high-interest student loans to lower market rates.

Lifetime Wealth. Iterating the budget constraint forward, we define lifetime wealth in present value at $t = 1$, evaluated at the equilibrium allocation, as:

$$W_1 \equiv \sum_{t=1}^{T} \left(\prod_{u=1}^{t-1} R_u^{-1} \right) [w_t \ell_t (1 - \tilde{\tau}_t) + G_t] \tag{J4}$$

where $1-\tilde{\tau}_t$ is the effective retention rate on labor earnings and G_t denotes transfers. For borrowers with $r_b > r$, future income is discounted more heavily, so W_1 is lower than if discounted at the market rate.

Elasticities. We define elasticities in terms of log-changes. Let $\lambda \equiv u'(c_1)$ denote the marginal utility of consumption in period 1; when the Euler equation holds, this pins down the marginal utility in all periods via $\lambda_t = \lambda \cdot \prod_{j=1}^{t-1} (\beta R_j)^{-1}$. The Frisch elasticity is the response of labor supply to the effective net wage $w_t(1 - \tilde{\tau}_t)$, holding λ constant:

$$\varepsilon^F \equiv \left. \frac{\partial \ln \ell_t}{\partial \ln[w_t(1 - \tilde{\tau}_t)]} \right|_{\lambda}$$

Given the isoelastic disutility function $v(\ell) = \xi \ell^{1+1/k}/(1+1/k)$, the first-order condition is $\xi \ell^{1/k} = \lambda w(1 - \tilde{\tau})$. Taking logs:

$$\frac{1}{k} \ln \ell = \ln \lambda + \ln[w(1 - \tilde{\tau})] + \text{const}$$

Differentiating while holding λ constant yields $(1/k)(d\ell/\ell) = d[w(1 - \tilde{\tau})]/[w(1 - \tilde{\tau})]$, so $\varepsilon^F = k$. The Hicksian (compensated) elasticity is the response holding lifetime utility U constant:

$$\varepsilon^H \equiv \left. \frac{\partial \ln \ell_t}{\partial \ln[w_t(1 - \tilde{\tau}_t)]} \right|_{U}$$

To derive ε^H for the static case, note that $\lambda = u'(c) = c^{-\sigma}$, so the FOC becomes $\xi\ell^{1/k} = c^{-\sigma}w(1-\tau)$. Let $s \equiv w\ell(1-\tau)/c$ denote the share of after-tax labor income in consumption; when non-labor income is present, $s < 1$. Taking logs and differentiating the FOC:

$$\frac{1}{k}\frac{d\ell}{\ell} = -\sigma\frac{dc}{c} + \frac{d[w(1-\tau)]}{w(1-\tau)}$$

Holding utility constant requires $u'(c)dc = v'(\ell)d\ell$, i.e., $c^{-\sigma}dc = \xi\ell^{1/k}d\ell$. Using the FOC, this gives $dc = w(1-\tau)d\ell$, so $dc/c = w(1-\tau)d\ell/c = s \cdot d\ell/\ell$. Substituting:

$$\left(\frac{1}{k} + \sigma s\right)\frac{d\ell}{\ell} = \frac{d[w(1-\tau)]}{w(1-\tau)} \quad \Rightarrow \quad \varepsilon^H = \frac{k}{1+k\sigma s} \tag{J5}$$

Since $\sigma, s > 0$, we have $\varepsilon^F = k > k/(1+k\sigma s) = \varepsilon^H$.[40]

J.1.2 General Identification Theorem

Consider a convex kink at earnings $y^* = w_1\ell^*$ in period 1, where the statutory marginal tax rate rises from τ^- to $\tau^+ = \tau^- + \Delta\tau$. Below the kink, taxes are paid contemporaneously, so $\tilde{\tau}^- = \tau^-$. Above the kink, the marginal increment $\Delta\tau$ is subject to delayed payment, giving $\tilde{\tau}^+ = \tau^- + Q_D\Delta\tau$ and hence $\Delta\tilde{\tau} = Q_D\Delta\tau$ (the standard case with no delay on the marginal increment corresponds to $Q_D = 1$). We assume the standard small-kink approximation conditions hold.[41]

To remain close to the assumptions in the existing conjectures that the bunching estimator identifies the Frisch elasticity in dynamic models, we assume that the agent only faces the kinked tax incentive once in their lifetime. In our empirical setting, the debt-conversion threshold only applies while in school, hence the kink is transitory by construction. More generally, this assumption is valid to the extent that deterministic or uncertain wage growth renders it unlikely that an individual finds itself in the bunching region of the same kink on multiple occasions.[42]

Theorem 1 (General Identification) *Assume the agent is at an interior solution for saving. Define the income effect parameter:*

$$\phi \equiv k\sigma \cdot \frac{(1-\tau^-)y^*}{W_1}$$

The bunching estimator identifies:

$$\varepsilon^{\text{bunch}} = Q_D \cdot \frac{k}{1+\phi} \tag{$\star$}$$

[40] This formula mirrors Bastani and Selin (2014), Section 6.

[41] Specifically, $\Delta\tau$ is small enough that log-linearizations are valid. The debt-conversion threshold involves a large statutory change ($\Delta\tau \approx 0.50$), which pushes the limits of this approximation. However, the effective *present-value* tax change faced by students is much smaller, $\Delta\tilde{\tau} = Q_D\Delta\tau \approx 0.1969 \times 0.50 \approx 0.10$. Moreover, bunching methods are primarily biased when the behavioral response Δy^* is large; given the small observed elasticities, this concern is mitigated.

[42] At the other extreme, one could assume that the bunching decision is not dynamic, as absent the kink, the agent would locate itself at the same exact earnings level each period. In this case, the decision reduces to a static one, in which the bunching estimator identifies the Hicksian.

Proof. Let state A be the counterfactual (linear tax at τ^-, agent earns $y^* + \Delta y^*$) and state B be the actual choice at the kink (agent earns y^*, facing effective marginal rate $\tilde{\tau}^+$ for any additional earnings). The labor FOC $\xi \ell^{1/k} = \lambda w(1-\tilde{\tau})$ applies with the relevant marginal rate in each state. Taking the ratio:

$$\left(\frac{\ell^A}{\ell^B}\right)^{1/k} = \frac{\lambda^A}{\lambda^B} \cdot \frac{1-\tau^-}{1-\tilde{\tau}^+}$$

Taking logs and using $\ln(\ell^A/\ell^B) \approx \Delta y^*/y^*$ (since $y = w\ell$) and $\ln(1-\tau^-) - \ln(1-\tilde{\tau}^+) \approx \Delta\tilde{\tau}/(1-\tau^-)$:

$$\frac{1}{k}\frac{\Delta y^*}{y^*} \approx \ln\left(\frac{\lambda^A}{\lambda^B}\right) + \frac{\Delta\tilde{\tau}}{1-\tau^-} \tag{J6}$$

We now derive the change in marginal utility $\ln(\lambda^A/\lambda^B)$. The Euler equation $c_t^{-\sigma} = \beta R_t c_{t+1}^{-\sigma}$ implies $c_t = c_1\Gamma_t$ where $\Gamma_t \equiv \prod_{j=1}^{t-1}(\beta R_j)^{1/\sigma}$. Substituting into (J4):

$$c_1 \underbrace{\sum_{t=1}^{T}\left(\prod_{j=1}^{t-1}\frac{(\beta R_j)^{1/\sigma}}{R_j}\right)}_{\equiv \mathcal{A}} = W_1 \tag{J7}$$

where $\mathcal{A}$ is the annuity factor. Reducing earnings by Δy^* lowers lifetime wealth by $\Delta W_1 = (1-\tau^-)\Delta y^*$ (gross income lost minus tax saved). Since $c_1 = W_1/\mathcal{A}$, the consumption response is $\Delta c_1 = \Delta W_1/\mathcal{A}$. Using $\lambda = c^{-\sigma}$:

$$\ln\left(\frac{\lambda^A}{\lambda^B}\right) = -\sigma\frac{\Delta c_1}{c_1} = -\sigma\frac{(1-\tau^-)\Delta y^*}{\mathcal{A}} \cdot \frac{\mathcal{A}}{W_1} = -\sigma\frac{(1-\tau^-)\Delta y^*}{W_1}$$

Rewriting in terms of percentage changes and using the definition of ϕ:

$$\ln\left(\frac{\lambda^A}{\lambda^B}\right) = -\frac{\phi}{k} \cdot \frac{\Delta y^*}{y^*}$$

Substituting into (J6) with $\Delta\tilde{\tau} = Q_D\Delta\tau$:

$$\frac{1}{k}\frac{\Delta y^*}{y^*} + \frac{\phi}{k}\frac{\Delta y^*}{y^*} = \frac{Q_D\Delta\tau}{1-\tau^-}$$

$$\frac{\Delta y^*}{y^*} \cdot \frac{1+\phi}{k} = \frac{Q_D\Delta\tau}{1-\tau^-}$$

The bunching estimator relates the behavioral response Δy^* (the earnings reduction of the marginal buncher) to the statutory tax change via $\frac{\Delta y^*}{y^*} = \varepsilon^{\text{bunch}} \cdot \frac{\Delta\tau}{1-\tau^-}$.[43] Solving for $\varepsilon^{\text{bunch}}$ gives the result. ■

[43] Since the behavioral response is driven by the effective change $\Delta\tilde{\tau} = Q_D\Delta\tau$, but the bunching formula uses the statutory change $\Delta\tau$, the factor Q_D appears in the identified elasticity.

J.1.3 Discussion of Identification Theorem

Recovering the Hicksian Elasticity. In the absence of either financial frictions or delayed taxation, $Q_D = 1$, and the theorem simplifies to $\varepsilon^{\text{bunch}} = k/(1+\phi)$. In the static case ($T = 1$), lifetime wealth reduces to single-period consumption, $W_1 = c$, so $(1-\tau^-)y^*/W_1 = s$ where s is the labor income share. Then $\phi = k\sigma s$ and $\varepsilon^{\text{bunch}} = k/(1+k\sigma s) = \varepsilon^H$, recovering the expression for the Hicksian in equation (J5), and confirms consistency with Saez (2010).

Chetty (2012) conjecture. In the life-cycle case with $\beta R_t = 1$ and constant earnings ($y_t = y$), the Euler equation implies flat consumption.

$$W_1 = \sum_{t=1}^{T} \prod_{s=1}^{t-1} R_s^{-1} c_t = c_1 \sum_{t=1}^{T} \beta^{t-1} = c_1 \frac{1-\beta^T}{1-\beta}, \tag{J8}$$

$$c_1 = y^*(1-\tau^-). \tag{J9}$$

Hence, $\phi = k\sigma \frac{1-\beta}{1-\beta^T}$. Thus,

$$\varepsilon^{\text{bunch}} = \frac{k}{1+k\sigma\frac{1-\beta}{1-\beta^T}} \to \frac{k}{1+k\sigma(1-\beta)} \quad \text{as } T \to \infty. \tag{J10}$$

As income effects are spread over T periods, bunching approximates the Frisch elasticity up to an $O(1/T)$ correction *and* a $\frac{1}{1+k\sigma(1-\beta)}$ "structural" correction term. If we assume that $k = 0.3$, $\sigma = 5$, and $\beta = 0.97$, then $\varepsilon^{\text{bunch}} = 0.2871$, which is quite close to the assumed Frisch elasticity, $k = 0.3$.

If we impose the further assumption that $\beta = R = 1$ (as opposed to our weaker assumption that $\beta R = 1$), we obtain a formalization of Chetty's (2012) conjecture since the structural correction term vanishes.

We further note that with quasi-linear preferences ($\sigma = 0$), income effects vanish ($\phi = 0$) and bunching identifies the Frisch elasticity k regardless of horizon and whether $\beta = 1$. In the quasi-linear case, the Hicksian equals the Frisch, hence the bunching estimator also identifies the Hicksian, ε^H.

Financial frictions and delayed taxation. Financial frictions and delayed taxation modify these benchmark results through two channels. First, delayed taxation with $Q_D < 1$ proportionally scales down the identified elasticity. This scaling applies equally whether bunching identifies Frisch or Hicksian: in the life-cycle limit $\phi \to 0$ and $\varepsilon^{\text{bunch}} \to Q_D \cdot k$, while in the static case $\phi = k\sigma s$ and $\varepsilon^{\text{bunch}} = Q_D \cdot k/(1+\phi) = Q_D \cdot k/(1+k\sigma s)$. Second, borrowing constraints reduce lifetime wealth W_1 by discounting future income more heavily. This raises the ratio $(1-\tau^-)y^*/W_1$, which increases ϕ and pushes the identified elasticity toward Hicksian. In the limit $(1-\tau^-)y^*/W_1 \to 1$, we have $\phi \to k\sigma$ and $\varepsilon^{\text{bunch}} \to Q_D \cdot k/(1+k\sigma)$. Financial frictions thus effectively shorten the planning horizon.

J.2 Numerical Illustrations

We calibrate the model to our empirical setting: $T = 40$ years, $\beta = 0.97$, $\sigma = 5$, $k = 0.3$.[44]

Application to de facto Delayed Taxation For Students. We assume students near the delayed tax threshold face $r = 0.02$, and a high borrowing rate, $R = 1 + r_b = 1.20$, for ten years.[45] We also assume that the tax is delayed for $D = 10$ years (average maturity on student loans). We can then write out the discount factor as

$$Q_D = \left(\frac{1+r}{1+r_b}\right)^{10} = \left(\frac{1.02}{1.20}\right)^{10} \approx 0.1969.$$

We compute after-tax life-time wealth as

$$W_1 = \sum_{t=1}^{5} \frac{G^{\text{stipend}}}{(1+r_b)^t} + \sum_{t=1}^{5} \frac{(1-\tau^-)y^*}{(1+r_b)^t} + \left[\sum_{t=6}^{10} \frac{y^g}{(1+r_b)^t} + \frac{1}{(1+r_b)^{10}} \sum_{t=11}^{40} \frac{y^g}{(1+r)^{t-10}}\right], \tag{J11}$$

where G^{stipend} is a student stipend given for 5 years, y^* is the (average) labor income for the next five years, and y^g is the after tax income starting in year six.

We set $G^{\text{stipend}} = 32500$, $y^* = 120\ 162$ (mean delayed tax kink), $y^g = 281\ 707$ (median after-tax income at $t+5$, see Table A.6), and $\tau^- = 0.25$. This yields $\phi = 0.0784$

Applying the theorem:

$$\varepsilon^{\text{bunch}} = Q_D \cdot \frac{1}{1+\phi} \cdot k = 0.1969 \cdot \frac{1}{1+0.0784} \cdot 0.3 = 0.0548. \tag{J12}$$

Income effects provide a modest correction (about 8%), and the difference between ε and k is primarily driven by the present-value adjustment, Q_D.

J.2.1 Comparing elasticities across samples when income effects differ

This subsection illustrates how we can back out differences in interest rates (r_b) across samples when income effects may differ across samples.

Notation for contrasting samples. We start with some notation that draws from the previous subsections. We write the estimated elasticity for some sample, i, as

$$\varepsilon^i = Q_D^i \frac{k}{1+\phi^i}, \tag{J13}$$

[44] We abstract from optimization frictions, which would scale bunching elasticities by a factor $\zeta < 1$ (see Section 5).
[45] Note that Q_D depends on the duration of binding borrowing constraints, not the tax payment delay per se; years where the agent discounts at the market rate contribute equally to numerator and denominator and thus cancel.

where ε^i should be considered the net-of-present-value-tax labor supply elasticity inferred from bunching at the delayed tax threshold. We write the discount factor as

$$Q_D^i = \left(\frac{1+r}{1+r_b^i}\right)^D = \left(\frac{1.02}{1+r_b^i}\right)^D, \tag{J14}$$

where D is the duration of the tax delay. We assume that r_b^i is constant during the D-year long period (alternatively: we consider r_b^i to be the geometric average.) The income-effect adjustment term is written

$$\phi^i = k\sigma \cdot \frac{(1-\tau^{-,i})y^{*,i}}{W_1^i}, \tag{J15}$$

where k and σ are considered constants. We consider W_1^i to be the sum of present-value incomes, given by

$$W_1^i = \sum_{t=1}^{5} \frac{G^i}{(1+r_b^i)^t} + \sum_{t=1}^{5} \frac{(1-\tau^{-,i})y^{*,i}}{(1+r_b^i)^t} + \left[\sum_{t=6}^{10} \frac{y^{g,i}}{(1+r_b^i)^t} + \frac{1}{(1+r_b^i)^{10}} \sum_{t=11}^{40} \frac{y^{g,i}}{(1+r)^{t-10}}\right], \tag{J16}$$

where G^i is a student stipend given for 5 years, $y^{*,i}$ is the annual labor income for the next 5 years, and $y^{g,i}$ is the after tax income starting in year six.

Ratio of estimands. The ratio of estimands across two samples, A and B, can be written

$$\frac{\varepsilon^A}{\varepsilon^B} = \frac{Q_D^A \frac{1}{1+\phi^A} k}{Q_D^B \frac{1}{1+\phi^B} k} = \frac{Q_D^A}{Q_D^B} \frac{1+\phi^B}{1+\phi^A}. \tag{J17}$$

Computing implied interest-rate wedges. We solve this through iteration. We assert that $r_b^A = 0.15$. Our iteration starts by first assume that $\frac{1+\phi^B}{1+\phi^A} = 1$. With $D = 10$, we then obtain that

$$\frac{1+r_b^B}{1+r_b^A} \approx \left(\frac{\varepsilon^A}{\varepsilon^B}\right)^{1/10}, \tag{J18}$$

which allows us to solve for r_b^B given $r_b^A = 0.15$. We then use these r_b^i values to compute W_1^A and W_1^B and then ϕ^A and ϕ^B. In the next iteration, we then use these parameters (ϕ^A and ϕ^B) and equation (J17) to solve for Q_D^A/Q_D^B again, and thus obtain another solution for r_b^B (still assuming $r_b^A = 0.15$). We keep iterating until the ratio $(1+r_b^B)/(1+r_b^A)$ changes by less than 0.0001. We calibrate using statistics from Table 2 and Appendix Table A.6.

Application: interest rate wedge, students above (A) and below (B) median liquidity. We parameterize by setting

$$\begin{aligned}
\varepsilon^A &= (1.84/120.162)/(0.4666/0.75) = 0.0246 && \text{(J19)}\\
\varepsilon^B &= (0.72/120.162)/(0.4666/0.75) = 0.0096, && \text{(J20)}\\
y^{*,A} &= 120162, && \text{(J21)}\\
y^{*,B} &= 120162, && \text{(J22)}\\
y^{g,A} &= 297419, && \text{(J23)}\\
y^{g,B} &= 269512, && \text{(J24)}\\
G^A, G^B &= 32150, && \text{(J25)}\\
r^A &= 0.15, && \text{(J26)}\\
\tau^{-,A}, \tau^{-,B} &= 0.25, && \text{(J27)}
\end{aligned}$$

The initial approximation (assuming no differences in the income-effects adjustment) yields an interest rate ratio of $(1+r^B)/(1+r^A) = (0.0246/0.0096)^{1/10} = 1.0987$. We then use our iteration procedure, iterating 3 times, to obtain a ratio of 1.0921 (see Table A.7 for the iteration steps). Under the assumption that $r_A = 0.15$, the interest rate difference is $1.0921 \times 1.15 - 1.15 = 10.59$ percentage points.

J.3 Income effects in bunching estimates: Summary statistics

Table A.6: Summary Statistics for Income-Effect Adjustments

This table presents summary statistics used for calculating income-effects adjustments for the bunching estimates for different samples. After-tax income is measured five years after the individual is observed in any of the bunching samples. Stipends are calculated by dropping students who are above the delayed tax threshold (since their stipends would be crowded by higher student loans). We otherwise use the same sample restrictions as in Table 2.

	p50
Delayed tax sample, above-median After-tax income, $t+5$	297,419
Delayed tax sample, below-median After-tax income, $t+5$	269,512
Delayed tax sample, all After-tax income, $t+5$	281,707
Delayed tax sample Stipends, t	32,150
Regular tax sample After-tax income, $t+5$	267,783
All students below delayed tax threshold Stipends, t	33,110

TABLE A.7: INCOME EFFECTS IN BUNCHING ESTIMATES: NUMERICAL SOLUTIONS

This table provides the iteration steps and solutions when accounting for differential income effects across bunching samples.

Iteration	r_b^B	W_1^A	W_1^B	ϕ^A	ϕ^B	$(1+r_b^B)/(1+r_b^A)$
	Iteration steps for above-median (A) versus below-median (B) sample					
0	0.26313	2,552,085	1,123,491	0.05297	0.12032	1.09837
1	0.25532	2,552,085	1,176,455	0.05297	0.11491	1.09158
2	0.25593	2,552,085	1,172,209	0.05297	0.11532	1.09211
3	0.25588	2,552,085	1,172,535	0.05297	0.11529	1.09207

J.4 Additional figures

FIGURE A.4: 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.

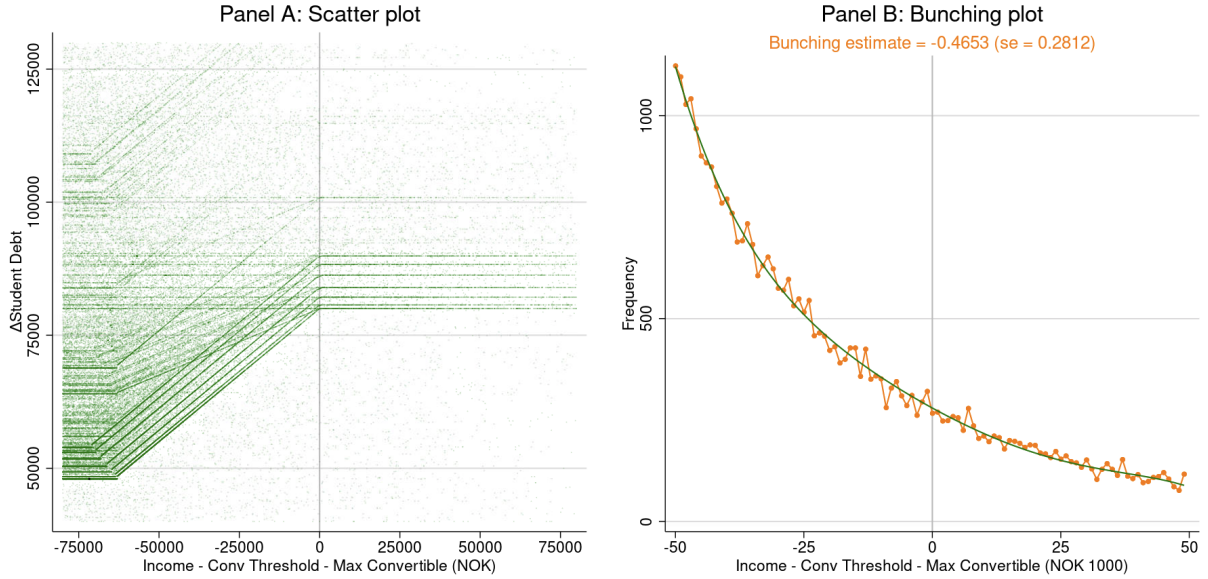
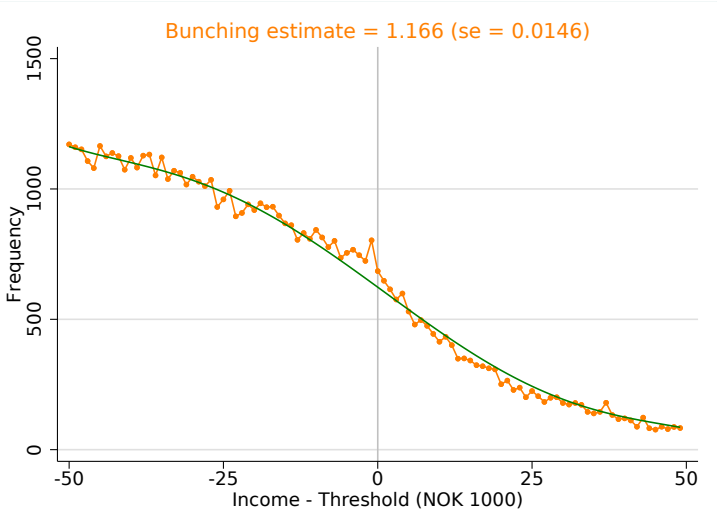


FIGURE A.5: BUNCHING AT DEBT-CONVERSION THRESHOLD FOR WORKERS WITH SALES AND HOSPITALITY OCCUPATIONS

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.



J.5 Individual-level bunching elasticities

We expand here on how individual level elasticities are constructed (see main text section 5.8).

In the standard bunching framework, b is 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

$$\hat{B} = \left(P[y_i \in BR] \cdot N^{sample} - P^{cf}[y_i \in BR] \cdot N^{sample}\right) \times \text{Bin width}. \tag{J28}$$

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.

$$\hat{b} = \frac{\left(P[y_i \in BR] \cdot N^{sample} - P^{cf}[y_i \in BR] \cdot N^{sample}\right) \times \text{Bin width}}{\frac{P^{cf}[y_i \in BR]}{N^{\text{BR Bins}}} \times N^{sample} \times \text{Bin width}}. \tag{J29}$$

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

$$\hat{b} = \frac{\left(P[y_i \in BR] - P^{cf}[y_i \in BR]\right)}{\frac{P^{cf}[y_i \in BR]}{N^{\text{BR Bins}}}}. \tag{J30}$$

From this we can define an individual level b_i as

$$b_i = \frac{\mathbb{1}[y_i \in BR] - P^{cf}[y_i \in BR]}{\frac{P^{cf}[y_i \in BR]}{N^{\text{BR Bins}}}}, \tag{J31}$$

which satisfies the property that

$$\widehat{\mathbb{E}}[b_i] = \hat{b}, \tag{J32}$$

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 43) and then further by the relative change in the after-tax wage.

K Computational Guide: Valuing Analytical Expressions

This appendix provides a self-contained guide for computing the analytical welfare decompositions in Corollary 1, Lemma 2, and Proposition 3 from numerical allocations. All formulas are evaluated at the benchmark economy ($\delta = 1$, $\tau_1 = \tau_2 = \tau$, $G_1 = G_2 = G$).

Inputs. For each agent type i, the solver provides allocations $(\ell_1^i, \ell_2^i, c_1^i, c_2^i, s^i)$ and parameters (w_1^i, w_2^i, π^i). The global parameters are $(\sigma, k, \xi, \beta, r, r_b, r_b^g, m_1, m_2, T, \tau, G)$.

K.1 Auxiliary variables

Utility derivatives. With CRRA–power utility (36):

$$u'(c) = c^{-\sigma}, \quad u''(c) = -\sigma\, c^{-\sigma-1}, \quad v'(\ell) = \xi\, \ell^{1/k}, \quad v''(\ell) = \frac{\xi}{k}\, \ell^{1/k-1}. \tag{K1}$$

Marginal return on savings. The marginal return $R'(s^i)$ and its derivative $R''(s^i)$ are computed from the smooth approximation (H2). An equivalent computation uses the Euler equation:

$$R'(s^i) = \frac{u'(c_1^i)}{\beta\, u'(c_2^i)}. \tag{K2}$$

Key ratios.

- After-tax wages: $\tilde{w}_t^i = w_t^i(1-\tau)$ for $t = 1, 2$.
- Incomes: $y_t^i = w_t^i \ell_t^i$.
- Effective delta: $\bar{\delta}^i = 1$ at $\delta = 1$.
- Private interest wedge: $\Delta_r^i = 1 - (1+r)^T / R'(s^i)$. Equals zero for savers ($s^i > 0$).
- Government interest wedge: $\Delta_{r,g}^i = 1 - (1+r)^T / R_g'(s^i)$, where R_g' uses government lending costs z^g in place of z.
- Conversion factor at $\delta = 1$: $\Gamma^i = \tau\, \Delta_r^i$.

R″ correction. Define

$$\Phi_R^i \equiv \frac{R''(s^i)\,u'(c_1^i)}{R'(s^i)\,u''(c_1^i)} = -\frac{R''(s^i)\,c_1^i}{\sigma\,R'(s^i)}. \tag{K3}$$

Generalized amplification factor. The amplification factor ι^i (eq. B7) can be computed via

$$\iota_0^i = 1 + \left(1 - \frac{[\tilde{w}_2^i]^2\,u''(c_2^i)}{v''(\ell_2^i)}\right)\frac{u''(c_1^i)}{\beta\,R'(s^i)^2\,u''(c_2^i)}, \tag{K4}$$

$$\iota^i = \frac{\iota_0^i + \Phi_R^i\,(\iota_0^i - 1)}{1 + \Phi_R^i\,(\iota_0^i - 1)}. \tag{K5}$$

When $R''(s^i) = 0$: $\Phi_R^i = 0$ and $\iota^i = \iota_0^i$.

K.2 Labor supply responses to τ_1

All responses are with respect to $d(1-\tau_1)$, holding τ_2 fixed, evaluated at $\tau_1 = \tau_2 = \tau$ and $\delta = 1$.

Income effects.

$$\frac{d\ell_1^i}{dG_1} = \frac{u''(c_1^i)\,\tilde{w}_1^i}{v''(\ell_1^i)\,\iota^i - (\tilde{w}_1^i)^2\,u''(c_1^i)}, \tag{K6}$$

$$\frac{d\ell_2^i}{dG_1} = \frac{\tilde{w}_2^i\,v''(\ell_1^i)}{\beta\,R'(s^i)\,\tilde{w}_1^i\,v''(\ell_2^i)}\left(1 - \Phi_R^i\,(\iota^i - 1)\right)\frac{d\ell_1^i}{dG_1}. \tag{K7}$$

Compensated substitution effects.

$$\left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c = \frac{u'(c_1^i)\,w_1^i\,\iota^i}{v''(\ell_1^i)\,\iota^i - (\tilde{w}_1^i)^2\,u''(c_1^i)}, \tag{K8}$$

$$\left(\frac{d\ell_2^i}{d(1-\tau_1)}\right)^c = \left(1 - \Phi_R^i\,(\iota^i - 1)\right)\left[\frac{\tilde{w}_2^i\,v''(\ell_1^i)}{\beta\,R'(s^i)\,\tilde{w}_1^i\,v''(\ell_2^i)}\left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c - \frac{v'(\ell_2^i)}{v''(\ell_2^i)\,(1-\tau)}\right]. \tag{K9}$$

K.3 Revenue effect parameters

Compensated substitution revenue.

$$X^i = \tau\,w_1^i\left(\frac{d\ell_1^i}{d(1-\tau_1)}\right)^c + \frac{\tau\,w_2^i}{(1+r)^T}\left(\frac{d\ell_2^i}{d(1-\tau_1)}\right)^c. \tag{K10}$$

Income effect on revenue.

$$\rho_1^i = \tau\,w_1^i\,\frac{d\ell_1^i}{dG_1} + \frac{\tau\,w_2^i}{(1+r)^T}\,\frac{d\ell_2^i}{dG_1}. \tag{K11}$$

K.4 Social welfare weights and shadow value

Shadow value of public funds.

$$\lambda = \frac{\sum_i \pi^i \left(u'(c_1^i) + \beta\, u'(c_2^i)\right)}{\sum_i \pi^i \left(1 + \frac{1}{(1+r)^T} - \rho_1^i - \rho_2^i\right)}, \tag{K12}$$

where ρ_2^i is defined analogously to ρ_1^i using $d\ell_t^i/dG_2$ (computed from the Jacobian system, see below).

Welfare weights.

$$g_1^i = \frac{u'(c_1^i)}{\lambda}, \qquad g_2^i = \frac{\beta\,(1+r)^T\, u'(c_2^i)}{\lambda}. \tag{K13}$$

For borrowers: $g_1^i - g_2^i = \Delta_r^i\, g_1^i$. For savers: $g_1^i = g_2^i$.

K.5 Timing correction Ψ^i

The timing correction $\Psi^i = (\Psi_1^i, \Psi_2^i, \Psi_s^i)$ accounts for the difference between the δ reform and a Γ-scaled τ_1 reform. It is defined by $\mathbf{J}^i \Psi^i = \Delta \mathbf{b}_\Gamma^i$ (eq. 31).

Jacobian. The 3×3 Jacobian $\mathbf{J}^i$ of the system (FOC_{ℓ_1}, FOC_{ℓ_2}, Euler) with respect to (ℓ_1, ℓ_2, s) at $\delta = 1$:

$$\begin{aligned}
J_{11}^i &= u''(c_1^i)\,(\tilde{w}_1^i)^2 - v''(\ell_1^i), & J_{13}^i &= -u''(c_1^i)\,\tilde{w}_1^i,\\
J_{22}^i &= u''(c_2^i)\,(\tilde{w}_2^i)^2 - v''(\ell_2^i), & J_{23}^i &= u''(c_2^i)\,\tilde{w}_2^i\,R'(s^i),\\
J_{31}^i &= u''(c_1^i)\,\tilde{w}_1^i, & J_{32}^i &= -\beta\,R'(s^i)\,u''(c_2^i)\,\tilde{w}_2^i,\\
J_{33}^i &= -u''(c_1^i) - \beta\,R''(s^i)\,u'(c_2^i) - \beta\,R'(s^i)^2\,u''(c_2^i). & &
\end{aligned} \tag{K14}$$

At $\delta = 1$, $J_{12}^i = J_{21}^i = 0$ (block-sparse).

RHS vector. The mechanical forcing term $\Delta \mathbf{b}_\Gamma^i$ at $\delta = 1$ (eq. B26):

$$\begin{aligned}
\Delta b_{\Gamma,1}^i &= -(1 - \Delta_r^i)\,u''(c_1^i)\,\tau\,y_1^i\,\tilde{w}_1^i,\\
\Delta b_{\Gamma,2}^i &= u''(c_2^i)\,(1+r)^T\,\tau\,y_1^i\,\tilde{w}_2^i,\\
\Delta b_{\Gamma,3}^i &= -\tau\,y_1^i\,(1 - \Delta_r^i)\left[u''(c_1^i) + \beta\,R'(s^i)^2\,u''(c_2^i)\right].
\end{aligned} \tag{K15}$$

Solution by block elimination. Since $J^i_{12} = J^i_{21} = 0$:

$$\Psi^i_s = \frac{\Delta b^i_{\Gamma,3} - \frac{J^i_{31}\,\Delta b^i_{\Gamma,1}}{J^i_{11}} - \frac{J^i_{32}\,\Delta b^i_{\Gamma,2}}{J^i_{22}}}{J^i_{33} - \frac{(J^i_{31})(J^i_{13})}{J^i_{11}} - \frac{(J^i_{32})(J^i_{23})}{J^i_{22}}}, \tag{K16}$$

$$\Psi^i_1 = \frac{\Delta b^i_{\Gamma,1} - J^i_{13}\,\Psi^i_s}{J^i_{11}}, \qquad \Psi^i_2 = \frac{\Delta b^i_{\Gamma,2} - J^i_{23}\,\Psi^i_s}{J^i_{22}}. \tag{K17}$$

Timing correction to behavioral revenue.

$$X^i_\Psi = \tau\, w^i_1\, \Psi^i_1 + \frac{\tau\, w^i_2}{(1+r)^T}\, \Psi^i_2. \tag{K18}$$

K.6 Income effects from period-2 transfers

The $d\ell^i_t/dG_2$ responses are needed for ρ^i_2 and λ. At $\delta = 1$, the RHS vector for a dG_2 perturbation is $\mathbf{b}^i_{G_2} = (0,\ -u''(c^i_2)\,\tilde{w}^i_2,\ \beta\, R'(s^i)\, u''(c^i_2))$. Block elimination gives:

$$\frac{ds^i}{dG_2} = \frac{\beta\, R'(s^i)\, u''(c^i_2) + \frac{J^i_{32}\, u''(c^i_2)\, \tilde{w}^i_2}{J^i_{22}}}{D^i_s}, \tag{K19}$$

$$\frac{d\ell^i_1}{dG_2} = \frac{-J^i_{13}}{J^i_{11}}\,\frac{ds^i}{dG_2}, \qquad \frac{d\ell^i_2}{dG_2} = \frac{-u''(c^i_2)\,\tilde{w}^i_2 - J^i_{23}\, ds^i/dG_2}{J^i_{22}}, \tag{K20}$$

where $D^i_s = J^i_{33} - J^i_{31}J^i_{13}/J^i_{11} - J^i_{32}J^i_{23}/J^i_{22}$ is the Schur complement. Then:

$$\rho^i_2 = \tau\, w^i_1\, \frac{d\ell^i_1}{dG_2} + \frac{\tau\, w^i_2}{(1+r)^T}\, \frac{d\ell^i_2}{dG_2}. \tag{K21}$$

K.7 Valuing Corollary 1: Decomposed effects of marginally introducing delayed taxation

The money-metric welfare effect (eq. 33) is decomposed as $\mathrm{DT}_{\text{total}} = \mathrm{DT}_1 + \mathrm{DT}_2 + \mathrm{DT}_3 + \mathrm{DT}_4$:

$$\mathrm{DT}_1 = \tau \sum_i \pi^i\, y^i_1\, (g^i_1 - g^i_2) \qquad \text{(consumption smoothing)}, \tag{K22}$$

$$\mathrm{DT}_2 = \sum_i \pi^i\, \tau\, \Delta^i_r\, X^i \qquad \text{(compensated substitution)}, \tag{K23}$$

$$\mathrm{DT}_3 = \tau \sum_i \pi^i\, y^i_1\, \left[\rho^i_1 - (1{+}r)^T\, \rho^i_2\right] \qquad \text{(income effects)}, \tag{K24}$$

$$\mathrm{DT}_4 = \tau \sum_i \pi^i\, y^i_1 \left(\frac{(1+r)^T}{R'_g(s^i)} - 1\right) \qquad \text{(fiscal cost of government lending)}. \tag{K25}$$

Note: $\mathrm{DT}_3 = \sum_i \pi^i\, \tau\, \Delta^i_r\, y^i_1\, \rho^i_1 + X^i_\Psi$, where the timing correction X^i_Ψ from (K18) cancels with the income effect correction η^i (see Section B.5), yielding the clean income term $\rho^i_1 - (1{+}r)^T \rho^i_2$. For

savers ($\Delta_r^i = 0$ and $\rho_1^i - (1+r)^T \rho_2^i = 0$), all behavioral terms vanish.

K.8 Valuing Lemma 2: Marginal welfare from government lending

From equation (34):

$$\frac{1}{\lambda}\frac{dW}{dx}\Big|_{\delta=1} = \underbrace{\sum_i \pi^i (g_1^i - g_2^i)}_{\text{Loan}_1\text{: smoothing}} + \underbrace{\sum_i \pi^i \left[\rho_1^i - (1+r)^T \rho_2^i\right]}_{\text{Loan}_2\text{: behavioral}} - \underbrace{\sum_i \pi^i \Delta_{r,g}^i}_{\text{Loan}_3\text{: fiscal cost}} . \tag{K26}$$

Only borrowers contribute to each term.

K.9 Valuing Proposition 3: Decomposing delayed taxation

The five terms in equation (36) are:

$$\text{(i)} = \tau\, \bar{y}_1 \cdot \frac{1}{\lambda}\frac{dW}{dx}\Big|_{\delta=1} \qquad \text{(lending effect)}, \tag{K27}$$

$$\text{(ii)} = \tau \sum_{i:s^i<0} \pi^i\, \Delta_r^i\, X^i \qquad \text{(reduced tax distortions)}, \tag{K28}$$

$$\text{(iii)} = \tau \sum_{i:s^i<0} \pi^i\, \Delta_r^i\, (y_1^i - \bar{y}_1)\, g_1^i \qquad \text{(loan targeting)}, \tag{K29}$$

$$\text{(iv)} = \tau \sum_{i:s^i<0} \pi^i\, (y_1^i - \bar{y}_1) \left(\rho_1^i - (1+r)^T \rho_2^i\right) \qquad \text{(income effect targeting)}, \tag{K30}$$

$$\text{(v)} = -\tau \sum_{i:s^i<0} \pi^i\, (y_1^i - \bar{y}_1)\, \Delta_{r,g}^i \qquad \text{(fiscal cost targeting)}, \tag{K31}$$

where $\bar{y}_1 = \sum_{i:s^i<0} \pi^i y_1^i / \sum_{i:s^i<0} \pi^i$ is the conditional mean income of borrowers.

K.10 Marginal Value of Public Funds

The MVPF is the ratio of willingness to pay (WTP) to net fiscal cost. For a δ reform:

$$\text{MVPF}_{\text{DT}} = \frac{\text{WTP}_{\text{DT}}}{\text{Cost}_{\text{DT}}}, \tag{K32}$$

where

$$\text{WTP}_{\text{DT}} = \tau \sum_i \pi^i\, \Delta_r^i\, y_1^i\, g_1^i, \tag{K33}$$

$$\text{Cost}_{\text{DT}} = -\tau \sum_i \pi^i\, \Delta_r^i\, X^i - \tau \sum_i \pi^i\, y_1^i \left[\rho_1^i - (1+r)^T \rho_2^i\right] + \tau \sum_i \pi^i\, y_1^i\, \Delta_{r,g}^i. \tag{K34}$$

The first term captures compensated substitution effects on revenue, the second is the income effect (loan-like), and the third is the mechanical fiscal cost. If $\text{Cost}_{\text{DT}} \leq 0$ (the reform generates net revenue), the MVPF is set to infinity.

For a subsidized loan (dx):

$$\text{MVPF}_{\text{Loan}} = \frac{\sum_i \pi^i (g_1^i - g_2^i)}{-\sum_i \pi^i \left[\rho_1^i - (1+r)^T \rho_2^i\right] + \sum_i \pi^i \Delta_{r,g}^i}. \tag{K35}$$

For an age-dependent tax cut ($d(1-\tau_1)$):

$$\text{MVPF}_{\text{AD}} = \frac{\sum_i \pi^i y_1^i g_1^i}{\sum_i \pi^i \left(y_1^i - y_1^i \rho_1^i - X^i\right)}. \tag{K36}$$