OPTIMAL INCOME TAXATION WITH PRESENT BIAS*

Benjamin B. Lockwood†

December 2019

Abstract

Work often entails up-front effort costs in exchange for delayed benefits, and mounting evidence documents present bias over effort in the face of such delays. This paper studies the implications for the optimal income tax. Optimal tax rates are computed for present-biased workers who choose multiple dimensions of labor effort, some of which occur prior to compensation. Present bias reduces optimal tax rates, with a larger effect when the elasticity of taxable income is high. Optimal marginal tax rates may be negative at low incomes, providing an alternative, corrective rationale for work subsidies like the Earned Income Tax Credit. (JEL: D04, H21)

*I thank Raj Chetty, Nathaniel Hendren, David Laibson, and Matthew Weinzierl for extensive discussions, and Charles Nathanson, Emmanuel Farhi, Louis Kaplow, Wojciech Kopczuk, Alex Rees-Jones, Dmitry Taubinsky, Glen Weyl, and seminar participants at Harvard, University of Pennsylvania, Duke, Georgetown, University of Oregon, University of Maryland, Cornell, the National Tax Association and the American Economic Association Annual Meetings, as well as the editor and anonymous referees for many helpful comments.

†University of Pennsylvania and NBER, ben.lockwood@wharton.upenn.edu.

1 Introduction

Present bias—the form of time-inconsistency in which consumers appear more impatient over immediate intertemporal tradeoffs than more distant ones—is often viewed as a form of misoptimization that accounts for a range of behavioral "mistakes," ranging from undersaving for retirement to exercising too little. Accordingly, a large literature emphasizes the potential for policies like forced pensions or retirement savings subsidies to protect against or correct such mistakes. More recently, a growing body of empirical evidence suggests that present bias may also affect decisions about labor supply. Simply put, work often entails up-front effort in exchange for delayed benefits. To the extent that immediate costs are inflated due to present bias, a worker may exert less effort than their "long run self" would prefer. As a result, present bias may affect the optimal design of policies relating not only to savings, but also to labor supply. This paper considers the implications for income taxation.

A natural question is whether the time delay between effort and compensation in the real economy is sufficiently long for present bias to be relevant. Two considerations suggest the answer is yes. First, empirical studies documenting present bias over labor effort find it operates at horizons of one to two weeks—the amount by which paychecks are delayed for most workers. If workers have some control over the number of hours worked, present bias may affect that choice.[1] Second, the compensation received at a particular time is often the result of many prior choices, some of which occur long in advance. Consider a sales representative whose effort consists of contacting prospective clients and following up on leads: much of that work occurs long before the realization of a sales commission. Even in the most prototypical hourly wage jobs, such as food service, workers can choose how much effort to put into the *quality* of their work, with the prospect of getting an eventual raise or promotion (or, if quality is low, raising the risk of dismissal). Such quality dimensions of effort may also be more easily adjustable—and thus more elastic to economic incentives—than the choice of hours. The empirical evidence of present bias over labor supply, discussed at length in section 3.1, suggests these forces are quantitatively important for labor supply behavior.

Motivated by these considerations, this paper presents a model of optimal taxation when workers are present-biased. Formally, the primary tax formulas can be interpreted to span a broad class of

[1] This may be particularly important for workers with flexible or alternative work arrangements, which constitute a growing share of the economy (Katz and Krueger, 2019).

reasons a policymaker might place less weight on effort disutility than an individual does.[2] However, I will focus on the mechanism of present bias both because it appears to be a prominent force in labor supply decisions, and because there is a substantial literature *quantifying* present bias across labor supply decisions and other domains, which facilitates numerical simulations of the optimal tax schedule.

The contributions of this paper are both theoretical and empirical. In the theory section, I generalize the benchmark model of optimal income taxation to allow for effort which is exerted by present-biased workers, prior to the time of compensation. I derive necessary and sufficient conditions to characterize the optimal tax in a simplified setting, as well as a necessary (first-order) condition for the optimal income tax under more general conditions with multiple dimensions of labor supply over time, representing effort along diverse channels such as work quality, on- and off-the-job training, and search effort, in addition to the usual choice of labor hours.

The theoretical results demonstrate that present bias tends to lower optimal marginal tax rates, consistent with the Pigouvian logic that uninternalized benefits call for a corrective subsidy. The results also highlight a number of more nuanced implications, including two key points for policy design. First, a higher elasticity of taxable income magnifies the optimal corrective component of the subsidy. This contrasts with the standard Pigouvian result that the optimal correction is insensitive to elasticity and is additively separable from the optimal redistributive tax. Second, although initial intuition might suggest that the delayed nature of tax bills and credits—which are typically paid at the end of the year—would be undermine their corrective strength, Section 2.4 shows this need not be the case. On the contrary, a modest delay is *optimal* in the presence of multidimensional labor effort, as it dampens the undesirable upward distortion of work subsidies for effort at the time of compensation.

Empirically, I draw from a wide range of evidence to calibrate present bias across the income distribution, which appears to be concentrated at low incomes. Using data on the income distribution and existing tax schedules, I calibrate the skill distribution accounting for present bias, and I simulate the optimal income tax for a range of normative preferences.

The quantitative results indicate that in practice the reduction in optimal marginal tax rates is

[2]Relatedly, this formal model spans settings in which workers place *too little* weight on the up-front costs of labor effort, for example by being a "workaholic" at the expense of one's family life. I thank an anonymous referee for pointing out this possibility.

concentrated at low incomes. The effect may be substantial: if redistributive motives are modest, marginal tax rates are optimally *negative* across a range of low incomes. Figure 1 plots the schedule of optimal marginal tax rates in the baseline model economy with and without accounting for present bias, under a modest degree of inequality aversion (see Section 3 for details). This result could provides a potential justification for negative marginal tax rates low incomes, like those arising under the Earned Income Tax Credit (EITC) in the U.S., which are suboptimal under many calibrated models of optimal taxation. To further explore this possibility, I perform an "inverse optimum" exercise, which backs out the social marginal welfare weights consistent with the existing EITC and income tax schedule, with and without accounting for present bias. Absent present bias, these weights exhibit a normative feature that is inconsistent with conventional assumptions in the optimal taxation literature: consumption for the poorest workers is valued *less* than for those in the middle of the income distribution. When accounting for present bias, however, social welfare weights decline monotonically with income, consistent with conventional normative assumptions.

This paper relates to two broad subfields in the optimal taxation literature. The first is optimal taxation with misoptimizing agents. Kanbur, Pirttilä and Tuomala (2006) surveys the literature on taxation with such "paternalistic" motives. More recently, Allcott, Lockwood and Taubinsky (2019) studies optimal commodity taxation with misoptimizing agents, while Gerritsen (2016) and Farhi and Gabaix (2019) study income taxation with behavioral agents who misoptimize labor supply, providing formulas for optimal tax rates in terms of general misoptimization wedges. This paper instead studies a specific source of misoptimization—present bias—which advances this literature in two ways. First, focusing on a known bias constrains the direction and magnitude of misoptimization, resulting in positive conclusions about the optimal tax system, including theoretical implications like the optimal timing results discussed above, which cannot be derived from a general model with misoptimization wedges. Second, the focus on present bias leverages the large empirical literature that estimates this bias, permitting simulations that give quantitative guidance the optimal tax system. In this respect, this paper is more similar to Spinnewijn (2015), which calibrates optimal unemployment insurance using data on mistaken beliefs about the probability of reemployment, and to Moser and Olea de Souza e Silva (2019), which characterizes optimal savings policies when agents undersave for retirement.

This paper also contributes to the literature on the optimality of negative marginal tax rates. In the canonical Mirrlees (1971) model of redistributive income taxation, negative marginal tax rates are suboptimal.[3] Later analyses explored the sensitivity of that result to positive and normative assumptions in the conventional model. Diamond (1980) and Saez (2002) influentially argued that marginal tax rates could theoretically be negative at low incomes if earnings responses are concentrated on the extensive margin—for example, if there are heterogeneous fixed costs of labor force participation. Yet quantitatively, that result appeared to be muted. Saez (2002) and Blundell and Shephard (2011) use models with discrete earnings levels to simulate optimal tax rates, finding low positive (or very slightly negative) marginal tax rates on the lowest positive earning type. Jacquet, Lehmann and Van der Linden (2013) refined this insight in a continuous model, showing that extensive margin effects generally call for a positive participation credit (a fixed amount paid to all labor force participants), with *positive* marginal tax rates at all positive incomes. Hansen (2018) further contributes by demonstrating that for sufficiently muted redistributive preferences, small negative marginal tax rates at low incomes may be optimal.

Other work has shown that marginal work subsidies may be justified by normative objectives which differ from those in the conventional model. Most plainly, negative marginal tax rates may be warranted if the policymaker's goal is to redistribute income *upward*—i.e., if marginal social welfare weights are rising with income (Stiglitz, 1982). Several papers have argued that such weights may arise from multidimensional heterogeneity (Cuff, 2000; Beaudry, Blackorby and Szalay, 2009; Choné and Laroque, 2010).[4] Fleurbaey and Maniquet (2006) shows how fairness considerations may generate welfare weights which increase with income in equilibrium. Drenik and Perez-Truglia (2018) provides empirical evidence for such views, while noting that such an objective could generate Pareto inefficient policy recommendations. The reasoning in this paper is not inconsistent with such normative objectives, but it demonstrates that negative marginal tax rates may be warranted even under the conventional assumption that policymakers wish to redistribute toward low earners.[5]

[3] This finding has been discussed extensively—see Seade (1977), Seade (1982), Hellwig (2007), and citations therein.

[4] Preference heterogeneity alone is not generally sufficient to generate negative marginal tax rates, however—see simulations in Lockwood and Weinzierl (2015), where optimal tax rates are lower in the presence of such heterogeneity, but remain positive.

[5] Two other proposed rationales for negative marginal tax rates have received some attention: (1) negative rates may be justified by "non-welfarist" objectives, e.g., if the government wishes to minimize poverty (Kanbur, Keen and Tuomala, 1994; Besley and Coate, 1992, 1995) or has preferences directly over the labor and leisure choices (Moffitt, 2006), or (2) work may have positive externalities, e.g., on children's outcomes (Dahl and Lochner, 2012).

The remainder of the paper is organized as follows. Section 2 presents a model of optimal taxation with present-biased workers, first in a simple case, then in a generalized setting with multiple dimensions of labor effort. Section 3 presents quantitative results, including calibrations of the optimal income tax, which draw on recent evidence of present bias over effort to calibrate present bias across the income distribution. Section 3.3 presents an inverse optimum exercise which computes redistributive preferences that are consistent with the existing EITC and income tax schedule, with and without accounting for present bias. Section 4 discusses implications for policy design, and Section 5 concludes.

2 Model

Although the canonical Mirrlees (1971) model of optimal income taxation contains a single dimension of labor effort (hours worked), that choice has long been understood to represent a more general setting with a broad array of choices that affect earnings, including training and search effort, some of which may occur well before compensation is received. This section first considers a model with a single dimension of labor supply which occurs prior to compensation, in order to derive necessary and sufficient conditions for the optimal tax. I then consider a more general model with multiple dimensions of labor supply, only some of which occur prior to compensation, in order to derive a more general optimal tax formula, and to consider the question of optimal tax timing.

2.1 A unidimensional model of present-biased workers

The economy consists of a population of individuals of measure one, indexed by i and distributed according to measure $\mu(i)$. Workers have utility $U(c, \ell) = u(c) - v(\ell)$, deriving weakly concave utility from consumption ($u' > 0$, $u'' \leq 0$), and additively separable convex disutility from labor ($v' > 0$, $v'' > 0$). Labor earnings z, which are observable to the government, are the product of a worker's labor effort and their (unobserved) effective wage: $z = \ell w(i)$.

Labor effort occurs prior to the time of compensation, and the worker is assumed to be time inconsistent, with β-δ quasi-hyperbolic discounting preferences (Laibson, 1997). In this simple case, I assume the exponential discount rate δ is equal to one. (This assumption is relaxed in the more general model with multidimensional labor supply below.) Therefore, from a long-run perspective—

prior to the time at which effort is exerted—the worker would like to maximize $\beta(u(c) - v(\ell))$, or, rescaling,

$$u(c) - v(z/w(i)). \tag{1}$$

I adopt the normative assumption that the policymaker agrees with this long-run perspective, and treats Equation (1) as the individual's *normative utility* function.[6]

At the time effort is chosen, the present-biased worker places full weight on labor effort disutility, but discounts the resulting future consumption utility by β, resulting in a *decision utility* function

$$-v(z/w(i)) + \beta u(c). \tag{2}$$

In this model, it does not matter whether workers realize they are present-biased (i.e., whether they are naive or sophisticated), since there is only a single period of unconstrained choice by the present-biased self. However, since the key source of misoptimization involves under-supplying labor, this model does rule out a setting in which workers are fully sophisticated about present bias and use binding commitment devices to fully correct that bias.

In keeping with the conventional Mirrleesian approach, I assume the policymaker can observe compensation z, but not labor supply or wages directly. The policymaker's problem is to select the tax function $T(z)$ which maximizes social welfare, equal to total normative utility, possibly subject to a weakly concave transformation $G(U)$:

$$\mathcal{W} = \int G(U(c(i), \ell(i)))d\mu(i), \tag{3}$$

where $c(i) = z(i) - T(z(i))$. This maximization is performed subject to the policymaker's budget

[6] Although this normative treatment of present bias is common, it is not universal—e.g., Bernheim and Rangel (2009) studies welfare analysis without taking a stand on which time-inconsistent preferences are "correct." Some suggestive support for the normative stance in this paper comes from subjective well-being evidence: Meyer and Sullivan (2008) study the change in consumption and labor supply of single mothers following the tax and welfare reforms of the late 1990s (which reduced lump-sum like benefits and inflated work subsidies), concluding "The significant drop in nonmarket time suggests that utility has fallen for those in the bottom half of the consumption distribution if this nonmarket time is valued at more than \$3 per hour." Yet several papers find that the subjective well-being of single mothers (relative to groups unaffected by these reforms) stayed constant or rose over this period (Ifcher, 2010; Herbst, 2013; Ifcher and Zarghamee, 2014), consistent with an undervaluation of the experienced utility returns to work in the pre-reform period.

constraint

$$\int T(z(i))d\mu(i) - E \geq 0, \tag{4}$$

where E denotes an exogenous revenue requirement, and subject to incentive compatibility constraints, which require that each agent's choice of $\ell(i)$ maximizes Equation (2) above.

2.2 Optimal taxes in a simple case

I first focus on the simple case of unidimensional labor supply and no income effects. In this case, the first-order (necessary) condition for optimal marginal tax rates can be written in terms of model primitives, and it is possible to specify sufficient conditions for this formula to represent the optimum.[7]

Specifically, for the purposes of this subsection I impose the following restrictive assumption:

Assumption 1. *The primitives of the planner's problem satisfy the following conditions:*

(a) Utility is quasilinear in consumption ($u(c) = c$), which rules out income effects.

(b) Ability w is distributed between $w_{min} \geq 0$ and $w_{max} < \infty$ with full support, and is continuously differentiable with density $f(w)$ and distribution $F(w)$.

(c) Present bias $\beta(i)$ is homogeneous conditional on ability $w(i)$, with $\beta(i)$ varying continuously with $w(i)$.

Since w and β vary jointly in this case (implying only a single dimension of heterogeneity), I simplify notation to write present bias and choice variables as functions of ability w rather than type i. I define $\zeta_\ell(w) = \frac{v'(\ell)}{v''(\ell)\ell}$ to denote the labor supply elasticity, and I define $\zeta_\beta(w) = \beta'(w)w/\beta(w)$, representing the elasticity of present bias with respect to ability. Then the following proposition characterizes the optimal income tax policy.

[7]The solution is expressed in primitives in the sense that the marginal tax rate schedule does not appear on the right-hand side, in contrast to the general expression in Proposition 2, where income elasticities depend on the tax curvature T''.

Proposition 1. *Under Assumption 1, the nonlinear income tax $T(z(w))$ satisfies the following expression at all points of differentiability:*

$$\frac{T'(z(w))}{1-T'(z(w))} = \mathcal{A}(w)\mathcal{B}(w) - \mathcal{C}(w), \tag{5}$$

with

$$\mathcal{A}(w) = \frac{1+1/\zeta_\ell(w)+\zeta_\beta(w)}{wf(w)} \tag{6}$$

$$\mathcal{B}(w) = \int_{x=w}^{w_{max}} (1-g(x))\, dF(x) \tag{7}$$

$$\mathcal{C}(w) = g(w)\left(1-\beta(w)\right), \tag{8}$$

where

$$g(w) = \frac{G'(U(c(w),\ell(w)))}{\int_{x=w_{min}}^{w_{max}} G'(U(c(x),\ell(x)))dF(x)}. \tag{9}$$

All proofs are provided in Appendix A. Expression (5) is *sufficient* to ensure optimality if the following conditions hold: (1) under the induced earnings schedule, each individual's globally optimal labor supply choice is characterized by the first-order condition $v'(\ell) = w\beta(w)(1-T'(z(w))$, and (2) a concavity condition is satisfied.[8] Although these conditions are not guaranteed, they are easy to check in any specific case, and they are satisfied in the numerical simulations in Section 3. If this solution would generate an earnings schedule which decreases over some region of ability, then marginal tax rates are discontinuous, implying that $T(z)$ is kinked, with a range of abilities bunching at the same level of earnings. In this case the first sufficiency condition is violated.

An expression analogous to (5) can be expressed in terms of the observable income distribution, in the style of Saez (2001). For an intuitive derivation, consider a tax reform which slightly raises the marginal tax rate by $d\tau$ in a narrow neighborhood ϵ around some income level z^*. This reform transfers $d\tau\epsilon$ from individuals earning more than z^* to the government, resulting in a "mechanical

[8]See equation (38) in the appendix for the exact condition. It is guaranteed if $(1-\beta(w))g(w)\left(\frac{G''(U(w))}{G'(U(w))}\cdot\frac{v'(\ell(w))^2}{\beta(w)}+1\right) < v''(\ell(w))$ and $2v''(\ell(w))+\ell(w)v'''(\ell(w)) > 0$. The first condition amounts to requiring that the marginal social welfare benefit from inducing a (utility compensated) increase in w's labor supply is declining. The second condition is identical to the concavity assumption in Mirrlees (1971) p. 186.

effect" identical to the one in Saez (2001):

$$dM = d\tau\epsilon \int_{z^*}^{\infty} (1 - g(s))h(s)ds, \tag{10}$$

where $g(z)$ is interpreted as the welfare weight as a function of income, and $h(z)$ is the income density, both of which are endogenous to the tax policy. The reform also generates a behavioral response among z^*-earners of $dz^* = -d\tau \cdot \varepsilon(z^*) \cdot \frac{z^*}{1-T'(z^*)}$, where $\varepsilon(z^*)$ denotes the local elasticity of taxable income with respect to the keep share $1 - T'(z^*)$. This adjustment generates a negative fiscal externality equal to $dz^* \cdot T'(z^*)$, and a corrective effect for the z^*-earner equal to $dz^* \cdot g(z^*)(1 - \beta(z^*))(1 - T'(z^*))$. This last effect is the key departure from Saez (2001), accounting for the first-order welfare loss to the z^*-earner from reducing earnings which are already suboptimally low due to present bias. The number of individuals subject to this behavioral response is $h(z^*)\epsilon$, implying that the total behavioral effect is

$$dB = -d\tau\epsilon \cdot \varepsilon(z^*)z^*h(z^*) \left[\frac{T'(z^*)}{1 - T'(z^*)} - g(z^*)(1 - \beta(z^*)) \right]. \tag{11}$$

At the optimal policy, this marginal reform has no first-order welfare benefit ($dM + dB = 0$), implying that the optimal tax satisfies

$$\frac{T'(z)}{1 - T'(z)} = \frac{1}{\varepsilon(z)h(z)z} \int_{z}^{\infty} (1 - g(s))h(s)ds - g(z)(1 - \beta(z)). \tag{12}$$

(See the proof of Proposition 2 for a more thorough derivation in terms of the income distribution.)

The expression in Proposition 1 is a generalization of Diamond (1998), which is identical except for the appearance of $\zeta_\beta(w)$ in $\mathcal{A}(w)$, and the final *corrective term*, $\mathcal{C}(w)$. The presence of $\zeta_\beta(w)$ in $\mathcal{A}(w)$ is unrelated to a corrective motive.[9] Rather, it reflects the effect of heterogeneous present bias on incentive compatibility constraints: if present bias is falling with ability at some $w^\dagger$ (so $\beta'(w^\dagger) > 0$), then high ability types are less tempted to mimic lower ability types by reducing effort, and thus marginal tax rates in the neighborhood of $w^\dagger$ are optimally higher. Thus the term $\zeta_\beta(w)$ is absorbed in the elasticity of taxable income in Equation 12.

[9]This can be seen by noting that if β is regarded as a normatively valid preference—eliminating any corrective motive—then the optimal tax satisfies $\frac{T'}{1-T'} = \mathcal{A}(w)\mathcal{B}(w)$; see proof in Appendix A.

The corrective term $\mathcal{C}(w)$, which appears via the Hamiltonian optimization in the proof, can be understood using economic intuition. In the standard Mirrlees (1971) model, the local adjustment of hours in response to higher marginal tax rates generates no first-order effect on individual utility due to the envelope theorem. However, when individuals under-supply effort due to present bias, they ignore a fraction $1-\beta$ of the marginal benefits from raising earnings, which are equal to $1-T'(z)$. Thus a small increase in earnings has a first-order effect on the individual's own utility of $(1-\beta(w))(1-T'(z(w)))$. Since this corrective benefit accrues to the individual in question, the effect is scaled by their welfare weight.

The result in Proposition 1 invites three key observations about the role of present bias which are particularly clear in this setting. The first is a "negative at the bottom" result:

Corollary 1. *If $w_{min} > 0$ and $v'(0) = 0$, in addition to the sufficiency conditions in Proposition 1, then there exists a $w^* > w_{min}$ such that $T'(z(w)) < 0$ for all $w < w^*$.*

This corollary, which contrasts the classic result that optimal marginal tax rates are everywhere nonnegative (Seade, 1982), is tightly related to finding that in a conventional model without present bias, the optimal marginal tax rate is zero at the bottom under these assumptions (Seade, 1977). That result is a special case of this corollary—the bottom skill level receives the optimal correction, which is zero in the standard model and negative in the presence of present bias.[10] Corollary 1 also shows that incorporating present bias is not isomorphic to adopting some alternative set of welfare weights, as there is no set of finite weights $g(w)$ which result in strictly negative marginal tax rates on the lowest earning type. Importantly, this result depends on a strictly positive lower bound to the ability distribution, and it may be quite local, so it will be important to examine the quantitative behavior of marginal tax rates across low incomes in the more realistic setting of the numerical simulations in Section 3.

The second observation is that the corrective term $\mathcal{C}(w)$ is proportional to $g(w)$, reflecting the planner's greater concern for correcting biases among individuals whose consumption is highly valued at the margin. This force leads a more redistributive planner (higher $g(w_{min})$) to favor a larger correction at low incomes. This observation highlights one way in which a bias-correction rationale

[10]Similar logic would yield an analogous result at the top of the income distribution: when the skill distribution has an upper bound, the well-known "zero at the top" result becomes "negative at the top." The zero at the top result requires the tax authority to know the highest income-earning ability with certainty—an assumption many find implausible, raising some question about the practical relevance of this implication of the model.

differs from a conventional (atmospheric) externality correction: a positive work externality would not call for a larger correction on individuals with higher welfare weights.

The third observation, perhaps least obvious, demonstrates that the strength of the optimal corrective subsidy increases with the labor supply elasticity.

Corollary 2. *Let $T'_{opt}(z(w))$ denote the schedule of optimal marginal tax rates accounting for present bias correction, and let $T'_{redist}(z(w))$ denote the optimal tax schedule which would be chosen by a purely redistributive policymaker who does not regard present bias as a mistake, but who has the same redistributive preferences at the optimum. The optimal corrective subsidy $T'_{redist}(z(w)) - T'_{opt}(z(w))$ is increasing in the size of the labor supply elasticity $\zeta_\ell(w)$.*

This result illustrates an important divergence between a corrective rationale for labor subsidies, and the extensive margin rationale proposed in Saez (2002), where a higher intensive elasticity *reduces* the size of the optimal labor subsidy.[11] The corollary holds fixed redistributive preferences at the optimum in order to highlight the effect of the elasticity alone on the optimal correction.

Corollary 2 also highlights the distinction between correcting labor supply misoptimization and correcting a conventional externality, where the well-known "additivity" result (Sandmo, 1975) implies that the optimal marginal tax rate is the sum of the optimal redistributive tax rate and a Pigouvian correction, which is insensitive to the elasticity of behavior. Intuitively, the distinction in this model comes from the fact that the uninternalized benefit from additional labor effort (in the form of future income) is itself subject to the income tax—the corrective effect in dB above is $g(z)(1-\beta(z))(1-T'(z))$, whereas an externality would be weighted by neither by $g(z)$ nor by $1-T'(z)$. By implication, the magnitude of the corrective benefit is larger when the marginal tax rate is low, as is the case when labor supply elasticity is high.

This model is highly simplified in several respects. First, in keeping with the Diamond (1998) model, it does not include an extensive margin of labor supply, wherein workers jump discontinuously from zero earnings to discretely positive earnings in responsive to a reduction in average tax rates. Appendix B considers an extension with heterogeneous fixed costs of work, in the style of

[11] This implication of a corrective motive also appears consistent with some arguments related to the EITC: the Clinton Administration emphasized the increase in single mothers' labor earnings after the 1993 EITC expansion as evidence of success and reason for further expansion (`https://clintonwhitehouse4.archives.gov/WH/EOP/nec/html/FurmanEITC000207.html`)

Saez (2002), and shows that although that modification introduces a new term in the tax formula, it does not alter the basic insights in this section.

Second, because this model has only a single period of consumption, it rules out the possibility that workers could borrow against their anticipated future labor income, perhaps alleviating their present bias. This restriction may be reasonable for low income workers who have little liquid wealth, where present bias is concentrated empirically (see Figure 2 and the corresponding discussion in Section 3). However, to explore the implications of access to borrowing, Appendix C considers an extension in which workers also consume during the period of labor effort, and can borrow against their future labor earnings. This extension doesn't alter the qualitative insights of the model above—even when present-biased workers can borrow, a strictly positive corrective work subsidy is optimal, and as a special case, if present bias along the saving dimension is also corrected (either through saving and borrowing policies or self control devices) then the optimal work subsidy is exactly the same as in this baseline model.

2.3 Optimal income taxes with multidimensional labor supply

I now extend the model above to allow for multiple dimensions of labor effort. Feldstein (1999) showed that the standard Mirrlees model can be extended to allow for multidimensional labor choice, using the elasticity of taxable income (rather than the hours elasticity) to quantify the tax distortion. Here I adapt the model of multidimensional labor supply choice in Chetty (2009) (Section 4.1), which formalizes the insight in Feldstein (1999), to allow for present-biased workers.

Earnings z are all received at a single time, but they depend on a vector of labor supply choices $\ell = \{\ell_j\}_{j=1}^{J} \in \mathbb{R}_+^J$ which occur at different times (ordered by j) leading up to the time of compensation. Following Chetty (2009), workers have strictly convex labor disutilities $v(\ell) = \sum_j v_j(\ell_j)$, and effective wages differ across dimensions of effort, so that worker i's wage vector is $w(i) = \{w_j(i)\}_{j=1}^{J}$. This model is restrictive in that labor supply disutilities are separable, and compensation depends separably on the dimensions of labor supply.

An individual's taxable earnings are $z = w \cdot \ell = \sum_j w_j \ell_j$. I assume the agent faces a smooth, nonlinear income tax $T(z)$, and that labor supply adjustments are continuous in response to marginal perturbations of the tax function.[12] In addition to allowing for multidimensional labor supply, this

[12]This assumption implies that only local incentive constraints bind at the optimum; as noted in Saez (2001), this

subsection extends the unidimensional model from above by relaxing Assumption 1. In particular, utility may be concave (allowing for income effects on labor supply), and individuals may have heterogeneous β conditional on w.

Let $\tau(j)$ denote the temporal distance between the time of labor effort ℓ_j and the time of consumption. A worker with time-consistent preferences and a discount factor δ would then prefer a labor supply vector which maximizes

$$U(c, \ell) = -\sum_{j=1}^{J} \delta^{-\tau(j)} v_j(\ell_j) + u(c), \tag{13}$$

with $c = w \cdot \ell - T(w \cdot \ell)$. This is the modified version of the expression for normative utility in Equation (1). Assuming an interior solution, the labor supply choice vector ℓ^* desired by a worker with time-consistent preferences satisfies the set of first-order conditions $v_j'(\ell_j^*)/w_j = \delta^{\tau(j)} u'(z^* - T(z^*))(1 - T'(z^*))$ for $j = 1, \ldots J$, where $z^* = w \cdot \ell^*$.

A present-biased individual will select a different labor supply vector, in which each dimension of labor effort maximizes decision utility from the perspective of the agent at the time that effort is exerted. In this setting with multidimensional labor effort, there is no single decision utility function analogous to Equation (2), as the worker has a different decision utility function for each time at which labor effort decisions are made. The present-biased worker's optimization problem when choosing present biased labor supply ℓ_j^{pb} prior to compensation can be written as follows, letting $\hat{\ell}_{j'}$ denote the worker's assumption about their other non-j labor supply choices (some of which lie in the future), and $\hat{z}_{-j} = \sum_{j' \neq j} w_{j'} \hat{\ell}_{j'}$ their anticipated income from those choices:

$$\ell_j^{pb} = \arg\max_{\ell_j} \left\{ -v_j(\ell_j) + \beta \left[\sum_{j' > j} \delta^{\tau(j) - \tau(j')} v_{j'}(\ell_{j'}) + \delta^{\tau(j)} u\left(\hat{z}_{-j} + w_j \ell_j - T(\hat{z}_{-j} + w_j \ell_j)\right) \right] \right\} \tag{14}$$

As is evident from this expression, in a setting with multiple dimensions of effort, an assumption must be made about how a present-biased individual forecasts income from labor choices at times other than the one in consideration. For the purposes of the derivation below, I impose a simple assumption:

condition can be ensured (when there is no bunching) by imposing the single-crossing property assumption on model primitives.

Assumption 2 (Accuracy of labor supply forecasts)**.** *When individuals make labor supply choices ℓ_j^{pb} to maximize Equation* (14)*, they set $\hat{\ell}_{j'} = \ell_{j'}^{pb}$ for all j'.*

In words, their perceptions about labor supply in other periods (which are relevant for determining total income) are correct, and do not depend on the current choice of labor supply. This assumption, which is adopted to simplify the analysis to follow, might seem a reasonable benchmark for several reasons. First, if the individual in question is sophisticated about their present bias, they will correctly anticipate their present-biased behavior for future labor supply decisions. Second, even if the individual is naive, if all labor supply decisions occur either at the time of compensation or at some horizon τ^* in advance of compensation (i.e., if $\tau(j) \in \{0, \tau^*\}$ for all j) then Assumption 2 holds automatically. This is true because potential mispredictions of future income stem from incorrectly predicting *future present-biased behavior*. Since effort exerted contemporaneously with compensation is not distorted by present bias, even naive individuals will correctly anticipate their behavior on those dimensions. If the remaining advance effort all occurs at horizon τ^*, then there is no remaining present-biased effort to mispredict. Third, the distortion generated by errors in forecasted income are proportional to the curvature of the utility function u and the tax function T. To the extent that the tax function and the utility of consumption are close to linear over the relevant range of income variation, forecasting errors will negligibly affect labor supply choice.

Invoking Assumption 2, the vector of present-biased labor supply choices ℓ^{pb} can be simply characterized by the following system of first-order conditions, which suppress the individual-specific index (i) for readability:

$$\begin{cases} v_j'(\ell_j^{pb})/w_j = \delta^{\tau(j)} u'(z^{pb} - T(z^{pb}))(1 - T'(z^{pb})) & \text{if } \tau(j) = 0 \\ v_j'(\ell_j^{pb})/w_j = \beta\delta^{\tau(j)} u'(z^{pb} - T(z^{pb}))(1 - T'(z^{pb})) & \text{if } \tau(j) > 0 \end{cases} \quad \text{for } j = 1, \ldots J, \tag{15}$$

with $z^{pb} = \sum_j w_j \ell_j^{pb}$. (Note that the labor supply choices thus represent a fixed point.)

The first-order conditions in Equation (15) illustrate that a present-biased agent under-supplies labor along only those dimensions of effort which occur prior to the time of compensation.

As before, I assume the policymaker maximizes the (possibly transformed) sum of individual normative utilities,

$$\mathcal{W} = \int \alpha(i) G(U(c(i), \ell(i))) d\mu(i). \tag{16}$$

which additionally allows for type-dependent Pareto weights $\alpha(i)$, permitting the assumption below that welfare weights at the optimum are constant conditional on earnings.

The optimal income tax which maximizes Equation (16) subject to the government budget constraint in Equation (4) and the conditions for individual choice in Equation (15) in this generalized setting can be characterized by a first-order condition. To write this expression concisely, I define the following notation:

- $g(i) = \frac{\alpha(i)G'(U)u'(c(i))}{\lambda}$: marginal social welfare weights, equal to i's marginal social value of consumption at the optimum, normalized by the marginal value of public funds λ
- $H(z)$ and $h(z)$: the distribution and density of incomes under the optimal tax.
- $\varepsilon(i) = \frac{dz(i)}{d(1-T')} \frac{1-T'}{z(i)}$: the compensated elasticity of taxable income with respect to the marginal "keep rate" $1 - T'(z(i))$.[13]
- $\eta(i) = -\frac{dz(i)}{dT}(1 - T')$: the income effect, representing the change in net earnings due to a small tax credit.
- $\phi(i) = \sum_{\{j|\tau(j)>0\}} w_j \frac{d\ell_j(i)}{d(1-T')} \Big/ \frac{dz(i)}{d(1-T')}$: the share of earnings response owing to labor supply adjustments prior to compensation.
- $\overline{a}(z^*) = \mathbb{E}[a(i)|z(i) = z^*]$: for any type-dependent statistic $a(i)$, I denote the income-conditional average value using "upper bar" notation.
- $\Sigma^{(z^*)}_{a,b} = \frac{\mathrm{Cov}[a(i),b(i)|z(i)=z^*]}{\overline{a}(z^*)\overline{b}(z^*)}$: the income-conditional covariance between any two type-dependent variables a and b (computed at the optimum), normalized by their income-conditional means. I further define $\Sigma^{(z^*)}_{a,b,c} = \frac{\mathbb{E}[a(i)b(i)c(i)|z(i)=z^*]-\overline{a}(z^*)\overline{b}(z^*)\overline{c}(z^*)}{\overline{a}(z^*)\overline{b}(z^*)\overline{c}(z^*)}$, the extension of this covariance-based definition to three variables (and likewise for more than three).

[13] All elasticities and income effects incorporate both the direct change in earnings due to the tax reform, and any additional adjustments due to the change in marginal tax rate as earnings adjust, due to curvature in T. See Jacquet, Lehmann and Van der Linden (2013) for a discussion of such "circular processes."

These statistics can be used to characterize a necessary (first-order) condition for the optimal marginal tax rate using the calculus of variations approach, and they suggest that the insights highlighted in the previous simple case carry through to this more general setting.

Using the notation above, and employing Assumption 2, we have the following proposition characterizing the optimal income tax.

Proposition 2. *The first-order condition for the optimal tax at some income z (where T is twice differentiable) can be written*

$$\frac{T'(z)}{1-T'(z)} = \mathcal{A}(z)\mathcal{B}(z) - \mathcal{C}(z), \tag{17}$$

with

$$\mathcal{A}(z) = \frac{1}{\overline{\varepsilon}(z)h(z)z}, \tag{18}$$

$$\mathcal{B}(z) = \int_{s=z}^{\infty}\left[1-\overline{g}(s)-\overline{\eta}(s)\left(\frac{T'(s)}{1-T'(s)}+\overline{g}(s)\overline{\phi}(s)(1-\overline{\beta}(s))\left(1+\Sigma^{(s)}_{1-\beta,\eta,g,\phi}\right)\right)\right]dH(s), \tag{19}$$

$$\mathcal{C}(z) = \overline{g}(z)\overline{\phi}(z)(1-\overline{\beta}(z))\left(1+\Sigma^{(z)}_{1-\beta,\varepsilon,g,\phi}\right). \tag{20}$$

As is standard for such first-order conditions, the terms in (17) are endogenous, and therefore the expression is not dispositive with respect to comparative statics. Nevertheless, such expressions are a useful guide to the forces which govern the optimal tax, and the comparative statics suggested by the expression are consistent with the numerical simulations presented in Section 3.

The optimal tax characterization in Proposition 2 clarifies a number of useful insights. First, the parallel structure to Proposition 1 is apparent, suggesting that the basic lessons and intuitions from that simple case extend to the context with multidimensional labor supply and heterogeneity. In particular, as before, present bias tends to reduce marginal tax rates, particularly when the marginal social welfare weight or earnings elasticity is high.

The additional insights from the multidimensional context are also apparent from the differences between the terms $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in Proposition 1 and Proposition 2. First, $\mathcal{B}$ now contains an additional term, proportional to $\overline{\eta}$, capturing both the fiscal externality and behavioral welfare effects from the earning adjustments due to income effects. Intuitively, raising marginal tax rates at z takes resources away from those with higher incomes, who therefore raise their earnings due to income effects. This adjustment is beneficial both through a fiscal externality (the government raises more revenue) and through a present bias correction (working more generates a first-order benefit for the individual in question). Those beneficial effects justify higher marginal tax rates at each z.

A second difference in Proposition 2, relative to Proposition 1, is the presence of the Σ terms, which incorporate effects of multidimensional heterogeneity.[14] If the statistics β, ε, η, g, and ϕ are mutually orthogonal conditional on income (perhaps a plausible benchmark assumption, until empirical research studies such patterns of heterogeneity more closely) then the Σ terms are zero and can be ignored. Otherwise, they intuitively illustrate the directional effect of income-conditional correlations. For example, if $1 - \beta$ and ε are highly correlated at a particular income, that implies the individuals most responsive to marginal work subsidies are those who are most biased, making corrective subsidies a more powerful tool and thus magnifying the optimal correction term $\mathcal{C}$.

A final, and potentially important, insight from the more general Proposition 2 is the presence of ϕ. The corrective term $\mathcal{C}$ (and the corrective component of the income effect term in $\mathcal{B}$) is multiplied by $\overline{\phi}$, the *share* of earnings adjustment which comes along labor supply margins chosen prior to compensation. Intuitively, if work subsidies primarily cause adjustments in labor supply choices contemporaneous with earnings (small ϕ), then corrective subsidies are of little use, since those contemporaneous labor effort dimensions do not require correction. On the other hand, if the most elastic labor supply adjustments are through choices made in advance, then subsidies are a more powerful corrective policy instrument.

2.4 Extension: optimal tax timing

One question which lies beyond the domain of the model thus far is the issue of optimal tax timing. The model as described assumes that taxes are levied (and subsidies paid) at the time compensation is received. This stands in contrast to some work subsidies in practice, such as the Earned Income Tax Credit in the U.S., which remits work subsidies in the form of a tax refund when taxes are filed during the following year. A full dynamic model is beyond the scope of this paper, but in this section

[14] Although individuals are multidimensionally heterogeneous in this setting, through their vector wages $w_j(i)$ and their present bias $\beta(i)$, since the policymaker faces only a single observed outcome (earnings), and has only a single income tax policy instrument, issues of multidimensional screening do not render the problem intractable.

I consider a simple extension to the previous two period reduced model, in which I allow subsidies and taxes to be delayed relative to the time income is received. (Note that since the planner observes only income and not labor supply by assumption, paying subsidies before compensation is received is infeasible.) This extension uncovers a novel and perhaps surprising result: when there are multiple dimensions of labor effort, only some of which are exerted contemporaneously with compensation, the targeting power of work subsidies can be sharpened by *delaying* their payment until after compensation is received.

To incorporate the possibility of delayed taxes as simply as possible, I now assume an additional period of consumption, after all labor is performed and compensation is paid, during which some portion of work subsidies are paid (or taxes levied). This period can be thought of as the date at which tax refunds are paid at the end of the year. To abstract from issues of intertemporal smoothing and saving behavior which are not of central focus, here I assume consumers live hand-to-mouth and have utility quasilinear in consumption, with exponential discount rate $\delta = 1$. As a result, it is unimportant whether other consumption or labor occurs during the additional final period.

The parameter $\psi \in [0, 1]$, set by the policymaker, is used to denote the share of net taxes withheld from one's paycheck, or, equivalently, the share of the work subsidies paid at the end of the year rather than up-front.

In this modified setting the individual's normative utility function becomes

$$U(c, \ell) = -\sum_{j=1}^{J} v_j(\ell_j) + w \cdot \ell - (1 - \psi)T(w \cdot \ell) - \psi T(w \cdot \ell). \tag{21}$$

The social welfare function (16) and budget constraint (4) remain unchanged. However, the first-order conditions for labor choice change, since even contemporaneous dimensions of labor effort are perceived to have delayed consequences due to delayed taxes. At the time the individual chooses any ℓ_j prior to compensation (i.e., for which $\tau(j) > 0$), their decision utility objective function is

$$\max_{\ell_j} \left\{ -v_j(\ell_j) + \beta \left[-\sum_{j'>j} v_{j'}(\ell_{j'}) + \underbrace{\hat{z}_{-j} + w_j\ell_j - (1 - \psi)T(\hat{z}_{-j} + w_j\ell_j)}_{\text{consumption at time of compensation}} - \underbrace{\psi T(\hat{z}_{-j} + w_j\ell_j)}_{\text{delayed tax/subsidy}} \right] \right\}. \tag{22}$$

When choosing labor supply exerted at the time of effort, this objective instead becomes

$$\max_{\ell_j}\left\{-v_j(\ell_j)+\underbrace{\hat{z}_{-j}+w_j\ell_j-(1-\psi)T(\hat{z}_{-j}+w_j\ell_j)}_{\text{consumption at time of compensation}}-\beta\underbrace{\psi T(\hat{z}_{-j}+w_j\ell_j)}_{\text{delayed tax/subsidy}}\right\}. \tag{23}$$

Together these objectives lead to a set of first-order conditions for individual labor supply choice that modify those in Equation (15):

$$\begin{cases} v_j'(\ell_j^{pb})/w_j=(1-\psi(1-\beta))(1-T'(w\cdot\ell^{pb})) & \text{if } \tau(j)=0 \\ v_j'(\ell_j^{pb})/w_j=\beta(1-T'(w\cdot\ell^{pb})) & \text{if } \tau(j)>0. \end{cases} \tag{24}$$

The policymaker's problem is now to select the nonlinear function T and the delay parameter $\psi\in[0,1]$ that maximizes social welfare, subject to the budget constraint and individual optimization. This modification generates the counterintuitive result that if workers are present-biased, it may be optimal to *delay* corrective subsidies until after the time at which their labor is compensated.

Corollary 3. *The policy* (T,ψ) *which maximizes* (16) *subject to the government's budget constraint* (4) *and individual optimization* (24) *satisfies the following conditions:*

1. *if* $\phi(i)=1$ *for all* i*, then the optimal tax and overall welfare are not sensitive to the share of delayed taxes* ψ*;*
2. *if* $\phi(i)<1$ *and* $\beta(i)<1$ *for some* i *with positive earnings, then the optimal policy features* $\psi>0$*: a strictly positive share of taxes is delayed at the optimum; and*
3. *if at the optimum* $g(i)(1-\beta(i))\leq 1$ *for all* i*, then the optimal policy features* $\psi=1$*: the optimal tax is fully delayed.*

Part 1 of the corollary states that if all elastic dimensions of effort occur prior to compensation, then it does not matter whether corrective subsidies are delayed or included with one's labor compensation. This result follows from the insensitivity of advance effort to ψ, and contrasts with what is perhaps a natural intuition that to the extent marginal work subsidies (like the EITC) are intended to corrective present bias over labor effort, they should be paid as close to the time of work as possible—for example, by being included with one's paycheck. In this case the choice of optimal tax timing may depend on considerations beyond the scope of labor supply correction, such as enforcement, liquidity, or income smoothing and forced savings.[15]

Parts 2 and 3 of the corollary highlight the potential benefits of a delay. Note that when work is subsidized at the margin, contemporaneous dimensions of effort are distorted *upward*. This may be optimal, relating to the result in O'Donoghue and Rabin (2006) that internality corrections are optimal even when some consumers are unbiased, since the distortion is initially second-order relative to the first-order corrective benefit for biased consumers. By delaying the subsidy, there is no loss in corrective power, but this undesirable distortion to contemporaneous effort is dampened, since a portion of the distorting subsidy is deflated by β. As a result, when $T' < 0$ the optimal share of delayed taxes is strictly positive. Part 3 states the condition under which the subsidy should be fully delayed.

Taken together, the components of Corollary 3 suggest that the logic of present bias does not, in itself, call for work subsidies which are included with one's paycheck. Indeed there may be a positive benefit from a modest delay, since delayed subsidies result in less upward distortion of contemporaneous dimensions of effort which are not subject to present bias. Note, however, that this delay need not be long to reap these benefits—just sufficiently delayed that the perceived subsidy benefits are discounted by β at the time of compensation.

3 Numerical analysis

This section considers the quantitative implications of present bias for optimal income taxation. Propositions 1 and 2 highlight the key sufficient statistics which are necessary to evaluate whether a candidate income tax is optimal, which fall into the four categories: (1) behavioral responses which have previously been estimated in the labor supply literature: the compensated elasticity of taxable income with respect to the marginal keep rate (denoted ε) and the income effect (denoted η); (2) the present bias parameter β, which has been estimated extensively in the behavioral economics literature; and (3) marginal social welfare weights g, which are often imposed exogenously or calculated from a utility function with some assumed curvature, but which have also been estimated in the inverse optimum literature (see Section 3.3).

[15]For an example of role of forced savings, see the model in the online appendix to Jones (2012).

Although these statistics are sufficient to evaluate the optimality of a status quo income tax, a structural model of behavior is necessary to estimate the optimal tax away from the observed equilibrium. To focus on the implications of present bias for the optimal tax, and to make these simulations comparable to existing results in the static optimal taxation literature, I calibrate these simulations around the unidimensional model in Section 2.1 with unidimensional heterogeneity. Specifically, I assume use the following specification for decision utility

$$U_i = -\frac{(z/w(i))^{1+k}}{1+k} + \beta(i)\left(z - T(z)\right). \tag{25}$$

Although this specification maps to a single dimension of labor supply choice consistent with the model in Section 2.2, it is isomorphic to multiple dimensions of labor supply with a constant elasticity of taxable income and $\phi = 1$, in which all earnings responses to tax changes are through labor supply choices made in advance of compensation.

I assume a fixed share of the population is disabled and has $w = 0$. Their income benefit is added as a component of the revenue requirement in the policymaker's budget constraint. (See Appendix E for details.) I further assume that present bias β is homogeneous conditional on w, though it may vary with w. The parameter k controls the labor supply elasticity.[16]

In this setup, the required structural parameters are the income elasticity parameter k, the ability distribution $F(w)$, and the mapping between w and present bias and welfare weights, $\beta(w)$ and $g(w)$. I use a baseline value of $1/k = 0.3$, close to the preferred intensive margin elasticity from Chetty (2012) of 0.33, and I also provide results for $1/k = 0.4$ and $1/k = 0.2$. The calibration of β is the primary new challenge for these simulations, and therefore I discuss that calibration in detail in the next subsection. After choosing structural values $\beta(w)$ and k, the structural ability distribution can be calibrated in the usual manner, by inverting the first-order condition for individual optimization to find the ability distribution which would generate the observed income distribution under the prevailing tax code. Details of the income distribution and status quo tax schedule calibration are discussed in Appendix E.

Finally, I employ a reduced-form representation of declining marginal social welfare weights $g(w)$, which can be interpreted either as declining α Pareto weights, or as arising due to concavity

[16] For comparability to Section 2, $\varepsilon(i) = \frac{1}{k + \frac{zT''}{1-T'}}$, so that at points where the income tax is locally linear, $\varepsilon = 1/k$.

of G (and therefore declining G') at the optimum. I adopt the conventional assumption that weights are everywhere declining as consumption increases; the exact patterns for the simulations are discussed below. For the planner's budget constraint, I impose an exogenous government revenue requirement of $5,000 per capita.

3.1 Calibrating present bias

Evidence of present bias over labor supply falls into three categories: preference reversals, demand for commitment devices, and discount rates over delayed compensation that are "too high."[17]

An example of a study which documents all three patterns in a field setting with high stakes incentivized labor supply decisions is Kaur, Kremer and Mullainathan (2015). Studying a year-long field experiment with payment contracts in a data-entry center, that paper finds that workers increase effort on or near payday, relative to their effort at a one-week horizon, suggesting substantial short-run discount rates. Workers also strictly prefer dominated commitment contracts, with demand concentrated among those exhibiting the highest discount rates over labor effort.

Also in a field setting, Mas and Pallais (2017) study workers' willingness to pay (through reduced wages) for various dimensions of job flexibility among job applicants at a U.S. call center. They find that about half of workers have a strictly negative willingness to pay for a job with a flexible number of weekly work hours, consistent with a desire for labor supply commitment.

Table 1 presents estimates of β from several other such papers. The checkmarks highlight two desirable features. The column labeled "low income/EITC" identifies papers whose subjects are drawn from a population of low earners, and in some cases EITC recipients, in the U.S. Such studies are useful for two reasons. First, they mitigate concerns about external validity, as their subjects resemble the population of particular interest for the simulations in this paper—those who are subject to negative marginal tax rates in practice. Second, estimates from low income populations mitigate the concern that monetary rewards are problematic for the estimation of present bias, since subjects can use their own funds to replicate (or undo) experimental variation in payoffs. Low income subjects are more likely to face liquidity constraints which prevent such arbitrage, perhaps explaining the substantial measured present bias even over monetary payoffs in

[17]For a discussion of why any measurable short-run discounting is evidence of present bias, see O'Donoghue and Rabin (2015), "Lesson #3."

those studies.[18]

The column labeled "effort" identifies studies which estimate β using intertemporal tradeoffs over effort tasks or labor supply, rather than money. These studies are particularly informative both because they avoid the shortcomings of monetary payoffs, and because the focus of this paper is labor supply, so to the extent that bias varies across domains, these studies identify the parameter of interest. All studies find estimates of β meaningfully (and statistically significantly) below one.

Figure 2 plots these values of β across the incomes in each study and, when possible, plots the relationship between β and income *within* studies. These calculations, the details of which are reported in Appendix D, are in some cases quite rough, and none of these studies was written with the primary goal of estimating present bias across incomes. Indeed, one implication of this paper is that the covariation of present bias with income is worthy of additional empirical research. Nevertheless, as shown in Figure 2, there is a strong positive correlation between β and income, both across studies and (when reported) within individual studies.

This relationship is consistent with a number of possible explanations. First, theory predicts that present-biased individuals endogenously exert less effort and therefore have lower earnings. Second, present bias likely reduces longer term human capital investments, leading to an inverse relationship between the bias wedge and underlying ability. Third, circumstances of material scarcity might *cause* greater present-bias (Mullainathan and Shafir, 2013). I remain agnostic about the mechanism for the relationship, effectively assuming that the plot in Figure 2 indicates a stable type-specific level of bias as a function of underlying ability.

Figure 2 also displays the best fit line estimated from the relationship between β and income across the studies reported in Table 1, excluding Kaur, Kremer and Mullainathan (2015) and Goda et al. (2015) (see Appendix D for details). I use this relationship for the structural calibration of β across ability in the simulations below, restricting to a maximum value of $\beta = 1$.

3.2 Optimal tax simulations

This section demonstrates the key finding previewed in Figure 1: if redistributive tastes are modest, then substantial negative marginal tax rates may be optimal at low incomes. More generally, present

[18] See Carvalho, Meier and Wang (2016) for evidence that liquidity constraints generate measurable present bias over monetary payoffs.

bias tends to lower optimal marginal tax rates across a large range of incomes.

For ease of computation and transparency, I encode redistributive preferences directly through declining marginal social welfare weights $g(w)$. For the baseline set of modest redistributive preferences, I select g-weights such that the lowest earners receive a weight 10% more than the median household, and top earners weighted by 40% less than the median, linearly interpolated between percentiles 0, 50, and 100. These weights are substantially less redistributive than those conventionally assumed in the optimal taxation literature (e.g., logarithmic utility over consumption). However there is other evidence that existing tax policies embodies more modest redistributive tastes than is conventionally assumed in that literature (see, for example, the inverse optimum literature cited in Section).

To explore the sensitivity of the optimal tax schedule to different assumptions, Figure 3 displays the schedule of simulated optimal marginal tax rates under six different specifications. Panel (a) reproduces the baseline calibration in Figure 1. Panel (b) displays optimal tax rates under logarithmic redistributive preferences, so that $g(w) = c(w)^{-1}$ at the optimum. Panels (c) and (d) use the baseline set of redistributive weights, with higher and lower labor supply elasticities. Panel (e) plots optimal tax rates assuming β is homogeneous across the income distribution, with a value equal to the population average in the baseline simulation, $\beta = 0.80$. Panel (f) assumes that one half of the labor supply response to tax changes comes through dimensions of labor effort which are not subject to present bias—i.e., in the notation of Proposition 2, $\phi = 0.5$.

These simulations highlight some key lessons for tax policy with present-biased workers. First, as discussed in the introduction, present bias tends to depress optimal marginal tax rates. If redistributive preferences are modest, like in the baseline specification, then marginal tax rates may be negative, and quantitatively similar to those which exist under the EITC. Second, as the log redistributive tastes case illustrates, if redistributive preferences are strong, marginal tax rates remain positive throughout the income distribution, although present bias still tends to reduce tax rates below what is optimal if the population is unbiased. This result may seem surprising in light of the fact that the correction term in Propositions 1 and 2 are weighted by the marginal social welfare weight, which is higher for lower earners under stronger redistributive preferences. That stronger corrective motive is outweighed, however, by the stronger desire to redistribute *across* low earners under log preferences, reflected by the high level of marginal tax rates under the rational optimum in the log case. These higher marginal tax rates are used to fund a larger lump sum (or basic income), equal to \$21,734, compared to \$4,535 in the less redistributive baseline case.

The third lesson from Figure 3 relates to the labor supply elasticity. A higher elasticity reduces optimal marginal tax rates in the present-biased optimum. In fact an elasticity of 0.4, higher than baseline but still well within the range of some empirical estimates, particularly from the macro literature (see Chetty (2012)) generates optimal marginal tax rates as low as -30%.[19] This result is the numerical illustration of Corollary 2, which demonstrates that a high labor supply elasticity magnifies the optimal corrective subsidy. This highlights the divergence between this comparative static and the effect of higher intensive margin elasticities for the the EITC-like subsidies in Saez (2002), where a higher intensive margin labor supply *reduces* the magnitude of marginal tax rates on the poorest workers.

On the other hand, if elasticities are low, as in Panel (c) of Figure 3, then present bias does not generate marginal work subsidies at all. Of course, although these simulations use a constant value for the elasticity, the elasticity may vary with income in practice. Thus if individuals at the bottom of the distribution are particularly elastic, subsidies may be justified even if elasticities are fairly low at higher points in the income distribution.

The primary lesson from Panel (e) is that the substantially negative marginal tax rates at low incomes in the baseline specification are mediated by the concentrated present bias in that portion of the income distribution. When instead present bias is homogeneous across incomes, as in Panel (e), the reduction in marginal tax rates is more evenly distributed across incomes.

3.3 Inverse optimum exercise

As shown in the previous section, one implication of present bias is that substantially negative marginal tax rates are optimal under some calibration specifications. This finding provides a potential qualitative justification for policies like the Earned Income Tax Credit. A complementary strategy to study this question is to ask what sorts of redistributive preferences appear consistent with the existing EITC, with and without accounting for present bias. This section adopts that

[19] In this simulation with quasilinear utility and fixed (finite) welfare weights at the bottom, it is possible for consumption to be negative at the optimum for some individuals. I therefore impose an additional constraint on the optimal tax that all incomes must be nonnegative. That constraint binds only in the high elasticity specification, generating high marginal tax rates at the very bottom.

approach, inverting the optimal policy simulation by taking the existing tax schedule as given and computing the redistributive preferences with which those policies are consistent. Since a primary goal of this exercise is to understand the potential role of present bias in rationalizing negative marginal tax rates, for this section the "existing policy" I consider will be the tax schedule facing prime-age households with children—i.e., those who qualify for the EITC when incomes are low.

This "inverse optimum" strategy is discussed in Saez (2001) (see p. 221 and footnote 21), and is implemented by Bourguignon and Spadaro (2012) for European countries and by Hendren (2019) and Lockwood and Weinzierl (2016) in the U.S. The inverse optimum procedure provides a reduced-form way to check whether existing policy generates redistributive weights which appear reasonable. Of course, the definition of "reasonable" is itself subjective, but two features are commonly thought to be sensible requirements in the optimal taxation literature, which typically takes a utilitarian perspective: welfare weights are positive at all incomes (Pareto efficiency) and weights are declining with income (redistribution toward lower skilled individuals). (For a discussion of the use of welfare weights for optimal taxation models with non-utilitarian objectives, see Saez and Stantcheva (2016).)

Figure 4 shows the key result: the implicit welfare weights consistent with the existing U.S. income tax and EITC under the usual assumption of perfect optimization exhibit a robust "unreasonable" feature: weights rise substantially with income across low earnings. Taken at face value, the weights suggest that current policy implicitly places greater value on a marginal dollar for middle earners than on a dollar in the hands of the poorest EITC-receiving households (typically working single mothers). However, if we instead perform this inversion exercise assuming misoptimization due to present bias, calibrated according to Figure 2, the resulting weights are monotonically declining.

These results capture the sense in which the EITC is difficult to reconcile with widespread normative assumptions under the conventional model. However under a calibrated degree of present bias, the EITC is consistent with conventional normative assumptions, since the dashed line in Figure 4 does entail a preference for redistribution across low earners, albeit a weak one.

4 Discussion

The preceding results impose a number of policy restrictions, and abstract from a number of potential complications, which are worth mentioning as potential topics of future research.

First, the treatment of optimal tax timing is limited by allowing only a single policy instrument: a nonlinear income tax. This restriction is useful for determining constrained optimal policies that avoid the complicated or impractical history-dependent features, which generally arise from full dynamic models in the mechanism design tradition. On the other hand, this constraint on policy instruments rules out some instruments which might be feasible and useful for targeting present-biased individuals, including finer adjustments to the current schedule of tax collections and refund payments. This also rules out other (potentially complementary) dynamic non-tax policy instruments, such as time limits for welfare benefits or in-kind transfers (see Fang and Silverman (2004) for a model of the optimal design of such policies with present biased workers).

Although a full dynamic model is beyond the scope of this paper, the results above do provide some guidance about the potential implications for different timing structures. One application of interest is the timing of EITC payments. The current EITC is paid in aggregate at the end of the tax year—a structure that has generated mixed reviews. On one hand, spreading the payment across more frequent installments would help smooth consumption across the year and potentially alleviate liquidity constraints. On the other hand, the lump sum nature of the current EITC provides a short-run forced savings mechanism, and recipients often use the large annual payment to invest in durable goods. Indeed, some anecdotal evidence suggests EITC recipients do not *want* to receive early EITC distributions (Halpern-Meekin et al., 2015), a finding consistent with the very low uptake of the "Advance EITC" option, which allowed for more frequent payments. (See Romich and Weisner (2000) for a discussion, and Jones (2010) for experimental evidence of low desire for the Advance EITC, although see also Andrade et al. (2017) for evidence that a modified advance program reduced food insecurity.) Corollary 3 in Section 2.4 suggests that it may be beneficial to levy taxes, or pay tax refunds, with a short lag—perhaps on the order of one month.

A second limitation is the omission of an explicit role for human capital. For the sake of simplicity and transparency, and to generate results comparable to the existing literature exploring the (sub)optimality of negative marginal tax rates, I restrict consideration to a static distribution of ability. In a sense, the model in this paper still allows for a reduced-form relationship between human capital and present bias by calibrating bias conditional on income. There is also some empirical evidence suggesting that the role of work subsidies in promoting human capital acquisition may be limited: individuals who randomly receive work subsidies do not experience *persistent* increases in income relative to those who do not (Card and Hyslop, 2005, 2009)—a finding inconsistent with the notion that such subsidies raise human capital via on-the-job training effects. Yet incorporating a fuller model of human capital acquisition and present bias remains a fertile area for additional research.

A third limitation is the assumption of perfectly competitive labor markets. Like much of the optimal taxation literature, I assume that workers are employed in a perfectly competitive labor market, and that labor demand is infinitely elastic. This assumption has been questioned by Rothstein (2010), who argues that the incidence of work subsidies falls partly on employers. Finitely elastic labor demand undermines the argument for an EITC relative to guaranteed minimum income with high marginal tax rates, since the latter regime tends to reduce labor supply, raising wages and total transfers from employers to employees. Also in this vein, Kroft et al. (Forthcoming) incorporate endogenous wages and unemployment (not all job seekers find jobs) using a sufficient statistics approach; their findings favor a negative income tax (rather than an EITC with negative marginal tax rates at low incomes) in a discrete model in the style of Saez (2002).

Finally, this model assumes that present bias is fixed at the individual level. In practice, biases may be mutable. One possibility, for example, is that exposure to the costs of present bias might lead individuals to improve their self control—a channel which would undermine the optimality of corrective subsidies that dampen such exposure. Another possibility, explored by Mullainathan, Schwartzstein and Congdon (2012) and Mullainathan and Shafir (2013), is that the conditions of poverty exacerbate behavioral biases. As with the case of human capital accumulation, this model would predict that temporary work subsidies should have persistent impacts on labor supply, inconsistent with the findings by Card and Hyslop cited above. Additionally, work by Carvalho, Meier and Wang (2016) suggests that although liquidity constraints exacerbate measured present bias over monetary payments, they do not affect present bias over labor effort—consistent with a stable degree of structural bias. Still, optimal taxation with endogenous biases is a promising avenue for further exploration.

5 Conclusion

As the study of optimal taxation begins to account for imperfect rationality and behavioral biases, a critical challenge is to quantify misoptimization accurately. This paper focuses on a particularly robust and well calibrated source of misoptimization—present bias—which, recent evidence suggests, generates substantial distortions to labor supply.

A model of optimal taxation with present bias generates new theoretical implications, including a "negative at the bottom" result and a surprising implication for optimal tax timing: if workers are present-biased and face multiple dimensions of labor choice, then it is beneficial to pay work subsidies with a delay.

A compilation of existing estimates of present bias provides strong evidence of such bias; more suggestive evidence indicates bias is heavily concentrated at low incomes. Although estimates of misoptimization will surely continue to improve, the consistency of results across methodologies provides some hope that misoptimization can be estimated with sufficient precision to provide clear guidance for policy design.

The implications for optimal tax policy depend on one's view of optimal redistribution. If redistributive tastes are modest, like those in the baseline calibrations in this paper, then optimal marginal tax rates may be substantially negative at low incomes—in the range of those generated by the EITC for households with children. In that case the policy implications of these results are clear: the existing EITC need not be reduced or greatly reformed—indeed, social welfare would rise if the EITC were extended to workers without children, and made more salient. On the other hand, if redistributive preferences are strong, in line with the welfare weights often assumed by optimal tax theorists (such as logarithmic utility of consumption) then negative marginal tax rates at low incomes appear to be suboptimal even in the context of present-biased workers. As a descriptive matter, the existing EITC schedule appears inconsistent with the conventional normative assumption of decreasing marginal social welfare weights in the standard model; a calibrated model of present bias resolves this inconsistency and gives rise to declining weights.

References

Allcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky. 2019. "Regressive Sin Taxes, with an Application to the Optimal Soda Tax." *Quarterly Journal of Economics*, 134(3): 1557–1626.

Andrade, Flavia Cristina Drumond, Karen Z Kramer, J Kale Monk, Andrew Jordan Greenlee, and Ruby Mendenhall. 2017. "Financial Stress and Depressive Symptoms: The Impact of an Intervention of the Chicago Earned Income Tax Periodic Payment." *Public Health*, 153: 99–102.

Augenblick, Ned, and Matthew Rabin. 2018. "An Experiment on Time Preference and Misprediction in Unpleasant Tasks." *Review of Economic Studies*, 86(3): 941–975.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working Over Time: Dynamic Inconsistency in Real Effort Tasks." *The Quarterly Journal of Economics*, 130(3): 1067–1115.

Beaudry, Paul, Charles Blackorby, and Dezsö Szalay. 2009. "Taxes and Employment Subsidies in Optimal Redistribution Programs." *American Economic Review*, 99(1): 216–242.

Bernheim, B Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *The Quarterly Journal of Economics*, 124(1): 51–104.

Besley, Timothy, and Stephen Coate. 1992. "Workfare Versus Welfare: Incentive Arguments for Work Requirements in Poverty-Alleviation Programs." *American Economic Review*, 82(1): 249–261.

Besley, Timothy, and Stephen Coate. 1995. "The Design of Income Maintenance Programmes." *Review of Economic Studies*, 62(2): 187–221.

Blundell, Richard, and Andrew Shephard. 2011. "Employment, Hours of Work and the Optimal Taxation of Low-Income Families." *Review of Economic Studies*, 79(2): 481–510.

Bourguignon, François, and Amedeo Spadaro. 2012. "Tax-Benefit Revealed Social Preferences." *The Journal of Economic Inequality*, 10(1): 75–108.

Card, David, and Dean R Hyslop. 2005. "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers." *Econometrica*, 73(6): 1723–1770.

Card, David, and Dean R Hyslop. 2009. "The Dynamic Effects of an Earnings Subsidy for Long-Term Welfare Recipients: Evidence From the Self Sufficiency Project Applicant Experiment." *Journal of Econometrics*, 153(1): 1–20.

Carvalho, Leandro S, Stephan Meier, and Stephanie W Wang. 2016. "Poverty and Economic Decision-Making: Evidence From Changes in Financial Resources at Payday." *American Economic Review*, 106(2): 260–284.

Chetty, Raj. 2009. "Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance." *American Economic Journal: Economic Policy*, 1(2): 31–52.

Chetty, Raj. 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica*, 80(3): 969–1018.

Chetty, Raj, John N Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review*, 103(7): 2683–2721.

Choné, Philippe, and Guy Laroque. 2010. "Negative Marginal Tax Rates and Heterogeneity." *American Economic Review*, 100(5): 2532–2547.

Cuff, Katherine. 2000. "Optimality of Workfare with Heterogeneous Preferences." *Canadian Journal of Economics/Revue canadienne d'économique*, 33(1): 149–174.

Dahl, Gordon B, and Lance Lochner. 2012. "The Impact of Family Income on Child Achievement: Evidence From the Earned Income Tax Credit." *American Economic Review*, 102(5): 1927–1956.

DellaVigna, Stefano, and M Daniele Paserman. 2005. "Job Search and Impatience." *Journal of Labor Economics*, 23(3): 527–588.

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F Schmieder. 2017. "Reference-Dependent Job Search: Evidence from Hungary." *The Quarterly Journal of Economics*, 132(4): 1969–2018.

Diamond, Peter. 1980. "Income Taxation with Fixed Hours of Work." *Journal of Public Economics*, 13(1): 101–110.

Diamond, Peter A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review*, 88(1): 83–95.

Drenik, Andres, and Ricardo Perez-Truglia. 2018. "Sympathy for the Diligent and the Demand for Workfare." *Journal of Economic Behavior & Organization*, 153: 77–102.

Fang, Hanming, and Dan Silverman. 2004. "On The Compassion of Time-Limited Welfare Programs." *Journal of Public Economics*, 88(7-8): 1445–1470.

Fang, Hanming, and Dan Silverman. 2009. "Time-Inconsistency and Welfare Program Participation: Evidence From the NLSY." *International Economic Review*, 50(4): 1043–1077.

Farhi, Emmanuel, and Xavier Gabaix. 2019. "Optimal Taxation with Behavioral Agents." *NBER Working Paper No. 21524*.

Feldstein, Martin. 1999. "Tax Avoidance and the Deadweight Loss of the Income Tax." *Review of Economics and Statistics*, 81(4): 674–680.

Fleurbaey, Marc, and Francois Maniquet. 2006. "Fair Income Tax." *Review of Economic Studies*, 73(1): 55–83.

Gerritsen, Aart. 2016. "Optimal Taxation When People Do Not Maximize Well-Being." *Journal of Public Economics*, 144: 122–139.

Goda, Gopi Shah, Matthew R Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff. 2015. "The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings." *NBER Working Paper No. 21482*.

Halpern-Meekin, Sarah, Kathryn Edin, Laura Tach, and Jennifer Sykes. 2015. *It's Not Like I'm Poor: How Working Families Make Ends Meet in a Post-Welfare World.* University of California Press.

Hansen, Emanuel. 2018. "Optimal Income Taxation With Labor Supply Responses at Two Margins: When Is an Earned Income Tax Credit Optimal?" *Working Paper.*

Hellwig, Martin F. 2007. "A Contribution to the Theory of Optimal Utilitarian Income Taxation." *Journal of Public Economics*, 91(7): 1449–1477.

Hendren, Nathaniel. 2019. "Efficient Welfare Weights." *NBER Working Paper No. 20351.*

Herbst, Chris M. 2013. "Welfare Reform and the Subjective Well-Being of Single Mothers." *Journal of Population Economics*, 26(1): 203–238.

Ifcher, John. 2010. "The Happiness of Single Mothers After Welfare Reform." *The BE Journal of Economic Analysis & Policy*, 11(1): 1–27.

Ifcher, John, and Homa Zarghamee. 2014. "The Happiness of Single Mothers: Evidence from the General Social Survey." *Journal of Happiness Studies*, 15(5): 1219–1238.

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden. 2013. "Optimal Redistributive Taxation With Both Extensive and Intensive Responses." *Journal of Economic Theory*, 148(5): 1770–1805.

Jones, Damon. 2010. "Information, Preferences, and Public Benefit Participation: Experimental Evidence from the Advance EITC and 401(k) Savings." *American Economic Journal: Applied Economics*, 2(2): 147–163.

Jones, Damon. 2012. "Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds." *American Economic Journal: Economic Policy*, 4(1): 158–185.

Jones, Damon, and Aprajit Mahajan. 2015. "Time-inconsistency and Saving: Experimental Evidence from Low-Income Tax Filers." *NBER Working Paper No. 21272.*

Kanbur, Ravi, Jukka Pirttilä, and Matti Tuomala. 2006. "Non-Welfarist Optimal Taxation and Behavioural Public Economics." *Journal of Economic Surveys*, 20(5): 849–868.

Kanbur, Ravi, Michael Keen, and Matti Tuomala. 1994. "Optimal Non-Linear Income Taxation for the Alleviation of Income-Poverty." *European Economic Review*, 38(8): 1613–1632.

Katz, Lawrence F, and Alan B Krueger. 2019. "Understanding Trends in Alternative Work Arrangements in the United States." *NBER Working Paper No. 25425*.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-Control at Work." *Journal of Political Economy*, 123(6): 1227–1277.

Kroft, Kory, Kavan J Kucko, Etienne Lehmann, and Johannes F Schmieder. Forthcoming. "Optimal Income Taxation with Unemployment and Wage Responses: A Sufficient Statistics Approach." *American Economic Journal: Economic Policy*.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *The Quarterly Journal of Economics*, 112(2): 443–478.

Laibson, David, Peter Maxted, Andrea Repetto, and Jeremy Tobacman. 2015. "Estimating Discount Functions With Consumption Choices Over the Lifecycle." *Harvard University Working Paper*.

Lockwood, Benjamin B., and Matthew Weinzierl. 2015. "De Gustibus Non Est Taxandum: Heterogeneity in Preferences and Optimal Redistribution." *Journal of Public Economics*, 124: 74–80.

Lockwood, Benjamin B., and Matthew Weinzierl. 2016. "Positive and Normative Judgments Implicit in U.S. Tax Policy, and the Costs of Unequal Growth and Recessions." *Journal of Monetary Economics*, 77: 30–47.

Martinez, Seung-Keun, Stephan Meier, and Charles Sprenger. 2017. "Procrastination in the field: Evidence from tax filing." *Working paper*.

Mas, Alexandre, and Amanda Pallais. 2017. "Valuing Alternative Work Arrangements." *American Economic Review*, 107(12): 3722–3759.

Meier, Stephan, and Charles D Sprenger. 2015. "Temporal Stability of Time Preferences." *Review of Economics and Statistics*, 97(2): 273–286.

Meyer, Bruce D, and James X Sullivan. 2008. "Changes in the Consumption, Income, and Well-Being of Single Mother Headed Families." *American Economic Review*, 98(5): 2221–2241.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, 38(2): 175–208.

Moffitt, Robert. 2006. "Welfare Work Requirements With Paternalistic Government Preferences." *The Economic Journal*, 116(515): F441–F458.

Moser, Christian, and Pedro Olea de Souza e Silva. 2019. "Optimal Paternalistic Savings Policies." *Federal Reserve Bank of Minneapolis Working Paper.*

Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity: Why Having Too Little Means So Much.* Henry Holt and Company.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J Congdon. 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Annual Review of Economics*, 4: 511–540.

O'Donoghue, Ted, and Matthew Rabin. 2006. "Optimal Sin Taxes." *Journal of Public Economics*, 90(10): 1825–1849.

O'Donoghue, Ted, and Matthew Rabin. 2015. "Present Bias: Lessons Learned and To Be Learned." *American Economic Review: Papers & Proceedings*, 105(5): 273–279.

Paserman, M Daniele. 2008. "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation." *The Economic Journal*, 118(531): 1418–1452.

Romich, Jennifer L, and Thomas Weisner. 2000. "How Families View and Use the EITC: Advance Payment Versus Lump Sum Delivery." *National Tax Journal*, 53(4): 1245–1265.

Rothstein, Jesse. 2010. "Is the EITC as Good as an NIT? Conditional Cash Transfers and Tax Incidence." *American Economic Journal: Economic Policy*, 2(1): 177–208.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2015. "Integrated Public Use Microdata Series: Version 6.0 [dataset]."

Saez, Emmanuel. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies*, 68(1): 205–229.

Saez, Emmanuel. 2002. "Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses." *The Quarterly Journal of Economics*, 117(3): 1039–1073.

Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized Social Marginal Welfare Weights for Optimal Tax Theory." *American Economic Review*, 106(1): 24–45.

Sandmo, Agnar. 1975. "Optimal Taxation in the Presence of Externalities." *The Swedish Journal of Economics*, 77(1): 86–98.

Seade, Jesus. 1977. "On the Shape of Optimal Tax Schedules." *Journal of Public Economics*, 7(2): 203–235.

Seade, Jesus. 1982. "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies*, 49(4): 637–643.

Seierstad, Atle, and Knut Sydsaeter. 1977. "Sufficient Conditions in Optimal Control Theory." *International Economic Review*, 367–391.

Spinnewijn, Johannes. 2015. "Unemployed but Optimistic: Optimal Insurance Design With Biased Beliefs." *Journal of the European Economic Association*, 13(1): 130–167.

Steuerle, C. Eugene. 2015. "Congressional Testimony: Marginal Tax Rates and 21st Century Social Welfare Reform."

Stiglitz, Joseph E. 1982. "Self-Selection and Pareto Efficient Taxation." *Journal of Public Economics*, 17(2): 213–240.

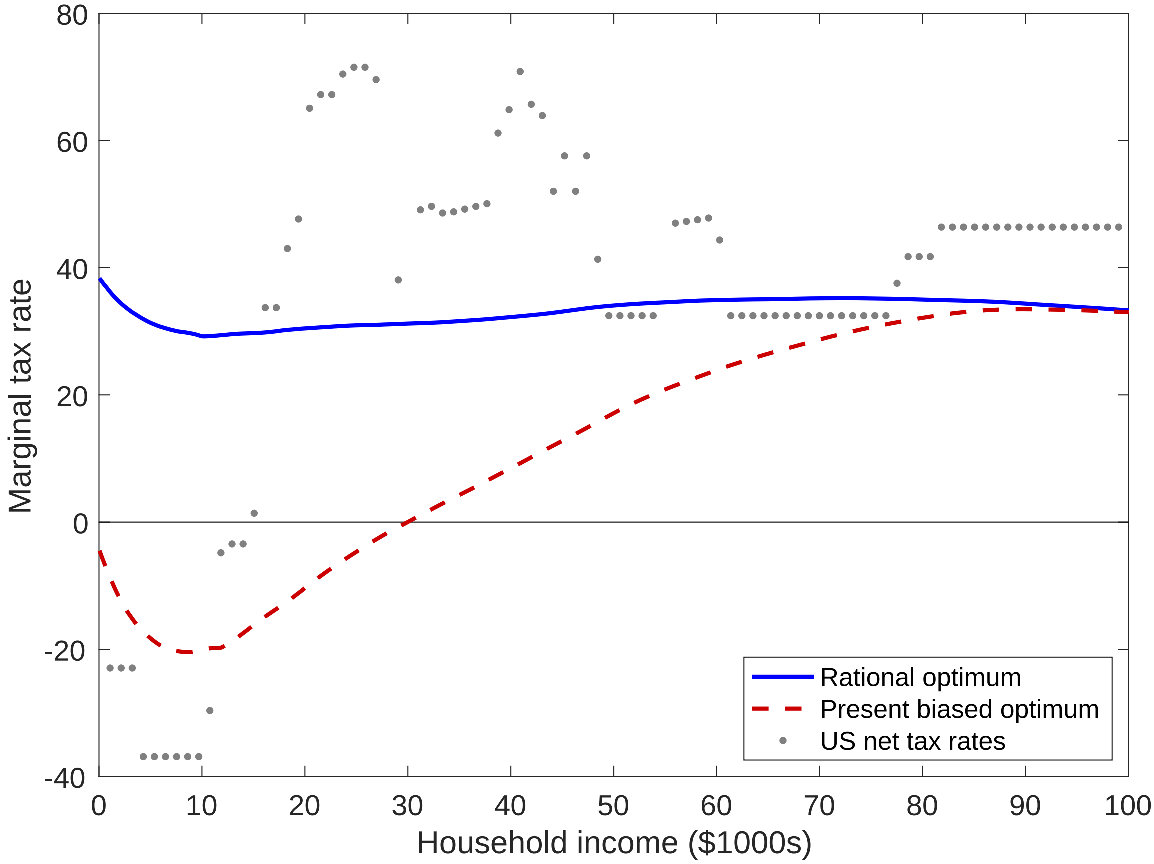
Tables and Figures

Table 1: Empirical Evidence of Present Bias

	implied β	low income/EITC	effort
Augenblick, Niederle and Sprenger (2015)	0.89		✓
Augenblick and Rabin (2018)	0.83		✓
DellaVigna et al. (2017)	0.58	✓	
Fang and Silverman (2009)	0.34	✓	✓
Goda et al. (2015)	1.01		
Jones (2010)	0.34	✓	
Kaur, Kremer and Mullainathan (2015)	0.71		✓
Martinez, Meier and Sprenger (2017)	0.92	✓	
Meier and Sprenger (2015)	0.69	✓	
Laibson, Repetto and Tobacman (2015)	0.50	✓	
Paserman (2008)	0.65	✓	

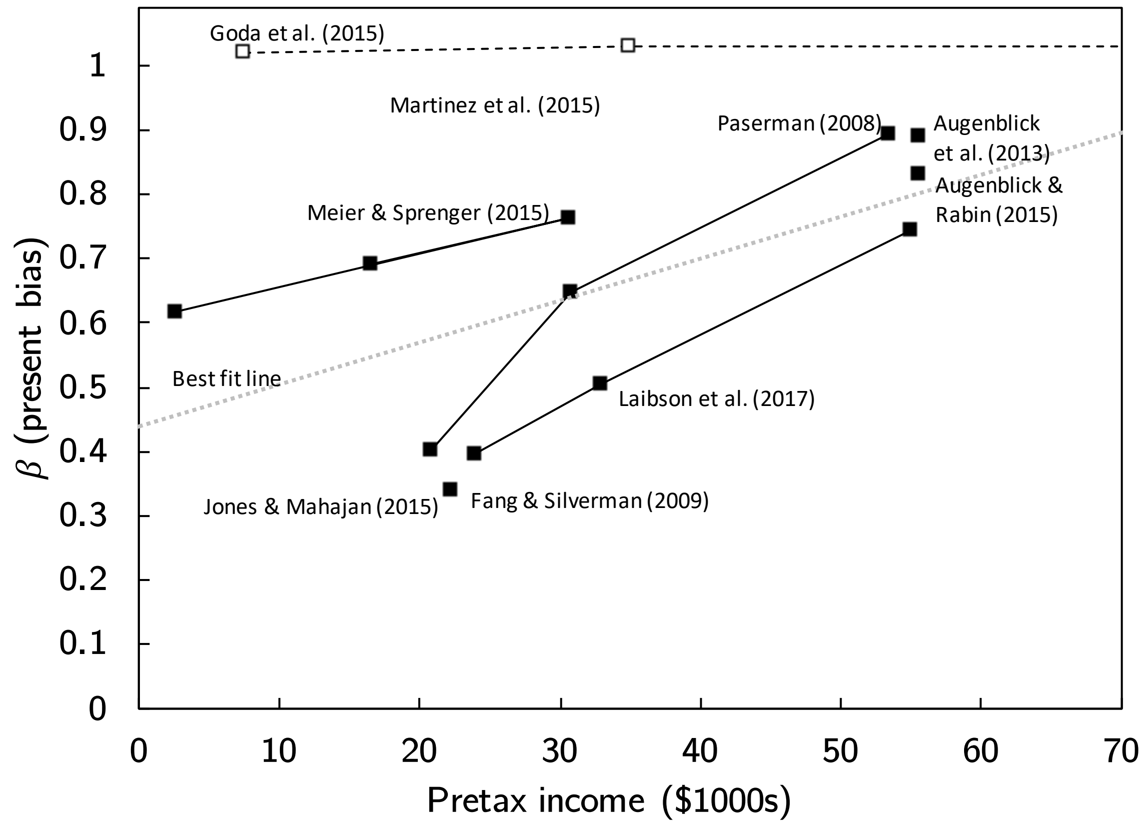
This table reports several estimates of the present bias β parameter conditional on income. See Appendix D for details.

Figure 1: Simulated optimal marginal tax rates with and without present bias.



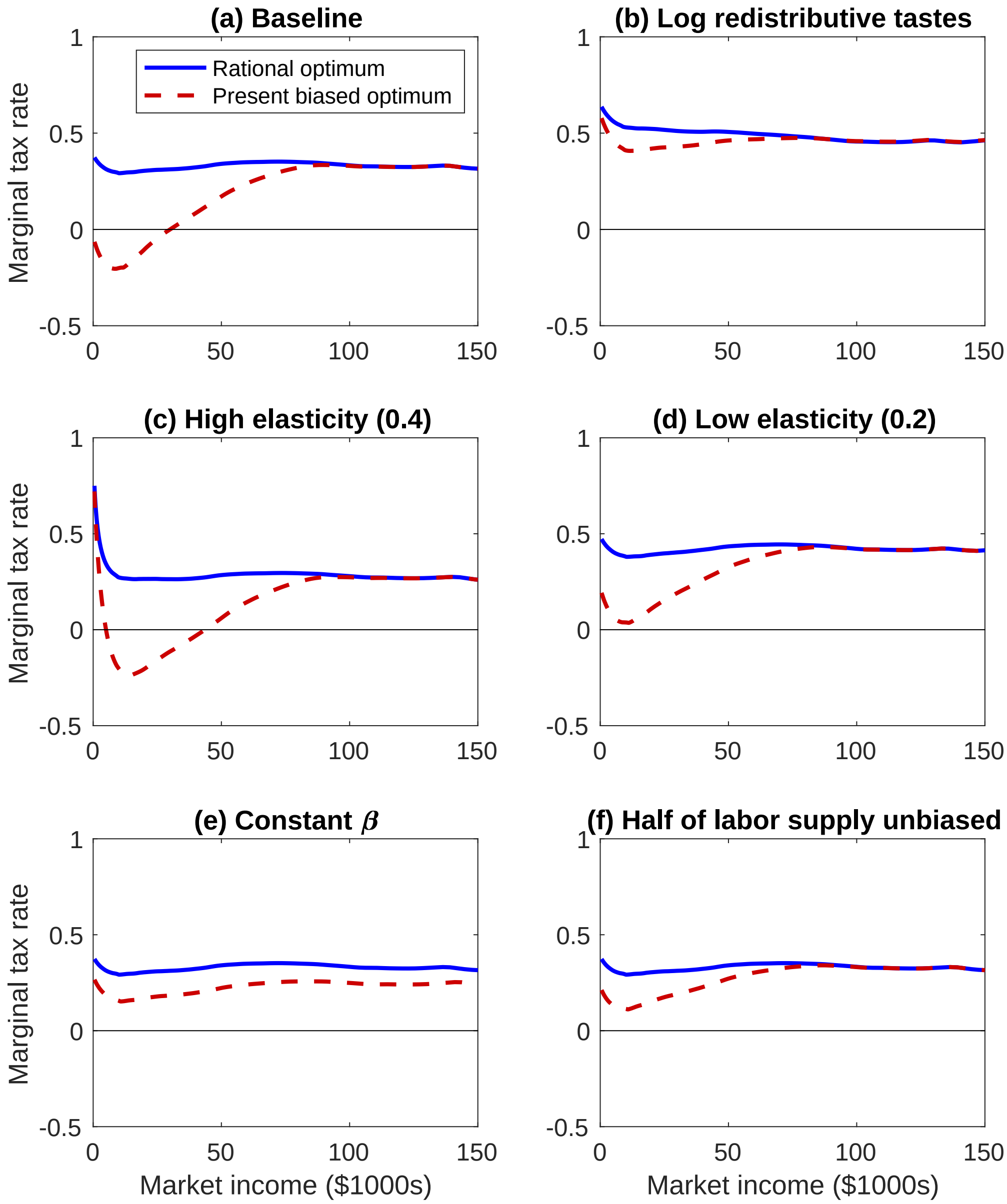
This figure displays simulated optimal tax rates under modest redistributive preferences, with and without accounting for present bias. Net marginal tax rates, including the phaseout of universal benefits, for a representative EITC-receiving household are plotted in gray (see details in Appendix E). The details of the model economy are discussed in Section 3, where the relationship between redistributive preferences and marginal tax rates is discussed at length.

Figure 2: Estimated relationship between income and present bias parameter β.



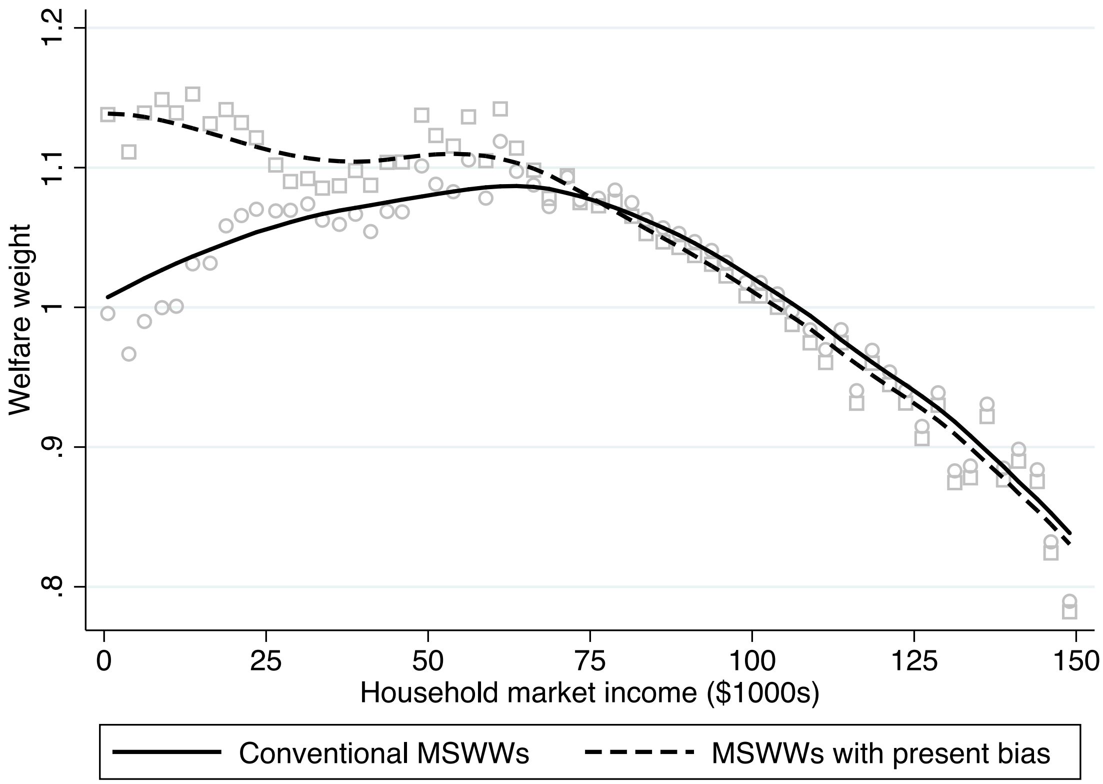
This figure plots estimates of β across income from several papers. The dotted "best fit line" is used in simulations for schedule of present bias across the skill distribution. See Appendix D for details.

Figure 3: Optimal income tax simulations.



This figure displays optimal simulated income taxes under six different specifications. Panel (a) plots the baseline specification, which uses a labor supply elasticity of 0.3 and modest redistributive preferences (bottom earners have a welfare weight 10% above the median, top earners have a welfare weight of 40% below the median, linearly interpolated). Panel (b) uses logarithmic redistributive preferences (see text for details). Panels (c) and (d) employ higher and lower labor supply elasticity values, with the same modest redistributive weights as in the baseline. Panel (e) assumes a homogeneous value of β across the income distribution, equal to the population average in the baseline. Panel (f) assumes that one half of the labor supply response to tax reforms is along dimensions of effort which are not subject to present bias.

Figure 4: Welfare Weights Implied By Existing EITC



This figure plots the welfare weights implicit in US policy under the conventional assumption of perfect optimization, and under calibrated present bias. Weights are computed by generating a smoothed income density based on 5th-order polynomial regression, then computing weights locally within $2500 income bins. Lines are generated using kernel regression with a bandwidth of $10,000. See text and Appendix E for details.

Appendix

Appendix A Proofs

Proof of Proposition 1

Proof. Let $\mathcal{V}(w)$ denote the rescaled decision utility function as perceived by type w's self at the time labor supply is chosen:

$$\mathcal{V}(w) := w\ell(w) - T(w\ell(w)) - \frac{v(\ell(w))}{\beta(w)}. \tag{26}$$

(Any exponential discount factor δ can be absorbed by rescaling v.) By assumption, at points of differentiable T the individual's global optimum satisfies the first-order condition

$$1 - T'(w\ell(w)) = \frac{v'(\ell(w))}{\beta(w)w}, \tag{27}$$

and so we have

$$\mathcal{V}'(w) = \frac{\ell(w)v'(\ell(w))}{\beta(w)w} + \frac{v(\ell(w))\beta'(w)}{\beta(w)^2}. \tag{28}$$

Normative utility is equal to

$$\mathcal{V}(w) + \left(\frac{1-\beta(w)}{\beta(w)}\right) v(\ell(w)), \tag{29}$$

and, in a modified version of the standard optimal control setup, we can take $\mathcal{V}(w)$ as the state variable and $\ell(w)$ as the control variable, writing the problem as

$$\max \int_{w_{min}}^{w_{max}} G\left(\mathcal{V}(w) + \left(\frac{1-\beta(w)}{\beta(w)}\right) v(\ell(w))\right) f(w)dw \tag{30}$$

subject to the (appropriately rewritten) budget constraint with required revenue E:

$$\int_{w_{min}}^{w_{max}} \left(\mathcal{V}(w) + \frac{v(\ell(w))}{\beta(w)}\right) f(w)dw \leq \int_{w_{min}}^{w_{max}} w\ell(w)f(w)dw - E. \tag{31}$$

and

$$\mathcal{V}'(w) = \frac{\ell(w)v'(\ell(w))}{\beta(w)w} + \frac{v(\ell(w))\beta'(w)}{\beta(w)^2}. \tag{32}$$

Letting λ denote the multiplier on the budget constraint in (31), and letting $m(w)$ denote the multipliers on the constraint in (32), the Hamiltonian for this problem is

$$\mathcal{H} = \left[G\left(\mathcal{V}(w) + \left(\frac{1-\beta(w)}{\beta(w)}\right) v(\ell(w))\right) - \lambda\left(\mathcal{V}(w) + \frac{v(\ell(w))}{\beta(w)} - w\ell(w)\right)\right] f(w) + \\ m(w)\left(\frac{\ell(w)v'(\ell(w))}{\beta(w)w} + \frac{v(\ell(w))\beta'(w)}{\beta(w)^2}\right). \tag{33}$$

The usual solution technique requires

$$m'(w) = -\frac{\partial \mathcal{H}}{\partial \mathcal{V}} = \left(\lambda - G'\right) f(w). \tag{34}$$

Maximizing $\mathcal{H}$ with respect to $\ell(w)$, we have

$$\begin{aligned}\left(-G' \times \left(\frac{1-\beta(w)}{\beta(w)}\right) v'(\ell(w)) + \lambda\left(\frac{v'(\ell(w))}{\beta(w)} - w\right)\right) f(w) \\ = m(w)\left(\frac{v'(\ell(w)) + \ell(w)v''(\ell(w))}{\beta(w)w} + \frac{v'(\ell(w))\beta'(w)}{\beta(w)^2}\right).\end{aligned} \tag{35}$$

Using the fact that $m(w_{max}) = 0$ (no distortion at the top) we have

$$m(w) = \int_{w_{max}}^{w} m'(w)dw = -\int_{w}^{w_{max}} m'(w)dw = \int_{w}^{w_{max}} \left(G' - \lambda\right) f(w)dw. \tag{36}$$

Substituting into (35) and rearranging yields

$$\frac{T'}{1-T'} = \frac{1}{f(w)}\int_{x=w}^{w_{max}} \left(1 - \frac{G'}{\lambda}\right) f(x)dx \left(\frac{1 + \frac{\ell(w)v''(\ell(w))}{v'(\ell(w))}}{w} + \frac{\beta'(w)}{\beta(w)}\right) - \left(\frac{G'}{\lambda}\right)(1-\beta(w)). \tag{37}$$

Then substituting in the expressions (from the text) for the elasticities of labor supply and present bias yields the expression in Proposition 1.

The expression for λ is derived by noting that the shadow value of public funds must equal the social welfare generated by a uniform marginal increase in consumption.

Sufficiency: the necessary condition in (37) could fail to be sufficient to characterize the optimum for three reasons. First, the individual's first-order condition in (27) may fail to characterize the individual's global optimum choice, for example because of nonlinearities in the resulting tax function. Mirrlees (1971) notes that this sufficient condition is implied by the single-crossing property and by monotonicity—income strictly increasing with type at the optimum. Second, the optimal tax schedule may feature points of non-differentiability, in which case the condition in (37) does not apply. Both possibilities are ruled out by the first sufficiency condition stated after Proposition 1: each individual's globally optimal labor supply choice at the under the optimal tax is given by their first-order condition. The third reason (37) could fail to be sufficient is if the Hamiltonian is not concave. As shown in Seierstad and Sydsaeter (1977), a sufficient condition in this setting is for $\mathcal{H}$ to be concave in ℓ at the optimum. Twice differentiating (33) with respect ℓ yields this concavity condition:

$$\begin{aligned}\frac{\partial^2 \mathcal{H}(w)}{\partial \ell(w)^2} = \left\{(1-\beta(w))\left[G'' \cdot \left(\frac{v'(\ell(w))}{\beta(w)}\right)^2 + G' \cdot \left(\frac{v''(\ell(w))}{\beta(w)}\right)\right] - \lambda\frac{v''(\ell(w))}{\beta(w)}\right\} + \\ m(w)\left[\frac{2v''(\ell(w)) + \ell(w)v'''(\ell(w))}{\beta(w)w} + \frac{v''(\ell(w))\beta'(w)}{\beta(w)^2}\right] < 0.\end{aligned} \tag{38}$$

Because $m(w) < 0$ and $v'' > 0$, sufficient conditions for this inequality to hold are for the term in braces to be negative and for the second term in brackets to be positive—conditions equivalent to the concavity conditions listed in footnote 8. □

The optimal tax condition without corrective motive

This result derives the optimal tax in the event that the policymaker has identical redistributive preferences—meaning that the weights $g(w)$ are the same at the optimum—but does not regard present bias to be a mistake, instead viewing it as a justified source of preference heterogeneity about labor disutility. The derivation proceeds identical to the Hamiltonian method above for Equations (27) – (28). However, the policymaker agrees with the decision-making agent's perceived utility function, and thus normative utility is simply $\mathcal{V}(w)$. Therefore the policymaker's problem is now

$$\max \int_{w_{min}}^{w_{max}} \hat{G}\left(\mathcal{V}(w)\right) f(w) dw, \tag{39}$$

which replaces (39). Since the modification to normative utility introduces a level change in utility, the concave transformation $\hat{G}$ is modified to allow for identical final redistributive weights $g(w)$ at the optimum.

Proceeding as before, with this nonpaternalistic objective, the Hamiltonian for the modified problem is

$$\mathcal{H} = \left[\hat{G}\left(\mathcal{V}(w)\right) - \hat{\lambda}\left(\mathcal{V}(w) + \frac{v(\ell(w))}{\beta(w)} - w\ell(w)\right)\right] f(w) + \\ m(w)\left(\frac{\ell(w)v'(\ell(w))}{\beta(w)w} + \frac{v(\ell(w))\beta'(w)}{\beta(w)^2}\right). \tag{40}$$

Maximizing $\mathcal{H}$ with respect to $\ell(w)$ gives

$$\lambda\left(\frac{v'(\ell(w))}{\beta(w)} - w\right) f(w) = m(w)\left(\frac{v'(\ell(w)) + \ell(w)v''(\ell(w))}{\beta(w)w} + \frac{v'(\ell(w))\beta'(w)}{\beta(w)^2}\right), \tag{41}$$

and replacing $m(w) = \int_w^{w_{max}}\left(\hat{G}' - \hat{\lambda}\right) f(w)dw$ yields

$$\frac{T'}{1-T'} = \frac{1}{f(w)}\int_{x=w}^{w_{max}}\left(1 - \frac{\hat{G}'}{\lambda}\right) f(x)dx\left(\frac{1 + \frac{\ell(w)v''(\ell(w))}{v'(\ell(w))}}{w} + \frac{\beta'(w)}{\beta(w)}\right). \tag{42}$$

Finally, imposing the assumption that redistributive preferences remain the same (so that $\hat{G}'/\lambda = g(w)$ as in Proposition 1), and employing the definitions of $\zeta_\ell(w)$ and $\zeta_\beta(w)$ from the text, we find $\frac{T'}{1-T'} = \mathcal{A}(w)\mathcal{B}(w)$, with $\mathcal{A}$ and $\mathcal{B}$ defined as in Proposition 1.

Proof of Corollary 1

Proof. By assumption, all individuals' earnings satisfy the first-order condition $v'(\ell) = w\beta(w)(1 - T'(z(w))$ at the optimum. Consider the limit

$$\lim_{w\downarrow w_{min}}\left\{\frac{1 + 1/\zeta_\ell(w) + \zeta_\beta(w)}{wf(w)}\int_{x=w}^{w_{max}}\left(1 - g(x)\right) dF(x)\right\} = \\ \frac{1 + 1/\zeta_\ell(w_{min}) + \zeta_\beta(w_{min})}{w_{min}}\left(\lim_{w\downarrow w_{min}}\frac{1}{f(w)}\int_w^{w_{max}}\left(1 - g(x)\right) f(x)dx\right), \tag{43}$$

where equality follows because all constituent functions are continuous in w. If $\lim_{w \downarrow w_{min}} f(w) > 0$ (the limit as w approaches w_{min} from above) then the limit evaluates to zero. If instead $\lim_{w \downarrow w_{min}} f(w) = 0$, the limit in parentheses can be evaluated, using l'Hôpital's rule, as

$$\lim_{w \downarrow w_{min}} \frac{(g(w) - 1)\, f(w)}{f'(w)}, \tag{44}$$

which in turn evaluates to zero. Therefore

$$\lim_{w \downarrow w_{min}} \left\{ \frac{T'(z(w))}{1 - T'(z(w))} \right\} = -g(w_{min})\,(1 - \beta(w_{min}))\,, \tag{45}$$

which, by assumption that $\beta(w)$ is bounded below 1, implies the marginal tax rate on the lowest earners is negative and bounded away from zero. By continuity of $\frac{T'(z(w))}{1-T'(z(w))}$ in w, there is thus a range of w sufficiently close to w_{min} which face negative marginal tax rates. □

Proof of Corollary 2

Proof. As shown in the preceding two appendix sections, $\frac{T'_{opt}}{1-T'_{opt}} = \mathcal{AB} - \mathcal{C}$ and $\frac{T'_{redist}}{1-T'_{redist}} = \mathcal{AB}$. Therefore, the optimal marginal labor subsidy S is given by

$$S(z(w)) = T'_{redist}(z(w)) - T'_{opt}(z(w)) \tag{46}$$

$$= \frac{1}{1 + \frac{1}{\mathcal{AB}}} - \frac{1}{1 + \frac{1}{\mathcal{AB} - \mathcal{C}}}. \tag{47}$$

The only term which depends on the labor supply elasticity is $\mathcal{A}$, and thus we have

$$\frac{\partial S(z(w))}{\partial \zeta_\ell(w)} = \frac{\partial S(z(w))}{\partial \mathcal{A}} \frac{\partial \mathcal{A}(w)}{\partial \zeta_\ell(w)} \tag{48}$$

The term $\frac{\partial \mathcal{A}}{\partial \zeta_\ell}$ is unambiguously negative. The term $\frac{\partial S}{\partial \mathcal{A}}$ is equal to $\frac{\mathcal{B}}{(T'_{redist}+1)^2} - \frac{\mathcal{B}}{(T'_{opt}+1)^2}$, and so since the corrective term $\mathcal{C}$ is positive, this derivative is negative overall. Therefore both terms in Equation (48) are negative, implying the subsidy rises with the size of the elasticity ζ_ℓ.

□

Proof of Proposition 2

Proof. Consider a reform to the optimal income tax which slightly raises the marginal tax rate by $d\tau$ in a narrow range of width ϵ around some income level z^*, where the optimal tax is assumed to be continuous and twice differentiable. The first-order effects of this reform can be decomposed into a mechanical effect dM (through raised revenue and a reduction in welfare), a *local* behavioral effect dL through the behavioral responses of individuals who earn z^*, and an *inframarginal* effect dI through the behavioral responses of individuals who earn more than z^*. Each of these terms represents the derivative of social welfare (through the channel in question) with respect to $d\tau$, taking the limit as $\epsilon \to 0$, normalized by the value of public funds. At the optimum, the sum of these effects must equal zero. (I normalize the size of each effect by the magnitude of the infinitesimal reform, $d\tau\epsilon$, since that term is common to each effect and cancels when the sum is set to zero.)

The mechanical effect, identical to that in Saez (2001), is straightforward:

$$dM = \int_{z^*}^{\infty} (1 - \overline{g}(s)) dH(s). \tag{49}$$

The local behavioral response effect on the intensive margin is composed of a fiscal externality $\mathbb{E}\left[\frac{dz(i)}{dT'(z(i))}\middle| z(i) = z^*\right] h(z^*)T'(z^*)$ and a welfare effect,

$$\mathbb{E}\left[G'(U_i)\sum_{j=1}^{J} \frac{\partial U_i}{\partial \ell_j(i)} \frac{\partial \ell_j(i)}{dT'(z(i))}\middle| z(i) = z^*\right] h(z^*), \tag{50}$$

where U_i is normative utility as in (13), evaluated at the optimal choices of consumption and labor supply.

Note that

$$\frac{\partial U_i}{\partial \ell_j(i)} = -\delta^{-\tau(j)} v_j'(\ell_j(i)) + u'(c(i))w_j(i)(1 - T'(z(i))). \tag{51}$$

From individual optimization, for all j such that $\tau(j) = 0$, (51) is equal to zero. For all other j,

$$-\delta^{-\tau(j)} v_j'(\ell_j(i)) + \beta(i)u'(c(i))w_j(i)(1 - T'(z(i))) = 0, \tag{52}$$

implying

$$\frac{\partial U_i}{\partial \ell_j(i)} = (1 - \beta(i))u'(c(i))(1 - T'(z(i))) \quad \text{for } j \text{ s.t. } \tau(j) > 0. \tag{53}$$

Therefore the local behavioral welfare effect in (54) can be rewritten

$$-\mathbb{E}\left[g(i)(1 - \beta(i))\phi(i)\varepsilon(i)|z(i) = z^*\right] h(z^*)z^* \tag{54}$$

Combining this welfare effect with the fiscal externality from the local behavioral response yields the total local intensive margin effect:

$$dL = -\overline{\varepsilon}(z^*)z^*h(z^*)\left[\frac{T'(z^*)}{1 - T'(z^*)} + \overline{g}(z^*)\overline{\phi}(z^*)(1 - \overline{\beta}(z^*))\left(1 + \Sigma^{(z^*)}_{1-\beta,\varepsilon,g,\phi}\right)\right]. \tag{55}$$

There are also inframarginal behavioral responses due to the increased level of taxes for individuals with earnings above z^*, through income effects. This inframarginal effect dI can be decomposed into a fiscal externality component, $\mathbb{E}\left[\frac{dz(i)}{dT(z(i))}\middle| z(i) = z\right] h(z)T'(z)$ for each $z > z^*$, while the welfare effect is

$$\int_{z^*}^{\infty} \mathbb{E}\left[G'(U_i)\sum_{j=1}^{J} \frac{\partial U_i}{\partial \ell_j(i)} \frac{\partial \ell_j(i)}{dT(z(i))}\middle| z(i) = s\right] dH(s) = \int_{z^*}^{\infty} \mathbb{E}\left[g(i)\phi(i)\eta(i)(1 - \beta(i))|z^*(i) = s\right] dH(s). \tag{56}$$

Here we have used the fact that $\sum_{\{j|\tau(j)>0\}} w_j \frac{d\ell_j(i)}{d(1-T')} \Big/ \frac{dz(i)}{d(1-T')} = \sum_{\{j|\tau(j)>0\}} w_j \frac{d\ell_j(i)}{dT} \Big/ \frac{dz(i)}{dT}$: the share of labor response to a tax perturbation which comes through labor choices prior to compensation is the same whether that perturbation concerns marginal tax rates or levels. The proof of this equivalence follows from the first-order condition for ℓ_j^{pb}, which can be written

$$\frac{v'(\ell_j^{pb})}{w_j \delta^{\tau(j)} B(j)} = u'(z - T(z))(1 - T'(z)), \tag{57}$$

where $B(j) = 1$ if $\tau(j) = 0$ and $B(j) = \beta$ if $\tau(j) > 0$. Consider a perturbation to the tax code dT which results in a vector of labor supply adjustments $d\ell_j^{pb}$. Employing (57), these changes satisfy

$$\frac{v''(\ell_j^{pb})}{(w_j \delta^{\tau(j)} B(j))^2} \frac{d\ell_j^{pb}}{dT} = K, \tag{58}$$

where K is the total derivative of the right side of (57) with respect to dT. Rearranging (58) gives

$$w_j \frac{d\ell_j^{pb}}{dT} = K \frac{w_j (w_j \delta^{\tau(j)} B(j))^2}{v''(\ell_j^{pb})} \tag{59}$$

Summing these equations over the j such that $\tau(j) > 0$, divided by the sum across all j, one finds that the term K cancels, so the share ϕ does not depend on the particular marginal source of the tax perturbation.

Combining these effects yields

$$dI = -\int_{z^*}^{\infty} \overline{\eta}(z) \left[\frac{T'(z)}{1 - T'(z)} + \overline{g}(z)\overline{\phi}(z)(1 - \overline{\beta}(z)) \left(1 + \Sigma^{(z)}_{1-\beta,\eta,g,\phi}\right) \right] dH(z). \tag{60}$$

Using these terms, the first-order condition for the optimal tax policy requires $dM + dL + dI = 0$. Rearranging yields the expression in Proposition 2. □

Proof of Corollary 3

The proof of Part 1 of the proposition follows immediately from the fact that the choice of earnings dimensions with $\tau(j) > 0$ does not depend on ψ. When $\phi(i) = 1$ for all i, effort is the only determinant of earnings, and the Part 1 of the proposition is implied.

Proof of Part 2. Consider the optimal tax when $\psi = 0$, which is characterized by Proposition 2, and suppose ψ is raised slightly, by $d\psi$. Any individuals with $\phi(i) = 1$ are insensitive to the change and can be ignored. Individuals with $\phi(i) < 1$ and $\beta(i)$ choosing some $\ell_{j'}$, where $\tau(j') = 0$, perceive the tax to be reduced by $d\psi(1 - \beta(i))T'(z(i))$, and therefore raise $\ell_{j'}$ by $d\psi(1 - \beta(i))\frac{d\ell_j(i)}{d(1-T')}$, which (beginning from $\psi = 0$) has no first-order effect on welfare due to the envelope theorem, yet has a strictly positive fiscal externality, implying that the total first-order effect on social welfare of slightly raising ψ is strictly positive, proving the proposition. □

Proof of Part 3. Extending the logic in Part 2, consider raising ψ by $d\psi$, but beginning from some $\psi > 0$. If the effect on social welfare of raising ψ remains positive at $\psi = 1$, the proposition is proved. There is still no effect on advance labor effort, so the reform generates a response in earnings through contemporaneous effort alone equal to $dz(i) = d\psi(1 - \beta(i))\frac{T'}{1-T'}(1 - \phi(i))z(i)\varepsilon(i)$. This behavioral response generates a fiscal externality equal to $dz(i)T'(z(i))$. However when $\psi > 0$, the envelope theorem no longer holds, so there is also a first-order effect on welfare, equal to

$-dz(i)g(i)T'(z^*(i))\psi(1-\beta(i))$. Combining the two effects, the total effect from i's hours response is $dz(i)T'(z(i))\left(1-\psi g(i)(1-\beta(i))\right)$, which is nonnegative for $\psi=1$ if and only if $g(i)(1-\beta(i)) \leq 1$. If that inequality holds for all i, then fully delayed taxes ($\psi = 1$) are optimal, proving the proposition. $\square$

Appendix B Optimal tax condition with a participation margin

A participation margin can be added to the model in Section 2.1 by adding heterogeneous fixed costs of work $\chi(i)$ to the individuals' utility function. Since such costs are conventionally understood as the fixed costs of labor effort at the time work is performed (such as transportation and child care), so that the normative utility function in Equation (1) is replaced by

$$u(c) - v(z/w(i)) - \chi(i), \tag{61}$$

and the decision utility function in Equation (2) is replaced by

$$-v(z/w(i)) - \chi(i) + \beta u(c). \tag{62}$$

In this model with discontinuous jumping, I use the perturbation approach to derive the first-order condition for the optimal marginal tax rate, wherein a small marginal tax rate increase of $d\tau$ is imposed in a narrow income band of width ϵ around some income level z^*, as in the proof of Proposition (2) above. The mechanical effect dM and the the local behavioral effect dL are the same as in the proof of Proposition 4, but with $\phi = 1$ and the covariances (Σ terms) equal to zero.

Unlike in the proof of Proposition (2), however, this marginal reform also generates discontinuous jumping from agents who experience an increase in the tax level—those earning more than z^*. This effect can be written in terms of the participation tax rate, $\bar{T}(z) = \frac{T(z)-T(0)}{z}$, and the participation elasticity, $\rho(z) = -\frac{dh(z)}{dT(z)}\frac{(z-T(z))+T(0)}{h(z)}$.

The measure of z-earners who leave the labor force is equal to $-d\tau\epsilon\frac{\rho(z)h(z)}{(1-\overline{T}(z))z}$. Letting dP denote the behavioral responses on the participation margin, the fiscal externality due to the participation response is

$$dP_F(z^*) = -\int_{z^*}^{\infty} \rho(z)\left(\frac{\overline{T}(z)}{1-\overline{T}(z)}\right)h(z)dz. \tag{63}$$

There is also an internality term from this discrete participation margin. A marginal worker on the participation margin has

$$-v(z_p(i)/w(i)) + \beta(u(z_p(i) - T(z_p(i))) - u(-T(0))) = \chi(i), \tag{64}$$

where $z_p(i)$ denotes the earnings that i's present-biased self selects conditional on participating. As a result, when a marginal worker enters the workforce, that generates a change in welfare from the policymaker's (or long-run self's) perspective of

$$(1-\beta)(u(z_p(i) - T(z_p(i))) - u(-T(0))). \tag{65}$$

To incorporate this effect into our perturbation formula, it is useful to define the *extensive margin welfare weight*, the change in welfare (per dollar) from an increase in consumption equal to $z - T(z)$

for an unemployed individual i

$$g_{ext}(i) = \frac{G(u(z_p(i) - T(z_p(i))) - G(u(-T(0)))}{z_p - T(z_p) + T(0)} \cdot \frac{1}{\lambda}. \tag{66}$$

Further, I let $\mathcal{I}_{ext}(z)$ denote the set of individuals indifferent between earning z and exiting the labor force, and I define $\overline{g}_{ext}(z)$ and $\overline{\beta}_{ext}(z)$ to be the average values of $g_{ext}(z)$ and $\beta_{ext}(z)$ over the set $\mathcal{I}_{ext}(z)$. Then the welfare internality effect from the behavioral response on the participation is

$$dP_W(z^*) = -\int_{z^*}^{\infty} \rho(z)\overline{g}_{ext}(z)(1 - \overline{\beta}_{ext}(z))h(z)dz. \tag{67}$$

Combining these, and incorporating the into the full optimal tax expression, the first-order condition for the optimal income tax satisfies

$$\frac{T'(z)}{1 - T'(z)} = \mathcal{A}(z)\mathcal{B}(z) - \mathcal{C}(z), \tag{68}$$

with

$$\mathcal{A}(z) = \frac{1}{\overline{\varepsilon}(z)h(z)z}, \tag{69}$$

$$\mathcal{B}(z) = \int_{s=z}^{\infty} \left[1 - \overline{g}(s) - \overline{\eta}(s)\left(\frac{T'(s)}{1 - T'(s)} + \overline{g}(s)(1 - \overline{\beta}(s))\right)\right] dH(s) \tag{70}$$

$$- \int_{s=z}^{\infty} \rho(s)\left[\frac{\overline{T}(z)}{1 - \overline{T}(z)} + \overline{g}_{ext}(z)(1 - \overline{\beta}_{ext}(z))\right] dH(s) \tag{71}$$

$$\mathcal{C}(z) = \overline{g}(z)(1 - \overline{\beta}(z)). \tag{72}$$

where the key difference is the appearance of the integral on line (72) in $\mathcal{B}(z)$. This is again an endogenous first-order condition, so again one cannot make general statements about comparative statics, but notice that the presence of the bias term $1 - \beta_{ext}(z)$ enters negatively, suggesting that present bias further depresses marginal tax rates (relative to an economy without present bias) through the presence of an extensive margin responses. Intuitively, workers misoptimize on two dimensions—working too little due to present bias, conditional on working, and also leaving the labor force too eagerly.

Appendix C Extension: when workers can borrow and save

The baseline model from Section 2 includes only a single period of consumption, which occurs at the time of labor compensation. As a result, workers cannot borrow against their future earnings at the time labor supply decisions are made. This extension relaxes that assumption to illustrate the effect of access to borrowing and saving on the degree of present bias over labor supply.

The corrective terms $\mathcal{C}(w)$ and $\mathcal{C}(z)$ in Propositions 1 and 2 arise because a present-biased agent chooses a level of labor supply other than the one which maximizes normative utility. This misoptimization can be quantified using a *misoptimization wedge*:

$$\gamma_i = 1 - \frac{v'(\ell(i))/w(i)}{u'(c(i))(1 - T')}. \tag{73}$$

This wedge corresponds to the difference between the consumer's marginal rate of substitution from labor to consumption, and their price ratio, $\frac{1}{w(1-T')}$. In this paper's model of present-bias with hand-to-mouth consumers, we simply have $\gamma_i = 1 - \beta_i$. More generally, however, the bias wedge γ_i could be substituted in place of $1 - \frac{v'(\ell(i))/w(i)}{u'(c(i))(1-T')}$ in the proof of Proposition 2 to reach line (54), with γ_i replacing $1 - \beta_i$.

To understand the implications of borrowing for the optimal tax formula, we can focus on the effect of borrowing on the misoptimization wedge γ_i, taking into account the agent's endogenous adjustment of borrowing or saving. Here I extend the unidimensional model in Section 2.2 to allow for consumption during both periods, denoted c_1 and c_2. (For the remainder of this section, all variables refer to a given agent, so indexing by i is suppressed.)

I assume the agent begins period 1 with some endowed resources I, and that she can save an amount s (possibly negative) at an interest rate r. Let s denote (possibly negative) net savings, so that the agent's choice can be written as the pair (ℓ, s)—a combination of labor effort and net saving.

In this modified setting, the labor misoptimization wedge γ quantifies the wedge between the utility costs of labor effort, and the resulting consumption benefits of additional net income during period 2, accounting for any endogenous adjustment in savings. Therefore, it's helpful to define *net income* earned from labor during period 2: $y = w\ell - T(w\ell)$, so that the misoptimization wedge analogous to (73) is

$$\gamma = 1 - \frac{\left(-\frac{\partial U}{\partial \ell}\right)}{\left(\frac{dU}{dy}\right)} \cdot \frac{1}{w(1-T')}, \tag{74}$$

where $\frac{dU}{dy}$ incorporates any endogenous adjustment of savings in response to a change in y.

The agent's period 1 decision utility function in this modified setup is

$$u(c_1) - v(\ell) + \beta u(c_2), \tag{75}$$

with $c_1 = I - s$ and $c_2 = (1+r)s + w\ell - T(w\ell)$. Normative utility is

$$u(c_1) - v(\ell) + u(c_2). \tag{76}$$

The corrective effect of income taxes will turn out to depend on whether there are already policies in place that correct present bias in the savings domain. Therefore, I will consider two possibilities separately—the case of no corrective savings policy, and the case of optimally corrective savings policy.

Case 1: No Corrective Savings Policies. Suppose that the agent can costlessly borrow and save between periods 1 and 2, implying there is no present bias correction (either due to actions of the policymaker or the long-run self) to the level of savings s. Since these time periods are understood to be fairly short—consistent with the delay between labor effort and compensation—this setup assumes the relevant real market interest rate r is zero. Then the total effect on normative utility of a change in net earnings y, allowing for any endogenous adjustment of savings s, is

$$\frac{dU}{dy} = u'(c_2) + \frac{\partial s}{\partial y}\left(-u'(c_1) + u'(c_2)\right). \tag{77}$$

The agent chooses savings s to satisfy the first-order condition $u'(c_1) = \beta u'(c_2)$. Moreover, let $\mathcal{M} = 1 + \frac{ds}{dy}$, denoting the marginal propensity to consume during period 2 out of marginal net earnings from labor. Then Equation (77) can be rewritten

$$\frac{dU}{dy} = u'(c_2) - (1-\mathcal{M})(1-\beta)u'(c_2) \tag{78}$$

$$= (\mathcal{M} + (1-\mathcal{M})\beta)u'(c_2). \tag{79}$$

Substituting this into Equation (74), the misoptimization wedge can be written

$$\gamma = 1 - \frac{v'(\ell)}{(\mathcal{M} + (1-\mathcal{M})\beta)u'(c_2)} \cdot \frac{1}{w(1-T')}. \tag{80}$$

The agent's first-order condition for choice of labor supply implies $v'(\ell) = \beta u'(c_2)w(1-T')$, so we can rewrite Equation (80) as

$$\gamma = 1 - \frac{\beta}{(\mathcal{M} + (1-\mathcal{M})\beta)}. \tag{81}$$

This demonstrates that in the presence of borrowing, the misoptimization wedge depends on the marginal propensity to consume, $\mathcal{M}$, during the second period. Note that as $\mathcal{M}$ approaches 1, γ approaches $1-\beta$, as in the baseline model from Section 2. As $\mathcal{M}$ approaches 0, γ goes to 0, implying no misoptimization wedge on the labor supply dimension. For increasing and strictly concave u and v, as assumed here, $\mathcal{M}$ lies strictly between 0 and 1.[20] Therefore, although the degree of optimal present bias correction may be reduced in an environment with unconstrained borrowing and saving, in general *some* degree of correction will still be optimal, with the misoptimization wedge γ_i from (81) replacing the term $1-\beta_i$ in Proposition 2.

Case 2: With Corrective Savings Policies Now suppose that the ability to borrow or save between periods 1 and 2 is itself subject to some corrective policies. Such policies could take many forms. A policymaker might subsidize saving directly, for example. Or the long-run self may take actions which change the effective cost of borrowing for the short run self. For example, consider a setting in which a consumer can access both low-cost liquidity, via reduced retirement contributions or a home equity line of credit, and high-cost liquidity, via credit card borrowing or a payday loan. If the low-cost liquidity options require some advance action, such as adjusting one's automatic retirement contributions or submitting a credit application, then the cost of borrowing for the short-run self effectively lies in the hands of the long-run self. As such, we could view the cost of

[20] Differentiating the FOC for savings, $-u'(I-s) + \beta u'(s+y) = 0$, with respect to y, yields

$$\frac{\partial s}{\partial y}\left(u''(c_1) + \beta u''(c_2)\right) + \beta u''(c_2) = 0, \tag{82}$$

implying

$$\frac{\partial s}{\partial y} = -\frac{1}{1 + \frac{u''(c_1)}{\beta u''(c_2)}} \tag{83}$$

and

$$\mathcal{M} = \frac{1}{1 + \frac{\beta u''(c_2)}{u''(c_1)}}. \tag{84}$$

Since $-\infty < u''(c) < 0$, we have $0 < \mathcal{M} < 1$.

short-run borrowing, r, as a variable which is controlled (at least partially) by the long-run self. Finally, the quantity of savings s may be directly constrained by either public policies (such as restrictions on payday loan availability) or the long-run self (such as automatic contributions to retirement plans).

As the specific mechanisms of corrective savings policies are beyond the scope of this appendix, I refrain from imposing any particular structure and instead assume that there is some policy which raises the amount of saving s (or, equivalently, lowers the amount of borrowing $-s$) relative to what the period 1 self would choose under costless borrowing. The strength of this policy can be quantified by the savings wedge r^{corr}, the compensated tax on borrowing (subsidy on saving) which would lead the short-run self to choose that level of saving:

$$r^{corr} = \frac{u'(c_1)}{\beta u'(c_2)} - 1. \tag{85}$$

Note that $r^{corr} = 0$ corresponds to Case 1 above—no corrective saving policies. Alternatively, since the long-run self and policy make prefer to set $u'(c_1) = u'(c_2)$, an optimally corrective borrowing policy would result in $r^{corr} = 1/\beta - 1$.

In the presence of corrective saving policies, the step from Equation (77) to (78) is not quite right. Using the definition of r^{corr}, we instead get

$$\frac{dU}{dy} = u'(c_2) - (1 - \mathcal{M})(1 - \beta(1 + r^{corr}))u'(c_2). \tag{86}$$

As a result, the misoptimization wedge analogous to Equation (81) is

$$\gamma = 1 - \frac{\beta}{(\mathcal{M} + (1 - \mathcal{M})\beta(1 + r^{corr})}. \tag{87}$$

If $1+r^{corr} = 1/\beta$, corresponding to the optimally corrective savings policy, then $\gamma = 1-\beta$, as in the baseline version of the model with hand-to-mouth consumers. More generally, this illustrates that corrective work subsidies are a complement to corrective savings and borrowing policies, rather than a substitute for them. Intuitively, this result reflects the intuition that in the presence of corrective savings policies, reducing labor supply is effectively an untaxed means of borrowing present utility against future consumption. To the extent that other policies are already correcting present bias along observable dimensions of saving and borrowing, labor subsidies which correct this otherwise unobserved borrowing channel are welfare-improving.

Appendix D Details of Table 1 and Figure 2

Table 1 cites several papers which estimate present bias (or substantial short-run discounting) in contexts which are informative for the calibration of β in Section 3. Figure 2 relates these estimates to incomes, when possible. This appendix discusses these sources, and describes the construction of Table 1 and Figure 2.

Augenblick, Niederle and Sprenger (2015) and Augenblick and Rabin (2018) both analyze laboratory experiments with college students at UC Berkeley, who are asked to make decisions about real effort tasks. Augenblick, Niederle and Sprenger (2015) presents a lab experiment in which student participants face a fixed amount of effort to be performed within a given period. Individuals without commitment devices exhibit an apparent discount rate of about 11% per week. If the individuals had time-consistent preferences (with no discounting) beyond one week, this would suggest a misoptimization wedge of 0.89—this is the estimate reported in Table 1. Augenblick and Rabin (2018) estimates β explicitly from effort-for-money choices at various time horizons. Both studies find evidence of commitment demand, which is correlated with individual-specific measures of present bias. These papers do not study the relationship between β and any measure of income. Figure 2 places both estimates at \$55,535, the annual income for graduates of Berkeley after 10 years,[21] according to the US Department of Education's College Scorecard.[22]

Kaur, Kremer and Mullainathan (2015) measures the labor supply responses of employees in an Indian data entry center who were exposed to a number of treatments during a year-long experiment. Two findings are of particular interest. First, workers generated more output on paydays, with production rising smoothly over the weekly pay cycle as payday approached. This "payday effect" suggests a daily discount factor of about 5%. Their results are inconsistent with a strict β δ model of quasi-hyperbolic discounting, as effort rises smoothly as payday approaches, rather than jumping upward discretely. Because the time horizon in Kaur, Kremer and Mullainathan (2015) is shorter than in Augenblick, Niederle and Sprenger (2015) or Augenblick and Rabin (2018), they need not be inconsistent, if individuals have a daily discount rate of 0.05 for the upcoming week, with no discounting thereafter. Interpreted as such, Kaur, Kremer and Mullainathan (2015)'s results suggest a β equal to implied discount factor at a one week horizion, equal to discount factor of $\left(\frac{1}{1.05}\right)^7 = 0.71$, which is the value reported in Table 1. Like Augenblick and Rabin (2018), Kaur, Kremer and Mullainathan (2015) finds demand for commitment which is correlated with individual-specific present bias. Since comparisons between incomes among Indian data center workers and US EITC recipients are difficult (and since the authors do not report annual earnings) I exclude this estimate from Figure 2.

Meier and Sprenger (2015) presents a field experiment wherein EITC filers in Boston are given choices between intertemporal tradeoffs between monetary payments at different horizons. Subjects exhibit greater impatience at shorter horizons, suggestive of present bias. The estimated β for the full sample is 0.69; that is the value reported in 1. Monetary tradeoffs (as opposed to effort tradeoffs) may generate upward-biased estimates of β (understating the degree of present bias) if payments are not immediately converted into consumption (i.e., in the presence of saving or borrowing). If individuals are liquidity constrained, however, monetary payments may be consumed promptly, consistent with this paper's substantial measured present bias in this study. Moreover, an advantage of Meier and Sprenger (2015), relative to the preceding experimental studies, is that it studies precisely the population of interest for understanding the implications of present-biased behavior for low-income work subsidies: low income EITC recipients in the US. This is one of the few studies which reports the covariation of β with income—the paper finds a strong positive correlation, with β rising by about 0.05 for every \$10,000 of income.[23] To generate an approximate plot of β estimates across income for 2, I plot the overall average estimate of β (0.69) at the sample's mean income of \$16,603. In addition, I plot incomes approximately one standard deviation above and below the mean income (where I use \$14,000 to approximate the standard deviation, see Table 1 of that paper), with corresponding values of β computed using their linear best fit estimate of 0.05 per \$10,000 of income.

Martinez, Meier and Sprenger (2017) uses the pattern of tax filing among low-income tax filers to estimate the degree of procrastination in this population. That paper finds that the observed filing patterns cannot easily be matched by a calibrated model with exponential discounting, but

[21] All incomes are converted into 2010 dollars using the CPI-U.

[22] See `https://collegescorecard.ed.gov/school/?110635-University-of-California-Berkeley`.

[23] One possible confound is that higher income EITC recipients may be less liquidity constrained, and may therefore exhibit less present bias over money payments. This possibility points to the value of further effort-based present bias experiments on populations with heterogeneous incomes.

can be matched quite well by a model with present bias. I adopt that paper's highest likelihood specification, Table 8 column (4), for which $\beta = 0.92$. The average income in their population is \$17,000.

Jones and Mahajan (2015) conducts a field experiment designed to measure time inconsistency among low-income tax filers, allowing them to deposit funds in a liquid or illiquid account, with either immediate or delayed payments for doing so. I adopt their preferred value of $\beta = 0.34$, with an average income of \$17,600 in their population.

Laibson et al. (2015) uses the method of simulated moments to perform a calibration using data on income, wealth, and credit use. They report β computed separately for three partitions of education: those who did not finish high school, those who completed high school but not college, and those who completed college (with β values of 0.40, 0.51, and 0.74, respectively). The Bureau of Labor Statistics reports average weekly incomes within each of these education bins;[24] corresponding to annual incomes of \$23,939, \$32,868, and \$54,907, respectively. These income values are used to plot the points for Laibson et al. (2015) in Figure 2.

Paserman (2008) estimates a structural model of job search with quasi-hyperbolic preferences using the NLSY, extending the approach of DellaVigna and Paserman (2005) which finds strong evidence of present-biased search behavior. The paper reports β estimated separately for three partitions of the wage distribution: the bottom quartile of the wage distribution, the middle half, and the top quartile, with estimates of 0.40, 0.65 (averaging the lognormal and normal specifications) and 0.89, respectively. I convert the mean re-employment weekly wage for each group into annual incomes of \$20,822, \$30,717, and \$53,409; these are the income values used to plot the points in Figure 2.

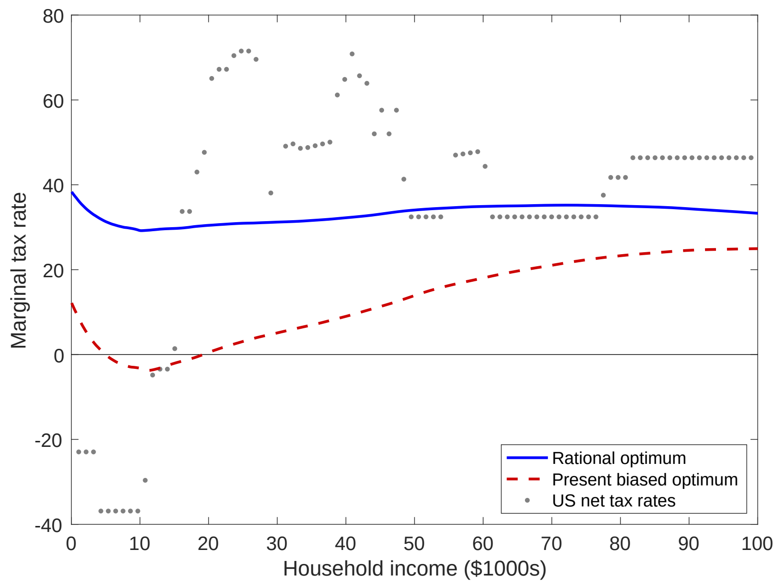
Fang and Silverman (2009) calibrates a quasi-hyperbolic model of model of welfare takeup and labor supply using data from the NLSY. The resulting estimate of β is 0.34, and the sample has an average income \$22,179.

DellaVigna and Paserman (2005) estimates a model of reference-dependent job search effort using Hungarian administrative data. The paper finds that admitting quasi-hyperbolic preferences (with $\beta < 1$) improves the fit significantly. The reported best-fit estimate of β is 0.58; this value is reported in Table 1.

Finally, Goda et al. (2015) measures time inconsistency using a series of hypothetical questions fielded to respondents using the American Life Panel and the Understanding America Survey. The results are outliers relative to the preceding results, in that they find no present bias on average ($\beta = 1.03$)—although there is substantial variation in β which is correlated with retirement savings—and no variation in present bias across incomes (see Figure B.4). In contrast, the paper finds very high exponential discounting at the annual horizon ($\delta = 0.71$). Although it not clear why these results differ from the others, it is notable that unlike the other studies cited here, the discounting questions were unincentivized. Because of these inconsistencies, I omit this study in the baseline best fit calibration of present bias across incomes (Figure 2). If the the calibration instead includes Goda et al. (2015), the resulting optimal tax rates are higher at low incomes, though they become slightly negative between \$5,000 and \$15,000 in income (see Figure 5).

[24] See `https://www.bls.gov/emp/ep_table_001.htm`.

Figure 5: Simulated optimal marginal tax under alternative present bias calibration.



This figure is identical to Figure 1, except that the present bias schedule is calibrated including the Goda et al. (2015) results in Figure 2.

Appendix E Calibration details

For the simulations in Section 3.2, I draw the income from the the 2010 Current Population Survey, restricted to households with positive total income, and I use kernel density estimation to calibrate the density across incomes.[25] I assume present bias is skill-specific, with the profile plotted in Figure 2. The first-order condition for effort choice can then be inverted to compute the implicit skill distribution. The first-order condition depends on the individual's marginal tax rate, which is estimated from CPS and NBER's TAXSIM. Specifically, I use TAXSIM's estimated net federal marginal tax rate, including employer and employee portions of payroll taxes, based on wage income, number of dependents, marital status, and age. I average this value across individuals at each level of income, and I construct an approximate implicit marginal tax rate from the phaseout of benefits using CPS data by performing a kernel regression of the value of food stamps and welfare income on market income, then differentiating the resulting schedule. I use a bandwidth of $2000 for the computation of marginal tax rates, and $5000 for the density estimation, where a greater degree of smoothing is useful for generating smooth schedules of simulated optimal tax rates.

To account for agents with very low ability levels, while avoiding complications of imperfect screening and optimal disability insurance, I assume that disability status is observable to the tax authority and that the required revenue for disabled individuals is exogenously given.[26] I assume the exogenously determined benefit payment for disabled individuals is $7,500, equal to average Social Security income in this age group in the Current Population Survey. Thus disability insurance effectively contributes to the government's revenue requirement.

[25] All data comes from University of Minnesota's IPUMS database (Ruggles et al., 2015).

[26] Specifically, I assume that 2% of individuals are disabled and unable to work altogether, consistent with the share of respondents in CPS between ages 25 and 55 with positive SSI income.

The optimal simulated tax schedule is plotted in Figure 1, along with estimated U.S. marginal tax rates for a representative EITC-qualifying household. Specifically, status quo tax rates are plotted for a single parent with two children residing in Colorado in 2015, computed using $1000 intervals. Calculations include the phaseout of universally available benefits: SNAP, Medicaid, CHIP, and ACA premium assistance credits. Estimates were computed by Eugene Steuerle and Caleb Quakenbush for congressional testimony.

Appendix F Details of Inverse Optimum Calculation

As described in the text, to focus on the implicit normative preferences consistent with the existing Earned Income Tax Credit, I use a sample different from the one in the benchmark economy of Section 3, though I continue to draw data from the CPS. Specifically, I use the 2015 March CPS, restricted to households with 2 children, and I calibrate marginal tax rates using TAXSIM, as well as the phaseout benefits from CPS. (All figures are reported in 2010 dollars, adjusted using the CPI-U.) I restrict to households in which the respondent is the head of household and is between the ages of 25 and 55. I further restrict to households with two children and with positive total family income. A continuous income distribution is constructed by discretizing the income space into $2500 bins and using a fifth order polynomial regression on the number of households in each bin to generate a smooth density with a continuously differentiable derivative. The schedule of marginal tax rates is drawn from the National Bureau of Economic Research's TAXSIM model. To compute the marginal tax rate at each point in the income distribution, I submit data on year, filing status, wage earnings (attributing family income outside the respondent's earnings to the non-responding spouse, for married respondents) and the number and age of dependent children to TAXSIM, which provides an effective marginal tax rate on additional earnings, accounting for credits and deductions. I include the marginal tax rate from payroll taxes (both the employer and employee portions). I then compute approximate implicit marginal tax rates from the phaseout of benefits by performing a local polynomial regression of benefits (the sum of food stamps, housing subsidies, heat and energy subsidies, and other welfare income) on a measure of market income (total family income less any social security, unemployment, and welfare income). The local derivative is interpreted as the implicit marginal tax rate from phaseouts, which is added to the marginal tax rate from TAXSIM. I then average these marginal tax rates within each $2500 bin, and use these averages to compute marginal social welfare weights following the approach in Hendren (2019).

A strength of the inverse optimum approach is that it permits a more detailed representation of the complexities of the actual economy. Since this approach entails only a local inversion of the first-order condition for optimal taxes, it does not require a structural model of earnings responses to non-local tax reforms. As a result, it is possible to incorporate a more detailed calibration of elasticities, including non-constant elasticities of taxable income and positive labor force participation elasticities (see Appendix B for an extension of the model in Section 2.2 to that setting).

I assume an elasticity of 0.33 at middle and high incomes, the preferred value in Chetty (2012), which lies well in the range of other estimates. For the elasticities at low incomes, I draw from evidence drawn specifically from the EITC-receiving population. Chetty, Friedman and Saez (2013) estimate intensive margin elasticities of 0.31 and 0.14 in the phase-in and phase-out regions of the EITC, respectively, identified by differences in knowledge of (and, by assumption, responses to) the EITC across geographic regions.

I follow the elasticity calibration assumptions in Hendren (2019), which performs an inverse optimum calculation using the universe of tax records. These entail an intensive margin elasticity of 0.31 in the phase-in region of the EITC, and 0.14 in the phase-out region, based on estimates from

Chetty, Friedman and Saez (2013), and an elasticity of 0.3 at higher incomes. Therefore I set the elasticity to be 0.31 for households with less than \$10,000 in earnings, and 0.14 for households with earnings of \$30,000. Also like Hendren, I assume an intensive elasticity of 0.3 for incomes above the EITC eligibility threshold (about \$50,000 for a married family with two children in 2015). To avoid sharp breaks in the distribution, I interpolate linearly across the transition from \$10,000 to \$30,000, and from \$30,000 to \$70,000. Finally, as in Hendren (2019) I assume a participation elasticity of 0.09 for households who receive the EITC, and zero at higher incomes, with a transition interpolated between \$30,000 and \$70,000.

The resulting marginal social welfare weights are plotted in Figure 4, both under the conventional assumption of no misoptimization, and under the assumption that individuals are present-biased. Plotted points represent the weight computed locally in \$2500 income bins, while the line plots the smoothed relationship. (Because the income distribution and the implied skill distribution extend to zero, the Seade (1977) result of "zero distortion at the bottom" does not apply.)

As shown by the dashed line in Figure 4, this unconventional feature disappears when a calibrated degree of present bias is incorporated into the calculation of welfare weights. Weights are substantially higher than 1 at the bottom of the distribution, and they decline monotonically with income. Although strict monotonicity is sensitive to the choice of smoothing bandwidth and precise assumptions about the patterns of elasticities over income, the main result that welfare weights are sharply increasing with income at the bottom under conventional asssumptions—but not after accounting for present bias—is quite robust.