Essays in Optimal Taxation

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

Essays in Optimal Taxation

A dissertation presented

by

Benjamin B. Lockwood

to

The Committee for the PhD in Business Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

April 2016

Dissertation Advisor:

Professor Matthew Weinzierl

Author:

Benjamin B. Lockwood

Essays in Optimal Taxation

Abstract

Policy often differs from the recommendations of theoretical optimal tax models in substantial and enduring ways. Such differences are sometimes surely because policy is suboptimal; however they may also be driven by alternative objectives which shape policy in practice, but which do not appear in the benchmark theoretical model. This dissertation considers three cases of such alternative objectives. The first chapter supposes that work subsidies like the Earned Income Tax Credit may be justified by corrective considerations, rather than the usual redistributive rationale for income taxation, if people are present biased and some benefits from work are delayed. The second chapter explores the role of income taxes in directing talented individuals into professions which are beneficial for the rest of society, such as teaching or medical research, and away from professions with negative externalities. The third paper considers the common concern that "sin taxes" on harmful goods—such as cigarettes or soda—are regressive, by incorporating redistributive concerns into a model of optimal corrective commodity taxation.

Contents

List of Tables

List of Figures

x

Acknowledgments

Introduction

Tax policy often differs from the recommendations of benchmark models of optimal taxation in substantial and enduring ways. Such differences are sometimes surely because policy is suboptimal; however they may also be driven by alternative objectives which shape policy in practice, but which do not appear in the benchmark theoretical model. This dissertation considers three cases of such alternative objectives.

My first essay studies the implications of present bias for the optimal income tax schedule. Work often entails up-front costs in exchange for delayed benefits—one must search for a job before becoming employed, paychecks are typically delayed by a few weeks, and a promotion may come only after months or years of extra effort—and mounting evidence documents present bias over labor supply in the face of such delays. I derive expressions for optimal tax rates as a function of observable elasticities and present bias, conditional on income. Present bias lowers optimal marginal tax rates, with a larger effect when the elasticity of taxable income is high. I calibrate the model using both existing estimates of present bias and a new estimate of residual present bias using subjective well-being trends following US welfare reforms in the 1990s. All evidence suggests bias is concentrated at low incomes. Numerical simulations show that for modest redistributive preferences, optimal marginal tax rates are substantially negative across low incomes, comparable to those under the Earned Income Tax Credit in the US. Yet the model also generates novel results about optimal tax timing, with implications for improving the schedule of EITC payments.

My second essay considers an alternative rationale for income taxes. Taxation affects the allocation of talented individuals across occupations by blunting material incentives and

thus relatively magnifying the non-pecuniary benefits of pursuing a "calling." Estimates from the literature suggest that high-paying professions pursued by these individuals have negative externalities while low-paying professions have positive externalities. As a result, a calibrated model indicates that total wealth is maximized by subsidies on middle class incomes and realistic tax rates on the rich. This result is robust to many uncertain features of the environment, though depends crucially on externality estimates and substitution patterns across professions, both of which deserve greater empirical study.

My third essay considers a policy question: how should governments design corrective "sin taxes"—commodity taxes intended to correct overconsumption of harmful goods—in the presence of both consumer mistakes and large wealth inequality? A common objection to such taxes is that many such taxes are regressive, falling largely on low income consumers. This paper extends the literature on "internality taxes"—taxes intended to correct overconsumption due to consumer misoptimization—by studying the interaction between corrective and redistributive motives, and shows how these motives can either dampen or amplify each other. We derive general, elasticity-based formulas for optimal taxes, and we show that the optimal tax can be computed as a function of a few estimable sufficient statistics: the price elasticity of demand, and the covariances between consumer bias, demand elasticities, and consumers' incomes. We also extend our analysis to the optimal use of nonsalient tax instruments and to the role of persuasive advertising, such as graphic warning labels, and we present conditions under which such unconventional policy instruments are strictly superior to corrective taxes. Quantitatively, we apply our model to cigarette taxes, and we use numerical simulations to trace out optimal tax policy as a function of consumer bias, the strength of redistributive motives, and the demand elasticity. We show that for the range of elasticities typically documented in empirical work, the optimal cigarette tax remains positive, but is significantly dampened by redistributive concerns.

Chapter 1

Optimal Income Taxation with Present Bias

1.1 Introduction

How should income tax policy account for the fact that people make mistakes? A growing body of research extends models of optimal taxation to allow for misoptimization of a general form, quantified by "misoptimization wedges". But estimating these wedges is challenging, since misoptimization cannot generally be inferred directly from observed choices. As a result, policy implications often require a deep understanding of the *way* in which people misoptimize. This paper focuses on a particularly robust and well measured form of misoptimization—present bias. I draw from the large literature on time inconsistency, wherein people grow more impatient over intertemporal tradeoffs as they grow near, and I adopt the common normative assumption that such impatience is a mistake (hence present "bias").

The motivation for focusing on present bias is threefold. First, recent evidence suggests present bias generates substantial labor supply distortions. Workers exert greater effort as payday approaches, impatient individuals take longer to find work, and students put

off unpleasant tasks to the last minute.[1] Since work often entails such delayed benefits in practice—whether due to job search, coarse pay periods, or infrequent promotions—accounting for that reality is of first-order importance in understanding the tax policy implications of misoptimization generally. Second, a sizeable empirical literature documents the magnitude of the present bias, including some evidence about how bias varies across income levels—a key determinant of optimal tax design—making plausible empirical calibrations more feasible than for other less well estimated sources of bias. Third, the specificity of the benchmark model of present bias provides discipline. Adding that behavioral model to the standard model of optimal taxation generates testable empirical predictions which are consistent with otherwise puzzling patterns in the labor supply literature.

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 present biased workers. I first assume that workers are naive about their bias, in which case the misoptimization wedge maps directly to their "structural" degree of present bias (conditional on income). I use optimal control theory to derive necessary and sufficient conditions for the optimal tax in this case, including a "negative at the bottom" result. I also allow for substantial multidimensional heterogeneity—including differences in compensation delays—with novel and surprising implications for the optimal *timing* of tax payments. In particular, if some individuals are paid up-front, then it is beneficial to *delay* the collection of taxes and the payment of work subsidies. This result has implications for improving the timing of EITC payments—a topic of substantial policy discussion.

The second part of the theory section considers the possibility that workers are sophisticated and can mitigate their bias by signing labor supply commitment contracts with employers. In this case, optimal taxes depend on the *residual uncorrected* bias at each income. If contracts are subject to a limited liability constraint preventing payments from workers to

---

[1]See Kaur, Kremer and Mullainathan (2015), DellaVigna and Paserman (2005), and Augenblick, Niederle and Sprenger (2013), respectively, and Section 1.4 for an extensive discussion of existing empirical evidence.

firms, this residual bias may remain positive even for sophisticated workers, particularly at low incomes. This model generates novel positive predictions—a U-shaped pattern of elasticities of taxable income, and an inverse U-shaped pattern of income effects—which appear consistent with patterns in the empirical labor supply literature.

Empirically, I draw on a wide range of evidence to estimate the misoptimization wedge due to present bias, conditional on income. I first review existing experimental evidence and structural estimations to calibrate the underlying (structural) degree of present bias across incomes. I then present an analysis—new to this paper—which estimates *residual* bias using trends in subjective well-being after the 1990s welfare reforms in the US. All evidence suggests bias is concentrated at low incomes.

The implications of present bias for optimal income tax policy are striking. In the final sections of the paper, I present numerical calibrations of optimal taxation for a range of parameters and redistributive preferences. Most notably, when redistributive motives are modest, optimal marginal tax rates are negative across a substantial range of low incomes. This finding suggests that corrective motives may provide a novel rationale for marginal work subsidies, such as the US Earned Income Tax Credit (EITC). Such policies are common across countries and enjoy strong bipartisan support among policy makers, and among economists.[2] Yet under the conventional assumption that the goal of income taxation is to redistribute resources to low skilled individuals, such policies are suboptimal.[3] Although the existence of negative marginal tax rates need not imply their optimality, this paper revisits this debate and shows that negative marginal tax rates may indeed be consistent

---

[2]Among policy makers, for example, President Obama and Paul Ryan have both advocated expanding the EITC for households without children.

[3]Optimal income taxes generally feature a large guaranteed minimum income, combined with *positive* marginal tax rates at all levels of earnings. Some have argued that marginal work subsidies may be optimal if labor supply responses are concentrated on the extensive margin (i.e., if workers respond to a tax increase by exiting the labor force rather than continuously reducing earnings)—see Diamond (1980) and Saez (2002b). However numerical simulations suggest this rationale may be quantitatively limited—the benchmark simulations in Saez (2002b), for example, calls for a marginal tax rate of 10% on the first $4,000 of earnings, and much higher rates thereafter. Moreover, recent research has shown that such models generate a "participation credit"—a fixed sum paid to all labor force participants—combined with positive marginal tax rates at all positive incomes (Jacquet *et al.*, 2013).

with an optimal tax schedule.

To preview the quantitative implications of present bias, Figure 1.1 plots the schedule of optimal marginal tax rates in the baseline model economy with and without accounting for present bias. Results are plotted for a modest degree of inequality aversion (see Section 1.5 for details), which tends to reduce marginal tax rates. Even in this case, marginal tax rates at low incomes are quite high under the rational optimum, generating a sharp divergence between that policy and effective marginal tax rates in the US (plotted here for households with two children). Stronger redistributive preferences exacerbate this divergence. However after accounting for present bias, optimal marginal work subsidies can match or exceed those generated by the EITC.[4]

Relation to the literature

This paper relates to a number of subfields in the optimal taxation literature.

Optimal taxation with misoptimizing agents. Farhi and Gabaix (2015) give the topic a general treatment, characterizing results in optimal taxation in terms of "behavioral sufficient statistics" (misoptimization wedges), without regard to the source of misoptimization. In independent research, Gerritsen (2015) characterizes optimal income taxation with abstract misoptimization, and attempts to measure status quo labor supply misoptimization using survey data.[5] This paper takes an alternative (and complementary) approach. By focusing on a specific model of misoptimization, I am able to generate new testable hypotheses and concrete policy recommendations. In this respect, the present paper is more similar

---

[4]The income tax here represents the integrated tax and transfer schedule, including programs such as welfare and Food Stamps. Kaplow (2007) notes that the phaseouts of such programs could offset the negative statutory marginal tax rates of the EITC, yielding positive net marginal tax rates. However analyses by the Congressional Budget Office (2012) and the Center on Budget and Policy Priorities (2014), as well as Kotlikoff and Rapson (2009), find net marginal tax rates remain negative—generally between -10% and -25%, for low income households with children.

[5]Gerritsen (2015) also uses subjective well-being data to estimate status quo labor supply misoptimization, adopting the approach, pioneered by Di Tella *et al.* (2001), of regressing subjective happiness reports on covariates and using the ratio of coefficients to estimate a marginal rate of substitution. Although that paper lacks quasi-experimental variation, raising concerns about omitted variable bias, it too argues that labor is undersupplied at low incomes, consistent with the findings in Section 1.4.
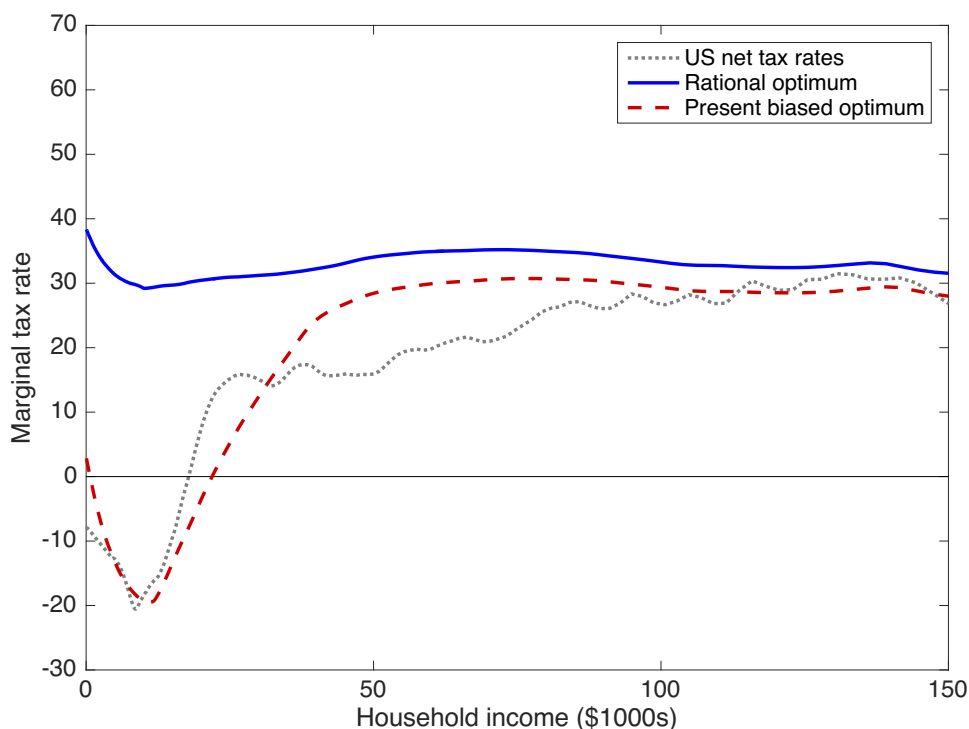
Figure 1.1: *Simulated optimal marginal tax rates with and without present bias under modest redistributive preferences, compared to approximate net tax rates for a family with children in the US. The calibration of present bias is discussed in Section 1.4. The details of the model economy are discussed in Section 1.5, where the relationship between redistributive preferences and marginal tax rates is discussed at length.*

to Spinnewijn (2014), which calibrates optimal unemployment insurance using data on mistaken beliefs about the probability of reemployment, and to Allcott and Taubinsky (forthcoming) and Lockwood and Taubinsky (2015), which perform quantitative calibrations of optimal taxation with consumers who misoptimize when consuming particular goods, such as those with delayed energy savings or uninternalized health consequences.

Optimality of negative marginal tax rates.   As noted above, the reasoning in this paper presents a novel rationale for negative marginal tax rates. In the benchmark model of redistributive income taxation laid out by Vickrey (1945a) and Mirrlees (1971b), negative marginal tax rates are suboptimal (this finding has been discussed extensively—see Seade (1977), Seade (1982), Werning (2000), Hellwig (2007), and citations therein). A number of later analyses explored the sensitivity of this result to positive and normative assumptions

in the conventional model. Diamond (1980) and Saez (2002b) influentially argued that marginal tax rates could 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. Saez (2002b) and Blundell and Shephard (2011) use models with discrete earnings levels to simulate optimal tax rates, finding low (or slightly negative) marginal tax rates on the lowest positive earning type. Jacquet *et al.* (2013) extended this insight to 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—a structure quite different from the current EITC in the US, which features negative marginal tax rates on earnings up to $13,000.

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 tax authority's goal is to redistribute income *upward*—i.e., if marginal social welfare weights (the social value of marginal consumption for households with a given level of earnings) are rising with income. This point was originally made in a discrete context by Stiglitz (1982); a number of recent papers have argued that such weights may arise from multidimensional heterogeneity, in particular preference heterogeneity (Cuff, 2000; Beaudry *et al.*, 2009; Choné and Laroque, 2010).[6] Similarly, Fleurbaey and Maniquet (2006) show how fairness considerations may generate welfare weights which increase with income. Drenik and Perez-Truglia (2014) provide empirical evidence for such views, while noting that such an objective could generate Pareto inefficient policy recommendations. Although the reasoning in this paper is not inconsistent with such normative objectives, these findings demonstrate that negative marginal tax rates may be warranted even under the conventional assumption that policy makers wish to redistribute to low earners.[7]

---

[6]Preference heterogeneity is not 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.

[7]Two other proposed rationales for negative marginal tax rates have received somewhat less attention. First, negative rates may be justified by "non-welfarist" objectives, for example if the government wishes to minimize poverty (Kanbur, Keen and Tuomala, 1994; Besley and Coate, 1992, 1995) or has preferences directly over the

The remainder of the paper is organized as follows. Section 1.2 presents a baseline model of optimal taxation with naive present biased workers and derives conditions characterizing the optimal income tax, including optimal tax rates and the optimal timing of tax collections, with a proof that delayed taxes can be optimal if some employers pay their workers up-front. Section 1.3 extends the environment to allow for sophisticated workers and commitment contracts. Section 1.4 estimates present bias conditional on income, drawing on several recent studies with a range of methodologies, as well as a new estimation of misoptimization using subjective well-being trends after the EITC expansion in the mid 1990s. Section 1.5 presents numerical simulations of the optimal policy in the baseline economy, and the resulting welfare gains from accounting for present bias. Section 1.6 performs an "inverse optimum" exercise to compute the implicit redistributive preferences which would rationalize existing policy under the conventional (perfect optimization) model and allowing for present bias. Section 1.7 discusses implications for policy design, as well as several important limitations. Section 1.8 concludes.

1.2 Optimal taxation with present bias: naive workers

The economy consists of a population of individuals of measure one, indexed by i and distributed according to measure $\mu(i)$. Utility is a separable function of consumption c, labor effort z/w_i (where z denotes pre-tax income, and $w_i > 0$ denotes ability), and fixed costs of work χ_i, where

$$U_i(c,z) = u(c) - v(z/w_i) - \chi_i \cdot 1\{z > 0\}, \qquad (1.1)$$

labor and leisure choices of poor individuals (Moffitt, 2006). Second, work may have positive externalities (or poverty negative externalities, e.g., Wane (2001))—indeed, a more recent literature has demonstrated that labor force participation may have positive effects on children's outcomes (Dahl and Lochner, 2012; McGinn, Ruiz Castro and Lingo, 2015). This justification is not inconsistent with the present bias rationale in this paper—indeed the distinction between present bias and positive externalities on children is a blurry one. Although I will focus on shorter term benefits when calibrating present bias in Section 1.4, to the extent that these positive externalities are further undervalued, I will underestimate the size of optimal work subsidies.

I assume increasing, weakly concave utility of consumption ($u' > 0$ and $u'' \leq 0$), and decreasing and strictly concave utility of production ($v' > 0$ and $v'' > 0$), and I normalize $v(0) = 0$.

The policy maker's problem is to select the income tax function $T(z)$ which maximizes social welfare—a weighted sum of total utility, with individual-specific "Pareto weights" given by α_i, and I use $z(i)$ to denote i's earnings:[8]

$$\mathcal{W} = \int \alpha_i U_i(z(i) - T(z(i)), z(i)) d\mu(i),$$ (1.2)

subject to a budget constraint

$$\int T(z(i)) d\mu(i) - E \geq 0,$$ (1.3)

where E denotes an exogenous revenue requirement.

I then introduce a simple modification to this otherwise conventional setup: consumption occurs in the period following labor effort, and consumers are present biased—they discount utility in future periods by β_i, so that i's earnings satisfy

$$z(i) = \arg\max_z \left\{ -v(z/w_i) - \chi_i \cdot 1\{z > 0\} + \beta_i u(z - T(z)) \right\}.$$ (1.4)

Here β_i can be interpreted as in a β, δ model of quasi-hyperbolic discounting (Laibson, 1997), where the long-run discount factor δ has been normalized to 1.

1.2.1 Necessary and sufficient conditions for optimal taxes in a simple case

I first focus on an instructive simplified case, motivated by Diamond (1998a), 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.

Specifically, I impose the following restrictive assumptions:

---

[8]In this constrained setting, (1.2) is isomorphic to a planner who maximizes a convex transformation of individual utility, $\int \Psi(U_i) d\mu(i)$, where the weights are set to equal the marginal social value of utility, $\alpha_i = \Phi'(U_i)$, evaluated at the optimum.

1. Quasilinear utility: $u(c) = c$.

2. Heterogeneous ability distributed between $w_0 \geq 0$ and $w_1 < \infty$.

3. Homogeneous fixed costs: $\chi_i = \bar{\chi}$ for all i.

4. Constant α_i conditional on ability w_i.

5. Constant β_i conditional on ability w_i.

Notationally, I let $F(w)$ denote the distribution of ability, and I abuse notation to write Pareto weights and present bias as functions of ability w rather than type i: $\alpha(w)$ and $\beta(w)$. I let $e_\ell(w) = \frac{v'(z/w)w}{v''(z/w)z}$ denote the labor supply elasticity evaluated at the optimal choice of earnings for ability w, and $e_\beta(w) = \beta'(w)w/\beta(w)$, the elasticity of present bias with respect to ability. Then the following proposition characterizes the optimal income tax policy.

Proposition 1.1. *Under assumptions 1–4, the nonlinear income tax $T(z)$ which solves (1.2)–(1.4) must satisfy the following expression at all points of differentiability:*

$$\frac{T'(z(w))}{1 - T'(z(w))} = \frac{1 + 1/e_\ell(w) + e_\beta(w)}{wf(w)} \int_w^{w_1} \left(1 - \frac{\alpha(x)}{\lambda}\right) dF(x) - \left(\frac{\alpha(w)}{\lambda}\right)(1 - \beta(w)),$$

where $\lambda = \int_0^\infty \alpha(w)dF(w)$. Moreover, this expression is sufficient to characterize the optimum if the schedule of earnings which satisfies the resulting first-order condition, $\frac{v'(z(w)/w)}{w} = \beta(w)(1 - T'(z(w)))$, is strictly increasing in w.

Proof. See Appendix A.1. □

If instead this schedule would generate decreasing $z(w)$, then marginal tax rates are discontinuous—$T(z)$ is kinked—with a range of abilities bunching at the same level of earnings.

The expression in Proposition 1.1 is a generalization of Diamond (1998a), which is identical except for the appearance of $e_\beta(w)$, and the final *corrective term*, $-\left(\frac{\alpha(w)}{\lambda}\right)(1 - \beta(w))$. Intuitively, the former effect accounts for the fact that if bias is decreasing with ability (so $e_\beta(w) > 0$) high ability types are less tempted to imitate low ability agents, and thus

11

marginal tax rates are optimally higher. The corrective term, which appears via the Hamiltonian optimization in the proof, can also be derived using economic intuition. Unlike in the conventional model, if individuals are biased then the "decision utility" marginal rate of substitution is not equal to the "experienced utility" marginal rate of substitution, and the welfare benefit from inducing a small increase in earnings is proportional to the ratio of the former to the latter. When misoptimization is due to present bias, this ratio is simply the share of marginal benefits from labor supply which are not internalized in decision utility when labor supply is set. Since a share β of delayed compensation is internalized, the remaining share $1 - \beta$ quantifies misoptimization, and appears in the corrective term. This quantification motivates the following formal definition:

Definition 1.1. *The "misoptimization wedge" for an individual i earning z is equal to*

$$1 - \frac{v'(z/w_i)/w_i}{u'(z - T(z))(1 - T'(z))}.$$

This wedge can loosely be thought of as the additional marginal tax rate which would induce an unbiased individual to choose z.

The corrective term has three notable features. First, the correction is negative when $\beta(w) < 1$—lower tax rates induce greater labor supply and thus help correct present bias— and it is larger in magnitude when $\beta(w)$ is far below 1. This overturns the familiar result that $T'(z) \geq 0$ for all z (Seade, 1977, 1982), since the corrective term may outweigh the preceding (nonnegative) term, generating negative marginal tax rates. Second, the term is weighted by $\frac{\alpha(w)}{\lambda}$, reflecting the planner's greater concern for correcting biases among individuals whose welfare is highly valued. Third, and perhaps least obviously, for a given Pareto weight and degree of bias, the corrective term has a constant effect on $\frac{T'(z)}{1-T'(z)}$, implying that the correction is larger when $T'(z)$ is low in the conventional (rational) model. Thus any feature of the standard model which reduces tax rates also inflates the size of the corrective term. For example, a higher elasticity of labor supply $e_\ell(w)$ tends to increase the size of the correction. This implies that if marginal tax rates are negative due to the corrective term, a higher elasticity of labor supply will make them more negative still. This result contrasts

with other proposed rationales for negative marginal tax rates, where a greater labor supply elasticity tends to *reduce* the magnitude of negative rates. This relationship is explored in more quantitative detail in Section 1.5.

Proposition 1.1 also amends the "zero at the bottom" result from Seade (1977, 1982), as captured by the following corollary:

Corollary 1.1. *If $\beta(w_0) < 1$ and $z(w_0) > 0$, then $T'(z(w_0)) < 0$.*

In words, if all individuals work, and if the lowest ability workers are biased, then optimal tax rates are negative on the lowest earners. The proof follows immediately from evaluating the expression in (1.1) at w_0, which, since $z(w_0) > 0$, holds even at w_0. This result highlights the distinction between this setup and flexible marginal social welfare weights, as there is no schedule of Pareto weights $\alpha(w)$ which would justify $T'(z(w_0)) < 0$ in a model without bias.

1.2.2 Multidimensional heterogeneity

I now relax the simplifying assumptions 1–4 to derive a more general expression for optimal marginal tax rates which permits income effects and multidimensional heterogeneity across w_i, χ_i, and β_i. This derivation employs the calculus of variations, as in as in Saez (2001a), to express the first-order condition in terms of observables, which are endogenous to the tax system. Although more general, this approach does not permit a simple characterization of sufficient conditions, and thus I will focus on the necessary (first-order) condition and assume that it is also sufficient for optimal policy for the remainder of the paper. (I will, however, confirm that the resulting tax function generates a schedule of income which strictly increases with ability in the numerical simulations in Section 1.5.)

The necessary condition for the optimal income tax can be derived by considering a reform to the optimal tax code which slightly raises marginal tax rates by $d\tau$ in a small band ϵ around an income level z^*. Let $\mathcal{I}(z') = \{i | z(i) = z'\}$, representing the set of people with earnings z' at the optimum. In what follows, it will be useful to distinguish between endogenous parameters, such as the compensated elasticity of taxable income ε, for a given

13

individual at the optimum, denoted $\varepsilon(i)$, and the parameter's average value across all individuals with a given income, denoted $\bar{\varepsilon}(z') = \frac{\int_{\mathcal{I}(z')} \varepsilon(i)d\mu(i)}{h(z')}$. For brevity the dependence of endogenous parameters on the tax code, though acknowledged, is notationally suppressed.

The relevant observables, defined loosely here and formally in Appendix A.2, are the income density $h(z)$, the compensated elasticity of taxable income $\varepsilon(i) = -\frac{dz(i)}{dT'(z(i))}\frac{1-T'(z(i))}{z(i)}$, the income effect $\eta(i) = -\frac{dz(i)}{dT(z(i))}(1 - T'(z(i)))$, and the participation elasticity $\rho(z) = -\frac{dh(z)}{dT(z)}\frac{(z-T(z))+T(0)}{h(z)}$. I follow Jacquet *et al.* (2013) in defining these responses to include any circularities due to the curvature of $T(z)$, which leads the marginal tax rate to change as earnings are locally adjusted.

The welfare effect of this perturbation also depends on the redistributive preferences of the planner (or, isomorphically, the perceived concavity of individual utility) which is encoded using standard *marginal social welfare weights*—written as a function of income— defined to be the marginal social welfare from additional consumption for an individual i earning $z(i)$:

$$g(i) = \frac{\alpha_i u'(z(i) - T(z(i)))}{\lambda}.$$

As before, λ denotes the marginal value of public funds.[9] If the α_i weights are constant across individuals, marginal social weights are the same for all individuals with the same income. This has the implication that the government does not seek to redistribute across individuals with a given income who have different combinations of present bias and ability—a plausible restriction given the potential difficulties of disentangling the the two sources of income variation. Thus more generally, even if the the α_i weights vary across individuals, it may be natural to assume that the weights are selected to yield constant marginal social welfare weights conditional on income at the optimum. Therefore in the derivations to follow, I invoke the following assumption, which turns out to simplify the expressions and render them more transparent.

---

[9]One feasible reform raises revenue by reducing all individuals' consumption so as to lower everyone's utility by a constant amount. Thus at the optimum the marginal value of public funds must equal $\left[\int \left(\frac{1}{\alpha_i u'(z(i)-T(z(i)))}\right) d\mu(i)\right]^{-1}$.

Assumption 1.1. *Marginal social welfare weights $g(i)$ and $g_{ext}(i)$ are the same for all individuals with a given income.*

For notational consistency, I'll use $\overline{g}(z^*)$ to denote the marginal social welfare weight on individuals earning z^*, though in light of Assumption 1.1, this is simply equal to the common $g(i)$ of all individuals earning z^*.

The tax reform in question has a number of effects. First, it mechanically raises $d\tau\epsilon \int_{z^*}^{\infty} h(z)dz$ in funds (collected from workers with earnings above z^*) at a welfare loss of $d\tau\epsilon \int_{z^*}^{\infty} \overline{g}(z)h(z)dz$, for a combined mechanical effect of

$$dM = d\tau\epsilon \int_{z^*}^{\infty} (1 - \overline{g}(z))h(z)dz.$$

Second, it induces a change in earnings of $dz^* = -d\tau \frac{\varepsilon(z^*)z^*}{1 - T'(z^*)}$ among the $\epsilon h(z^*)$ individuals who experience a change in marginal tax rates. This alters tax revenues, resulting in a *fiscal externality substitution effect* of

$$dS_F = -d\tau\epsilon h(z^*)\varepsilon(z^*)z^* \left(\frac{T'(z^*)}{1 - T'(z^*)} \right).$$

The change in earnings also has a direct effect on the welfare of z^*-earners, since they misoptimize due to present bias—a "welfare internality." The change in welfare from the local earnings adjustments due to the reform through the substitution effect is equal to

$$dS_W = -d\tau\epsilon h(z^*)z^* \mathbb{E}\left[\varepsilon(i)g(i)\left(1 - \beta_i\right) | i \in \mathcal{I}(z^*) \right].$$

The formal derivation is provided in Appendix A.3. The economic intuition behind this expression is straightforward: as in the preceding subsection, only a fraction β_i of the benefits of work are internalized by the individual—thus the remaining share $1 - \beta_i$ represents the welfare internality which is not taken into account. This expression can be simplified, and rendered more transparent, by invoking Assumption 1.1 to factor out the marginal social welfare weight, and by defining

$$\sigma_\varepsilon(z^*) = Cov\left[\frac{\varepsilon(i)}{\overline{\varepsilon}(z^*)}, \frac{1 - \beta_i}{1 - \overline{\beta}(z^*)} \middle| z(i) = z^* \right],$$

15

the covariance between the relative magnitude of elasticity and relative bias among z^*-earners. Then dS_W can be rewritten as

$$dS_W = -d\tau\epsilon h(z^*)z^*\overline{g}(z^*)\overline{\varepsilon}(z^*)\left(1-\overline{\beta}(z^*)\right)\left(1+\sigma_\varepsilon(z^*)\right). \tag{1.5}$$

The term $\sigma_\varepsilon(z^*)$ highlights the key role of heterogeneity. If more biased individuals earning z^* are also more responsive to the tax reform (high $\sigma_\varepsilon(z^*) > 0$) then the corrective term is inflated.[10] As shown formally in Appendix A.2, if $v(\cdot)$ has an isoelastic functional form, then the compensated elasticity of taxable income is constant across individuals with the same level of earnings, and thus the covariance is zero.

A third effect of the reform is an intensive margin change in earnings due to an income effect, equal to $dz = -d\tau\epsilon\frac{\eta(z)}{1-T'(z)}$, among individuals earning more than z^*. If leisure is a normal good, η is negative—individuals raise their earnings in response to their increased tax burden. The resulting *income effect fiscal externality* is

$$dI_F = -d\tau\epsilon\int_{z^*}^{\infty}\eta(z)\left(\frac{T'(z^*)}{1-T'(z^*)}\right)h(z)dz.$$

As with the substitution effect, the income effect also generates a welfare internality. I denote the covariance between the relative income effect and bias—analogous to $\sigma_\varepsilon(z^*)$ above—as

$$\sigma_\eta(z^*) = Cov\left[\frac{\eta(i)}{\overline{\eta}(z^*)}, \frac{1-\beta_i}{1-\overline{\beta}(z^*)}\bigg|z(i) = z^*\right].$$

which, again employing Assumption 1.1, can be written

$$dI_W = -d\tau\epsilon\int_{z^*}^{\infty}\overline{g}(z)\overline{\eta}(z)\left(1-\overline{\beta}(z)\right)\left(1+\sigma_\eta(z)\right)h(z)dz.$$

As with the substitution effect, $\sigma_\eta(z)$ is equal to zero if $v(\cdot)$ has an isoelastic form.

Finally, the reform generates a change in labor force participation as some individuals with $z > z^*$ drop out of the labor force in response to the increase in taxes. These effects can be written in terms of the participation tax rate $\overline{T}(z) = \frac{T(z)-T(0)}{z}$. The measure of z-earners

---

[10]This is analogous to the observation in Allcott and Taubinsky (forthcoming), in the domain of commodity taxes, that the relevant measure of bias for policy is the average bias weighted by the elasticity of demand response of marginal consumers.

who leave the labor force is equal to $-d\tau\epsilon\frac{\rho(z)h(z)}{(1-\overline{T}(z))z}$, so the resulting fiscal externality is

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

To write the corresponding welfare internality in the case of this discrete change, 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,

$$g_{ext}(i) = \frac{\alpha_i \left(u(z - T(z)) - u(-T(0)) \right)}{z - T(z) + T(0)} \cdot \frac{1}{\lambda},$$

and to let $\mathcal{I}_{ext}(z^*)$ denote the set of individuals indifferent between earning z^* and exiting the labor force at the optimum.[11] Then, as shown in Appendix A.3, we have

$$dP_W = -d\tau\epsilon \int_{z^*}^{\infty} \rho(z)\overline{g}_{ext}(z) \left(1 - \overline{\beta}_{ext}(z) \right) h(z)dz.$$

At the optimum this reform must generate no first-order increase in welfare, implying that $dM + dS_F + dS_W + dI_F + dI_W + dP_F + dP_W = 0$.

This assumption holds if the marginal social utility of consumption depends only on one's level of consumption. This is the case if the α weights are constant, or more generally, if the weights are assumed to be selected so that $g(i)$ is constant conditional on consumption at the optimum. (Thus this assumption need not rule out the special case of quasilinear utility of consumption with welfare weights declining with income.)

Then using the notation $\overline{\beta}_{ext}(z^*) = \mathbb{E}[\beta_i | i \in \mathcal{I}_{ext}(z^*)]$, denoting the average degree of present bias among those indifferent on the extensive margin, the optimal income tax is characterized implicitly by the following proposition.

Proposition 1.2. *At the optimum, if Assumption 1.1 holds, then the optimal income tax satisfies*

---

[11]Formally,

$\mathcal{I}_{ext}(z^*) = \left\{ i | v'(z^*/w_i)/w_i = \beta_i u'(z^* - T(z^*)))(1 - T'(z^*)) \cap -v(z^*/w) - \chi_i + \beta_i u(z^* - T(z^*)) = \beta_i u(-T(0)) \right\}.$

the following condition at all points of differentiability:

$$\frac{T'(z^*)}{1 - T'(z^*)} = \frac{1}{h(z^*)\overline{\varepsilon}(z^*)z^*} \times$$

$$\left\{ \int_{z^*}^{\infty} (1 - \overline{g}(z))h(z)dz - \int_{z^*}^{\infty} \overline{\eta}(z) \left(\frac{T'(z)}{1 - T'(z)} + \underbrace{\overline{g}(z)\left(1 - \overline{\beta}(z)\right)\left(1 + \sigma_\eta(z)\right)}_{I_W \text{ correction}} \right) h(z)dz \right.$$

$$\left. - \int_{z^*}^{\infty} \rho(z) \left(\frac{\overline{T}(z)}{1 - \overline{T}(z)} + \underbrace{\overline{g}_{ext}(z)(1 - \overline{\beta}_{ext}(z))}_{P_W \text{ correction}} \right) h(z)dz \right\} - \underbrace{\overline{g}(z^*)(1 - \overline{\beta}(z^*))(1 + \sigma_\varepsilon(z^*))}_{S_W \text{ correction}}.$$

$$\text{(1.6)}$$

This expression resembles the standard first-order condition for optimal tax rates with participation and income effects, with the addition of the terms involving β, which result from individual misoptimization. Although the parameters in (1.6) are endogenous to the tax code, they provide a transparent guide to the effect of present bias on tax rates.

The *substitution correction* (S_W) reduces marginal tax rates at z^* to induce z^*-earners to raise earnings, thereby correcting their tendency to underwork. The *participation correction* (P_W) reduces marginal tax rates at z^* to lower the level of taxes at higher earnings, which induces consumers to enter the labor force at those incomes. The *income effect correction* (I_W) works in the opposite direction—it tends to raise the optimal marginal tax rate (recall that η is nonpositive) since raising the level of taxes on higher earners induces them to beneficially increase their labor supply through an income effect. In the absence of participation and income effects, and if $\sigma_\varepsilon = 0$, this expression is equivalent to the simpler representation in Proposition 1.1.

Although this result provides guidance about the relationship between present bias and optimal tax rates, it also highlights the importance of estimating present bias *conditional on income*—a question taken up in Section 1.4.

1.2.3 Optimal tax timing

The two period model in Section 1.2.2 is useful for its simplicity, but since consumption and labor each occur in only a single period, it rules out some policy instruments, such as tailoring the *timing* of tax payments. This is an area of substantial policy interest, discussed more in Section 1.7.

To compare alternative timing regimes, this subsection considers an infinite horizon model in which the optimal steady state income tax is characterized by the same first-order condition as in Proposition 1.2. This setup generates two surprising insights. First, if all compensation is delayed, then tax delays are approximately irrelevant. (Irrelevance holds exactly under $\beta\delta$ quasi-hyperbolic discounting with $\delta = 1$, and otherwise holds approximately.) This finding contrasts with the common intuition that since present biased individuals discount the future, work subsidies should be paid as soon as possible for maximal corrective effect. Intuitively, taxes can subsidize earnings only at or after the time of compensation—but not before—implying that they are discounted by approximately the same amount regardless of their timing. Second, if some employers compensate their workers up-front (at the time of work) then it is beneficial to *delay* work subsidies. These results highlight the advantage of adopting a precise behavioral model for misoptimization, as they are not apparent from a generic misoptimization wedge.

Steady state model with delayed compensation

I begin with the case in which all compensation is delayed by a single period and taxes are levied at the time income is earned. Individuals are again indexed by i and choose earnings $z^t(i)$ each period t. As before, the planner selects a nonlinear income tax, and the problem is assumed to give rise to a steady state in which each individual i earns $z^*(i)$ each period, and the income tax $T(z)$ is constant over time. Individuals choose earnings to maximize the quasi-hyperbolically discounted stream of future period utilities, taking their own steady

19

state labor supply in other periods as given, so that $z^*(i)$ satisfies

$$z^*(i) = \arg\max_z \left\{ U_i(z^*(i) - T(z^*(i)), z) + \beta_i \left[\delta U_i(z - T(z), z^*(i)) + \sum_{t=2}^{\infty} \delta^t U_i(z^*(i) - T(z^*(i)), z^*(i)) \right] \right\}.$$

Here δ represents the rational (exponential) discounting factor, as in the standard β, δ model (Laibson, 1997). Period length is selected to reflect the length of the "present," which receives full psychological weight—in light of Kaur *et al.* (2015), this might be a week or two.[12] Thus period length is sufficiently short that there is likely very little rational discounting, and so I will use the approximation $\delta = 1$ in what follows. Although the stream of future utility does not converge for exactly $\delta = 1$, the corresponding first-order condition remains well-defined:

$$v'(z^*(i)/w_i)/w_i = \beta_i u'(z^*(i) - T(z^*(i)))(1 - T'(z^*(i))), \tag{1.7}$$

and therefore we can define behavior as that which arises in the limit as $\delta \to 1$.

The government is assumed to maximize steady state period utility,

$$\mathcal{W} = \int \alpha_i U_i(z^*(i) - T(z^*(i)), z^*(i)) d\mu(i), \tag{1.8}$$

equivalent to maximizing the exponentially discounted stream of welfare in each period in the limit as $\delta \to 1$. I rule out government borrowing, so that the budget constraint requires that steady state taxes sum to zero.

Note that Equations (1.7) and (1.8) are identical to the individual first-order condition and the planner's objective function in the two period model from Section 1.2. Therefore letting ε, η, and ρ denote the compensated elasticity, income effect, and participation elasticity of steady state income with respect to steady state taxes (defined as in Appendix A.2) then Proposition 1.2 characterizes the optimal steady state tax policy without modification.

---

[12]This model can also be written to allow for more general time inconsistent discounting functions, although the key insights remain the same.

The benefits of delayed taxes

This setup permits tractably modeling taxes which are paid with a delay. Suppose first that all compensation is delayed by one period. Notationally, let $T_j(z)$ denote a tax that is levied with a delay of j periods. Then we can immediately verify the following proposition.

Proposition 1.3. *If all compensation is delayed and if $\delta = 1$, then the optimal income tax $T_j(z)$, and the resulting level of social welfare, is independent of j.*

Proof. The first-order condition for optimization under a tax with a delay of j periods is

$$v'(z^*(i)/w_i)/w_i = \beta_i \delta u'(z^*(i) - T(z^*(i)))(1 - \delta^j T'(z^*(i))).$$

Invoking the assumption that $\delta = 1$, this equation is identical to (1.7), and thus the problem has the same solution and resulting social welfare, regardless of j. □

This result should be understood as a limiting case, as there are likely to be other (unmodeled) costs of levying taxes with long delays, such as administrative or enforcement costs. In that case, this proposition should be understood to show that, contrary to a common intuition, the costs of delayed taxes are small and are not driven by present bias.

Now consider the possibility that compensation delays are heterogeneous, and that the tax authority, who observes only compensation and not effort, cannot distinguish between income paid with with different delays. (For simplicity, suppose delays vary across individuals, and are outside their control.) By the logic of the preceding proof, such heterogeneity is irrelevant if all compensation is delayed by *at least* one period: delayed compensation is discounted by β_i regardless of the length of delay.

However if some individuals are paid up-front (a delay of zero), then their behavior differs from that of other individuals with delayed compensation, with implications for optimal taxes, in two respects. First, their labor supply is unbiased, so the government has no desire to encourage greater labor supply. However the planner is willing to tolerate some oversupply of labor among these unbiased individuals (beginning from the Mirrlees optimum, the resulting welfare costs are initially second order) in order to achieve the

21

benefits of correcting bias among workers with delayed compensation. Second, unlike workers with delayed compensation, present biased workers with up-front payments *are* sensitive to tax delays. Subsidies paid (or taxes levied) up front are "in the present," and hence receive full weight, whereas delayed subsidies occur "in the future" and thus are discounted by β_i, muting their effect on labor supply response. Thus by employing delayed subsidies, the planner can dampen the inefficient overprovision of labor among unbiased workers who are paid up-front. In the notation of Proposition (1.2), delayed taxes induce a positive covariance between bias and the elasticity of taxable income ($\sigma_\varepsilon > 0$). Intuitively, delayed taxes are a more targeted corrective instrument, and the size of the corrective term is magified.

To formalize this result, suppose that a fraction ϕ of subsidies are delayed to the period following the corresponding income, with the remainder $1 - \phi$ paid at the time of compensation. Then we have the following proposition.

Proposition 1.4. *Starting from the optimal contemporaneous tax, if some present biased individuals are compensated at the time of work, social welfare is strictly rising with ϕ.*

Proof. Begin with the optimal $T_0(z)$, and consider the steady state effect on social welfare of a tax reform in which a fraction ϕ of taxes are levied with a delay of one period. Let π denote the fraction of individuals receiving up-front compensation. By the logic of the proof for Proposition 1.3, the behavior of the $1 - \pi$ individuals with delayed compensation is insensitive to ϕ. The first-order condition for steady state earnings $z^*(i)$ for an individual who is compensated up-front is

$$v'(z^*/w_i)/w_i = u'(z^* - T(z^*))(1 - (1 - \phi)T'(z^*) - \phi\beta_i T'(z^*))$$

$$= u'(z^* - T(z^*))[1 - T'(z^*)(1 - \phi + \phi\beta_i)],$$

where I have suppressed the dependence of z^* on i for brevity. The welfare effect of i's change in earnings $\frac{dz^*(i)}{d\phi}$ is

$$\frac{dz^*(i)}{d\phi} \left[\frac{u'(z^* - T(z^*))(1 - T'(z^*)) - v'(z^*/w_i)/w_i}{\lambda} + T'(z^*) \right] = \frac{dz^*(i)}{d\phi} T'(z^*) \left[1 - \phi g(i)(1 - \beta_i) \right].$$

Moreover, for sensible elasticities and income effects, $\frac{dz^*(i)}{d\phi}$ has the same sign as $T'(z^*(i))$—an increase in ϕ mutes the effect of taxes on earnings behavior, raising incomes if taxes are positive and lowering them if taxes are negative. Therefore the welfare effect from a change in i's earnings has the same sign as $1 - \phi g(i)(1 - \beta_i)$. Thus when $\phi = 0$, this effect is strictly positive. $\qquad\square$

The economic intuition behind this result is straightforward: beginning from optimal contemporaneous taxes, a perturbation toward delayed taxes has no first-order effect on welfare (since workers with up-front compensation are unbiased) but has a positive first-order effect on the budget constraint—reducing earnings for those receiving marginal subsidies and raising it for those paying positive marginal tax rates—and thus the reform is beneficial.

More generally, if the government chooses discretely between contemporaneous taxes and a delayed tax, we have the following corollary:

Corollary 1.2. *If $g(i)(1 - \beta_i) < 1$ for all i, then a delayed tax is strictly preferable to a contemporaneous one.*

The proof follows directly from the preceding one, which shows that if $g(i)(1 - \beta_i) < 1$, then the welfare change from raising ϕ remains positive even when $\phi = 1$, i.e. when all taxes are delayed.[13] This condition is easily satisfied under the schedules of welfare weights used in the simulations in Section 1.5, and those computed under the inverse optimum exercise in Section 1.6, for the range of estimates of β_i in Section 1.4.[14]

The desirability of delayed taxes is closely related to the role of the term σ_ε in Proposition 1.2. Intuitively, the planner favors tax instruments which induce a greater corrective response among individuals who are biased. When some individuals are paid up-front, and are

---

[13]Note that this condition is a sufficient condition—and quite a conservative one—since it implies that the welfare change from raising ϕ remains positive at $\phi = 1$. However even if this change were negative, e.g., if the optimal ϕ lies between 0 and 1, then it is possible that $\phi = 1$ dominates $\phi = 0$, in which case the planner would prefer a fully delayed tax to a fully contemporaneous one.

[14]Since sections 1.5 and 1.6 have maximal marginal social welfare weights 1.2 or less, then $g(i)(1 - \beta_i) < 1$ provided that $\beta_i > 0.2$—a condition which holds for all the estimates in Section 1.4.

thus unbiased, work subsidies induce them to overwork inefficiently—a second-order cost (starting from the Mirrlees optimum with no present bias correction) that the planner tolerates in exchange for the first-order corrective benefits from subsidies for present biased workers. By employing delayed taxes, the planner reduces such inefficient overworking, effectively allowing for more targeted corrections. This both raises welfare and raises the size of the corrective term in the expression from Proposition 1.2, implying larger work subsidies (or tax rate reductions).

Finally, although this exposition has been discussed in terms of different individuals, the same logic could be extended to apply *within* individuals who face multiple sources of earnings.[15] Intuitively, delayed taxes allow the government to encourage effort on tasks with delayed benefits, which are pursued too little (due to present bias), while inducing less overprovision of effort on tasks with up-front payments.

1.3   Sophisticated workers and commitment contracts

I now allow for the possibility that workers are sophisticated and can sign private commitment contracts with firms to mitigate present bias. For simplicity, I use the finite-horizon model in this case, although the results extend to the infinite horizon model discussed in the previous subsection. Commitment contracts are modeled by allowing for a "period 0" during which an individual and firm sign an "employment contract" which maps individual production y to the compensation paid by the firm, $z(y)$. In period zero, the individual discounts both utility from consumption and disutility of labor by β, so there is no present bias wedge between the two.

In some cases, the ability to sign contracts eliminates the need for corrective taxation. For example, if individuals have full information about their labor/leisure tradeoff and there are no restrictions on contracts, then this result holds trivially: the individual's "self 0" signs a contract promising the privately optimal level of labor effort, on penalty of

---

[15]Formally, this extension requires two additional assumption in order to align exactly with the preceding proofs: quasilinear utility of consumption, and separable disutility of labor effort from different income sources.

sufficiently large punishment, and the misoptimization wedge is zero. The model in this section relaxes the assumption of perfect information in period zero and of an unrestricted contracting space, thereby demonstrating circumstances in which corrective taxation is optimal even in the presence of private contracting. Nevertheless, it must be stressed that the optimality of corrective taxation is considerably more fragile when individuals are sophisticated and contracting is feasible than in the baseline model from Section 1.2. As will be seen in this section, the addition of contracting does not alter the necessary condition for optimal tax rates in Proposition 1.2, provided that $1 - \beta$ is understood to represent the degree of *uncorrected* present bias (i.e., the misoptimization wedge). Rather, this setup alters the interpretation of the parameters therein, conditional on income—for example, the misoptimization wedge is zero at high incomes, and the elasticity of taxable income is non-constant.

1.3.1 Model with commitment contracts

Utility is defined as in 1.1. In period 0, individuals have the option to sign a contract mapping individual production y to compensation z—in which case they are bound by that contract in periods 1 and 2—or to forego a contract and work without commitment as in 1.2. Firms are assumed to exist in a competitive labor market, so that all surplus is captured by workers, and firms can commit fully to contracts. Firms face a "hiring cost" or "vacancy cost" κ from creating and filling a position with a worker. Such costs were irrelevant in the previous section, since the deduction of such costs from workers' compensation could be subsumed into the fixed costs of work χ without loss of generality. For consistency, however, in this section I will assume that even workers who do not commit, and who flexibly choose to generate a product of y in period 1, receive compensation of $z = y - \kappa$ in period 2.

I assume that ability w_i and present bias β_i are known to both the firm and worker in period 0, either because the relationship is repeated, or because the firm can verify the worker's reputation (costs of which contribute to κ). Fixed costs $\chi_i \in \left[\underline{\chi}, \overline{\chi}\right]$ are assumed to be subject to some uncertainty in period 0, which resolves in period 1. Let $F_i(\chi)$ denote the

distribution of these costs, conditional on information in period 0, from which χ_i (private to i and unverifiable) is drawn in period 1; $F_i(\chi)$ is known to the worker and the firm.

Individual production $y(i)$ is selected by "self 1" and, if chosen freely, would be too low from the perspective of self 0. However self 0 can choose a labor commitment contract—which maps production y to compensation z—in order to alter self 1's incentives and potentially achieve a more desirable outcome. Thus the optimal contract maximizes the utility of the "principal," self 0, subject to a participation constraint imposed by firms (nonnegative profits), and an incentive compatibility constraint imposed by the "agent," self 1. Moreover, contracts are restricted by a limited liability constraint, which requires that $z(y) \geq \underline{z}$ for all y. For simplicity, I will assume $\underline{z} = 0$, so that firms cannot impose fines on workers, no matter their level of labor supply.

Since the task at hand is to model the optimal contract for a given individual, dependencies on i are suppressed. For notational convenience, let $\tilde{u}(z) = u(z - T(z))$ (embedding the tax function, which is taken as given), and let $\tilde{v}(y) = v(y/w)$. In what follows, I assume that $T(z)$ is convex or not "too concave," so that $\tilde{u}(z)$ is a concave function, and I assume that $\tilde{u}(0) > -\infty$. Formally, a contract can be written as a mechanism mapping realized fixed cost to production and compensation: $Y(\chi)$, $Z(\chi)$. The optimal contract for a given individual solves

$$\max_{Y,Z} \int_\chi \left[-\tilde{v}(Y(\chi)) - \chi \cdot 1\left\{Y(\chi) > 0\right\} + \tilde{u}(Z(\chi)) \right] dF(\chi) \tag{1.9}$$

subject to self 1's incentive compatibility constraint

$$- \tilde{v}(Y(\chi)) - \chi \cdot 1\left\{Y(\chi) > 0\right\} + \beta\tilde{u}(Z(\chi)) \geq$$
$$- \tilde{v}(Y(\chi')) - \chi \cdot 1\left\{Y(\chi') > 0\right\} + \beta\tilde{u}(Z(\chi')) \quad \forall \chi, \chi', \tag{1.10}$$

and subject to the firm's participation (IR) constraint of nonnegative profits,

$$\int_\chi (Y(\chi) - Z(\chi)) dF(\chi) - \kappa \geq 0, \tag{1.11}$$

26

and subject to the limited liability constraint on contracts:

$$Z(\chi) \geq 0 \quad \forall \chi. \tag{1.12}$$

The selection of the optimal contract can be simplified by way of the following lemma:

Lemma 1.1. *The optimal contract is characterized by a threshold $\chi^* \in \left[\underline{\chi}, \overline{\chi}\right]$, with $Y(\chi) = 0$ and $Z(\chi) = Z_0$ for all $\chi > \chi^*$, and with $Y(\chi) = Y^*$ and $Z(\chi) = Z^*$ for all $\chi < \chi^*$.*

Proof. See Appendix A.4. $\qquad\square$

Employing Lemma 1.1, the selection of the optimal contract amounts to the optimal selection of the parameters Y^*, Z^*, Z_0, and χ^*. Thus the optimal contract can be rewritten:

$$\max_{Y^*, Z^*, Z_0, \chi^*} F(\chi^*) \left[-\tilde{v}(Y^*) - \mathbb{E}\left[\chi | \chi < \chi^*\right] + \tilde{u}(Z^*) \right] + (1 - F(\chi^*))\tilde{u}(Z_0) \tag{1.13}$$

subject to nonnegative firm profits,

$$F(\chi^*)(Y^* - Z^*) - (1 - F(\chi^*))Z_0 - \kappa \geq 0, \tag{1.14}$$

and to incentive compatibility for self 1,

$$-\tilde{v}(Y^*) - \chi^* + \beta\tilde{u}(Z^*) \geq \beta\tilde{u}(Z_0), \tag{1.15}$$

and limited liability,

$$Z_0 \geq 0. \tag{1.16}$$

This setup gives rise to three distinct contracting regions, characterized by the following proposition:

Proposition 1.5. *For a given β and distribution $F(\chi)$, different levels of ability give rise to three distinct contract regions:*

1. *Full commitment. Above some ability level w^h, constraints (1.15) and (1.16) are both slack, and all individuals work. Everyone is paid $Y^* = Z^* - \kappa$, and the misoptimization wedge is zero.*

27

2. *No commitment. Below some ability level w^ℓ, no feasible contract is preferable to the outside option of working without commitment in period 1. In this case no contract is signed, and in period 1 individuals either work at self 1's preferred labor supply with a misoptimization wedge of $1 - \beta$, or do not work at all.*

3. *Limited commitment. At intervening levels of ability, bias is partially corrected, and the misoptimization wedge lies between 0 and $1 - \beta$.*

Proof. See Appendix A.5. □

1.3.2 A numerical example

To illustrate the dynamics of the model above, consider a simple numerical example with quasilinear utility of consumption, $u(c) = c$ and isoelastic disutility of labor effort necessary to generate earnings $v(y, w) = \frac{(y/w)^{(1+1/e)}}{1+1/e}$, where e is the elasticity of labor supply (with respect to a change in the tax keep rate) in a conventional model. As in Section 1.5, I set this elasticity equal to 0.3, and I assume a flat tax with $T'(z) = 0.3$ for all z. I assume a population with heterogenous ability and present bias, with w and β distributed independently and uniformly over $[\underline{w}, \overline{w}] \times [\underline{\beta}, \overline{\beta}]$. (See appendix A.6 for details of this example.)

Implications for labor supply behavior Before turning to the implications for optimal policy, it's worth noting two sharp predictions about the income response to tax reforms across the income distribution. First, the compensated elasticity of taxable income is U-shaped over the region where some individuals have limited commitment. Under the functional forms in this example, the income elasticity (compensated and uncompensated) is e for individuals with full commitment and no commitment, and it is zero for individuals with limited commitment. Second, the income effect exhibits an inverse U-shape over the income distribution. In this example, quasilinearity of utility from consumption implies that income effects are zero among individuals with full commitment and no commitment.

Figure 1.2 plots the average compensated elasticity of taxable income and the income effect as a function of income. The income effect, unlike the compensated elasticity, is not

constant conditional on income, and the black line smoothes the simulated results using kernel regression. Mean income effects at each simulated income grid point are plotted by gray points in Figure 1.2. I also plot a kernel regression to smooth this relationship (with a bandwidth of 0.7), which illustrates the inverse U-shape relationship between the income effect and income.

While modeling labor supply in the presence of limited commitment contracts is not the central focus of this paper, these patterns are consistent with two puzzling patterns from the labor supply literature. First, some evidence suggests elasticities are U-shaped. For example, Chetty, Friedman and Saez (2013) finds an elasticity of 0.31 in the phase-in region of the EITC, and an elasticity of 0.14 in the phase-out region. Many estimates place the elasticity at higher incomes at rather higher values, suggesting a U-shape (see, e.g., Chetty (2012a), which favors an overall value of 0.33). Second, Meyer (2010) notes the puzzling absence of a reduction in hours worked among single mothers in the phase-out region following the EITC expansions in the 1990s. Such individuals experienced both an increase in benefits and an increase in marginal tax rates, suggesting that income and substitution effects should reduce hours worked—yet hours appear to be stable or to have increased slightly among this group. However if the income effect is positive and the compensated elasticity is low, as suggested by the middle region of the the income distribution in Figure 1.2, then such a reform might generate no change in hours, or even a positive change. Of course, more work is required to determine whether limited commitment contracts are indeed responsible for these patterns, but it is perhaps noteworthy that this simple model generates predictions consistent with observed patterns which otherwise appear puzzling.
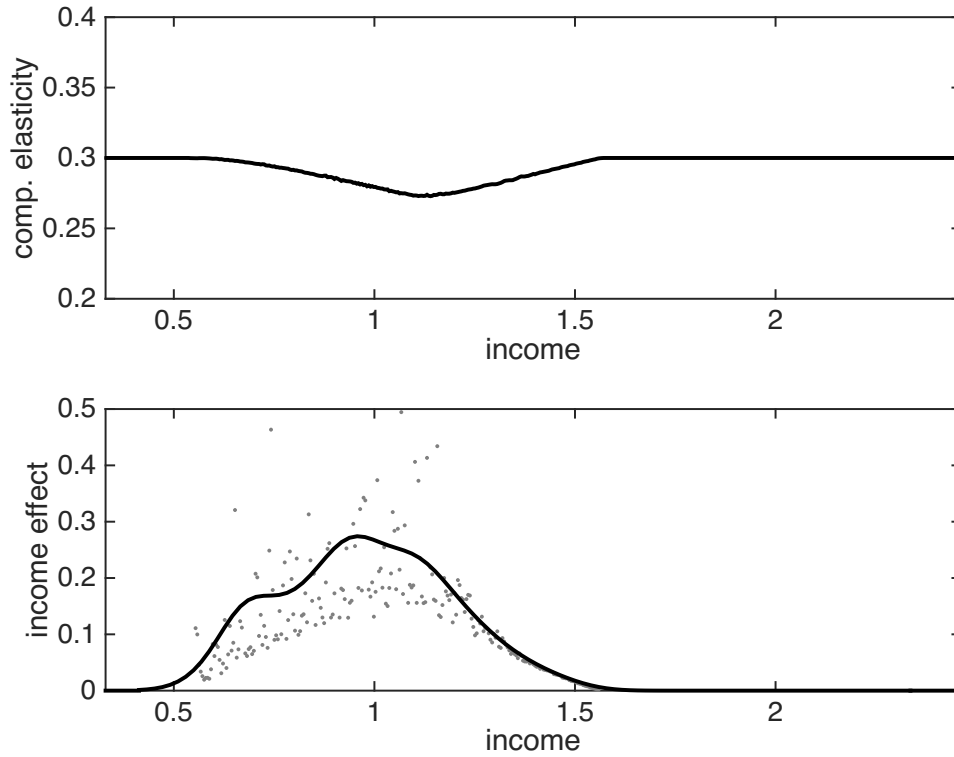
Figure 1.2: *This figure displays the compensated elasticity of taxable income (top panel) and the income effect (bottom panel) as a function of income. Income effects are not continuous in income (individuals with similar incomes but different combinations of β and w have different income effects), so the bottom panel plots both the income effects in the simulated population (the gray points) as well as a kernel regression (the black line) to show the smoothed inverse U-shape pattern.*

The misoptimization wedge. In the context of this model, the misoptimization wedge can be viewed as the residual or "uncorrected" present bias. For individuals in the "no commitment" region, the wedge is simply $1 - β$. For individuals with full commitment, the wedge is zero. And for individuals with limited commitment, the wedge lies between 0 and $1 - β$, reflecting the residual bias that remains after imposing the limited commitment contract.

Figure 1.3 plots the average misoptimization wedge conditional on income for this numerical example. Although the specific shape of the declining wedge in Figure 1.3 is the result of the assumed uniform joint distribution between ability and bias, the general shape—high at low incomes, declining to zero at high incomes—is a robust feature of this model.

30
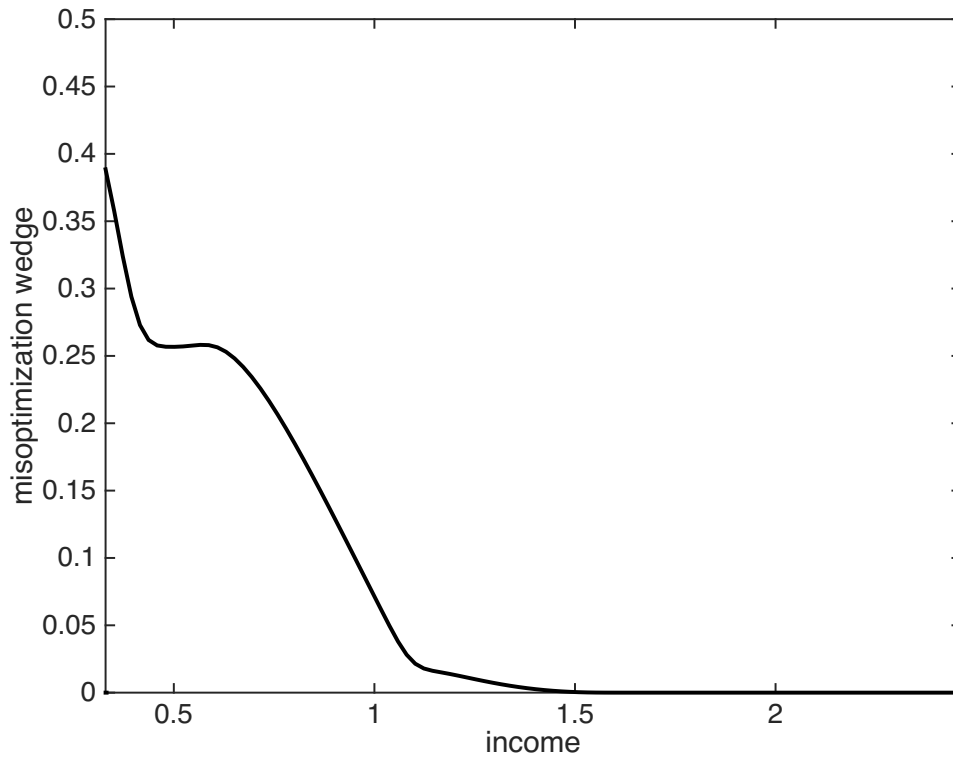
Figure 1.3: *This figure displays the average misoptimization wedge conditional on income. The figure has again used a kernel regression to generate a smooth function of average bias by income.*

This figure highlights a number of features which will prove relevant for policy design in the next section. First, the wedge is quite large at the bottom of the income distribution, both because all such low earners lack labor commitments, and because within that group, individuals sort on income so those with the greatest bias earn the lowest incomes. At moderately low incomes, where all workers are still uncommitted, the bias plateaus at the mean level of $1 - \beta$. The wedge then declines rapidly with income, as the share of workers with limited commitment and full commitment rises rapidly over this income range. Above the maximum income earned by workers without commitment, the wedge is quite small. This reflects the fact that although limited commitment is imperfect, most workers in the limited commitment range have largely mitigated their present bias, leaving little room for additional correction in this range. This suggests that the importance of the limited commitment individuals may be largely through their impact on elasticities and income

effects, rather than for corrective policy. Finally, the wedge is equal to zero at higher incomes where all individuals are in the full commitment region. Thus in this region the familiar optimal taxation results, such as the first-order condition in Saez (2001a) and the optimal top tax rate formula, apply without modification. In this sense, effects of present bias on optimal policy are most pronounced at low incomes, where workers are marginally attached (or tempted to become marginally attached) to the labor force, even if substantial present bias exists among higher income individuals.

1.3.3 Implications of commitment contracts for optimal policy

Although the structural model in this section is more complicated than that discussed in Section 1.2, the expression for marginal tax rates given in Proposition 1.2 remains approximately correct, provided that the income-conditional bias wedge and income-conditional elasticities and income effects from Figures 1.2 and 1.3 are inserted in the formula. The reason this is an approximation is that the formula requires independence between behavioral responses and the bias wedge, conditional on income, and although this is correct for the compensated elasticity, the income effect is larger for more biased individuals in the partial commitment contract region. (Intuitively, greater bias corresponds implies a greater need to adjust commitment to appease self 1's temptation to quit.) This inflates the benefit of targeting income to those who are partially committed. This effect is rather limited, however, by the small size of the misoptimization wedge in the limited commitment region (see Figure 1.3). Thus the expression in Proposition 1.2 remains a close approximation of the optimal tax in the face of commitment—and is exactly right at low incomes where no workers are bound by commitment.

Therefore the primary effect of incorporating private contracts is to account for the effect of commitment on the misoptimization wedge, and for the effect of contracts on the shape of elasticities. This highlights the importance of correctly measuring the "uncorrected" present bias conditional on income. In the next section, I present empirical evidence on the degree of structural present bias across income levels, as well as the extent of uncorrected residual

bias relevant for the determination of optimal policy.

1.4 Calibrating present bias

Measuring the misoptimization wedge is a key challenge in generating policy recommendations from the model in the previous section. Unlike elasticities, misoptimization wedges cannot be estimated directly from responses to income tax reforms (succinctly: revealed preference does not identify misoptimization) so evidence on their magnitudes must be drawn from other, sometimes unconventional, sources.

1.4.1 Estimates of structural present bias

Several recent studies quantify present bias. Table 1.1 reports estimates from a variety of such papers—details of their methodologies are discussed in Appendix A.7. I distinguish between two identification strategies: "choice-based" and "structural".[16] The choice-based approach identifies an environment in which choices are trustworthy (i.e., believed to reflect true utility) and assumes that similar preferences obtain in alternative environments where choices are believed to be suspect. In the context of present bias, choices over distant intertemporal tradeoffs are taken to be trustworthy; if they exhibit greater patience than choices over immediate tradeoffs, the difference identifies present bias.

Papers in the second category use structural models to estimate present bias, typically employing maximum likelihood methods to estimate a model with quasi-hyperbolic discounting. Further assuming that true utility is discounted time-consistently (exponentially), the estimate of β provides a structural estimate of the misoptimization wedge.

Table 1.1 also highlights papers with two desirable features. First, the column labeled "low income/EITC" indicates papers whose subjects are low earners, and in some cases EITC recipients, in the US. Such studies are useful for two reasons. First, by they mitigate concerns about external validity, as their subjects resemble the population of interest for

---

[16]These are two of the strategies for welfare analysis with behavioral models discussed in Chetty (2015); the third approach, direct measurement of utility, is employed below, in Section 1.4.2.

Table 1.1: *Estimates of structural present bias*

	low income/ EITC	labor supply	β estimate avg	β estimate low inc.	β estimate high inc.
Choice-based					
Augenblick, Niederle and Sprenger (2013)		✓	0.89	—	—
Kaur, Kremer and Mullainathan (2015)		✓	0.82	—	—
Jones and Mahajan (2015)	✓		0.34	—	—
Meier and Sprenger (2015)	✓		0.78	0.75	0.92
Structural					
Laibson, Repetto and Tobacman (2007)			0.70	—	—
Paserman (2008)		✓	0.65	0.40	0.89
Fang and Silverman (2009)	✓	✓	0.34	—	—
DellaVigna *et al.* (2015)		✓	0.59	—	—

Notes: Estimates of present bias conditional on income from a variety of sources and methodologies. See text for details.

the simulations in this paper. Second, monetary rewards are known to be problematic for the estimation of present bias, since subjects can use their own funds to replicate (or undo) experimental variation in payoffs. However 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 those studies.

Second, the column labeled "labor supply" identifies papers which are estimated using intertemporal tradeoffs over 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 below one. Moreover, the two studies which permit the estimation of present bias across income levels—Meier and Sprenger (2015) and Paserman (2008)—suggest that bias is concentrated among low earners. (See Figure 1.5 in the

next section for a graphical depiction of this relationship.) However the measures of present

bias in Table 1.1 are not necessarily the ones relevant for policy. The choice-based and

structural estimation methods in these studies identify the underlying ("structural") degree

of present bias—β_i in the notation from Sections 1.2 and 1.3)—rather than residual bias

after accounting for possible labor commitments in equilibrium. If workers are naive about

their bias, as in the model from Section 1.2, then these measures are identical. However if

workers are sophisticated and can sign commitment contracts, then the structural β_i will be

lower than the residual bias.

1.4.2 Estimation of residual bias

This section presents an alternative method for measuring present bias—one which which

identifies residual bias after commitment—using a strategy that measures a correlate of

utility directly (subjective well-being) in response to an induced change in one's leisure-

consumption bundle. Recall that residual bias is equal to the ratio of the decision utility

marginal rate of substitution (MRS) to the experienced utility MRS. This approach draws

from two sets of evidence to estimate each MRS separately.

The intuition behind this strategy is illustrated graphically in Figure 1.4. The figure

plots the space of consumption-labor bundles, which has upward sloping indifference

curves. When agents are present biased, decision utility undervalues consumption, and thus

indifference curves are steeper for decision utility than for experienced utility. At a given

consumption bundle, for example point A in the figure, the ratio of the slope of decision

utility to the slope of experienced utility quantifies the misoptimization wedge, and is equal

to $1/\beta$. The decision utility MRS can be measured using standard price theory logic: if A is

freely chosen (subject to the individual's budget constraint), then the decision utility MRS at

A is equal to the net wage, which is observable.

Measuring the experienced utility MRS is more difficult. Suppose the agent's budget

constraint is altered by a tax reform, after which the agent is observed to choose point

B, and suppose further that the individual's level of experienced utility was known to be
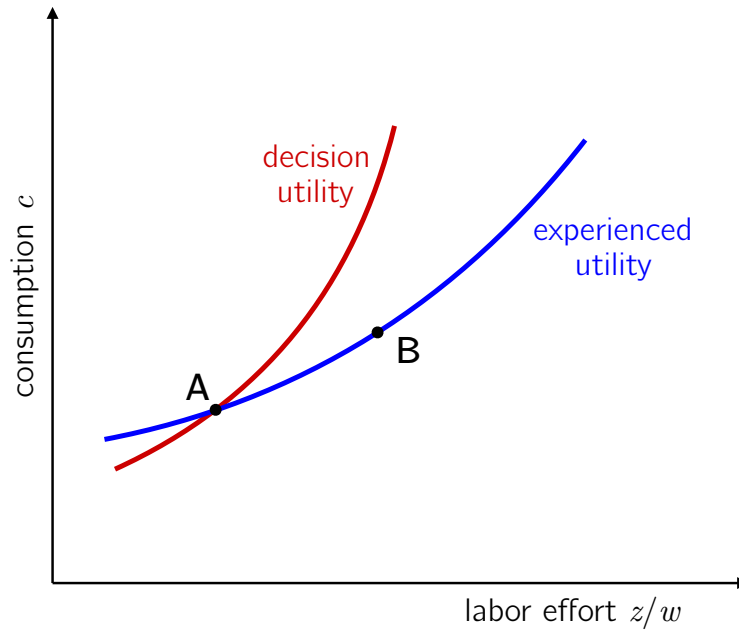
Figure 1.4: *This figure illustrates the identification strategy used to estimate residual present bias. Points A and B denote the consumption-labor bundles of low income single mothers before and after the 1990s welfare reforms in the US. The decision utility MRS at point A is estimated using the net wage prior to the reform. The experienced utility MRS at point A is approximated by the slope of the line connecting points A and B, employing the fact that the data reject a decline in reported subjective well-being among low income single mothers over the reform period.*

the same at points A and B. Then by implication, A and B lie on the same experienced utility indifference curve, and the slope of the line connecting the points approximates the experienced utility MRS at point A. This slope can be computed as the ratio of the change in labor effort (hours worked) to the change in consumption.

This reasoning also illustrates why this method provides an estimate of *residual* bias, rather than structural bias. If point A is the result of a decision process that involves a commitment device which partially corrects structural bias, then the individual's "self 1 indifference curve" passing through A will be steeper even than the decision utility indifference curve plotted in the figure. Estimates of structural bias, as in the preceding section, measure the ratio of the slope of that steeper indifference curve to the experienced utility indifference curve, implying a greater degree of bias than is actually present after commitment. In contrast, this section estimates residual bias by using variation generated by

the 1990s welfare reform in the US, with points A and B representing the consumption-labor bundles of low income single mothers before and after the reform. I perform the estimation separately for each of the bottom five income deciles of single mothers, effectively treating each decile as a single individual.

It's worth stressing several important limitations of this exercise. First, the assumption that the decision utility MRS equals the pre-reform net wage assumes that workers were at an interior optimum. If they were constrained from working as much as they desired prior to the reform, the net wage may overestimate decision utility MRS. The fact that Meyer and Sullivan (2008) find labor supply increased in response to the reform may provide some indication that higher labor supply choices were feasible. Although economic growth in the US during the late 1990s may have relaxed such constraints during that period, Meyer and Sullivan (2008) find similar labor supply increases when comparing the 1993–1995 period to either the 1997–2000 period or the 2001–2003 period, the latter of which should capture the ensuing economic contraction of the 2001 recession.

Second, this approach assumes that the consumption and labor supply measures examined by Meyer and Sullivan (2008) capture the variables relevant for experienced utility. This assumption is accurate if expenditures and hours worked are perfectly measured, if non-work time is spent enjoying leisure, and if agents live hand-to-mouth without saving— or, more precisely, if these potential confounds are not differentially active before vs. after the reforms. However if single mothers began saving more after the reform, for example, then total disposable income may have increased more than is indicated by the rise in consumption measured by Meyer and Sullivan (2008), biasing downward our estimate of the experienced utility MRS. Finally, this strategy is subject to the usual caveats associated with subjective well-being data.

At the same time, this estimation is a conservative estimate of bias in two important respects. First, the straight line connecting A and B is actually an overestimate of the experienced utility MRS at A, due to curvature of the indifference curve. Given the large measured change in consumption and hours worked due to the reform (about $1,500 and 550

hours, respectively) this curvature may not be trivial. Second, the calculation above assumes that experienced utility was known to be equal at points A and B. As will be seen below, evidence suggests that reported subjective well-being actually *increased* for low income single mothers after the reform.[17] This reasoning suggests the calculations below represent a lower bound for residual bias. On the other hand, the studies of subjective well-being trends (cited below) are not able to identify trends separately within each decile—further motivating the use of a conservative assumption on the change in well-being.[18]

In spite of these limitations, it is also worth emphasizing the general difficulty of measuring residual bias—yet since this parameter is ultimately relevant for policy, even coarse estimates provide useful information, and help test whether the estimates of structural bias appear to wildly different. Additionally, despite the notable limitations, the present exercise avoids some concerns present in previous attempts to quantify bias using subjective well-being data. For example, Gerritsen (2015) estimates the experienced utility MRS without a source of quasi-experimental variation, and notes concerns about omitted variable bias.[19] Still, as subjective well-being data improves, employing this approach with larger data sets and across other natural experiments will no doubt generate more precise estimates of bias.

I now turn to the data sources used to implement this estimation strategy. The 1990s welfare reforms imposed time limits and restrictions on "lump sum"-like benefits, such as Aid for Families with Dependent Children, while augmenting work subsidies for single parents via an expanded EITC. Meyer and Sullivan (2008) examine the effects of the reform using data from the Consumer Expenditure Survey from 1993 to 2003, finding that both labor supply (annual hours worked) and consumption increased substantially for single

---

[17]The data reject a decrease in single mothers' subjective well-being with $p < 0.01$ across all baseline specifications.

[18]Ifcher (2010) finds larger relative increases in reported subjective well-being when the sample is restricted to single mothers with only a high school education or below, suggesting the rises in SWB are unlikely to be concentrated especially at higher incomes within these groups. This provides some assurance that the operative assumption of an increase in SWB from point A to B is, if anything, especially likely to hold for lower income deciles, for whom residual bias is measured to be greatest.

[19]For example, if unobservable temporary depression spells reduce subjective well-being and one's willingness to work, it may appear that low labor supply *causes* lower happiness, spuriously suggesting misoptimization.

mothers over this period. They compute the resulting implicit net wage by dividing the increase in consumption by the increase in hours spent working. The implicit net wages for the bottom five deciles of the consumption distribution among single mothers are $2.11, $2.69, $3.21, $4.67, and $6.75 (in 2010 dollars). These figures correspond to the slopes of the lines connecting points A and B for each income decile. As the authors note, these findings suggest that if single mothers valued their time near their market wages (which were substantially higher), the reform made them worse off.

The second set of evidence is data on the evolution of subjective well-being among single mothers over this time period. Ifcher (2010) examines the evolution of reported subjective well-being among single mothers with children (who generally qualify for the EITC) relative to single childless women and to single men, before and after the substantial EITC expansion in 1996. The paper uses data from the General Social Survey, which since 1972 has asked respondents "Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?" Using a difference-in-differences analysis, the paper finds that self-reported happiness increased among single mothers after the welfare reforms of the 1990s, relative to both single women and single men without children. Consistent with these findings, Herbst (2013) uses data from a different survey with a shorter history (since 1985) but with a larger survey sample and state identifiers, and concludes that the reforms in the 1990s raised subjective well-being among single mothers. Boyd-Swan, Herbst, Ifcher and Zarghamee (2013) and Ifcher and Zarghamee (2014) also fail to find the decrease in reported happiness among single women that might be predicted from the findings in Meyer and Sullivan (2008).

Invoking the assumptions, noted above, that (1) single mothers working in the pre-reform period were at an interior optimum and (2) the subjective well-being evidence indicates that single mothers' happiness weakly increased within each consumption decile after the reforms, then the implicit net wages computed from Meyer and Sullivan (2008) represent upper bounds for the experienced utility MRS.

To compute the net wage of the bottom five consumption deciles of single mothers

during the pre-reform period, I again turn to the CPS. To align with Meyer and Sullivan (2008), I use the sample of women respondents to the 1993–1995 March survey waves, restricted to women who are not married with a spouse present, who have at least one child, and who have positive wage income and total household income. I first construct measures of market income (total household income less welfare, unemployment insurance, social security and disability income) and net income (market income less federal income and payroll taxes net of credits, plus food stamps, and welfare income). I then divide the sample into deciles sorted by net income, to represent the deciles of single mothers (by consumption) in Meyer and Sullivan (2008), and I compute the net wage within each decile. The net wage is simply the individual's gross wage—equal to their self-reported wage, plus the employer portion of payroll taxes—multiplied by one minus the effective marginal tax rate. I estimate the effective net marginal tax rate by performing a 5th order polynomial regression of net income on market income, then predicting the local marginal tax rate at each level of market income. This approach appears to accord with the expected pattern of changes in net marginal tax rates due to the 1990s welfare reforms—the estimated net marginal tax rate on the bottom decile prior to the reform is 46%, whereas the same calculation on the analogous sample from the 2001–2003 March survey waves yields a negative net marginal tax rate.

The resulting estimates of bias are displayed in Figure 1.5, plotted along with the estimates of structural bias across income for the papers which permit such an estimation. As with estimates of structural present bias, residual bias appears to be heavily concentrated at low incomes. This is consistent with the prediction from the model with commitment in Section 1.3 that commitment contracts are most difficult to sustain among low earners, and thus that residual bias will be greatest in that region.

One distinction between the estimates of structural bias and residual bias is that the latter may encompass misoptimization for reasons other than present bias. If so, it may nevertheless be preferable to use estimates of residual bias, since the relevant statistic for optimal taxation is the *total* misoptimization wedge across all sources of bias. On the other
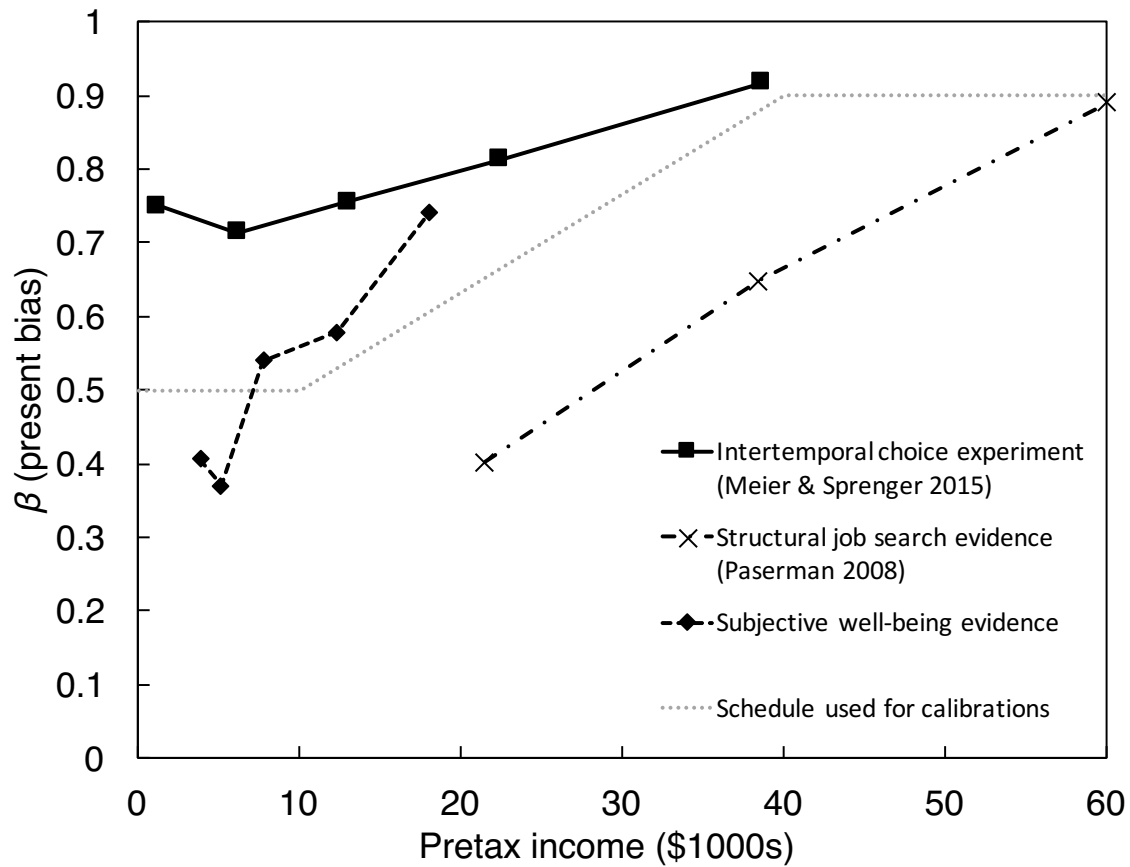
Figure 1.5: *Estimated relationship between income and present bias parameter β from three sources (see text for details). The dotted line shows the approximation used for the simulations below.*

hand, the similar shape of the schedules in Figure 1.5 suggests that present bias may be the primary source of bias relevant for labor supply misoptimization.

1.4.3 Implications for the misoptimization wedge

Figure 1.5 displays the resulting schedule of implied values of β for each of the data sources that identify bias across incomes. In each case, β is substantially below 1 at the bottom of the income distribution, and rises with income. The gray dotted line in Figure 1.5 shows the schedule of β across incomes used for the simulations to follow. For those with annual market incomes below \$10,000, I use a value of $\beta = 0.5$—approximately the average of the

beta computed at the lowest incomes in each data source. For middle and high earners, I set $\beta = 0.9$. (The average value of β among the highest earners in each data source is 0.85.) The appropriate range of transition between $\beta = 0.5$ and $\beta = 0.9$ is rather uncertain—I interpolate linearly between market incomes of \$10,000 and \$40,000, as shown in Figure 1.5.

This relationship is consistent with a number of possible explanations. First, theory predicts that present-biased individuals endogenously select lower earnings. Second, present bias likely reduces 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; Shah, Mullainathan and Shafir, 2012). I remain agnostic about the mechanism for the relationship, effectively assuming that the plot in Figure 1.5 indicates a stable type-specific level of bias as a function of underlying ability.

1.5 Numerical analysis

In this section I present the details of the simulated economy used to generate the tax policy displayed in Figure 1.1. I also explore the effects of alternative normative objectives and modeling assumptions, where I show that the optimality of negative marginal tax rates like those under the EITC depends on the planner having fairly modest redistributive motives, especially across low incomes. In Subsection 1.5.3 I show that although the optimal policy accounting for present bias is quite different from that with rational agents, the welfare gains from accounting for bias are relatively modest—a dollar-equivalent gain of less than 1% of median income. A reform accounting for bias tends to benefit the "working class" (those between the 20th and 50th percentiles of the income distribution), while lowering the welfare of those in the bottom quintile.

1.5.1 The model economy

The primitives of the simulated economy consist of a specification of individual preferences, the policy maker's redistributive tastes, and the distribution of types (skills and biases). I assume individual utility is quasilinear in consumption, with $u(c) = c$, and disutility of

labor supply is isoelastic, with $v(z/w) = \frac{(z/w)^{1+1/e}}{1+1/e}$. This form imposes two simplifications relative to the general form of individual utility in (1.1). First, it rules out income effects—a common assumption in the theoretical optimal tax literature (see, for example Diamond (1998a) and Saez (2002b)) which is consistent with empirical findings (Gruber and Saez, 2002; Saez *et al.*, 2012a). Second, it rules out discontinuous jumping from the extensive margin. I impose this restriction for two reasons. First, as shown by Jacquet *et al.* (2013), the addition of discontinuous jumping from the extensive margin does not generally rationalize negative marginal tax rates—rather it tends to generate a discontinuity in the tax function at zero, with marginal tax rates that look remarkably similar to the standard intensive margin model. So little is lost for the modeling of marginal tax rates, and some simplicity is gained, by omitting such discontinuities. Second, in practice many individuals whose responses are concentrated on the extensive margin tend to enter and exit the labor force multiple times per year. Indeed, data from the Current Population Survey indicate that even the lowest income decile of single mothers work an average of over 500 hours per year. When the extensive margin frequency is much higher than the tax period, it behaves more like an intensive margin of labor supply, and so I use an intensive margin as the benchmark model. I assume a fixed share of the population is disabled and has $w = 0$. To avoid complications of screening and optimal disability insurance, I assume that disability status is observable to the planner and that the required revenue for disabled individuals is exogenously given.[20]

I use a constant labor supply elasticity of 0.3—a value in the middle of empirical estimates (Chetty, 2012a; Saez *et al.*, 2012a). Note that this parameter is similar to the compensated elasticity of taxable income from Section 1.2, however it this does not account for the curvature of the tax code. Thus the expression for optimal taxes in terms of this labor supply parameter would depend on the virtual income density (as discussed in Saez (2001a)) rather than the actual income density as in Proposition 1.2.

---

[20]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. I assume the exogenously required income for disabled individuals is $7,500, equal to average SSI income in this age group in CPS. Thus disability insurance effectively contributes to the government's exogenous revenue requirement.

I assume individuals are present biased according to the calibration from Figure 1.5. This provides a mapping from income to ability, via the individual's first-order condition for labor supply choice. Therefore I can infer the ability distribution from the income distribution from the Current Population Survey in 2010. Importantly, I infer the ability distribution taking account of the assumed degree of present bias, so that each policy represents the optimum given the *observed* income distribution and elasticities. I restrict to households with positive total family income (thereby excluding those with declared business and farm losses), and I use kernel density estimation to calibrate the density across incomes. The first-order condition depends on the individual's marginal tax rate, which is estimated from CPS and NBER's TAXSIM.[21]

The planner's policy objective is as in (1.2). Marginal social welfare weights are simply equal to the α-weights, normalized to have a mean of one. Note that this is isomorphic to the specification of a concave transformation of individual utility, e.g. $\log(U(c, z, w))$, where the α-weights are set equal to the marginal utility of consumption at the optimum. For this reason, I'll adopt the conventional assumption that weights are declining as consumption increases. For the planner's budget constraint, I impose an exogenous government revenue requirement of $7,250 per capita, approximately equal to the net revenues currently raised by the federal income tax.

1.5.2 Optimal income taxes

This section demonstrates the key finding previewed in Figure 1.1: if redistributive tastes are modest, then negative marginal tax rates on par with those generated by the EITC may be approximately optimal.

For the baseline set of modest redistributive preferences, I select α-weights such that the

---

[21]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 then construct an approximate implicit marginal tax rate from the phaseout of benefits using CPS data by performing a kernel regression of the value food stamps and welfare income on market income and differentiating the resulting schedule. This is also how the tax schedule in Figure 1.1 is computed, restricted to households with 2 children. 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.
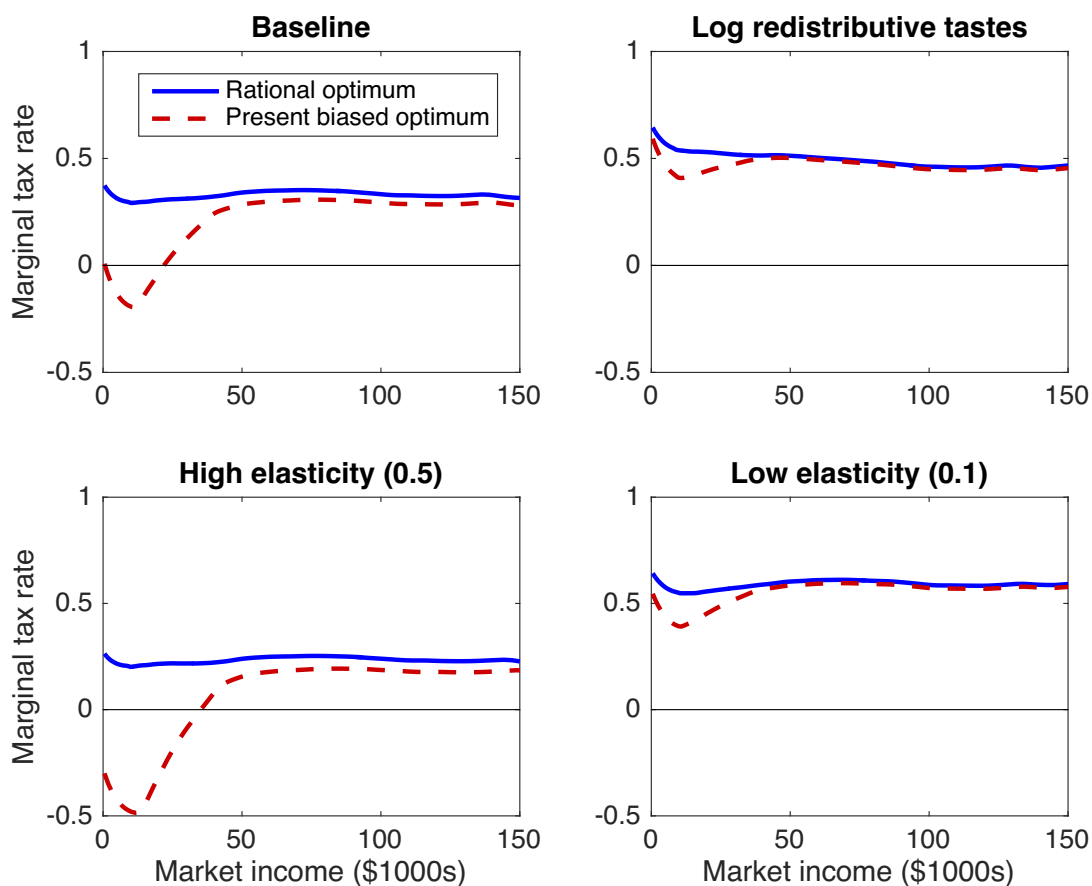
Figure 1.6: *Optimal income taxes for alternative calibrations. The baseline uses a labor supply elasticity of 0.3 and modest redistributive preferences (bottom earners have a welfare weight 10% above the median, top earners have a welfare weight of 40% below the median, linearly interpolated). The second panel uses logarithmic redistributive preferences (see text for details), while the bottom two panels use the same modest redistributive preferences with alternative higher and lower values of labor supply elasticities.*

lowest earners receive a weight of 10% more than the median household, and top earners weighted by 40% less than the median, linearly interpolated across intervening percentiles. 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 the tax code embodies such modest redistributive tastes, e.g. through rather low tax rates at the top of the income distribution Lockwood and Weinzierl (2016).

Figure 1.6 displays the schedule of optimal marginal tax rates for four simulations. The top left panel reproduces the baseline calibration in Figure 1.1. The top right panel shows

optimal taxes with logarithmic redistributive preferences, so that $\alpha(w) = U(c, z, w)^{-1}$ at the optimum. The bottom two panels use the baseline set of redistributive weights, with alternatively higher and lower labor supply elasticities.

These simulations highlight some key lessons for tax policy with present biased workers. First, as highlighted in the introduction, negative marginal tax rates like those under the EITC may be justified if redistributive preferences are modest. Second, as the log redistributive tastes case illustrates, if redistributive preferences are strong, marginal tax rates remain positive throughout the income distribution, although the inclusion of bias still tends to reduce tax rates below the optimum for unbiased individuals. This result may seem surprising in light of the fact that the bias term in Proposition 1.2 is 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. Since the corrective term operates on the term $\frac{T'}{1-T'}$, when marginal tax rates are high (close to one) in the rational case, this fraction is very large, and even a substantial corrective term has little effect.

The third lesson from Figure 1.6 relates to the labor supply elasticity. A higher elasticity has a strong effect on reducing optimal marginal tax rates in the present biased optimum. In fact an elasticity of 0.5, higher than baseline but still well within the range of some empirical estimates, particularly from the macro literature (see Chetty (2012a)) generates optimal marginal tax rates as low as -50%—subsidies much greater than the net rates generated by the EITC. It's worth stressing the divergence between this comparative static and the effect of higher intensive margin elasticities for the the EITC-like subsidies in Saez (2002b), 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 the lower right panel of Figure 1.6, then present bias does not generate marginal work subsidies at all. Of course, although these
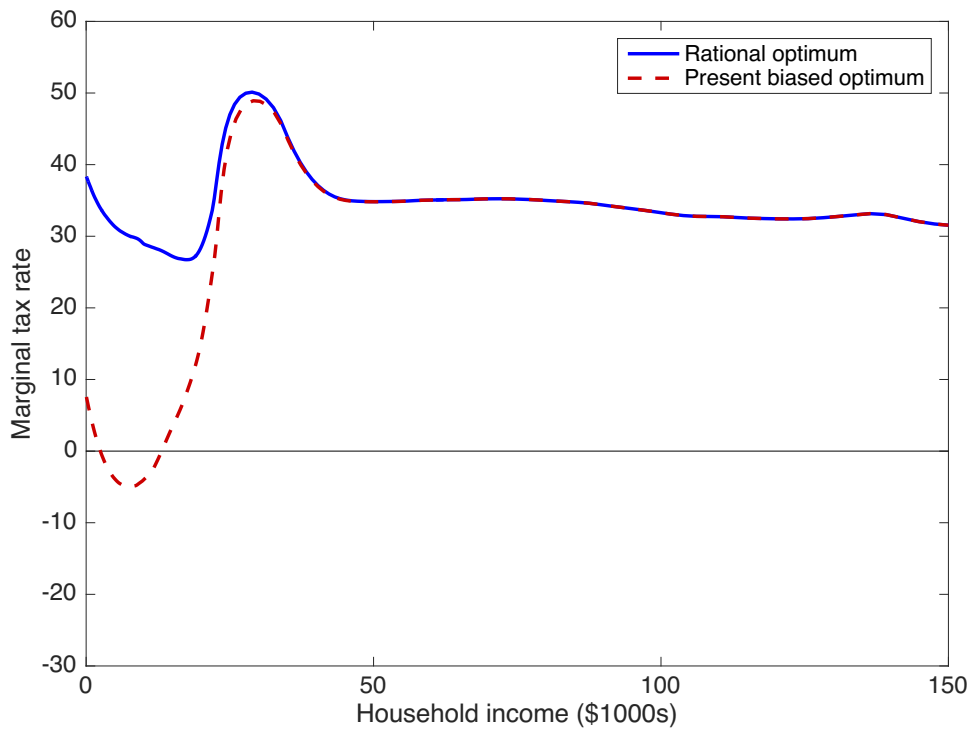
Figure 1.7: *Approximation of optimal marginal tax rates with sophisticated agents and commitment contracts.*

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 responsive to subsidies, they may be justified even if elasticities are fairly low at higher points in the income distribution.

To give a sense of the quantitative implications of contracting for the optimal income tax, Figure 1.7 modifies the simulation in the baseline economy in two ways. First, it is assumed that only a fraction of individuals are uncommitted at each level of income. I calibrate this share using the fraction of working individuals who work less than full time (2080 hours annually) in CPS. This is a coarse assumption, in the sense that some part time workers are likely committed to their privately optimal level of labor supply, while some who work 2080 hours or more may nevertheless be uncommitted along more flexible dimensions of effort (e.g., in exerting extra effort for a promotion). Still, this can represent a rough approximation of commitment in that many individuals who lose their jobs and spend some time searching for employment (thus, who are uncommitted at the margin)

supply fewer than 2080 hours annually.[22] Unsurprisingly, in part due to the mechanical relationship between hours worked and income, the share uncommitted by this measure is much higher at low incomes (over 90%) than at high incomes, where it falls to around 30%. The second change is to assume a U-shaped pattern of elasticities—rather than a constant value of 0.3, I assume an elasticity of 0.3 for individuals with earnings below $10,000 or above $40,000 (under the status quo US income tax), with a U-shape between these incomes, falling to 0.2 at $30,000, with spline smoothing in between.

It is worth stressing the illustrative nature of this calibration; it is ad-hoc in the sense that it assumes the share of workers and the elasticity is type-specific, rather than arising endogenously with a structural model of contracts. Nevertheless, the pattern of tax rates in Figure 1.7 is informative in two respects. First, marginal tax rates remain negative for a range of low incomes, although they are more muted than in the baseline specification in the baseline calibration. The reduction in the magnitude of the subsidy comes from the effective decrease in bias at low incomes, since a portion of low earners are now assumed to be setting labor supply correctly, and are in fact distorted by the corrective subsidies aimed at the biased individuals. Second, the U-shaped pattern of elasticities generates a hump in optimal marginal tax rates for the "working class"—those with incomes between $25,000 and $50,000. This suggests that the higher marginal tax rates in the phase-out region of the EITC—typically regarded as a necessary if undesirable feature of the credit by its proponents—may in fact be an optimal feature of the tax system.[23] This result is necessarily speculative in light of the substantial uncertainty about the correct model of firm and worker commitment contracts, but given the empirical evidence supporting a U-shaped pattern of elasticities, this possibility deserves greater exploration.

---

[22]Results look similar if I instead use individuals who work fewer than 52 weeks as a proxy for the share with uncommitted labor supply.

[23]As evident from the dotted line in Figure 1.7, these high phase-out tax rates are sufficiently diffuse across families with different incomes and tax situations that they do not appear in the smoothed net tax schedule.

1.5.3 Welfare gains

This section considers the welfare gains from accounting for the existence of present bias—that is, for reforming from the "rational optimum" to the "present biased optimum" in Figure 1.6. The results are displayed in Table 1.2, which reports the welfare gains that result from reforming from the "rational optimum" to the "present biased optimum" in each of the specifications from Figure 1.6. The gains are measured both in dollars and as a percentage of aggregate consumption.

Two features of these results stand out. First, the absolute size of the gains are modest—approximately $100 per household in the baseline specification. This finding is perhaps surprising in light of the large change in the tax code evident from Figure 1.6. One way to understand this "flatness" of welfare over the tax reform is through the reallocation of welfare across individuals. The reform to the optimal tax indeed raises efficiency by correcting the underworking of low earners, but to do so it sharply reduces the lump sum grant—from $19,500 to $6,900 in the baseline calibration. This reduces the welfare of the lowest earning individuals. However the low marginal tax rates mean that middle and upper-income individuals face a lower tax burden, and thus see their welfare rise. Thus while it is correct to say that in this framework marginal work subsidies "help" low earners overcome their bias by adding a corrective motive, when that motive is incorporated into the integrated optimum, it does not generate a net benefit for low earners relative to a policy which ignores present bias. (The implications of this conclusion for US policy may be rather limited, however, as the counterfactual optimum without present bias entails a far larger lump sum grant than exists in the US.)

The second notable feature of the welfare gains in Table 1.2 is the wide *variance* in the gains across calibrations—the dollar gain in the high elasticity calibration is over ten times larger than that in the low elasticity case. In general, welfare gains are larger when the elasticity of taxable income is high, and when redistributive motives across low earners are modest, since the efficiency gains from large marginal work subsidies on low incomes are particularly large in that case. This finding reinforces the importance, already prominent in

49

Table 1.2: *Welfare gains from accounting for present bias*

	Dollars (per capita)	% of aggregate consumption
Baseline	$105	0.14%
Log redistributive preferences	$19	0.03%
High elasticity (0.5)	$243	0.3%
Low elasticity (0.1)	$12	0.02%

Notes: Table displays estimated welfare gains from implementing optimal tax, accounting for present bias, relative to a policy which appears optimal if the policy maker does not account for bias.

the optimal taxation literature, of correctly measuring the elasticities at various points in the income distribution.

1.6 The "inverse optimum" approach and the EITC

The previous results have characterized the optimal tax schedule taking as given the redistributive preferences of the planner, through the marginal social welfare weights $g(z)$. Yet in reality there is great uncertainty about the appropriate set of redistributive preferences to apply for these calculations. One approach is to simply stipulate a schedule of marginal social welfare weights, with the caveat that the results are sensitive to them. This is the approach adopted, for example, by Saez (2001a, 2002b) and Jacquet *et al.* (2013).

In this section I adopt an alternative approach, inverting the optimal policy simulation by taking existing policy as given and computing the redistributive preferences with which those policies are consistent. This strategy, implemented by Bourguignon and Spadaro (2012) for European countries and further explored by Hendren (2014) and Lockwood and Weinzierl (2016), provides a reduced-form way to check whether existing policy generates weights which appear reasonable. Of course, the definition of "reasonable" is itself controversial in this context, but two features are commonly thought to be sensible requirements in the optimal policy literature: weights are positive at all incomes (Pareto efficiency) and weights are declining with income (redistribution toward lower abilities). In this section, I show

that the implicit weights in the EITC population which arise under the usual assumption of perfect optimization exhibit a robust "unreasonable" feature: weights rise substantially with income across the bottom quarter of the income distribution. Taken at face value, the weights suggest that current policy implicitly places greater value on a marginal dollar for the median earner than on a dollar in the hands of the poorest EITC-receiving households, typically working single mothers. I compare these weights to those which arise implicitly from a model which allows for misoptimization due to present bias.

A secondary strength of the inversion approach is that it permits a more detailed representation of the complexities of the actual economy. Since this approach involves only an inversion of the local first-order condition for optimal taxes, it does not require a structural model of earnings responses to non-local tax reforms. This is particularly beneficial if individuals can sign labor commitment contracts as in Section 1.3, since a structural model of that commitment process may be infeasible. As a result, it is possible to incorporate a rather more detailed calibration of elasticities, including participation elasticities and non-constant compensated elasticities of taxable income. This section can therefore be viewed as a sort of robustness check for the result, from Section 1.5, that after accounting for present bias, the EITC is consistent with plausible redistributive weights, even under a more detailed calibration of the economy, whereas the redistributive motives implied if one does *not* account for bias appear rather unconventional.

1.6.1 Calibrating the income distribution and tax schedule

In the interest of computing the weights generated by the EITC, I focus on the set of families affected by the credit. Therefore I use a sample different from the one in the benchmark economy of Section 1.5, though I continue to draw data from the CPS. The income distribution is drawn from CPS, using the March survey waves for the years 2001–2010. (All figures are reported in 2010 dollars, adjusted using the CPI-U.) I restrict to nonfamily householders and to households for whom the respondent is the head of household, between the ages of 25 and 55, and I restrict to households with two children

and with positive total income. I drop families with income below $30,000 who do not receive the EITC (23% of all households in that earnings range). A continuous income distribution is constructed by discretizing the income space into $1000 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, 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 average these marginal tax rates within each $1000 bin, and use these averages to compute marginal social welfare weights.

1.6.2 Parameter assumptions

In addition to the income distribution and schedule of marginal tax rates, the optimal tax condition depends on the compensated elasticity of taxable income, the income effect, and the participation elasticity, as well as the misoptimization wedge, conditional on income.

Beginning with compensated elasticities, I assume an elasticity of 0.3 at middle and high incomes—a central value in the range of existing estimates, and near the preferred value in Chetty (2012a) of 0.33 to reconcile existing micro and macro estimates in the presence of optimization frictions. 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 regions. Therefore I assume a compensated elasticity of 0.31 for households with incomes below $13,000 (the approximate upper bound of the phase-in region) and 0.14 for those with incomes between $13,000 and $40,000.

For the participation elasticity, Saez (2002b) performs calibrations using estimates of 0,

0.1, and 0.5 for the bottom half of the population (and zero for the top half). Chetty *et al.* (2013) estimates an elasticity of 0.19 among EITC recipients. To avoid a discontinuous drop in the participation elasticity conditional on income, I interpolate (linearly) between $\varepsilon = 0.2$ at the bottom of the income distribution, declining to $\varepsilon = 0$ at an income of $40,000$. I further assume that income effects are zero throughout the distribution (Gruber and Saez, 2002; Saez *et al.*, 2012a).

As in Section 1.5, I assume the schedule income-conditional present bias is as reflected in Figure 1.5.

1.6.3   Implicit welfare weights on EITC recipients in the United States

The resulting marginal social welfare weights are plotted in Figure 1.8, both under the conventional assumption of no misoptimization, and under the assumption that individuals are present biased. In each case, the plotted points represent the weight computed locally for each percentile of the income distribution, while the line plots the smoothed relationship, using kernel regression with a bandwidth of $5000.

The schedule of weights computed without present bias are strikingly unconventional in two respects. First, they rise over a substantial range of low incomes, peaking at about $30,000 in annual earnings. Since these EITC recipients are primarily single mothers, this result indicates that according to the conventional model, policy is designed to benefit single mothers at the low middle range of the income distribution much more than very poor working mothers—a result at odds with conventional normative assumptions. Similarly, the demographic homogeneity within this group of EITC recipients suggests that multidimensional heterogeneity, as in Choné and Laroque (2010), are an unlikely explanation for the pattern.

The second unconventional feature of the welfare weights computed under the conventional model is their low *level* on the poorest EITC recipients. Indeed, the weights suggest that policy has a strong preference for allocating marginal consumption to median earners (with welfare weights above 1.1) than to those with the very lowest incomes.

53
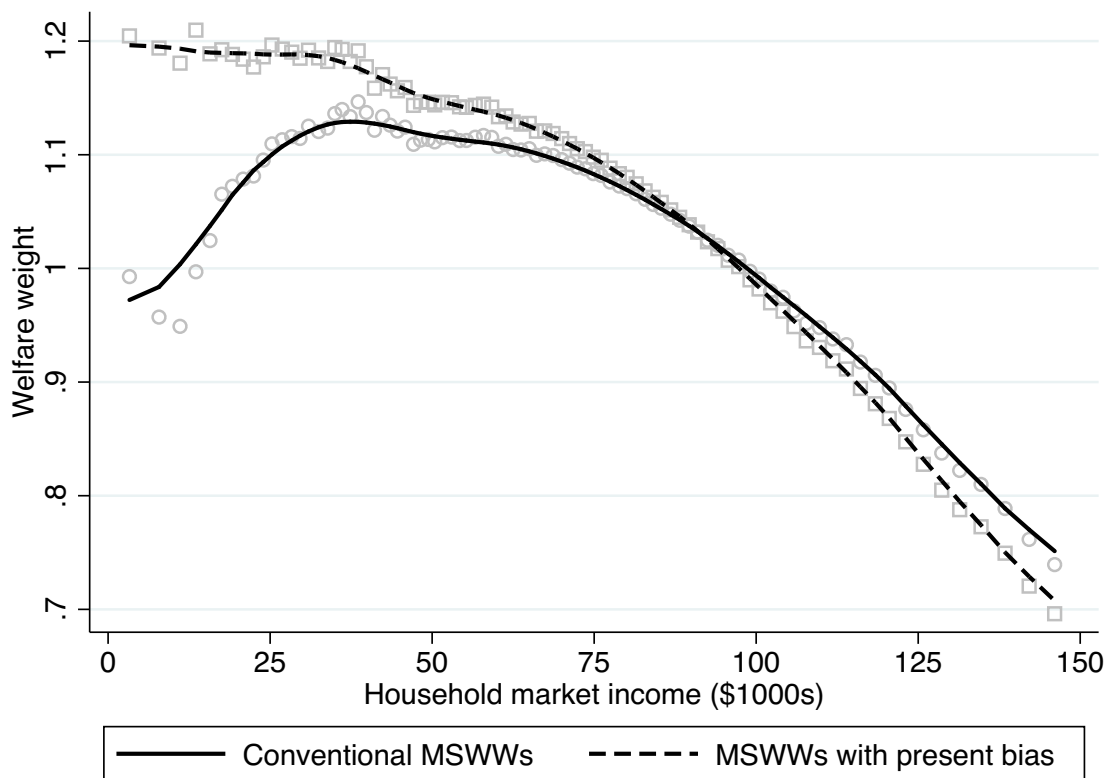
Figure 1.8: *Welfare weights implicit in US policy under the conventional assumption of perfect optimization, and under calibrated present bias. Weights are computed by smoothing the income density using a fifth-order polynomial regression, then computing weights locally within each percentile of the income distribution. Lines are generated using kernel regression with a bandwidth of $5000.*

As shown by the dashed line in Figure 1.8, these unconventional features disappear when a calibrated degree of present bias is incorporated into the calculation of welfare weights. Specifically, weights are substantially higher than 1 at the bottom of the distribution, and decline monotonically with income.[24]

These results capture the sense in which, accounting for present bias, the EITC is consistent with modest (but conventionally shaped) redistributive preferences. The dashed line in Figure 1.8 does not entail a strong desire to redistribute across low earners—indeed

---

[24]A third possibly unconventional feature, common to both schedules, is the rather high welfare weight on high earners, even though many familiar utility functions have the marginal utility of consumption declining to zero as income grows large. The analysis of revealed redistributive preferences for high earners is not the focus of this paper, but see Lockwood and Weinzierl (2016) for a discussion.

those with the lowest incomes have weights only slightly above the median—but the weights are nevertheless decreasing, consistent with the conventional objective of redistributing toward lower ability individuals.

1.7 Discussion

1.7.1 Implications for policy design

The analysis from the preceding sections allows the government only a single policy instrument: a nonlinear income tax. This restriction is useful for deriving policy implications that can realistically be implemented, avoiding complicated or unrealistic history-dependent policies that might arise from a full dynamic model. On the other hand, this constrained environment ignores some possible instruments which might be realistically feasible and useful for tailoring policies more specifically for present biased individuals. Here I speculate about two such instruments.

Timing of EITC payments. The current EITC is paid in aggregate at the end of the tax year—a structure that has generated mixed reactions. On one hand, spreading the credit across more frequent installments would help smooth consumption across the year and potentially alleviate liquidity constraints. Many EITC recipients carry substantial credit card balances, for example, and more frequent EITC payments would provide additional liquidity and could reduce costly interest expenses. 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—saving for which might otherwise prove difficult. Indeed, anecdotal evidence suggests EITC recipients do not *want* to receive distributed payments (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.)

The model in Section 1.2.3 suggests that it is beneficial to levy taxes with a lag—thus

the delayed nature of the present EITC is not as detrimental for combatting present bias as initial intuition might suggest. However that model recommends a rather different structure than the current one, very lumpy schedule of EITC payments. Rather than paying subsidies at the end of each year, subsidies should be smoothed with a small, constant delay—perhaps on the order of one month. (Indeed, Kaur *et al.* (2015) suggests substantial discounting occurs even within one week.)

How can this policy implication be squared with the seemingly desirable forced savings function of the current EITC? First note that existence of present bias is also consistent with the low observed uptake of the Advance EITC, and the lack of interest in up-front payments. As is well appreciated, theory predicts that sophisticated present biased agents will demand forced savings mechanisms. By paying the credit as an annual lump sum, the current EITC provides such a mechanism.[25] Nevertheless, whatever the merits of forced savings vehicles, there is no obvious benefit from bundling it with the corrective subsidy of the EITC. An unbundled policy would still allow sophisticated individuals to benefit from the EITC's corrective subsidy while choosing to divert a flexible share of income to a forced savings vehicle. On the other hand, naive individuals, who would not use the forced savings vehicle (and who might by repelled by that aspect of the current EITC) would, under an unbundled policy, nevertheless benefit from the corrective work subsidy.

This result sheds light on a number of policy discussions which have proposed reforming the EITC to provide more frequent payments. The Center for Economic Progress is currently exploring such a reform in its Chicago Periodic EITC Payment Pilot, wherein EITC payments are distributed quarterly rather than annually[26] and presidential candidate Marco Rubio has proposed replacing the EITC with a wage enhancement program which would effectively include marginal work subsidies directly in recipients' paychecks.[27] The logic in this paper

---

[25]Many EITC recipients make use of refund anticipation loans, at high implicit interest rates, to avoid the typical three to six week processing delay. If the "long run self" chooses the EITC Advance enrollment, while the "short run self" decides whether to take out such a loan, both low EITC Advance uptake and widespread loan use are consistent with a model of sophisticated present bias.

[26]See http://www.economicprogress.org/content/rethinking-eitc for details.

[27]See http://taxfoundation.org/blog/marco-rubio-proposes-replacement-earned-

suggests the former approach is likely to provide a more targeted and efficient correction of present bias, while also providing EITC-receiving households with greater liquidity throughout the year than the present annual EITC.

Benefits waiting periods. A second possible dimension of fine tuning policy involves the timing of "lump sum" like benefits, $-T(0)$ in the notation of the models from Sections 1.2 and 1.3. In practice, these payments typically vary across programs. One feature that may reduce their distortions, particularly in contexts where workers may be tempted to quit and immediately draw benefits, is a required "waiting period" before benefits can be claimed. Under the model with commitment from Section 1.3, for example, a larger grant $-T(0)$ undermines productive commitments between firms and workers—however if the grant were delayed, so that self 1 discounts it by β, then it would appear a less tempting alternative. (Of course, such a waiting period must be weighed against the harm of delaying benefits for recipients.)

This may provide an economic rationale for the substantial delays associated with qualifying for programs such as disability insurance. Although this paper abstracted from many complexities of disability insurance by assuming that disabled workers could be screened perfectly, if screening is imperfect, then such delays may help prevent present biased workers from enrolling inefficiently.

1.7.2 Limitations

This paper abstracts from several important complications. Here I discuss several simplifying assumptions and their implications.

No human capital acquisition. Although I allow for delayed benefits from labor effort, I assume these benefits are separable from later labor supply decisions. This effectively rules out considerations of human capital accumulation. I make this assumption for a number of

income-tax-credit.

reasons. The first is pragmatic: the complexities of adequately accounting for human capital acquisition in optimal taxation are substantial, even without incorporating misoptimization (Stantcheva, 2014). 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. Second, the model in this paper still allows for a reduced-form relationship between human capital and present bias by calibrating bias conditional on income. Indeed, results in Section 1.4 suggest that bias is concentrated among those with lowest ability, as would be expected if some component of ability variation is due to human capital in which present biased individuals underinvest. Third, there is some empirical evidence that 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.

To the extent that human capital effects are important for the design for optimal work subsidies, the directional implications are ambiguous. On one hand, some human capital is surely acquired on the job, raising the delayed benefits of work and likely inflating the size of optimal subsidies. On the other hand, human capital acquisition might other at times be a *substitute* for work with even more delayed benefits—for example, one may need to forego work to attend college. In that case work subsidies could exacerbate bias by discouraging human capital acquisition (see Shah and Steinberg (2015) for an application). This latter possibility may generate a rationale for exempting college-age individuals from EITC eligibility.

Perfectly competitive labor markets. In keeping with 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, Kucko, Lehmann and Schmieder (2015) incorporate endogenous wages and unemployment (not all job seekers find jobs) using a sufficient statistics approach. Their empirical results 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 (2002b). Considerations of inelastic labor demand are beyond the scope of this paper, though they represent a natural extension, particularly for exploring the implications for the firm side of the economy with endogenous commitment contracts.

Fixed present bias. I assume throughout the model and calibrations that although present bias may vary with ability, it is fixed within individuals. 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 *et al.* (2012) and Mullainathan and Shafir (2013), is that the conditions of poverty exacerbate behavioral biases. This possibility could be accommodated by writing bias as a decreasing function of consumption—a modification which would favor raising degree of overall redistribution. 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, recent work by Carvalho *et al.* (2014) 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 present bias is a promising avenue for further exploration.

1.8 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

attempts progress by focusing on a particularly robust and well understood source of misoptimization—present bias—which, recent evidence suggests, generates substantial labor supply distortions.

A theoretical 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 some individuals are paid up-front, then it is beneficial to levy taxes with a delay. A model with sophisticated agents and endogenous commitment contracts generates novel predictions for labor supply patterns—a U-shaped schedule of elasticities, and an inverse U-shaped schedule of income effects—which are consistent with patterns in the empirical literature.

A compilation of existing estimates of present bias provides strong evidence of such bias at a structural level. To allow for the possibility that bias is corrected by commitment contracts, I present new estimates of "residual bias" based on trends in subjective well-being responses following the 1990s welfare reforms in the US. Estimates of structural bias and residual bias are similar, and suggest 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 the negative marginal tax rates generated by the EITC may be close to optimal. In that case the policy implications of these results are clear: the existing EITC should not be reduced or greatly reformed—indeed, it might beneficially be extended to workers without children, and made more salient as a wage subsidy. 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 widely noted fact that tax rates on high earners appear to embody substantially weaker

redistributive concerns than conventional utility functions provides some indication that present bias may be a reasonable explanation for the shape of the existing EITC.

Chapter 2

Taxation and the Allocation of Talent[1]

2.1 Introduction

The allocation of talented individuals across professions varies widely over time and space.[2] If, as Baumol (1990) and Murphy *et al.* (1991) argue, different professions have different ratios of social to private product, these differences in talent allocation across societies have important implications for aggregate welfare. Recent evidence strongly suggests such externalities not only exist but are large (Murphy and Topel, 2006; French, 2008). In this paper, we quantitatively evaluate the impact of non-linear income taxation on the allocation of talent, and we compute the tax schedule that maximizes aggregate (Pigouvian) welfare.

Our analysis adds to a growing literature (Philippon, 2010; Piketty *et al.*, 2014; Rothschild and Scheuer, 2014, Forthcoming) that emphasizes the role of income taxation in responding to externalities of some activities. We extend this literature—and the perturbation approach used more generally to derive optimal taxes (Saez, 2001b)—by incorporating a discrete,

---

[1]Co-authored with Charles Nathanson and E. Glen Weyl

[2]According to Goldin *et al.* (2013), more than twice as many male Harvard alumni from the 1969-1972 cohorts pursued careers in academia and in non-financial management as pursued careers in finance. Twenty years later, careers in finance were 50% more common than in academia and were comparable to those in non-financial management. Murphy *et al.* (1991) document that in 1970 among the 91 countries studied by Barro (1991), the 25th percentile of the share of college students studying engineering equals 3.8% and the 75th percentile equals 14.31%; the 25th percentile of the share of college students studying law equals 2.7% and the 75th percentile equals 11.2%.

long-run "allocative" elasticity that governs talented workers' choice of profession. This margin of labor supply is distinct from both the standard short-run intensive margin of effort emphasized in the literature above and the extensive margin of exiting the labor force studied by Saez (2002c).

In this allocative framework, workers make a long-term choice between well-paying professions and lower-paying "callings" that offer higher non-pecuniary benefits. Higher marginal tax rates incent workers to "follow their passion" by reducing the relative after-tax pecuniary compensation of the more lucrative professions. To the extent that better-paying professions generate negative (or less positive) externalities, raising marginal tax rates can generate social welfare gains from the movement of workers into socially productive professions. Because individuals might switch into a number of professions—each generating different externalities and tax revenues—when taxes rise, the full set of substitution patterns of individuals across professions becomes critical to determining optimal taxes. As we highlight theoretically in Section 2.2 and with a simple example in Section 2.3, because they involve discrete jumps in income, these substitution patterns make the first-order approach of Mirrlees (1971a) invalid.

The core of this paper is therefore a structural model of profession choice that imposes strong restrictions on substitution patterns in order to estimate how the allocation of talent would change under different income-tax regimes. The key inputs into our estimation are the distributions of income within different professions, the elasticities of labor supply on both the intensive and allocative margins, and the aggregate externalities on society from each profession. We take the externality estimates from the economics literature, which suggests these externalities are, although highly uncertain, likely to be huge and quite heterogeneous.[3] Our main findings for optimal policy are the following:

1. The optimal income tax features top rates of about 35%, which are close to the existing top rates in the year from which we draw data (2005). This positive top rate induces

---

[3]Murphy and Topel (2006) estimate that medical research generates a positive externality of more than 15% of GDP, whereas French (2008) calculates the financial profession's income includes 1.4% of GDP in rent-seeking.

long-term migration of talented workers to professions in which they earn less income but produce more externalities.

2. Although the optimal non-linear, profession-general tax rates differ significantly from 0, they achieve only small welfare gains (1.2%) relative to *laissez-faire*. By contrast, profession-targeted policies can achieve much more. We show an optimal linear subsidy to research professions achieves more than 40 times the welfare gains of our baseline optimal tax.

3. The key features of the optimal nonlinear income tax are robust to the details of how externalities accrue—which professions affect output in which others, and whether the externalities are linear or have diminishing returns to scale. Our results *are* sensitive to the magnitude of externalities we assume, especially in the research and management professions, and to the nature of allocative substitution across professions. This sensitivity suggests these understudied patterns are crucial to determining optimal tax policy.

Throughout this paper, we analyze only efficiency, rather than redistributive, gains from taxation. We focus on pure efficiency maximization because it highlights as sharply as possible the role of substitution patterns. In particular, efficiency maximization implies (see Section 2.2.2) the *total* elasticity of taxable income, which is so crucial in the canonical Vickrey (1945b) redistributive framework, is irrelevant for deriving the optimal tax schedule. Only the relative importance of the allocative and intensive margins impacts optimal tax rates.

Another reason to restrict attention to efficiency is to probe the explanatory power of the "Just Desserts Theory" of Mankiw (2010) that taxation should ensure individuals receive their social contribution. Perhaps surprisingly, we show such a theory is able to account for the broad outlines of existing US income taxation. We also believe this paper's framework is a useful tool for organizing, comparing, informing and potentially reconciling views of optimal taxation outside of economics, which many have argued is an important goal of

applied welfare economics (Gul and Pesendorfer, 2008; Weinzierl, 2014; Saez and Stantcheva, 2016).

In addition to our focus on allocation and efficiency, our analysis departs from the literature in several other ways. First, the allocative margin we introduce allows workers to choose between earning distinct income levels in different professions. By contrast, Rothschild and Scheuer (2014, Forthcoming) assume a concave utility function of continuous effort in different activities, which rules out changes in income for marginal workers who switch activities. Second, we allow for positive externalities, not just rent-seeking as in Rothschild and Scheuer, and these positive externalities turn out to be the largest quantitative contributors to our results. Third, our analysis is primarily quantitative. The present model incorporates a large number of professions and is estimated using a variety of data sources; previous literature on how income taxation should respond to externalities has involved primarily qualitative, illustrative models. Finally, in contrast to Piketty *et al.* (2014) and Rothschild and Scheuer, we abstract from any role taxes may have on the allocation of time within a profession across activities of different merit, assuming a homogeneous externality created by all output of a given profession.

2.2   A Model of Optimal Income Taxation with Externalities

All formal proofs of results and omitted derivations in this section appear in Appendix B.1.

2.2.1   Statement of the problem

A mass 1 of individuals work in n professions. Each worker is characterized by a $2n$-dimensional type $\theta = (a, \psi)$, where $a = (a_1, ..., a_n)$ is a vector of profession-specific productivities and $\psi = (\psi_1, ..., \psi_n)$ is a vector of the non-pecuniary utility the worker receives in each profession. The distribution of types θ among workers is given by a non-atomic and differentiable distribution function f with full support on a convex and open $\Theta \subseteq \mathbb{R}^{2n}$.

Labor supply consists of allocative and intensive margins. On the allocative margin, each worker chooses exactly one of the n professions to enter; we denote the profession

65

choice of a worker of type θ by $i(\theta)$.[4] The intensive margin consists of a choice of hours $h_i(\theta)$ to work in profession i, where $h_i(\theta) \geq 0$ for all i and θ. Because each individual works in only one profession, $h_i(\theta) = 0$ for $i \neq i(\theta)$.

Following Rothschild and Scheuer (2014, Forthcoming, henceforth RS), we assume externalities in this economy operate through production. For all profession pairs i, j, output in profession j can affect the productivity of workers in profession i. These relationships are summarized through nonnegative functions $E_i(Y_1, ..., Y_n)$ as in RS.[5]

The private product of a worker in i is linear in hours worked h_i. Hence, the private product of worker θ in profession i coincides with that worker's income and is given by

$$y_i(\theta) = a_i(\theta) h_i(\theta) E_i(Y_1, ..., Y_n)$$

where $Y_j = \int_\Theta y_j(\theta) f(\theta) d\theta$ is the total output in profession j. When E_i does not depend on Y_j, profession j exerts no externality on profession i. An economy without externalities corresponds to the case in which all E_i are constant.

Worker utility is linear in after-tax income, non-pecuniary utility ψ, and an hours cost function $\phi(\cdot)$ for which $\phi'(\cdot), \phi''(\cdot) > 0$:

$$U(\theta) = y_{i(\theta)}(\theta) - T(y_{i(\theta)}(\theta)) - \phi(h_{i(\theta)}) + \psi_{i(\theta)}(\theta), \tag{2.1}$$

where $T(\cdot)$ is the tax schedule set by the government. This specification abstracts from income effects, as does much of the recent literature on optimal taxation (Diamond, 1998b). This set-up is particularly convenient in our setting, because introducing income effects without adding a redistributive motive would require departing from the simple utilitarian welfare criterion we employ. In our specification, the cost of effort and the non-pecuniary benefit or cost of a profession are additively separable, thereby ruling out richer interactions

---

[4]As shown below, the choice of i for each θ is unique in the equilibrium we consider.

[5]A slight difference exists between our setup and that considered by RS. Whereas our externalities depend on the total *output* of each profession, theirs depend on the hours worked in each profession. Thus, in our model, externalities from profession i to profession j directly amplify the externalities of j, whereas in RS, the i externalities do so only indirectly by drawing labor into j. Appendix B.2 shows that the baseline quantitative results of our paper do not differ greatly when this alternate specification is used.

between intensive and allocative labor-supply decisions.

We assume the functional form $\phi(h) = h^{1/1+\sigma}/(1/1+\sigma)$, which leads all workers to have the same, constant intensive elasticity of labor supply σ. Each worker takes as given the tax schedule $T(\cdot)$ and the profession outputs $Y_1, ..., Y_n$, and then chooses a profession $i^*(\theta)$ and hours $h^*_{i^*(\theta)}(\theta)$ to maximize utility. To capture the case in which a worker is indifferent between multiple professions, we let $I^*(\theta)$ denote the *set* of professions that maximize the utility of a type-θ worker. When $|I^*(\theta)| > 1$, the worker chooses $i^*(\theta) \in I^*(\theta)$ randomly, with all type-θ workers making the same choice $i^*(\theta)$.[6] We denote the total utility, income, and non-pecuniary utility at the optimal profession and hours choices by $U^*(\theta)$, $y^*(\theta)$, and $\psi^*(\theta)$, respectively. We simplify notation by defining $h^*(\theta) = h^*_{i^*(\theta)}(\theta)$, and also let $U^*_i(\theta)$ and $y^*_i(\theta)$ denote the utility and income resulting from maximizing utility conditional on $i^*(\theta) = i$.

The government must finance a net expenditure of R, and chooses a tax schedule $T(\cdot)$ that maximizes total worker utility while raising this revenue:

$$T = \arg\max_{\overline{T}} \int_\Theta U^*(\theta)f(\theta)d\theta \ \Big| \ \int_\Theta \overline{T}(y^*(\theta))f(\theta)d\theta \geq R.$$

In our estimation of the optimal income tax in Section 2.4, we focus on bracketed tax systems that are messy to characterize analytically because they lead to "bunching" of workers with different productivity at the same income. For expositional clarity and comparability with existing literature, in this section, we follow Mirrlees (1971a) and Saez (2001b) in restricting attention to tax schedules for which an interior solution for hours always exists and is smooth.

Assumption 2.1. *The government considers only tax schedules T whose second derivative exists, and such that for all incomes y, $T'(y) < 1$ and*

$$\frac{yT''(y)}{1-T'(y)} > -\frac{1}{\sigma},$$

---

[6]The specific choice of $i^*(\theta)$ for such workers is irrelevant for aggregates like total utility and total output because the measure of indifferent workers equals zero—the dimension of the set of indifferent workers is smaller than the dimension of the set of all workers.

where σ is the elasticity of labor supply.

As shown in Appendix B.1.1, any solution to the worker's first-order condition for hours is a strict local maximum (due to a negative second-order condition) when this inequality holds. As a result, the hours choice admits a unique maximum. Intuitively, uniqueness can fail when T' decreases too quickly because a worker can have an interior solution for hours at a low income and high tax rate as well as an interior solution for hours at a high income and low tax rate.

Given the quantitative focus of this paper, we follow RS in assuming the existence of a unique Hicksian stable competitive equilibrium of the economy; our necessary conditions for optimization are valid only for tax schedules that induce such an equilibrium.

2.2.2 The government's first-order condition

The tax schedule T consists of a lump-sum tax T_0 paid by all workers, and a marginal tax schedule $T'(\cdot)$. These two aspects of the tax schedule uniquely determine T by the formula

$$T(y) = T_0 + \int_0^y T'(\overline{y}) d\overline{y}. \tag{2.2}$$

The government chooses T_0 and $T'(\cdot)$ to maximize worker utility while raising revenue R.

The equilibrium allocation of output $Y_1^*, ..., Y_n^*$ depends on $T'(\cdot)$ and not on T_0. Indeed, workers' intensive labor-supply choices depend on $T(\cdot)$ only through $T'(\cdot)$. And their profession choices depend on level differences in utility across professions, which remain constant—due to quasi-linear utility—as the common lump sum grant T_0 changes. This invariance condition means the optimal marginal tax schedule $T'(\cdot)$ cannot depend on R.

Lemma 2.1. *The optimal* marginal *tax schedule $T'(\cdot)$ is independent of the revenue requirement R.*

Due to Lemma 1, we ignore the revenue requirement in this paper and focus on the choice of the optimal marginal tax schedule $T'(\cdot)$.

To derive the optimal $T'(\cdot)$, we follow the intuitive perturbation approach to calculus of variations pioneered in economics by Wilson (1993) and in optimal income taxation by

Saez (2001b). Suppose the government slightly raises the marginal tax rate $T'(y)$ by dT' for incomes between y and $y + dy$, and rebates the additional revenue to workers through lowering T_0. This perturbation leaves the total revenue raised by the tax unchanged, but could raise or lower utility by leading workers to adjust their labor supply. At the optimum $T'(\cdot)$, the resulting change to utility is 0.

Raising $T'(y)$ leads to both intensive and allocative labor-supply changes. On the intensive margin, workers for whom $y^*(\theta) = y$ lower their hours $h^*(\theta)$. We denote the set of these workers by $\Theta(y) = \{\theta \mid y^*(\theta) = y\}$, and the set of such workers in profession i by $\Theta_i(y) = \{\theta \mid y^*(\theta) \text{ and } i^*(\theta) = i\}$. The tax change also lowers the level of after-tax income by $dT'dy$ of all workers earning above y. Therefore, the tax change induces profession switching for workers who are indifferent between a profession in which they earn more than y, and a profession in which they earn less. We denote the set of such workers by

$$\Theta_S(y) \equiv \left\{\theta \;\middle|\; \text{there exist } i_l, i_h \in I^*(\theta) \text{ such that } y^*_{i_l}(\theta) < y < y^*_{i_h}(\theta)\right\},$$

and we denote their measure by $f_S(y)$.

The perturbation to $T'(\cdot)$ causes additional, secondary labor-supply changes. Due to externalities operating through the E_i, the intensive and allocative margin adjustments just described change the productivity in all professions, leading all workers to modify their labor supply. Sufficient statistics that we term *externality ratios* capture the resulting changes to aggregate utility. The externality ratio e_i of profession i equals

$$e_i \equiv \frac{\partial}{\partial Y_i} \int_\Theta U^*(\theta) f(\theta) d\theta,$$

where the partial derivative denotes the cumulative effect on welfare through changes in the E_j that result from a change in Y_i. Thus, the externality ratio of a profession gives the marginal externality of a dollar earned in that profession. It can be positive or negative. When a profession causes no externalities, $\partial E_j/\partial Y_i \equiv 0$ for all j, so the externality ratio equals 0.

This ratio is a central yet subtle object in our analysis, so we describe its meaning and

derivation in some detail. A change in Y_i induces a series of subsequent changes. The direct effect of a change in Y_i is to alter the productivity in all professions. These productivity changes lead to adjustments in labor supply on both the intensive and allocative margins: workers choose to work more or less and also may choose different professions entirely. These labor-supply responses change the output Y_j in each profession, inducing another round of adjustments in labor supply, which beget yet another round of adjustments, and so on. Externality ratios solve the fixed-point problem that captures this infinite series of labor-supply adjustments. The solution is local to the equilibrium under consideration. In Appendix B.1.1, we explicitly solve this problem to express the e_i in terms of the Jacobian of the externality function E at the equilibrium $(Y_1^*, ..., Y_n^*)$ and the full set of labor-supply responses.

The average externality ratio of workers earning y is

$$ e(y) = \frac{\sum_{i=1}^{n} e_i \int_{\Theta_i(y)} f(\theta) d\theta}{\int_{\Theta(y)} f(\theta) d\theta}. $$

Using the definition of worker utility (2.1) and the revenue requirement $\int_\Theta T(y^*(\theta)) f(\theta) d\theta = R$, we write the government's objective function as

$$ \int_\Theta U^*(\theta) f(\theta) d\theta = -R + \int_\Theta \left(y^*(\theta) - \phi(h^*(\theta)) + \psi^*(\theta) \right) f(\theta) d\theta. \tag{2.3} $$

The government maximizes the integral on the right: total income less disutility from labor plus non-pecuniary utility from work. We calculate how the perturbation to $T'(\cdot)$ at y changes this integral. We first consider the change from intensive margin labor-supply adjustments. Then we separately consider how allocative margin adjustments change utility, and finally we present the first-order condition that combines these effects.

Intensive margin

Consider a worker in profession i for whom $y^*(\theta) = y$. Denote the *wage* of this worker by $w_i(\theta) = a_i(\theta) E_i(Y_1^*, ..., Y_n^*)$. The hours for this worker are determined by $h^*(\theta)^{1/\sigma} = w_i(\theta) (1 - T'(w_i(\theta) h^*(\theta)))$. For h near $h^*(\theta)$, the relationship $T'(w_i(\theta) h) = T'(y) + w_i(\theta)(h -$

$h^*(\theta))T''(y)$ holds to the first order. Using this first-order expansion, we totally differentiate the hours equation with respect to $T'(y)$ to find that

$$dh^*(\theta) = \frac{\sigma h^*(\theta)}{1 - T'(y) + \sigma y T''(y)} dT'.$$

This intensive-margin response directly changes the type-θ worker's contribution to (2.3), and also alters the income of other workers through an externality. The direct effect is $(w_i(\theta) - \phi'(h^*(\theta)))dh^*(\theta)$. Because $\phi'(h^*(\theta)) = w_i(\theta)(1 - T'(y))$, the direct effect reduces to $T'(y)w_i(\theta)dh^*(\theta)$. To uncover the externality, note $dY_i^* = w_i(\theta)dh^*(\theta)$, so the externality equals $e_i w_i(\theta)dh^*(\theta)$. We sum the direct and externality effects on utility across all workers earning y to obtain the complete change in the government's objective from intensive margin adjustments. Let $f(y) = \int_{\Theta(y)} f(\theta)d\theta$; the mass of workers earning between y and $y + dy$ is $f(y)dy$. The complete intensive-margin change in the government objective from the perturbation to the tax schedule is

$$\partial^{int} \int_{\Theta} U^*(\theta) f(\theta) d\theta = \frac{\sigma y f(y)}{1 - T'(y) + \sigma y T''(y)} \left(T'(y) + e(y)\right) dT' dy. \tag{2.4}$$

Allocative margin

Consider a worker for whom $\theta \in \Theta_S(y)$. This worker is indifferent between a profession i_h in which she earns $y_{i_h}^*(\theta)$, and a profession i_l in which she earns $y_{i_l}^*(\theta)$, with $y_{i_l}^*(\theta) < y < y_{i_h}^*(\theta)$. The tax perturbation decreases the after-tax income, and hence utility, in i_h by $dT'dy$ while leaving utility in i_l unchanged. As a result, the worker switches from i_h to i_l.

This switch directly changes the value of the government's objective function (2.3) by $y_{i_l}^*(\theta) - \phi(h_{i_l}^*(\theta)) + \psi_{i_l}^*(\theta) - \left(y_{i_h}^*(\theta) - \phi(h_{i_h}^*(\theta)) + \psi_{i_h}^*(\theta)\right)$. By the envelope theorem (because the worker receives the same utility in i_l and i_h), this difference equals the fiscal externality $T(y_{i,l}^*(\theta)) - T(y_{i_h}^*(\theta))$. We define the average proportional tax change from switching workers by

$$\Delta_T(y) \equiv \int_{\Theta_S(y)} \frac{T(y_{i_h}^*(\theta)) - T(y_{i_l}^*(\theta))}{y} \frac{f(\theta)}{f_S(y)} d\theta.$$

The worker's switch from i_h to i_l also changes the government's objective function

71

through externalities. The worker's presence in profession i increases Y_i^* by $dY_i^* = y_i^*(\theta)$, so the total externality of a worker's presence in i is $e_i y_i^*(\theta)$. The change in externalities from switching from i_h to i_l is therefore $e_{i_l} y_{i_l}^*(\theta) - e_{i_h} y_{i_h}^*(\theta)$. We define the average proportional externality change from switching workers by

$$\Delta_e(y) \equiv \int_{\Theta_S(y)} \frac{e_{i_h} y_{i_h}^*(\theta) - e_{i_l} y_{i_l}^*(\theta)}{y} \frac{f(\theta)}{f_S(y)} d\theta.$$

Recall these externalities incorporate *all* of the indirect, general equilibrium effects of production in a profession.

We sum the direct and externality effects on utility across all switching workers to obtain the complete change in the government's objective from allocative margin adjustments. Because the change in the relative income of i_h and i_l is $dT'dy$, the completely allocative margin change in the government objective from the perturbation to the tax schedule is

$$\partial^{all} \int_{\Theta} U^*(\theta) f(\theta) d\theta = y f_S(y) \left(\Delta_T(y) + \Delta_e(y) \right) dT'dy. \tag{2.5}$$

In Saez's (2002c) analysis of the effect of taxes on the extensive margin of labor supply, the analogous expression depends only on the aggregate taxes paid by each individual and the density of workers indifferent to exiting the labor force.

Total first-order condition

The government's first-order condition holds when the intensive margin change (2.4) and allocative margin change (2.5) to the government's objective resulting from the tax perturbation sum to 0. Because we arbitrarily chose the income y at which $T'(\cdot)$ was perturbed, the first-order condition holds for all y. Proposition 2.1 produces the first-order condition by adding (2.4) and (2.5) and then dividing by $y f(y) dT'dy$.

Proposition 2.1. *The optimal tax schedule T for the government satisfies the equation*

$$0 = \underbrace{\frac{\sigma f(y)}{1 - T'(y) + \sigma y T''(y)} \left(T'(y) + e(y) \right)}_{intensive} + \underbrace{f_S(y) \left(\Delta_T(y) + \Delta_e(y) \right)}_{allocative}$$

for all incomes y. Here σ is the elasticity of labor supply, $e(y)$ is the average externality ratio of output for workers earning y, $f(y)$ is the measure of workers earning y, $f_S(y)$ is the measure of workers indifferent between earning above y in one profession and below y in another, $\Delta_T(y)$ is the average proportional difference in taxes between the two professions for such workers, and $\Delta_e(y)$ is the average proportional difference in externalities between the two professions for such workers.

The optimal tax T is Pigouvian, because it offsets externalities on both the intensive and allocative margins. Without externalities, $e(y)$ and $\Delta_e(y)$ globally equal 0, in which case the optimal tax given by Proposition 1 is lump sum ($T' \equiv 0$). We build further intuition by considering the intensive and allocative margins separately.

When only the intensive margin is present, the optimal tax satisfies $T'(y) = -e(y)$. In this case, the marginal tax rate exactly equals the average negative externality ratio at each income level. RS refer to this tax as the "Pigouvian" correction because it appears in a model with only an intensive margin. In particular, the weight of this effect in the total first-order condition scales with $\sigma f(y)$, the product of the intensive labor-supply elasticity and the number of individuals subject to this elasticity. The greater this product is, the more closely the optimal tax satisfies $T'(y) = -e(y)$.

Conversely, the optimal tax in the presence of just the allocative margin satisfies $\Delta_T(y) = -\Delta_e(y)$ for all y. In this case, taxes offset gross changes in negative externalities from workers switching professions. The weight of this effect scales with $f_S(y)$, the measure of the workers who switch profession around y. The more sensitive profession choices are to income differences, the greater $f_S(y)$ becomes and the more closely the optimal tax satisfies $\Delta_T(y) = -\Delta_e(y)$.

Note the optimal tax is related only to the *relative size* of the allocative and intensive responses,

$$\frac{f_S(y)\left[1 - T'(y) + \sigma y T''(y)\right]}{\sigma f(y)},$$

and not to the level of these responses. For example (assuming a linear tax for the moment), suppose σ and f_S doubled so that the size of both the intensive and allocative responses were twice as large. This doubling would have no impact on optimal taxes, in sharp contrast

to the standard Vickrey model whereby a redistributive state is constrained in its ability to extract revenue by the overall elasticity of taxable income.

2.3 An Example with Three Professions

This section builds quantitative intuition in closed form for the full calibration in a simple example that captures the key features of the data and our estimation. In particular, we use Proposition 1 to calculate the optimal top tax rate, $\lim_{y \to \infty} T'(y)$, for the optimal T. This rate measures the marginal tax rate the top earners face (although possibly only at extremely high incomes) and is similar to the object explored in Saez (2001b) and Saez *et al.* (2012b). Given our focus on the allocation of talented individuals, many of whom earn very high incomes, this limiting rate seems particularly relevant in our context.

2.3.1 Specification and optimal top tax rate

Three professions exist: U, H, and L. Some fraction of the workers are "unskilled" and are restricted to U. The remaining workers are "skilled" and choose between H and L.[7] For each skilled worker, productivity $a_H(\theta)$ in H exceeds productivity $a_L(\theta)$ in L by a constant multiple $r^{1/(1+\sigma)}$, where $r > 1$, which leads in equilibrium to income that is higher in H than in L by a factor of r. Above some level $\bar{a}_i$, the distribution of a_i in each profession is Pareto, with conditional probability distribution $\Pr(a_i(\theta) \geq a \mid a_i(\theta) \geq \bar{a}_i) = (\bar{a}_i/a)^{\alpha(1+\sigma)}$ for some $\alpha > 0$; in equilibrium, the Pareto exponent for the *income* distribution will equal α. For skilled workers, non-pecuniary utility ψ_i of working in $i = H$ or L is distributed as $\psi_i \mid a \sim \beta^{-1}[(a_L^{1+\sigma} + a_H^{1+\sigma})/2](\bar{\psi}_i + F_\psi)$, where the $\bar{\psi}_i$ are constants and F_ψ is a standard Gumbel distribution given by $F_\psi = e^{-e^{-\psi}}$. The term $(a_L^{1+\sigma} + a_H^{1+\sigma})/2$ is a normalization to ensure non-pecuniary utility is of the same order of magnitude as income, and $\beta > 0$ is a parameter we call the *allocative sensitivity*. Output in U causes no externality, whereas H and L output both affect productivity in U. Thus, E_L and E_H are equal to 1, whereas $E_U(Y_L, Y_H)$

---

[7]Formally, $\psi_H(\theta) = \psi_L(\theta) = -\infty$ for the unskilled workers and $\psi_U(\theta) = -\infty$ for the skilled workers.

increases in Y_L and decreases in Y_H, so $e_H < 0 < e_L$.

This specification broadly matches the data we present in Section 2.4. In our baseline analysis, engineering, teaching, and research professions cause positive externalities, whereas law and finance lead to negative externalities. We find that, in the upper tail of the income distribution, the incomes in the first set of professions are lower than those in the second.

The present specification allows us to explicitly calculate the optimal top tax rate in the special cases in which only the intensive or allocative labor-supply margin operates. We first analyze the intensive optimal top tax rate. From Proposition 1, this rate satisfies $\tau_{int} = -\lim_{y \to \infty} e(y)$. Hence,

$$\tau_{int} = -(s_H e_H + s_L e_L),$$

where e_H and e_L are the externality ratios and s_i is the share of workers at top incomes in profession i.[8] This tax is more positive when the share s_H of top earners in H is higher and when the negative externality e_H is larger in magnitude. Conversely, the intensive optimal top tax rate is less positive when s_L is larger and when e_L is greater. The rate τ_{int}, dubbed the "Pigouvian correction" by RS, is optimal when profession choices are fixed.

The allocative optimal top rate looks quite different from τ_{int}. From Proposition 1, $\Delta_T(y) + \Delta_e(y) = 0$ for high incomes at this rate. These difference terms are determined by the relative income for the same skilled worker in H and L, rather than by the distribution of workers earning any given income. Because $y_i^*(\theta) = a_i^{1+\sigma}(\theta)(1 - T'(y_i^*(\theta)))^\sigma$, $y_H^*(\theta) = r y_L^*(\theta)$ at high incomes.[9] The parameter r equals the ratio of income in H to income in L for a skilled worker. Therefore, each switching worker's contribution to $\Delta_T(y)$ is $\tau(r-1)y_L^*(\theta)$ and to $\Delta_e(y)$ is $(re_H - e_L)y_L^*(\theta)$, where τ is the top tax rate. The optimum sums these to 0, and is

$$\tau_{all} = -\frac{re_H - e_L}{r - 1}.$$

---

[8]Formally, $s_i = \lim_{y \to \infty} \int_{\Theta_i(y)} f(\theta)d\theta / \int_{\Theta(y)} f(\theta)d\theta$.

[9]This statement requires $\lim_{y \to \infty} T'(y) < 1$ or $\sigma = 0$.

Intuitively, this rate equals the change in negative externalities from a switching worker divided by the change in that worker's income. Although it is the allocative margin analogue of RS's Pigouvian correction, it often behaves very differently quantitatively. In particular, τ_{all} is unambiguously positive because L produces positive externalities and H causes negative externalities ($e_L > 0 > e_H$). This result stands in contrast to τ_{int}, which could be positive or negative.

The size of τ_{all} is greater when e_H or e_L is greater in magnitude. Unlike τ_{int}, τ_{all} depends not on share of the population in H and L but on r, the ratio of income in H to L for a given worker. Simple differentiation shows it to be strictly decreasing in r so long as $e_H < e_L$. To see this relationship between τ_{all} and r dramatically, note that as $r \to 1$, a switching worker is indifferent between working in H and L and both yield the same income and therefore tax revenue. However, a switch to L increases social welfare by $e_L - e_H$ times the worker's income, so the τ_{all} becomes arbitrarily large to compensate this discrete change in externalities with a discrete change in tax revenue accrued over a very small difference in incomes.

The true optimal tax τ^* combines the logic of both τ_{int} and τ_{all} and is always strictly between these two rates, as we show in Appendix B.1.2.

2.3.2 Calibration

To calculate τ^*, we need values of α, σ, β, r, e_H, e_L, $\overline{\psi}_H - \overline{\psi}_L$, and the share of workers that are skilled. We take these values from the data used in the estimation in the next section and thus discuss our calibration choices only briefly here and expand on this discussion in Appendix B.3. We also explore the different values of τ^* generated by a reasonable range of the parameters.

We first set α, the Pareto parameter for the tail of the US income distribution, to 1.5 based on the ratio of the total income earned by the top 1% of the US income distribution to the 99th percentile of the income distribution. We set $\sigma = 0.24$ and $\beta = 1.5$ based on our estimation in the next section, where we try to match the elasticity of income with respect

to the tax rate (Chetty, 2012b) and the concurrent growth in relative finance wages and employment from 1980 to 2005 (Philippon and Reshef, 2012). To calculate r, we compare the incomes in H and L at the same percentiles of the profession-specific distributions. Because productivity in H and L are perfectly correlated, a worker in the 99th percentile of H incomes will also be in the 99th percentile of L incomes. We use a 99th percentile income in H (finance and law) of \$1,900,000 and in L (engineering, research and teaching) of \$400,000 based on a weighted-average of our profession-specific income-distribution estimations across the professions that make up H and L.

To calculate the externality ratios e_H and e_L, we take a weighted average of approximate externality ratios of the professions constituting each of H and L. As we discuss in Appendix B.3, dividing a profession's aggregate spillover by its aggregate income provides an accurate approximation of its externality ratio. In Section 2.4, we estimate these aggregate spillovers by drawing on the economics literature, and we calculate the aggregate incomes using data on profession-specific income distributions and worker counts from the Bureau of Labor Statistics (BLS) and the Internal Revenue Service (IRS). These figures yield approximate externality ratios of -0.33 for finance, -0.10 for law, 0.15 for engineering, 9.28 for research, and 2.28 for teaching. An average using weights proportional to the representation of these professions at high incomes in the data then yields $e_H = -0.24$ and $e_L = 2.67$. Although these ratios are endogenous to the tax structure, we take them as fixed for the purposes of this calibration exercise. In the exercise in Section 2.4 we allow the externality ratios to depend on the tax structure.

Finally, we set $\overline{\psi}_H - \overline{\psi}_L$ to match the share of workers in H and L, given the data and the tax rate in 2005. According to Bakija *et al.* (2012), 7.2% of the top 1% of earners in 2005 were in L and 22.3% were in H.

Using these parameters and externality ratios, we calculate the optimal top tax rate to be $\tau^* = 0.24$. Relative to a *laissez-faire* tax rate of 0, τ^* induces 11% more of skilled workers subject to the tax rate to choose the lower-paying but higher-externality profession L. To break down the top tax rate, we calculate τ_{int} and τ_{all} at the optimum. When $\tau = \tau^*$,

$s_H = 0.18$ and $s_L = 0.08$, leading to an intensive optimal tax rate of $\tau_{int} = -0.17$. Thus, the intensive optimal rate is negative, even though the total optimal rate is positive. The negative τ_{int} results because the order-of-magnitude higher externalities from L overwhelm the negative externalities from H, as H has only three times greater representation at high incomes. By contrast, $\tau_{all} = 1.03$, confiscating more than all of the marginal income of top earners. The total optimum τ^* balances the intensive and allocative optima at a rate of 0.24. This rate is reasonably close to the top tax rate of 0.37 that we calculate in Section 2.4.

Table 2.1: *Optimal Top Tax Rate for Different Parameter Values*

	r	σ	β	e_H	e_L
Half of Baseline	26%	34%	15%	20%	20%
Baseline	24%	24%	24%	24%	24%
Double Baseline	19%	15%	30%	31%	28%

Notes: This table reports the optimal top tax rate τ^* in the three-profession example defined in Section 2.3. In each column, we hold all but the header parameter constant, varying that parameter to 50%, 100%, and 200% of its baseline value and reporting the optimal top tax rates. The baseline parameters are $r = 4.7$, $\sigma = 0.24$, $\beta = 1.5$, $e_H = -0.24$, and $e_L = 2.67$.

Table 2.1 reports the sensitivity of τ^* to changes in the parameters. For each parameter, we recalculate τ^* using values at half and double the baseline, while holding the other parameters constant.[10] The results confirm the intuition discussed above. Higher values of r lower optimal rates, as profession switching generates smaller positive externalities relative to lost tax revenue when r is greater. Higher values of σ lower optimal rates, as a greater σ makes the intensive margin more important, and the intensive optimal tax rate is negative. Similarly, a greater β increases the optimal top rate as it makes the allocative margin more important. Finally, higher absolute values of the externalities increase the optimal top tax rate by increasing the efficiency gains from switches. Raising the negative externality in H has a greater impact than raising it in L, despite the much greater magnitude of the

---

[10]Thus, as σ moves, the income ratio r and the Pareto parameter α for income stay constant. These quantities are calibrated to match observed data, so we do not want them to change as σ moves. Our specification allows r and α to stay constant as σ moves, by involving σ in the distribution of skilled productivity.

externality in L. Intuitively, a greater externality in H raises both τ_{all} and τ_{int}, but increasing the positive externality of L lowers τ_{int} while raising τ_{all} resulting in conflicting effects on $\tau^{\star}$.

2.4   Empirical Strategy

In this section, we specify the richest version of the full model that we believe we can credibly estimate and we fit it to data from the United States in 2005. Appendix B.1.3 includes all derivations and proofs, and Appendix B.4 summarizes additional empirical details.

2.4.1   Specification

Our specification of f is as follows. We separate the professions into n skilled professions $i = 1, ..., n$ and 1 low-skilled profession, which we index by $i = 0$. An exogenous share s_0 of the workers are "low-skill" and always choose $i^*(\theta) = 0$, because $\psi_i(\theta) = -\infty$ for $i > 0$. The remaining $1 - s_0$ of the workers are "skilled" and choose only among the skilled professions because $\psi_0(\theta) = -\infty$.

Each worker's productivity a_i in i is drawn from a profession-specific distribution F_i^a. We specify the correlation structure of productivity draws for skilled workers by a *Gaussian copula*:

$$f(a_1, ..., a_n) = f_{0,\Sigma}^{\mathcal{N}}\left(\Phi^{-1}(F_1^a(a_1)), ..., \Phi^{-1}(F_n^a(a_n))\right),$$

where Φ is the CDF of a unidimensional standard normal and $f_{\mu,\Sigma}^{\mathcal{N}}$ is the PDF of a multivariate normal with mean μ and covariance matrix Σ. This specification preserves the marginal productivity distributions F_i^a (i.e., $f\mid_{a_i} = f_i^a$ for all i), but allows correlation specified by Σ. We assume that all the diagonal elements of Σ are 1 and all off-diagonal elements are equal to some ρ which governs the correlation of productivity between every distinct pair of professions. When $\rho = 1$, productivity across professions is perfectly correlated so that workers are characterized by a single "talent" parameter that determines their percentile in

79

each profession's productivity distribution. Smaller values of ρ allow sorting on comparative advantage, in which the workers who choose i are those who are most productive in i relative to the other professions, as suggested by the empirical work of Reyes *et al.* (2013) and Kirkebøen *et al.* (Forthcoming).

Conditional on the productivity vector $a = (a_1, ..., a_n)$, each preference ψ_i is drawn independently from the distribution

$$\psi_i \sim \beta^{-1} \left(\frac{1}{n} \sum_{j=1}^{n} a_j^{1+\sigma} \right) (\overline{\psi}_i + F_\psi),$$

where $\overline{\psi}_i$ is a constant and F_ψ is a standard Gumbel distribution given by $F_\psi = e^{-e^{-\psi}}$, generating a standard logit discrete-choice model among individuals with a given ability. The normalization by productivity keeps professional choice scale-invariant with respect to income.[11] Thus, we can interpret β as the *allocative sensitivity*, with higher β indicating greater elasticity of profession choice to changes in relative incomes across professions and thus to taxation. The constants $\overline{\psi}_i$ determine the average relative attractiveness of each profession i; more workers enter i when $\overline{\psi}_i$ is higher.

Our specification of each externality function E_i has the form

$$E_i(Y_0, ..., Y_n) = \prod_{j=0}^{n} (1 + \epsilon_{i,j} Y_j^{\gamma}).$$

γ captures the returns to scale of the externalities; $\gamma = 1$ implies externalities are linear in output; lower values of γ lead to diminishing marginal returns. $\epsilon_{i,j}$ captures the targeting of externalities across professions emphasized by RS. For our estimation, we reduce the dimensionality of these coefficients according to the specification

$$\epsilon_{i,j} = \delta_{i,j} \epsilon_j,$$

where $\delta_{i,j} \in \{0, 1\}$ if $i \neq j$. We thus restrict externalities coming from profession j to be uniform in magnitude across all professions i on which profession j has any impact. The $\epsilon_{i,i}$

---

[11]This property holds exactly when taxes are linear, as only relative income $y_i / \sum_j y_j$ matters in (2.9) when $T'(\cdot)$ is constant.

remain unrestricted, allowing independence of the own externalities from those on other professions.

2.4.2 Identification

This section discusses the identification of f and E. The empirical inputs into our estimation are the existing tax schedule T_{2005}, the distributions of income $f_0^y, ..., f_n^y$, the population shares in each profession $s_0, ..., s_n$, and the marginal social products $\partial Y / \partial Y_0, ..., \partial Y / \partial Y_n$ of output in each profession.[12] These inputs come from data we describe in Section 2.4.3. For the moment, we take the parameters σ, β, ρ, γ, and the matrix $\{\delta_{i,j}\}$ as given, postponing discussion of their selection until Section 2.4.3. The outputs of the present estimation are $f_0^a, ..., f_n^a, \overline{\psi}_1, ...\overline{\psi}_n$, and $\epsilon_0, ..., \epsilon_n$.

First, we calculate the aggregate income in each profession and in the economy. For each i, $Y_i = s_i \int_0^\infty y f_i^y(y) dy$, and $Y = \sum_{i=0}^n Y_i$.

Next, we calculate the externality coefficients $\epsilon_0, ..., \epsilon_n$ using the aggregate income data and the marginal social product measures $\partial Y / \partial Y_j$. As we define it, this derivative gives the cumulative increase in the economy's output from a unit increase in output in j, holding labor supply constant in the entire economy. The change to Y_j can be thought of as coming from a small shock to productivity in that profession. As with the externality ratios, the marginal social product includes feedback effects: an increase in Y_j alters output of all professions, inducing further changes to output in the economy and so on. As we show in Appendix B.1.3, the marginal social product equals

$$\frac{\partial Y}{\partial Y_j} = 1'(I - J)^{-1} 1_j, \tag{2.6}$$

where $1 = (1, ..., 1)'$, $1_j = (0, ..., 1, ..., 0)'$ with 1 in just the jth spot, I is the identity matrix,

---

[12]Rather than use the true non-linear value of T_{2005}, we use a linear approximation in which the marginal tax rate is constant ($T'_{2005} = 0.3$). The true tax schedule T_{2005} features discontinuous marginal rates. Therefore, in a model such as ours in which primitives are smooth and workers are fully optimizing, bunching would result in the income distributions. Because empirical income distributions are smooth, we cannot fit underlying skill distributions to the empirical income distributions using the true T_{2005}. Using the linear version allows us to fit the skill distributions. A number of optimal tax papers take a similar approach, including Saez (2001b, 2002c).

and J is the quasi-Jacobian matrix

$$J = \left\{ \frac{Y_i}{Y_k} \frac{\gamma \delta_{i,k} \epsilon_k Y_k^{\gamma}}{1 + \delta_{i,k} \epsilon_k Y_k^{\gamma}} \right\}_{i,k}.$$

Note that when externalities are absent from the economy, $J = 0$ so $\partial Y / \partial Y_j = 1$ for each j: marginal social product coincides with marginal private product. Equation (2.6) delivers $n + 1$ equations in the $n + 1$ unknowns $\epsilon_0, ..., \epsilon_n$, allowing us to solve for these parameters numerically.

The subsequent step is to infer the empirical productivity distributions $\widetilde{f}_i^a$ that appear in the data. Selection of workers across professions determines these distributions, and hence the $\widetilde{f}_i^a$ differ from the underlying productivity distributions f_i^a we eventually estimate. The following equation delivers a one-to-one mapping between the productivity a_i of a worker in i and her income y_i:

$$a_i = y_i^{\frac{1}{1+\sigma}} \left(1 - T'_{2005}(y_i)\right)^{-\frac{\sigma}{1+\sigma}} E_i(Y_0, ..., Y_n)^{-1}. \tag{2.7}$$

We define $y_i(a_i)$ to be the unique value of y_i that solves this equation given a_i.[13] Then

$$\widetilde{f}_i^a(a_i) = y_i'(a_i) f_i^y(y_i(a_i)). \tag{2.8}$$

No selection occurs into or out of the low-skilled profession $i = 0$, so $f_0^a = \widetilde{f}_i^a$.

The penultimate step is to calculate the relative utility $\widetilde{u}_i(a)$ of working in i for a skilled worker with productivity vector a, ignoring profession-preference utility ψ. The relative utility $\widetilde{u}_i(a)$ determines the share of workers with productivity a who choose to work in profession i. It is defined as $\widetilde{u}_i(a) = (U_i^*(\theta) - \psi_i(\theta)) / \left(n^{-1} \sum_j a_j^{1+\sigma}\right)$, where the productivity component of θ equals a. Appendix B.1.3 derives the following closed-form expression for relative utility:

$$\widetilde{u}_i(a) = \frac{y_i(a_i) - T_{2005}(y_i(a_i)) + \sigma \left(y_i(a_i) T'_{2005}(y_i(a_i)) - T_{2005}(y_i(a_i))\right)}{(1+\sigma) n^{-1} \sum_j y_j(a_j) (1 - T'_{2005}(y_j(a_j)))^{-\sigma} E_j(Y_0, ..., Y_n)^{-(1+\sigma)}}. \tag{2.9}$$

---

[13]That the solution to (2.7) is unique follows because the right side of (2.7) strictly increases in y_i due to Assumption 2.1.

Finally, we derive the conditional distribution of a_{-i} given a_i. We use this conditional distribution to back out the underlying productivity distributions f_i^a from the empirical distributions $\widetilde{f}_i^a$, which are affected by selection. Given $a_i = a$, the conditional distribution of a_{-i} follows a Gaussian copula. The $\Phi^{-1}(F_j^a(a_j))$ for $j \neq i$ are distributed as a multivariate normal with mean $\Phi^{-1}(F_i^a(a_i))\varrho$ and covariance $\Sigma_{-1} - \varrho'\varrho$, where Σ_{-1} is the top $(n-1) \times (n-1)$ block of Σ and $\varrho = (\rho, ..., \rho)$ is a $1 \times (n-1)$ vector. We now state the equations that allow us to identify underlying productivity f_i^a and profession preferences $\overline{\psi}_i$ from the data.

Lemma 2.2. *Given empirical population shares $s_0, s_1, ..., s_n$ and income distributions $f_1^y, ..., f_n^y$, the underlying productivity distributions $f_1^a, ..., f_n^a$ and profession-preference parameters $\overline{\psi}_1, ..., \overline{\psi}_n$ solve the n functional equations*

$$\frac{s_i \widetilde{f}_i^a(a_i)}{1 - s_0} = f_i^a(a_i) \int_{\mathbb{R}_+^{n-1}} \frac{e^{\beta \widetilde{u}_i(a) + \overline{\psi}_i}}{\sum_j e^{\beta \widetilde{u}_j(a) + \overline{\psi}_j}} f_{\Phi^{-1}(F_i^a(a_i))\varrho, \Sigma_{-1} - \varrho'\varrho}^{\mathcal{N}} \left(\Phi^{-1}(F_1^a(a_1)), ..., \Phi^{-1}(F_n^a(a_n)) \right) da_{-i}$$

(2.10)

for $1 \leq i \leq n$ and all $a_i > 0$. Here, $\widetilde{f}_i^a$ is the empirical productivity distribution in i calculated from (2.8) and $\widetilde{u}_i(a)$ is the relative utility of working in i for a worker with productivity vector a calculated from (2.9). These solutions uniquely determine the f_i^a and are unique up to constant for the $\overline{\psi}_i$.

We solve equation (2.10) using a numerical solver.

2.4.3   Data

Income distributions

We follow the classifications of Bakija *et al.* (2012), whose data we use, in partitioning all US workers into one low-skill profession, which we deem *Other*, and 11 high-skill professions: *Art* (artists, entertainers, writers, and athletes), *Engineering* (computer programmers and engineers), *Finance* (financial managers, financial analysts, financial advisers, and securities traders), *Law* (lawyers and judges), *Management* (executives and managers), *Medicine* (doctors and dentists), *Operations* (consultants and IT professionals), *Real Estate* (brokers, property managers, and appraisers), *Research* (professors and scientists), *Sales* (sales representatives and advertising and insurance agents), and *Teaching* (primary and secondary school teachers).

For each profession i, we calculate the share s_i of workers in that profession as well as the empirical distribution of pre-tax income f_i^y in 2005 using two sources of data and several parametric assumptions.

Data on the top of each income distribution come from income-tax filings reported to the IRS. The IRS uses the self-reported profession on personal tax returns (1040s) to assign each filer a Standard Occupation Code (SOC). Bakija *et al.* (2012) aggregate these codes into the 11 professions we use; we report this classification in Appendix B.4.1.[14]

Their unit of observation is a tax return, of which there are 145,881,000 in 2005.[15] They define the profession of a tax return as that of the primary filer, which is the filer whose social security number is listed first in the case of couples. Bakija *et al.* (2012) report the number of workers in each profession earning more than $280,000 and $1,200,000, as well as the average income of each group of workers above these thresholds.[16]

For each SOC, the BLS reports in the annual Occupational Employment Statistics (OES) database the number of workers as well as the 10th, 25th, 50th, 75th, and 90th income percentiles. The BLS produces the OES using surveys of non-farm establishments. Using these data, we calculate the number of workers in each profession by summing the number in each constituent SOC, and then calculate s_i as the share of all workers in each profession.[17]

_____

[14]One exception to Bakija *et al.* (2012)'s unique assignment of SOCs to professions concerns the SOCs for chief executives (11-1011) and general and operations managers (11-1021). Bakija *et al.* assign such workers to *Finance* if the industry of the employer listed on the W-2 is "Finance and Insurance" (NAICS code 52); they assign such workers to *Management* otherwise.

[15]Bakija *et al.* (2012) obtain this count from Piketty and Saez (2003), who report this number in an updated table at `http://elsa.berkeley.edu/~saez/TabFig2010.xls`.

[16]To be more precise, Bakija *et al.* (2012) report data that allow direct computation of these statistics. They report the share of tax returns in the top 1% and top 0.1% in each profession, and write that these income cutoffs are $280,000 and $1,200,000, respectively, in 2005 dollars. Because we know the number of tax returns, we can directly compute the number of workers in each profession earning more than each cutoff. Similarly, they report the share of aggregate reported income in the United States earned by workers in each profession in the top 1% and top 0.1% of the income distribution. The aggregate income number comes from Piketty and Saez (2003), who report it as $6,830,211,000,000 in the spreadsheet referenced in the previous footnote. Using this figure, we directly compute the total income of workers in each profession earning more than $280,000 and $1,200,000, and then divide by the counts to arrive at the average.

[17]To match Bakija *et al.* (2012)'s splitting of SOCs 11-1011 and 11-1021 into Finance and Management (see fn. 13), we use BLS data for SOC-NAICS pairs to split these SOCs into a category in which the NAICS = 52 and one in which the NAICS $\neq$ 52.

To calculate the profession-specific income distribution f_i^y we first assume that for $y \geq$ $1,200,000$, the income distribution is Pareto: $f_i^y(y) = \alpha_i m_i^{\alpha_i}/y^{\alpha_i+1}$. We can uniquely compute the parameters of the Pareto distribution using the mean income of workers earning above this threshold and the number of such workers, both of which are reported by Bakija *et al.* (2012). Next, we linearly interpolate f_i^y between $280,000 and $1,200,000, adding a break point at $580,000, the geometric average of these income cutoffs.[18] Finally, we solve for the income distribution below $280,000 under the parametric assumption that over this range, incomes within each profession follow a *Pareto-lognormal distribution* (Colombi, 1990). We denote the standard pdf of a Pareto-lognormal by $f_{\alpha,\mu,\nu}$. This smooth distribution approximates a lognormal with parameters μ and ν at low incomes and a Pareto with parameter α at high values and therefore does a good job of matching both the central tendency and upper tail of the income distribution.

Our precise parametric assumption is that $f_i^y(y) = A_i f_{\alpha_i,\mu_i,\nu_i}(y)$ for $y \leq$ $280,000. We choose A_i, α_i, μ_i, and ν_i to maximize the likelihood of observing the BLS data, conditional on f_i^y taking the form already estimated for $y \geq$ $280,000 and conditional on continuity at $y =$ $280,000. Specifically, for each profession i, the BLS partitions all workers in i into income bins. These bins can be written as $\{s_{i,k}, y_{i,k}^-, y_{i,k}^+\}$, where k indexes the constituent SOCs in i, and $s_{i,k}$ workers in i have incomes in $[y_{i,k}^-, y_{i,k}^+)$; $\sum_k s_{i,k} = s_i$, the total number of workers in i. Let $\widehat{f_i^y}$ denote the income distribution heretofore estimated for $y \geq$ $280,000. We use the following likelihood estimator to obtain the Pareto-lognormal parameters:

$$\widehat{A}_i, \widehat{\alpha}_i, \widehat{\mu}_i, \widehat{\nu}_i, = \arg \max_{A,\alpha,\mu,\nu} \sum_k s_{i,k} \log \left(F_i^y(y_{i,k}^+) - F_i^y(y_{i,k}^-) \right),$$

where F_i^y is the cdf corresponding to the pdf f_i^y, and the following constraints bind: $f_i(y) = A_i f_{\alpha_i,\mu_i,\nu_i}(y)$ for $y <$ $280,000, $f_i^y(y) = \widehat{f_i^y}(y)$ for $y \geq$$280,000, and $A_i f_{\alpha_i,\mu_i,\nu_i}($280,000$)$

[18]The break point adds a second degree of freedom in extending the pdf from $1,200,000 to $280,000. Using two degrees of freedom, we perfectly match the number of workers in this interval as well as their average income. Matching both statistics is critical for our analysis. The average income in this interval determines much of the aggregate spillover of each profession. The number of workers in the interval determines the average externality of workers earning these incomes, which matters for the optimal income tax at these incomes.

$$= \widehat{f_i^y}(\$280,000).^{19}$$

Table 2.2: *Summary Statistics of Estimated Professional Income Distributions*

	Population Share	Income Share	Median Income	99th Percentile Income
Art	1.0%	1.4%	$40,000	$497,000
Engineering	2.0%	3.9%	$73,000	$452,000
Finance	0.9%	4.3%	$85,000	$2,075,000
Law	0.4%	2.1%	$113,000	$1,627,000
Management	3.9%	12.9%	$83,000	$1,273,000
Medicine	0.5%	2.9%	$201,000	$1,346,000
Operations	2.4%	3.6%	$54,000	$368,000
Real Estate	0.3%	1.0%	$50,000	$1,393,000
Research	1.1%	1.8%	$62,000	$399,000
Sales	2.3%	3.3%	$48,000	$414,000
Teaching	3.2%	3.2%	$43,000	$126,000
Other	82.0%	59.7%	$29,000	$118,000

Notes: "Population share" is the fraction of the total workers in each profession, and "Income share" is the fraction of aggregate income earned by workers in each profession. "Median income" and "99th percentile income" are the 50th and 99th percentile incomes within each profession. The results describe the United States in 2005.

Table 2.2 reports summary statistics on the resulting distributions of income for each profession. Skilled professions comprise 18% of all workers, and skilled workers earn 40% of all income. The most populated skilled professions are management and teaching, and the least are real estate, law, and medicine. Substantial heterogeneity in income exists among the skilled professions. Median income ranges from $40,000 in art to $201,000 in medicine. Incomes vary even more at the 99th percentile. For instance, engineering and finance have similar median incomes, but the 99th percentile income in finance ($2,075,000) is more than four times greater than that in engineering ($452,000).

Figure 2.1 shows the allocation of workers across professions at each income. Although skilled workers account for only 18% of the total population, they comprise the majority

---

[19]The α_i estimated at this step need not equal the α_i estimated to fit the income distribution over $1,200,000.

Figure 2.1: *Distribution of workers at each income level: (a) all workers; (b) skilled workers. At each income y, the share of workers in profession i is $s_i f_i^y(y) / \sum_j s_j f_j^y(y)$, where s_i is the share of all workers in i and f_i^y is the probability density function for income in i. The results describe the United States in 2005.*

of high earners, as documented in Panel (a). Panel (b) details the composition of skilled workers at each income. At low incomes, the most common profession for skilled workers is art, a result resonant with the image of the "starving artist." Teaching, sales, and operations comprise most of the skilled lower middle class, whereas engineering and management are the largest group in the upper middle class. Nearly all wealthy skilled workers are in finance, law, management, and medicine, and the very wealthy work primarily in management and finance, with some also in law and real estate.

These income distributions by profession are determined in equilibrium by sorting as well as underlying income possibilities. In Appendix B.4.2, we graph, under our baseline assumption of no comparative advantage, the estimated underlying distributions of income at each skill level, from which individuals choose professions.

Preference and skill parameters

In our baseline analysis, we use a value of $\rho = 1$, which imposes a unidimensional skill distribution on the skilled workers and rules out sorting on comparative advantage. In the broad population and in the short term, this assumption is clearly problematic given the strong evidence of sorting into educational tracks based on comparative advantage shown empirically by Kirkebøen *et al.* (Forthcoming). However, reconciling a significant, long-term comparative advantage at the top end of the income distribution with the massive reallocations of talent (from the middle-class professions to law and finance) over time observed by Goldin *et al.* (2013) and Philippon and Reshef (2012) is difficult. The sorting patterns such a comparative advantage would create are counterintuitive. For example, they imply an upward productivity shock in finance will cause mean wages to fall in finance because those who switch in will primarily be workers without large profession-idiosyncratic ability draws. This pattern seems inconsistent with the influx of extremely high-skilled workers that accompanied the growth of the financial profession as documented quantitatively by Philippon and Reshef and discussed ethnographically by Patterson (2010).

We therefore focus on the admittedly very special case of general ability, because of the

more plausible sorting patterns it induces and because comparative advantage may be less extreme in the long term when educational curricula and long-term life goals of students may be adjusted. In the sensitivity analysis, we use a smaller value of $\rho = 0.75$ to explore the effects of comparative advantage on optimal tax rates.[20]

We then calibrate σ and β to match two moments of the distribution of income given the parameters and distributions estimated by Lemma 2, which in turn use σ and β. We iterate this step until we reach convergence on a fixed point. The first moment is the elasticity of total economy income with respect to 1 minus the tax rate. A vast literature (Saez *et al.*, 2012b) estimates this moment using tax reforms. Chetty (2012b) reviews this literature and favors a long-run value for this elasticity of 0.33. We adopt this value as our baseline, and experiment with 0.1 and 0.5 in the sensitivity analysis. To match the moment, we consider the response of aggregate income to a change in taxes, holding profession externalities constant but allowing workers' hours and professional choices to vary. Precisely, we compute $\partial \log Y / \partial \log(1 - T')$, where Y is total income and T' is a constant marginal tax rate. We numerically compute this derivative around the average empirical marginal tax rate T'_{2005}, holding each $E_i(Y_1, ..., Y_n)$ constant.

The second moment is the sensitivity of profession choice with respect to relative income, which helps tie down β, but has not been previously estimated in the literature to our knowledge. To calibrate this sensitivity, we exploit the secular growth in finance wages and employment between 1980 and 2005. As estimated by Philippon and Reshef (2012), the share of all workers in finance grew from 0.35% to 0.87% over this time, while the wages in finance relative to the rest of the (non-farm) economy grew from 1.09 to 3.62.[21] To match

---

[20]When we vary ρ to 0.75, we re-estimate the productivity distributions but we continue to use the values of σ and β estimated with $\rho = 1$. We do this because for $\rho = 0.75$ (and other similar values) we cannot find β to match the moment in (2.11). When comparative advantage is high, a positive productivity shock to finance actually *lowers* the relative wage in finance, because the shock attracts workers with low productivity to switch into finance. Thus, low ρ rules out a secular increase in finance employment and relative wages as a response to a productivity shock. Rather than try to model these increases differently, we simply hold σ and β constant as we vary ρ.

[21]These figures use the "other finance" subprofession defined by Philippon and Reshef (2012), because it is constructed similarly to our "finance" profession. The number of workers estimated by Philippon and Reshef (2012) in "other finance" in 2005 equals the number of workers we estimate in "finance" in 2005.

these trends, we study the marginal effect of a productivity shock to finance, which we model as a shock that multiplies each worker's productivity in finance by some constant $\bar{a}$.[22] The relative wage of finance equals $\widetilde{w}_i = \int_{\Theta_i} w_i(\theta)/s_i f(\theta) d\theta / \sum_{j \neq i} \int_{\Theta_j} w_j(\theta)/(1 - s_i) f(\theta) d\theta$, where i denotes the index of finance. The moment we match is the fraction

$$\frac{\partial s_i / \partial \bar{a}}{\partial \log \widetilde{w}_i / \partial \bar{a}}, \tag{2.11}$$

where each partial derivative is evaluated at $\bar{a} = 1$.[23]

Externalities

The identification of the externality parameters ϵ_i relies on three inputs: the returns to scale γ of each externality, the marginal social product $\partial Y / \partial Y_i$ of output in each profession, and the $\delta_{i,j}$ linkages.

The literatures we draw on provide no clear guidance on the returns to scale from the various externalities we consider. In our baseline analysis, we therefore choose $\gamma = 1$. The alternate values we use for sensitivity analysis are 0.5, 0.9, and 1.1, which allow us to explore the effects of diminishing and increasing returns to scale of the externalities. Similarly, we set $\delta_{i,j} = 1$ (uniform externalities) for all i and j as a baseline and consider alternate specifications in the sensitivity analysis.

To calculate the marginal social output from each profession, we draw on the literatures that estimate economy-wide externalities from various professions. Although we have done our best to faithfully represent the current literature, we emphasize that these estimates are *highly* uncertain extrapolations from heterogeneous and not easily comparable studies primarily aimed at different estimands than those we draw from them. The resulting

---

[22]Specifically, for i corresponding to finance, $a_i(\theta)$ is replaced by $\bar{a}a_i(\theta)$ for all θ.

[23]To obtain an empirical value for this moment, we must make an assumption about how frequently new workers replace incumbent ones. Our model is one of long-term professional choice, so s_i is best interpreted as the *flow* of workers into finance; the Philippon and Reshef (2012) data concern the stock. In our baseline analysis, we assume 5% of the worker stock is replaced each period. Appendix B.4.3 shows this assumption leads to a value of the above derivative of 0.01. In the sensitivity analysis, we use replacement rates of 3% and 10%, which lead to respectively higher and lower values of β.

estimates are listed in Table 2.3.

Table 2.3: *Aggregate Externalities by Profession: Baseline Estimates*

	Externality as share of income	Source	Method
Arts	0	—	—
Engineering	0.6%	Murphy *et al.* (1991)	Cross-country regression of GDP on engineers per capita
Finance	−1.4%	French (2008)	Aggregate fees for active vs. passive investing
Law	−0.2%	Murphy *et al.* (1991)	Cross-country regression of GDP on lawyers per capita
Management	0	Gabaix and Landier (2008)	Calibrated model indicating CEO pay captures managerial skill and firm characteristics
Medicine	0	—	—
Operations	0	Bloom *et al.* (2013)	Randomized experiment measuring effect of consultants on plant productivity
Real Estate	0	—	—
Research	16.7%	Murphy and Topel (2006)	Willingness-to-pay for longevity gains from medical research
Sales	0	—	—
Teaching	7.3%	Card (1999)	Returns to education in excess of teacher salaries
Other	0	—	—

Notes: This table reports the total externalities for each profession as a share of the total income in the economy, which is $6.3 trillion according to our estimates from Table 2.2. We calculate each externality using results from the listed papers; see the text for a description of how we map each paper's results to an aggregate externality figure. We use these externalities throughout the paper, except in the case of management and research. In sensitivity analysis, we explore the implications of a negative externality for management (taken from Piketty et al., 2014) and a smaller positive externality for research (taken from Jaffe, 1989).

To arrive at the marginal social product $\partial Y / \partial Y_i$, we divide each profession's total social product by its total private product.[24] The private product is given by the "income share" column of Table 2.2, and the social product is the sum of this private product and the externality given by Table 2.3. For example, the marginal social product of teaching equals

---

[24]This empirical ratio gives the average externality rather than the marginal one. However, some of the aggregate spillovers we take from the literature seem better interpreted as marginal effects (Murphy *et al.*, 1991; Chetty *et al.*, 2014). We believe simply dividing the social product by the private product to estimate the marginal externality is most transparent, rather than making further adjustments with the estimates from the literature.

$(3.2\% + 7.3\%)/3.2\% = 3.28$.

Given the high degree of uncertainty and inevitable subjectivity in these estimates, we devote the remainder of this section to briefly highlighting how we calculate the aggregate externalities in Table 2.3, with required calculations left to Appendix B.4.4. Our prior is that Coasian bargaining should eliminate externalities, so when these literatures do not offer a clear finding, we set the aggregate externality to 0. In the cases in which these literatures offer conflicting results, we adopt one value as a baseline and use an alternate value for sensitivity analysis.

Arts   Although some evidence, and a number of good theoretical arguments, suggest the arts generate some positive externalities, we are unable to find a plausible basis for estimating the magnitude of these externalities, and consequently assume 0 to be conservative.

Engineering   The only study we found of externalities from engineering is a cross-country ordinary least-squares regression by Murphy *et al.* (1991). They investigate the impact of the allocation of talent on GDP growth rates rather than on GDP levels. To be conservative and fit within our static framework, we interpret these impacts as one-time effects on the level of output rather than impacts on growth rates. We multiply their estimate of the GDP impact of an increase in the fraction of students studying engineering by the number of students studying engineering according to the OECD to obtain an externality of 0.6% of total income.

Finance   French (2008) estimates the cost of resources expended to "beat the market" by subtracting passive management fees from active management fees. Bai *et al.* (Forthcoming) show the informativeness of stock and bond prices (measured in their ability to predict earnings) has stayed constant since 1960, despite a vast growth of the finance profession documented by Philippon (2010). We therefore interpret the entirety of French (2008)'s estimates, which amount to 1.4% of total income in 2005, as negative externalities from finance.

Law   Murphy *et al.* (1991) estimate externalities from law in the same manner they calculate externalities from engineering, and we apply the same methodology to yield a -0.2% externality as a percent of total income. Kaplow and Shavell (1992) present several models of why the provision of legal advice may exceed the social optimum.

Management   Two strands in the literature offer competing views on the externalities of management. According to the first strand (Bertrand and Mullainathan, 2001; Malmendier and Tate, 2009), chief executive officer (CEO) compensation shifts resources from shareholders to managers in ways that do not actually reflect the CEO's marginal product. Piketty *et al.* (2014) argue that 60% of the CEO earnings elasticity with respect to taxes represents this rent-seeking behavior, implying the negative externalities from management are 7.7% of total income. The other half of the literature argues market forces can explain CEO compensation (Gabaix and Landier, 2008) and suggests that therefore externalities are 0. Most managers in our sample work at lower levels of firms where the problems of measuring marginal product highlighted by the critics of CEO compensation are less likely to apply, so we take the figure of 0 as our baseline and use the -7.7% figure in sensitivity analysis.

Medicine   We could find no literature estimating the externalities of (non-research) medicine and so set the externality to 0 to be conservative.

Operations   This profession is comprised of consultants and IT professionals. Bloom *et al.* (2013) conducted a field experiment to determine the causal impact of management consulting on profits. They interpreted their results as consistent with the view that consultants earn approximately their marginal product, and thus we assume no externality for consulting.

Real Estate   We could find no literature estimating the externalities of brokers, property managers, and appraisers and so set the externality to 0 to be conservative.

Research   Our baseline estimate for the externalities from research comes from the value of medical research, measured in terms of people's willingness to pay for the additional longevity this research makes possible. Murphy and Topel (2006) estimate the annual gains of medical research equaled 20% of GDP from 1980-2000. Traditional GDP accounting does not capture this externality, in contrast to our model, so we divide it by GDP augmented with this externality to obtain $\frac{.2}{1+.2} = 16.7\%$. Although this externality may be the largest externality from academia and science, this estimate is still conservative in assuming no gains accrue from other research fields.

An alternative measure of research externalities comes from the literature that calculates the social returns to R&D. Jones and Williams (1998) suggest the socially optimal amount of R&D activity is four times the observed amount, which we loosely translate into a three-times externality or 5.4% of GDP. A narrower benchmark for this externality focuses only on the externalities of universities to profits made by geographically proximate firms as studied in Jaffe (1989). We use his estimates to calculate a much smaller 2.7% externality, which we use as a lower-bound estimate in our sensitivity analysis.

Sales   Although an extensive theoretical literature argues the welfare effects of advertising can be positive or negative (Bagwell, 2007), we are not aware of any work attempting a comprehensive estimate of externalities, and therefore, as with medicine, we use an externality of 0.

Teaching   We calculate the social product of teaching as the impact of an additional year of schooling on aggregate earnings of all workers in the economy. The spillover from teaching is then this social product less the annual earnings of all teachers. As our estimate of the effect of a year of schooling on earnings, we use a 10.5% gain, which equals the midpoint of the numbers collected in Card (1999)'s review and also the estimate from Angrist and Krueger (1991). Because teachers earn 3.2% of economy income, we use a spillover from teaching of 7.3% of economy income.

We also compute the aggregate effect of teaching on earnings using Chetty *et al.* (2014)'s

measure of teacher quality and its long-run impact on eventual student earnings. We use the ratio of total teacher pay to its standard deviation in our data multiplied by the social product Chetty *et al.* (2014) estimate for a standard deviation in teacher quality to obtain an aggregate effect equal to 9.6% of economy income. This figure leads to a spillover of 6.4% of economy income. Given the similarity between the two spillover estimates and the fact that the estimate based on returns to schooling is more easily interpretable in the aggregate, we use the Card (1999) number as our estimate.

2.5 Results

Before investigating optimal taxes, considering the quantitative value of a leading force determining them is instructive: the externality ratio $e(y)$ in the equilibrium at the optimal tax schedule. We defined this externality ratio in Section 2.2 as the average marginal externality of income earned by those with income equal to y. Proposition 1 showed that in the special case when workers cannot switch professions, the optimal tax schedule satisfies $T'(y) = -e(y)$, thus setting marginal tax rates equal to the average negative externality ratio at each income level. We plot $-e(y)$ as the hashed line along with the optimal tax rates we discuss below in Figure 2.2. Absent the allocative labor supply margin, these two items in Figure 2.2 would be mirror images of each other. Interestingly, the results differ markedly from this benchmark.

2.5.1 Optimal taxes

Given the underlying skill distributions, preference parameters, and externalities we estimate, we numerically calculate the marginal tax schedule that maximizes social welfare. This procedure uses significant computational resources, so we restrict attention to schedules with eight brackets, with cutoffs at $25k, $50k, $100k, $150k, $200k, $500k, and $1m. This restriction clearly violates Assumption 1 but allows for direct optimization at reasonable cost.

Figure 2.2 presents the results. Optimal taxes (the solid line) begin with negative rates
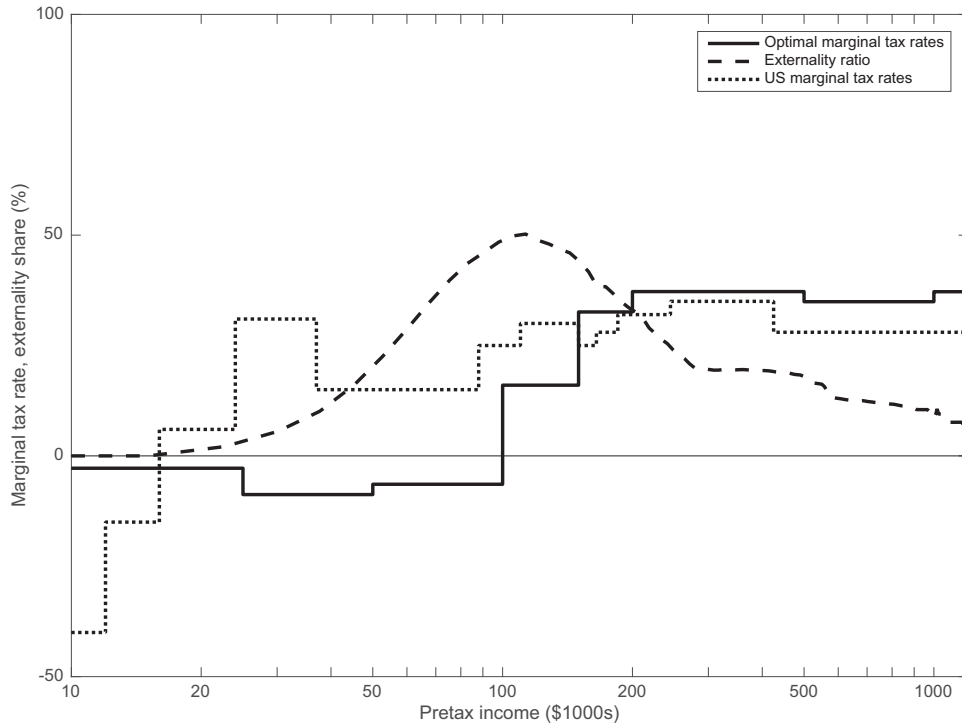
Figure 2.2: *Optimal marginal taxes for the United States in 2005. The externality ratio is the quantity $e(y)$, defined in Section 2.2 to be the average marginal externality of income earned by workers with income equal to y. The marginal tax rate displays the optimal tax rates over the eight brackets specified, given the data and baseline assumptions explained in Sections 2.4.1 and 2.4.2. US Marginal Tax Rates are taken from Figure 4 of CBO (2005) and denote the effective marginal tax rate for a married couple with two children in 2005, accounting for the Earned Income Tax Credit, the Alternative Minimum Tax, the phaseout of itemized deductions, the child tax credit, and personal exemptions.*

of about 6% on income up to $100,000 and then feature progressively increasing marginal rates after that. The top rate on income above $1m is 37%, and similar marginal rates hold for income above $150k in other brackets.[25]

To understand this tax schedule, consider the net tax liabilities of workers at different income levels relative to that of a worker with zero income. These *net* tax liabilities are all our optimal schedule identifies; the revenue requirement as explained by Lemma 1 solely

---

[25]Figure 2.2 presents a local maximum for marginal tax rates. We did find a second local maximum in which welfare was slightly higher ($22 per person). This alternative schedule is nearly identical to the one in Figure 2.2 except the marginal rates in the $150k-$200k bracket are much higher, over 95%. This optimum is likely an artifact of the way the brackets are constructed. It is present on only the smallest bracket (in log terms), and it disappears when we change the $150k-$200k bracket to $150k-$250k. For these reasons, we do not focus on this optimum. We report it in Appendix B.4.5.

determines the overall level of the tax schedule. Due to the negative rates that last until $100k, net tax liabilities are negative up to $138k, so that a worker earning $138k pays the same tax as a worker earning no income. Beyond this point, the marginal rate varies, but on average is about 35%. The smallest tax liability is for a worker earning $100k, who receives a net income subsidy of $6,100.

The top tax rates are close to the marginal tax rates the federal government in the United States has applied to top incomes since 1986; the 2005 US federal schedule of marginal rates is pictured in the small dashed lines. Thus, regarding tax rates on the rich, the model's recommendation matches the positive reality. Our model generates these optimal rates without any redistribution motive. The tax rates serve only to increase positive externalities and decrease negative ones.

The model's recommendations differ from policy at lower incomes. Empirically, rates below $100,000 are much higher than the model's negative optimal rates, both because statutory rates are higher (as depicted in Figure 2.2) and because benefits to the poor phase out as income increases over this range (CBO, 2005). The model prescribes negative rates on income all the way up to $100k, which is a much higher threshold than those used by income subsidies, such as the Earned Income Tax Credit.

To see most sharply the impact of the allocative margin, note that at high incomes, the externality ratio is positive but so are marginal tax rates. Researchers produce the positive externalities at these high incomes—although they constitute a small number of top earners, their externalities are extremely large relative to the negative externalities of law and finance. Yet despite the net positive externalities at high incomes, tax rates are still positive and large there because externalities are even higher at lower incomes. The top tax rates are positive to induce higher earners to switch to lower-paying professions that produce greater externalities.

2.5.2 Welfare gains and the allocation of talent

We now calculate the gains associated with taxation in our model with respect to two reference points: the empirical US economy in 2005, and a *laissez-faire* economy without any income tax. The latter comparison measures the general efficacy of the income tax for improving welfare, whereas the former provides the marginal improvement that could be obtained from changing the tax already in place. *Laissez-faire* serves as an informative benchmark for the additional reason that it is the optimal marginal tax schedule in our model when externalities are absent.

Table 2.4: *Welfare and the Allocation of Talent under Different Tax Regimes*

	Laissez-Faire	2005 US Data	Optimal Nonlinear Income Tax	Pre-Reagan Income Tax
	A. Per Capita Welfare Gains Relative to *Laissez-Faire*			
Levels	—	$312	$815	−$535
Percent	—	0.4%	1.2%	−0.8%
	B. Share of Skilled Workers in Each Profession			
Art	4.3%	5.1%	4.3%	5.4%
Engineering	12.4%	11.8%	13.3%	11.6%
Finance	5.7%	5.1%	5.3%	4.7%
Law	3.5%	2.7%	3.1%	2.3%
Management	24.0%	22.5%	24.4%	21.4%
Medicine	4.7%	2.9%	3.2%	2.1%
Operations	12.7%	13.3%	13.2%	13.9%
Real Estate	1.6%	1.7%	1.6%	1.8%
Research	5.9%	6.2%	6.2%	6.3%
Sales	10.6%	11.9%	10.8%	12.6%
Teaching	14.6%	16.7%	14.7%	17.8%

Notes: Each column reports results from simulating the estimated economy under different tax regimes. Laissez-Faire *denotes no marginal income tax, and "2005 US Data" denotes the approximate 2005 marginal tax schedule used to estimate the model. "Optimal Nonlinear Income Tax" refers to the optimal tax schedule shown in Figure 2.2, and "Pre-Reagan Income Tax" equals the 1980 US tax schedule.*

Panel A of Table 2.4 reports the results. Relative to *laissez-faire*, the optimal tax raises average utility by $815, or 1.2%. The tax achieves a smaller gain of 0.8% relative to the empirical economy, which is not surprising given the tax used to model the empirical

economy (a flat 30% tax) is close to the optimal tax we calculate. These gains are significant, but still small relative to the externalities calculated in Table 2.3. These large externalities—for instance, research at 16.7% of the economy—suggest a reallocation of talent to more productive professions could increase welfare by much more than the 1.2% achieved by the optimal income tax. Our findings that welfare gains are quite small are robust to all scenarios we consider in Table 2.5 except those with targeted subsidies to research; they never exceed 2.1%, and in some scenarios are smaller than 0.5%. The largest welfare gains come when the allocative margin is strongest (when individuals switch elastically across professions or the intensive-margin elasticity is small) and the smallest come when we assume a smaller externality of research.

A possible reason for the inefficacy of the income tax is that it induces little switching between professions, as workers' tax liabilities are independent of their professions. To investigate this idea, we calculate the allocation of talent under *laissez-faire* and under the optimal tax. Panel B of Table 2.4 reports the share of skilled workers in each profession in the data and in each of these two simulations. Relative to *laissez-faire*, the optimal tax decreases the share of workers in negative externality professions (finance and law) and increases the share in positive externality professions (engineering, research, and teaching). However, none of these changes are very large, and the broad allocation of talent stays the same. Relative to the status quo, the optimal tax primarily shifts individuals out of low-earning professions (e.g., art, sales and teaching) and into middle income professions (e.g., engineering and management). These changes result from the marginal rates in the status quo being much higher on the working and middle class than in the optimum. This reallocation does some good, mostly by raising incomes rather than externalities per unit income, but allocates workers out of teaching.

These results suggest that historical tax reductions are unlikely to have played a large role in the shifts in talent allocation. To confirm this hypothesis, we use the Tax Foundation's US Federal Individual Income Tax Rates history to simulate talent allocation and welfare under the 1980 ("Pre-Reagan") income-tax schedule. This schedule involves much higher rates

and a more progressive structure; it provides a more extreme departure from *laissez-faire* than the 2005 schedule. Welfare is substantially lower under the pre-Reagan rates relative to the status quo and is lower by 0.8% relative to *laissez-faire*. The allocation of talent under this schedule is shown in the final column of Table 2.4. As expected, the allocation of talent is only slightly different than *laissez-faire* under the pre-Reagan schedule.

2.5.3 Sensitivity to alternate assumptions

Table 2.5 reports the optimal tax rates under various alternate assumptions, which we now discuss.

Table 2.5: *Optimal Tax Rates and Welfare Gains Relative to* Laissez-Faire *for Different Assumptions*

	\$0-\$25k	\$25k-\$50k	\$50k-\$100k	\$100k-\$150k	\$150k-\$200k	\$200k-\$500k	\$500k-\$1m	\$1m+	Welfare Gain
				Tax Rate Bracket					
Baseline	−2.8%	−8.8%	−6.4%	**16.0%**	**32.6%**	**37.2%**	**34.9%**	**37.2%**	**1.2%**
A. Elasticities									
Low σ	−19.5%	11.5%	46.6%	71.7%	74.8%	61.7%	59.8%	65.0%	2.0%
High σ	−2.3%	−9.4%	−16.5%	0.4%	23.3%	32.3%	29.8%	32.4%	1.4%
Low β	−2.6%	−11.2%	−17.2%	2.3%	22.6%	30.8%	28.7%	31.6%	0.6%
High β	−2.6%	−6.2%	−0.1%	24.8%	37.5%	41.6%	39.8%	41.8%	1.8%
B. Externalities									
Smaller research externality	−1.4%	−0.7%	7.1%	20.6%	26.4%	24.8%	20.8%	24.4%	0.5%
Negative management externality	−2.3%	−5.7%	4.5%	31.2%	43.9%	46.2%	50.0%	60.3%	2.0%
Finance on finance	−2.8%	−10.4%	−8.7%	13.3%	28.9%	32.2%	28.0%	27.5%	0.8%
Engineering on engineering	−2.2%	−6.3%	−0.3%	18.4%	31.9%	36.3%	34.1%	36.3%	1.1%
Research on engineering	−1.2%	−1.0%	6.4%	19.5%	26.0%	24.2%	20.7%	23.3%	0.4%
$\gamma = 0.5$	−2.7%	−8.3%	−5.7%	15.8%	31.8%	36.1%	33.4%	35.7%	1.1%
$\gamma = 0.9$	−2.8%	−8.4%	−5.6%	16.9%	33.1%	37.4%	35.0%	37.3%	1.2%
$\gamma = 1.1$	−2.8%	−9.1%	−7.1%	15.4%	32.2%	37.1%	34.8%	37.1%	1.1%
Negative own externality	−1.6%	−3.7%	1.6%	19.9%	30.0%	29.8%	27.7%	30.8%	0.7%
C. Comparative Advantage									
$\rho = 0.75$	−3.2%	−12.5%	−19.5%	9.9%	32.6%	29.5%	16.0%	11.7%	1.1%
D. Tax Instruments									
Research subsidy (negative own externality)	−3.1%	−3.2%	4.3%	18.1%	30.9%	43.4%	28.8%	42.0%	37.3%
Research subsidy ($\gamma = 0.5$)	0.0%	0.0%	−0.1%	−1.1%	−3.1%	−0.1%	80.0%	80.0%	99.5%

Elasticities

We begin by varying our input for the elasticity of taxable income with respect to 1 minus the tax rate. We experiment with values of 0.1 and 0.5 (our baseline was 0.33). These inputs lead to estimated σ values of 0.01 and 0.4; our baseline estimate was 0.24. The estimated β changes only slightly to 1.47 and 1.42 relative to the baseline of 1.5. These changes are small because we are holding constant the separate moment that mostly determines β. Thus, this experiment alters the *relative* importance of the intensive-margin elasticity rather than just the overall elasticity that, as we noted in Section 2.2.2, plays no role in determining optimal tax rates under our theory.

Optimal tax rates are much higher with a lower σ, as shown by the rates above 70%, especially on the upper-middle-class range that makes the largest difference between being in the wealthy and middle-class professions. These high rates are consistent with the logic of our calibrated example; profession switching becomes more important relative to hours in determining optimal taxes when σ is small. The effect is more dramatic here, however, because the mixed sorting created by the richer substitution patterns in our analysis here means that the value of the allocative margin is smaller. Unless the intensive margin is very weak, it has a strong influence on optimal taxes, implying that weakening it significantly raises optimal taxes by leaving the weak allocative margin to determine tax rates uncontested. At the higher value of $\sigma = 0.4$, which is at the high end of estimates obtained from microeconomic studies (Chetty, 2012b), optimal rates are still progressive. The top rates are slightly smaller, and the negative rates on low earners are more extreme.

We next vary the profession-switching sensitivity β. As discussed earlier, we vary the assumed replacement rates of workers into finance in our calibration to 10% and 3% from the baseline of 5%. These alternate assumptions lead to values of β of 1.0 and 2.0 versus the baseline value of 1.5. Consistent with our argument that only the relative size of the intensive and allocative margin elasticities matters, the lower value of β gives results similar to the higher value of σ. The higher value of β moves toward results for the low value of σ, though not as dramatically, because it involves a much smaller change in the ratio of the

two forces (β increases by one third while σ fell by an order of magnitude). These results provide another quantitative confirmation that discrete profession switches are central to the progressive structure of taxes we find.

Externalities

We vary the externalities in numerous ways, given our substantial uncertainty over both their magnitude and functional form.

We begin with two specifications that change the magnitude of the externalities. The research externality is the largest externality. To investigate the degree to which this externality drives the results, we use a much smaller aggregate externality of 2.7% instead of 16.7% of economy income. As discussed earlier, this smaller number is calibrated from the literature on R&D externalities. This smaller research externality does indeed produce smaller top tax rates and higher rates for low-income workers, as can be seen in Table 2.5. The negative rates end earlier (at $50k), and rates on higher earners are lower (between 20% and 30%). But the basic structure of the tax system stays intact. Changing the management externality from 0 to an aggregate negative externality of 7.7% of economy income makes a much larger difference by making nearly all high-earning professions have negative externalities. Tax rates on income between $150k and $1m jump from about 35% to about 45%, and the tax rate on income over $1m rises to 60%. This result is analogous to our finding in Section 2.3.2 that raising the negative externality of the high-earning profession is more important than raising the positive externality of the low-earning profession. The range of externality magnitudes explored here—which reflects the opinions of various economists—generates as much variation in optimal tax rates as differences in the elasticity of taxable income.

We now alter the functional form of the externalities. First we consider what happens when the externality from finance falls entirely on itself by setting $\delta_{i,j} = 0$ for $i \neq j$ when j indexes finance. The resulting top tax rates, especially at the very top, are smaller than the baseline optimal rates. For instance, the rate above $1m falls from 37% to 27.5%. This

103

decline in rates is consistent with the theoretical results of RS, who show the social planner has little incentive to tax rent-seeking when the rent-seekers compete against each other, which is the case when finance externalities fall entirely on finance.

Next we alter the functional form to allow all engineering externalities to fall on engineering. This specification is motivated by industrial research clusters like Silicon Valley in which engineering firms create new ideas that enhance the productivity of other engineering firms (Saxenian, 2006). This specification leaves optimal tax rates essentially unchanged. We also consider a specification in which research externalities fall entirely on engineering. To be consistent with how we calibrate research externalities, we use the smaller externalities calibrated from the R&D literature for this exercise. Relative to the optimal rates under that calibration, the rates when research externalities fall entirely on engineering are largely unchanged. In principle, these different linkages could lead to larger rates by causing feedback effects that increase the net benefit of profession switching. This effect appears to be balanced by the lower aggregate externalities implied by Jaffe (1989)'s much lower externality estimates, suggesting that even his estimates, correctly interpreted, would lead to quite similar results.

Our baseline analysis assumed externalities were linear in output by setting the returns to scale parameter γ to 1. We explore the possibility of economies or diseconomies of scale in externalities by setting $\gamma = 0.5$ and 0.9. We also set $\gamma = 1.1$ to investigate the possibility of slightly increasing returns. None of these values materially change the optimal rates, although the low value of $\gamma = 0.5$ does slightly reduce top tax rates. Our tax schedule is sufficiently similar to the status quo that a linear approximation to externalities makes little difference to the results.

Finally, we consider congestion effects wherein the arrival of new workers lowers the productivity of existing workers in a given profession. We implement these congestion effects by assuming each dollar of private product in teaching or research raises the aggregate output of the profession by only 50 cents.[26] Optimal rates do diminish, but the effect is

---

[26]We choose $\delta_{j,j}$ for j corresponding to research and teaching so that the relevant diagonal entries in the

slight, with top rates falling from 37% to 31%. In contrast to the work of RS, we find the multi-profession nature of our economy likely significantly mitigates congestion effects. Negative externalities within a profession are only a small part of the overall impact of individuals migrating into a profession, compared to the impact of that profession on the broader economy.

Comparative advantage

We next consider the impact of allowing comparative advantage, which changes the patterns of substitution across professions. Absent comparative advantage, taxes induce shifts of the very skilled across fields. With comparative advantage, most substitution will occur among lower-ability individuals because higher-ability individuals will tend to have much lower ability in another field.

To explore this effect, we change ρ from 1 to 0.75. We draw from Kirkebøen *et al.* (Forthcoming) a sample statistic, which we call comparative advantage, to give a sense of the sorting caused by this lower value of ρ. For each skilled worker, define $i_1(\theta) = \arg\max_i F_i^a(a_i(\theta))$ to be the profession in which she is (relatively) most productive and $i_2(\theta) = \arg\max_{i \neq i_1(\theta)} F_i^a(a_i(\theta))$ to be the profession in which her (relative) productivity is second highest. The formula for comparative advantage is given by

$$
\sum_{\substack{1 \leq i,j \leq n \\ i \neq j}} Pr\left[i_1(\theta) = i, i_2(\theta) = j\right] \times
$$

$$
\left(\mathbb{E}\left[\log\left(\frac{y_i^*(\theta)}{y_j^*(\theta)}\right) \middle| i_1(\theta) = i, i_2(\theta) = j\right] - \mathbb{E}\left[\log\left(\frac{y_i^*(\theta)}{y_j^*(\theta)}\right) \middle| i_1(\theta) = j, i_2(\theta) = i\right] \right).
$$

This formula gives the average relative income premium of skilled workers in their most skilled profession. At $\rho = 0.75$, comparative advantage is equal to 0.4, representing an average premium of 40 log points of income, close to the figures observed empirically, though in a very different setting, by Kirkebøen *et al.*.[27] When $\rho = 1$, comparative advantage

---

quasi-Jacobian matrix J defined in Section 2.4.3 equal 0.5.

[27] They estimate comparative advantage using the field-of-study choice of students, as opposed to the ability

equals 0.

The tax schedule with comparative advantage features declining marginal rates for top incomes, with the rates at $150k-$200k similar to before but top tax rates much lower. The new rate on income above $1m is 11%, and the rate between $500k and $1m is 14%. The negative rates for low earners actually increase to a maximum of 21.4%. Comparative advantage makes profession switching unattractive to those earning very high incomes because they are likely to have high idiosyncratic incomes in their present profession. Thus, comparative advantage brings optimal rates for the wealthy closer to the (negative) intensive-margin optimum. Rates remain largely unchanged at middle incomes because individuals with low idiosyncratic ability may still substitute across professions.

The fact that comparative advantage changes the structure of taxes more than any other feature we analyze demonstrates the importance of profession-substitution patterns for optimal taxes.

Tax instruments

We argued the small gains from taxation result from an untargeted income tax struggling to precisely reallocate individuals. To explore targeted policies, we introduce a linear income tax (or subsidy) to supplement the non-linear income tax that the government can levy directly on research, which is the profession we estimate produces the strongest externalities.

Under our baseline assumptions these instruments can fail to have an optimum, so we modify the baseline parameters in two ways. First, we choose each $\delta_{i,i}$ so that the externality of teaching and research on themselves equals -0.1.[28]

---

levels, which they cannot observe. They also find an average value of 0.4, but their number is not closely analogous to ours because of the different definition, because their sample is limited to students on the margin between professions, and because they focus on fields of study rather than professions. We experimented with defining comparative advantage using profession choices but were not able to match their estimate even for $\rho = 0$. Sorting on non-pecuniary utility ψ significantly blunts comparative advantage given the β we have estimated.

[28]When these parameters change but the available tax instrument remains only a nonlinear income tax, the optimal rates stay close to the optimum in the baseline specification, changing to -2.5%, -5.1%, -0.3%, 19.8%, 32.0%, 32.8%, 30.4%, and 33.0%.

In this case, we find an optimal research tax of -392.1%, which would multiply salaries by four times even beyond their subsidized 2005 levels. An important part of this subsidy is to offset the negative effect on salaries of the crowding induced by the negative value of $\delta_{i,i}$. Table 2.5 reports the optimal income-tax rates accompanying the optimal subsidy, which hardly change from our baseline, even getting a bit higher at the top. Other professions still produce enough externalities that targeting research does not significantly change the picture. Furthermore, because all the negative effects of research fall onto research, the targeted subsidy can offset these burdens.

Welfare is much higher under the research subsidy. Relative to *laissez-faire*, welfare is 37.3% higher. The subsidy allocates 44.7% of skilled workers to research in the equilibrium, almost 10 times the baseline amount. Targeted support for certain key professions can thus greatly raise welfare, and a progressive income tax can still be optimal even in the presence of such targeting.

Another way of avoiding an explosive result is to impose diminishing returns in the production of externalities ($\gamma = 0.5$). Unlike within-research crowding, such diminishing returns do not diminish the private returns to research. They also have an equal effect on externalities in *all* professions, rather than specifically affecting research and teaching. This linearity of private returns makes much larger welfare gains possible, even with a less extreme subsidy. In particular, the optimal research subsidy is now 120%, a large number but much smaller than the previous case, and this subsidy achieves a *much* larger welfare gain of 99.5%. However, this smaller (if still very large) targeted research subsidy is accompanied by a radical change in the optimal income tax. As reported in Table 2.5, the optimal top tax rates are a nearly confiscatory 80%, the figure at which we capped rates for convergence; optimal unconstrained rates are likely higher. Rates below the top two are essentially 0.

Intuitively, a large targeted subsidy and confiscatory rates are two methods of inducing greater movement into professions, especially research and teaching, with large positive externalities. When teaching and research have negative externalities on themselves, targeted

subsidies are a more effective tool because they offset the reduction in private returns from negative self-externalities as untargeted taxes cannot. However, when the production of externalities merely produces decreasing returns, teaching and research remain competitive professions that near-confiscatory taxes can be effective in inducing individuals to enter. In particular, once a moderate subsidy has been applied to research, progressive untargeted taxes become a much more desirable tool because the subsidy raises the attractiveness of research, ensuring that most substitution out of high-earning professions occurs into research rather than into a field generating fewer positive externalities. Employing a smaller targeted subsidy and larger untargeted taxes is thus optimal because they also induce migration into teaching, which cannot be targeted.

2.6 Conclusion

This paper proposes an alternative framework for the optimal taxation of income relative to the standard redistributive theory of Vickrey (1945b) and Mirrlees (1971a). Income taxation acts as an implicit Pigouvian tax that is used to reallocate talented individuals from professions that cause negative externalities to those that cause positive externalities. Optimal tax rates are highly sensitive to which professions generate what externalities and to the labor-substitution patterns across professions. They do not depend on the overall elasticity of taxable income. Optimal taxes in our baseline calibration are not radically different from the US federal income tax schedule.

Our estimates of optimal tax rates depend crucially on several empirical objects whose value is highly uncertain. The first and most important of these are the externalities created by different professions. Our extrapolations from the cross-country regressions of Murphy *et al.* (1991) to determine the externalities of engineering and law are speculative at best, and could be greatly improved by further empirical analysis. For example, simple decomposition of legal activities between adversarial and compliance expenditures could already be useful. Kaplow and Shavell (1992) argue that an important component of an arms race exists in adversarial expenditures, whereas spending on compliance may be helpful in

ensuring rules are correctly implemented to avoid harmful externalities. Combining such an analysis with estimates of the impact of litigation on improving economic incentives could generate an account nearly as persuasive as that in Card (1999) and Chetty *et al.* (2014)'s estimates of the external effects of schooling. Similarly, output in engineering could be disaggregated into three components: new product development, where theory suggests imperfect appropriability creates positive externalities (Spence, 1976); operations, where externalities should be limited; and reverse engineering, where negative business-stealing externalities predominate (Hirshleifer, 1971).

The second uncertain empirical object is the pattern of labor substitution across professions. The closest evidence known to us comes from the causal impact on earnings of quasi-random assignment across fields of study at universities (Hastings *et al.*, 2013; Kirkebøen *et al.*, Forthcoming). But many of the professional choices studied in our paper are made conditional on a given undergraduate degree. Neither Hastings *et al.* (2013) nor Kirkebøen *et al.* (Forthcoming) identify substitution patterns in response to changes in material rewards. Studies of such substitution patterns are critical to determining optimal tax policy, but progress will likely require difficult-to-obtain long-term exogenous variation in professional wages.

Future research could relax the assumptions of our analysis in two ways. First, Piketty *et al.* (2014) and RS consider models in which individuals simultaneously engage in both rent-seeking and productive activities. By contrast, in our model, each unit of output from a profession causes the same externality. Yet the greatest benefit from reallocation might arise within professions. Take finance, for example. Hirshleifer (1971) argues that high-speed trading is oversupplied, whereas Posner and Weyl (2013) show long-term price discovery of large bubbles is just as likely to be undersupplied as innovative breakthroughs. Uniform income taxation, even by profession, is unlikely to be a sufficient tool to achieve such reallocation. Mechanisms that do are an exciting direction for future research.

Second, this paper assumes all statutory taxes are paid. Tax avoidance would substantially change the analysis, especially if avoidance is profession-specific. For example, if

financiers can avoid labor-income taxation by representing their income as capital income against which a lower rate is charged, income taxation might make finance more attractive rather than less. Incorporating avoidance considerations into our model is an interesting direction for future research.

Chapter 3

Regressive Sin Taxes[1]

3.1 Introduction

Tax policy is shaped by multiple social objectives that are not always aligned. One objective, emphasized by a growing body of work in behavioral public finance, is to address market failures due to consumer misoptimization. Governments tax goods such as cigarettes (Gruber and Kőszegi, 2004) and unhealthy foods (O'Donoghue and Rabin, 2006) that people are believed to consume too much, and create subsidies for goods such as energy-efficient products that people are believed to consume too little (Allcott and Taubinsky, forthcoming).[2] However, many such policies are regressive. Cigarettes are overwhelmingly consumed by low-income earners (Gruber and Kőszegi, 2004), while energy-efficiency subsidies are overwhelmingly taken up by high income earners (**?**). A government concerned with reducing income inequality – as in the large body of work in optimal tax theory initiated by Vickrey (1945a) and Mirrlees (1971b) – may thus face a tension between redistributive and corrective motives. In the absence of consumer mistakes (and externalities), goods that are preferred by low income earners – such as cigarettes – should be subsidized by a government attempting to reduce wealth inequality (Saez, 2002a).

---

[1]Co-authored with Dmitry Taubinsky

[2]In a (government-provided) health-insurance setting, **?** also suggest lowering copays for certain chronic-disease medication that people may not be adhering to due to psychological biases.

How should a government concerned with both wealth inequality and consumer mistakes decide what taxes or subsidies to set for cigarettes, unhealthy foods, or energy efficient products? Although the objectives of correcting consumer mistakes and reducing wealth inequality have mostly been analyzed in isolation, many pressing policy questions involve an interplay between these two motives. The aim of this paper is to provide a bridge between the behavioral economics literature characterizing policy in the presence of consumer mistakes and the optimal tax literature (e.g., Mirrlees, 1971b, Atkinson and Stiglitz 1976, Saez 2001, 2002) characterizing taxes in the presence of income inequality and redistributive motives. Our model nests the models considered in these two literatures, in particular building on and extending the crucial insights first advanced in Gruber and Kőszegi (2004).

The results in our paper can be categorized into five sets of contributions. The first is a simple "sufficient statistics" formula for the optimal tax rate on internality-producing goods which are consumed differentially across the income distribution. (For more on the sufficient statistics approach, see Saez 2001a and Chetty 2009 in public finance, and Chetty, Looney and Kroft, 2009; Mullainathan, Schwartzstein and Congdon, 2012; Allcott and Taubinsky, forthcoming; Farhi and Gabaix, 2015 in behavioral economics.) The formula shows that the optimal tax can be computed from a few high-level statistics: the price elasticity of demand for the good in question, and the covariances between the degree of overconsumption (or underconsumption), the elasticity of demand, and consumers' incomes. The formula also clearly indicates whether, and to what extent, redistributive concerns counteract or magnify the corrective role of sin taxes. Second, we supplement this formula with sharp comparative statics results about how the optimal sin tax changes with the size of consumer bias, the strength of redistributive motives, and the price elasticity of demand. Third, we reexamine the common intuition that internalities behave like targeted externalities. Although this intuition may be a reasonable quantitative guide in some circumstances, we demonstrate that the importance of "correlated heterogeneity"— for example the covariance between demand elasticity and consumers' incomes—is unique to internality taxes, and does not affect optimal externality taxes. Fourth, we demonstrate the applicability of our

framework by presenting calibrations of optimal cigarette taxes, accounting for the high concentration of smokers among the poor. We show that under the most plausible parameter ranges, the optimal cigarette remains positive, but is substantially reduced in the presence of redistributive concerns. Fifth, we consider alternative instruments such as nonsalient taxes or cigarette warning labels, and we show how redistributive concerns can render such instruments strictly superior to conventional fully salient taxes—results which do not generally hold in the absence of redistributive motives.

Our model builds on Saez's (2002a) extension of Atkinson and Stiglitz (1976) by considering an economy of consumers with heterogenous earnings abilities and tastes, who choose labor supply and a consumption bundle that exhausts their after-tax income. The policymaker chooses a non-linear income tax and a set of linear commodity taxes. But while the standard approach in optimal tax theory assumes that the planner is in agreement with consumers about what is best for them, we instead analyze the general case in which that assumption does not hold because of possible psychological biases such as present bias, limited inattention, or incorrect beliefs. This approach nests O'Donoghue and Rabin (2006) and Allcott and Taubinsky (forthcoming) as special cases, and allows us to generate robust economic insights that apply to the variety of psychological biases that policymakers have invoked to justify specific policies such energy-efficiency subsidies, incandescent lightbulb bans, and taxes on unhealthy foods.

Our baseline analysis in Section 3.2 shows that the interaction between corrective and redistributive motives are nuanced: the two motives can both oppose and amplify each other. On the one hand, redistributive motives directly oppose corrective motives when the taxes are regressive. Just as Atkinson and Stiglitz (1976) and Saez (2002b) show that goods preferred by low income consumers should be subsidized and those preferred by high earners should be taxed, our general framework suggests that redistributive motives can push against corrective motives. At the same time, redistributive motives can also amplify corrective motives: an inequality averse policymaker will be more concerned with money "left on the table" by low earners than by high earners. Thus in contrast to the Mullainathan,

113

Schwartzstein and Congdon (2012) and Allcott and Taubinsky (forthcoming) result for the case of no income inequality, we show more generally that it is not only the average marginal level of misoptimization that matters, but whether this misoptimization is concentrated on the low-end or the high-end of the income distribution. We show that the estimable degree of "bias concentration" consists of the covariance between income and bias, as well as the covariance between income and the price elasticity of demand. We then show that the extent to which redistributive motives oppose or amplify corrective motives is determined by the price elasticity of demand. The lower the elasticity, the more redistributive motives oppose corrective motives; the higher the elasticity the more redistributive motives amplify corrective motives.

In addition to deriving the optimal tax formulas, in Section 3.2 we also derive comparative statics results about how the sign and magnitude of the optimal tax depends on the size of the bias and the price elasticity of demand. These results are novel not only because they treat issues related to psychological biases, but also because such comparative statics results are often difficult to prove and thus rare in models that allow for a nonlinear income tax and general ability distributions. In addition to providing sharp qualitative guidelines for tax policy, these results also help relate our work to the fundamental insights first suggested by Gruber and Kőszegi (2004) about how the incidence of cigarette taxes on low-income earners is shaped by their price elasticity of demand.

We end Section 3.2 with two extensions. In Section 3.2.7 we elucidate the key conceptual differences between our analysis of internalities and the recent work on externalities in the Mirrlees model (**?**).[3] In Section 3.2.8 we consider a more general version of our model with income effects, i.e., in which exogenous wealth shocks can have meaningful impacts on consumption of the sin good. We show that in this case, the optimal commodity tax formulas we derive in the simpler setting carry forward, except with the addition of a new term corresponding to the fiscal externality from labor supply distortions caused by the

---

[3]While focusing on internalities rather than externalities, our work complements **?** by extending implications to preference heterogeneity, which appears to be an important factor in explaining cigarette consumption, for example, as discussed in 3.3.

commodity tax.

In Section 3 we apply this framework to a classic "sin good": cigarettes. We use data on smoking rates and elasticities by income to calibrate a model in which there are internalities from the health costs associated with smoking, and in individuals with lower income earning abilities prefer to smoke at higher rates than high earners. We compute the optimal cigarette tax for a range of degrees of behavioral bias. Consistent with the analytic results discussed in Section 3.2, we show that, relative to the optimal tax absent redistributive concerns, the optimal cigarette tax can be larger, smaller, or indeed negative (a cigarette subsidy) depending on the degree of bias and the elasticity of demand for cigarettes. For the most plausible parameter values, the optimal tax remains positive but substantially reduced.

Because our framework dispenses with the assumption that consumers optimize perfectly and respond only to prices, a second set of contributions in our paper is to consider non-standard policy instruments, and to non-standard responses to (seemingly) standard policy instruments. In Section 3.4.1 we consider the question of when the commodity tax should take the form of a salient commodity tax included in the price or a less salient commodity tax that a consumer only sees at the register. We show that the answer to this question depends on the covariance between earnings ability and attentiveness to the less salient tax. Restricting to the case of a regressive tax on sin goods, we first show that when all consumers are equally attentive to the less salient tax, the policymaker will never choose a less salient tax that is positive – he would instead prefer to use the more salient tax. And conversely, the policymaker should never use a salient subsidy for redistributive purposes – the less salient subsidy would be more efficient. We show, however, that the common intuition that corrective instruments should be maximally salient may be reversed when low-income earners are more attentive to the sales tax than high income earners. In a simplified version of our model, we show that for any level of bias, it can be optimal to tax with a less salient instrument if wealth inequality is sufficiently high and if high income consumers are sufficiently inattentive relative to low income earners.

In Section 3.4.2 we move on to study the efficiency persuasion tactics – such as graphic

images depicting the adverse consequences of smoking – and we show that these tactics are most efficient precisely when standard commodity taxes are most regressive. Generally, we show that in the presence of redistributive concerns, being able to use the persuasion campaign lowers the optimal corrective sin tax. In fact, we show that when persuasion is sufficiently powerful and its total social cost is sufficiently low, the optimal policy mix consists of a high level of persuasion and a "sin subsidy".

We conclude in Section 3.5 with a discussion of the behavioral statistics highlighted by our analysis. We argue that further empirical work on individual differences and correlated heterogeneity in behavioral biases is essential for providing robust guidelines for public policy.

3.2 Adding Redistributive Concerns to Models of Internality Taxes

3.2.1 Model set up

We begin with a simplified framework that adds redistributive concerns to the types of models considered in Gruber and Kőszegi (2004, henceforth GK), O'Donoghue and Rabin (2006, henceforth OR), Mullainathan, Schwartzstein and Congdon (2012, henceforth MSC), Allcott, Mullainathan and Taubinsky (2014, henceforth AMT), Allcott and Taubinsky (forthcoming, henceforth AT). These models consider a set up with quasilinear utility, and thus no effects of income on consumption of the good in question. Here, we will similarly begin with a situation in which changes in wealth do not change consumption of the sin good, and we show how adding redistributive concerns changes the basic conclusions. A key feature of the model considered in this section is that we are able to add redistributive concerns while keeping consumer behavior constant. The model here also complements the general treatment of the Ramsey model of taxation concurrently considered in Farhi and Gabaix (2015).[4]

---

[4]See also **?** for an analysis of optimal, government-mandated health insurance when consumers may misoptimize

We consider individuals with one-dimensional types indexed by $\theta \in \mathbb{R}$, distributed according to a cumulative distribution function F. In period 0, individuals choose earnings z, while in period 1 they choose a consumption bundle (c_1, c_2) subject to the budget constraint $c_1 + (p + t)c_2 \leq y$, where p is the relative price of c_2, t is the commodity tax on c_2, and y is after-tax income. In period 0, consumers choose according to a utility function $\hat{U}(c_1, c_2, z, \theta) = c_1 + \hat{u}(c_2, \theta) - \hat{\psi}(z, \theta)$, while in period 1 consumers choose consumption according to a utility function $U(c_1, c_2, \theta) = c_1 + u(c_2, \theta)$, with u not necessarily equal to $\hat{u}$. For concreteness, we assume that the consumer is sophisticated in the sense that in period 0 he correctly anticipates his period 1 behavior. However, our results would not change if we allowed for full or partial naivete.

The policy maker disagrees with the individual about the benefits of c_2, and believes the consumer should instead seek to maximize $V(c_1, c_2, z, \theta) = c_1 + v(c_2, \theta) - \psi(z, \theta)$. The policy maker sets a nonlinear income tax $T(z)$ and a linear commodity tax t to maximize total welfare given by $\int G(V(c_1, c_2, z, \theta))$ subject to the revenue constraint $\int (T(z(\theta)) + tc_2(\theta))dF \geq 0$, where $c_2(\theta)$ and $z(\theta)$ denote a type θ's choice of c_2 and z in the presence of the taxes $T(z)$ and t. The important difference from the standard optimal tax set-up is that the planner maximizes a weighted average of utility functions V that may not correspond to the utility functions U (or $\hat{U}$). Here, V denotes the utility function according to which consumers would choose if they were not subject to various psychological biases.

The difference between U (or $\hat{U}$) and V captures a variety of different possible psychological biases. We now provide some examples:

1. *Present bias.* As in O'Donoghue and Rabin (2006), suppose that consumers underweight the future health costs of c_2 (e.g, potato chips) because of present bias. Suppose that true utility from c_2 is given by $u(c_2) = m(c_2) - h(c_2)$, where h is the health cost realized in the future. Our framework can be seen as a reduced-form representation of a (slightly) more dynamic model in which consumers choose z in period 0, choose c_1 and c_2 in period 1, and experience the health cost in period 3. A consumer with present-biased, $\beta - \delta$, preferences (Laibson, 1997) will have a utility function $\hat{U} = \beta\delta(c_1 +$

$m(c_2)) - \beta\delta^2 h(c_2) - \psi(z, \theta)$ in period 0 and utility function $U = c_1 + m(c_2) - \beta\delta h(c_2)$ in period 1. The planner's utility function is $V = \delta(c_1 + m(c_2)) - \delta^2 h(c_2) - \psi(z, \theta)$.

2. *Limited attention* to certain attributes of c_2, as in Gabaix and Laibson (2006); DellaVigna (2009); **?**, and as documented by Allcott and Taubinsky (forthcoming) for energy-efficient appliances. Suppose that c_2 represents investment in energy-efficiency. Purchasing the more energy-efficient product generates immediate utility $m(c_2)$, but also generates energy-cost savings equal to $K(c_2)$. Inattentive consumers, however, underweight these future costs by some (potentially endogenous) $1 - \alpha$. Then $\hat{U} = c_1 + m(c_2) + \alpha K(c_2) - \psi(z, \theta)$ and $U = c_1 + m(c_2) + K(c_2)$.

3. *Incorrect beliefs*, as documented for food choices and energy-efficiency choices by Allcott and Taubinsky (2013); Attari *et al.* (2010); Bollinger *et al.* (2011). Consumers may simply have incorrect beliefs about certain attributes of c_2, such as its calorie content or its energy efficiency. In the setup in example 1, consumers may believe that the health costs are $\tilde{h}(c_2)$ rather than c_2. In the setup from example 2, consumers may believe that the energy efficiency is $\tilde{K}(c_2)$ rather than c_2.

4. Any combination of the above biases, as well as any other biases that allow consumer choice to be represented by a differentiable utility function.

In Section 3.2.8 we will consider the fully general case in which income levels can affect consumption of c_2 – we defer this more general model until later because the presence of income effects generates consumer behavior that is different from the behavior in the standard GK, OR, MSC, AMT, and AT models.

We assume that consumers' preferences are such that the choice of c_2 is interior and thus characterized by the first order condition $u_1(c_2, \theta) = p + t$. We also impose the mild technical condition that $\psi_1(0, \theta) > 0$. We assume that the optimal income tax gives rise to a one-to-one mapping between θ and z, and thus with some abuse of notation, we let $c_2(z, t)$ denote the consumption bundle chosen by a type θ individual who has after-tax income $z - T(z)$ and faces a commodity tax t. We let $H(z)$ denote the distribution of incomes at the

optimum, and we use $C_2(t) = \int c_2(z, t) dH$ to denote aggregate consumption of c_2.

For a type θ consumer, set $\gamma(c_2, \theta) = u_1(c_2, \theta) - v_1(c_2, \theta)$ to denote the bias in valuing the marginal utility from c_2 in period 1. Here too we abuse notation to let $\gamma(z, t)$ denote the bias of a consumer earning income z and consuming $c_2(z, t)$. Throughout the analysis in this section, we will follow AMT and AT in defining the *average marginal bias*:

$$\bar{\gamma}(t) = \frac{\int \frac{dc_2}{dt} \gamma(z, t) dH(z)}{\int \frac{dc_2}{dt} dH(z)}$$

Intuitively, if dC_2 is the marginal change in total consumption of c_2 when t is perturbed, $\bar{\gamma}$ is the average amount by which consumers over- or under-estimate the change in utility from that change in consumption. When no confusion results, we may eliminate the dependence on t for simplicity.

3.2.2 The case of no redistributive concerns

When G is linear, so that there are no redistributive concerns, the optimal tax system is straightforward:

Proposition 3.1. *Suppose that G is linear. Then the optimal tax schedule $(T^*(z), t^*)$ must satisfy $t^* = \bar{\gamma}(t^*)$.*

Proposition 3.1 replicates known "sin tax" results from GK, OR, MSC, AMT, AT. The proposition shows that when consumers overconsume some good, an optimal tax on that good should be positive and set at the level of the average amount of marginal misperception. Conversely, when $\gamma < 0$, so that consumers underconsume the good (as for the case of energy efficiency upgrades in AMT and AT), then the optimal policy is to subsidize the good at the average level of marginal misperception. Following MSC, AMT and AT, the proposition generalizes GK by allowing heterogeneity and mistakes other than present bias, and generalize OR by allowing for mistakes other than present bias. The OR result that taxing sin goods is optimal even if some people optimize perfectly is an immediate corollary of our more general characterization of the optimal tax system: if $\gamma(\theta) = 0$ for some types

but $\gamma(\theta) > 0$ for other types, then clearly $\bar{\gamma} > 0$ and thus a positive tax is always optimal.

Although Proposition 3.1 mostly replicates known results, there is one new insight contained in the proposition. Because we consider a two-period setting in which consumers first choose earnings z and then choose their consumption of c_1 and c_2, our framework allows for the possibility that a time-inconsistent consumer might alter his choice of before-tax income z because of the "mistakes" that he knows his period 1 self will make. This is in contrast to the static frameworks with fixed incomes considered in MSC, AMT, AT. Nevertheless, Proposition 3.1 shows that just like in the static frameworks in MSC, AMT and AT, the optimal commodity tax should still equal the marginal bias. Perhaps surprisingly, the period 0 preferences of the consumer do not alter the formula for the optimal commodity tax t.[5] The intuition behind this result will be made in clear in the derivations in the next subsection.

3.2.3 Adding redistributive concerns

We now move on to the case in which G is strictly concave, so that there are redistributive concerns. Consider a perturbation that increases the commodity tax t by dt. This reform has both mechanical effects and behavioral effects due to consumers' response to the tax change. The reform mechanically raises additional revenue equal to $C_2 dt$. Letting $g(z) = G'(V(z - T(z) - (p + t)c_2(z), c_2(z), z)/\lambda$ denote the social marginal welfare weight corresponding to a type earning income z, normalized by the marginal value of public funds λ, the corresponding mechanical welfare loss taxpayers is $-dt \int g(z)c_2(z,t)dH$. It also causes a loss in revenue as consumers substitute away from c_2 to c_1. Letting $dc_2(z,t)$ denote the change in c_2 purchased by a z-earner in response to the tax, the reduction in tax revenue from this behavioral response is equal to tdC_2. In neoclassical models, the substitution has no first-order effect on welfare due to the envelope theorem. Here, however, the planner disagrees with the consumer's choice of c_2, and thus the planner's marginal

---

[5]Of course, because the choice of z is still determined by $\hat{U}$, the nature of $\hat{U}$ might alter the optimal income tax, which then alters the social marginal welfare weights, and thus indirectly affects the optimal commodity t.

rate of substitution differs from the consumer's by γ. Thus the change in $c_2(z,t)$ generates a first-order welfare change equal to $-dt \int \gamma(z)g(z)\frac{dc_2(z,t)}{dt}dH$. Because this perturbation has no impact on the consumer's optimal choice of z (a consequence of our quasilinearity assumptions), the total welfare impact, normalized by the cost of public funds, is

$$\frac{dW}{dt} = \int \left(-g(z)c_2(z,t) - \gamma(z)g(z)\frac{dc_2(z,t)}{dt} + t\frac{dc_2(z,t)}{dt} + c_2(z,t) \right) dH(z) \qquad (3.1)$$

Now let $\zeta(z,t) = \frac{dc_2(z,t)}{d(p+t)}\frac{p+t}{c_2(z,t)}$ denote the elasticity of demand for c_2 and let $\bar{\zeta}(t) = \frac{dC_2(t)}{d(p+t)}\frac{p+t}{C_2(t)}$ denote the aggregate elasticity. (Again, we may omit the dependence on t.) Let $\omega(z,t) = \frac{\gamma(z)\zeta(z)}{\bar{\gamma}\bar{\zeta}}$ denote how biased and how elastic a person is relative to the average. Under the optimal policy, equation (3.1) is equal to zero, and the optimal tax satisfies the following first-order condition:

$$
\begin{aligned}
t &= \frac{\int \gamma(z)g(z)\frac{dc_2(z,t)}{dt}dH}{\int \frac{dc_2(z,t)}{dt}dH} + \frac{\int c_2(z,t)(g(z)-1)dH}{\int \frac{dc_2(z,t)}{dt}dH} \\
&= \frac{\int \gamma(z)g(z)\zeta(z,t)\frac{c_2(z,t)}{p+t}dH}{\int \frac{dc_2(z,t)}{dt}dH} + \frac{\int c_2(z,t)(g(z)-1)dH}{\int \frac{dc_2(z,t)}{dt}dH} \\
&= \bar{\gamma} + \frac{Cov_H[g(z),\gamma(z)\zeta(z,t)c_2(z,t)]}{\bar{\zeta}(t)\frac{C_2(t)}{p+t}} + \frac{Cov_H[g(z),c_2(z,t)]}{\bar{\zeta}(t)\frac{C_2(t)}{p+t}} \\
&= \underbrace{\bar{\gamma}\left[1 + \frac{Cov_H[g(z),\omega(z,t)c_2(z,t)]}{C_2(t)}\right]}_{\text{Corrective Benefits}} + \underbrace{(p+t)\frac{Cov_H[g(z),c_2(z,t)]}{\bar{\zeta}(t)C_2(t)}}_{\text{Regressivity Costs}} \qquad (3.2)
\end{aligned}
$$

where we have used the fact that $\int g(z)dH = 1$, which follows from our assumption that there are no income effects on labor supply. Set $\sigma(t) := Cov_H[g(z),\omega(z,t)c_2(z,t)]$ to denote the extent to which the average marginal bias is concentrated on the low income earners. Set $\rho(t) \equiv Cov_H[g(z),c_2(z,t)]$ to denote the regressivity of the tax. In terms of these two indices of concentration and regressivity, we find that the optimal tax t^* satisfies

$$t^* = \underbrace{\bar{\gamma}\left[1 + \frac{\sigma(t^*)}{C_2(t^*)}\right]}_{\text{Corrective Benefits}} + \underbrace{(p+t^*)\frac{\rho(t^*)}{\bar{\zeta}(t^*)C_2(t^*)}}_{\text{Regressivity Costs}} \qquad (3.3)$$

And under the simplifying assumption that γ and ζ are homogeneous, we have that $\sigma(t) = \rho(t)$ and thus that t^* must satisfy

$$t^* = \underbrace{\bar{\gamma}(t^*)\left[1 + \frac{\rho(t^*)}{C_2(t^*)}\right]}_{\text{Corrective Benefits}} - \underbrace{(p + t^*)\frac{\rho(t^*)}{|\zeta(t^*)|C_2(t^*)}}_{\text{Regressivity Costs}} \tag{3.4}$$

Focusing first on the simple case in equation (3.4) shows that the size of the corrective tax depends on two forces. The first component is the internality γ, but now scaled up or down by the extent to which consumption of c_2 is concentrated amongst the low income earners or the high income earners. When consumption of c_2 is concentrated amongst the low income earners, the planner is more concerned about the internality. This is reflected by $\rho > 0$, which inflates the corrective benefits. Conversely, when consumption of c_2 concentrated amongst the high income earners the planner is less concerned about the internality, which is reflected by $\rho < 0$.

The second component of (3.4) is the cost of imposing a regressive tax. When the low income earners prefer c_2, the tax is regressive and thus creates a redistributive cost, which again is reflected in $\rho > 0$. This force lowers the optimal tax and pushes it in the direction of being a subsidy when γ is close to zero. Conversely, when the high income earners prefer c_2, the tax is progressive: there is a redistributive gain from taxation of c_2. This force increases the optimal commodity tax and pushes it in the direction of being positive even when γ may be negative.

More generally, equation (3.3) shows that the corrective benefits depend not only on the extent to which consumption of c_2 is concentrated amongst low income earners, but also on the extent to which low income earners are more or less elastic than high earners. If the low income earners are particularly elastic to the tax, that further increases the gains from corrective taxation. This is captured in the first covariance term, which shows that it is not only the relationship between $g(z)$ and c_2 that matters, but also the relationship between $g(z)$ and $\omega(z)$.

Although (3.3) suggests that the optimal commodity tax t is a monotonic function of

bias, sharp comparative statics are not immediate because all terms in the formula in (3.3) are endogenous to the tax t. The optimal tax t^* is defined implicitly as the fixed point of the expression in (3.3), and thus care must be taken in characterizing how the tax varies with the primitives of the model. In the next two subsections, we formalize the intuitions suggested by equation (3.3) for "sin goods" that consumers overvalue (i.e. $\gamma > 0$) and "virtue goods" that consumers undervalue (i.e. $\gamma < 0$).

3.2.4 Comparative Statics for "Sin Goods"

Let t^{NR} denote the optimal tax when there are no redistributive motives, and let t^G denote the tax with redistributive motives, captured by the function G. We now characterize how t^G compares to t^{NR} as a function of bias for the case in which bias is positive, i.e., $\gamma > 0$, and a positive commodity tax is regressive, i.e., $\rho > 0$.

Proposition 3.2. *Suppose that G is strictly concave, that $\gamma(c_2, \theta) > 0$ for all c_2, θ, and that c_2 is decreasing in z. Holding constant U and ω, the following hold for an optimal tax policy (t^G, T^G):*

1. *Suppose that $\zeta(z, t)$ is bounded from above and bounded away from zero. Then there exists $\gamma^\dagger > 0$ such that $t^G < 0$ if $\gamma(c_2, \theta) < \gamma^\dagger$ for all c_2, θ.*

2. *Suppose $\{|\zeta(z, t)|\}_{z,t}$ and $\{|\omega(z, t)|\}_{z,t}$ are bounded away from zero. Then there exists a $\gamma^\dagger$ such that $t^G > 0$ if $\gamma(\theta) > \gamma^\dagger$ for all θ*

3. *Suppose that the conditions of (2) hold and that, moreover, $\omega(z, t)$ are nondecreasing in z for each t. Then for each t' there exist $\gamma^\dagger > 0$ such that $t^G > t'$ if $\gamma(\theta) > \gamma^\dagger$ for all θ.*

4. *Suppose that i) $\omega(z, t)$ is non-decreasing in z for each t ii) $\gamma(c_2, \theta) \equiv \gamma^\dagger \in \mathbb{R}^+$ iii) for each possible value of $\gamma^\dagger$, welfare is quasiconcave in t when the income tax is given by $T^G - (t - t^G)C_2$, iv) u_1 is bounded. Then $t^G > t^{NR}$ for high enough $\gamma^\dagger$.*

5. *In contrast, if $\sigma(t) < 0$ for all t then it always holds that $t^G < t^{NR}$.*

Part 1 of the Proposition says that under an innocuous regularity condition bounding elasticity ratios, the optimal sin tax may be negative even when bias is positive. The intuition

for this result can be obtained from (3.3) by setting $\bar{\gamma} = 0$. In verifying this result formally, the proof in the appendix shows that it is necessary to establish that $\bar{\gamma}/\rho \to 0$ as $\bar{\gamma} \to 0$ at the optimal tax policy.

Part 2 shows that under another set of mild regularity conditions, the optimal sin tax is positive for sufficiently high bias. And part 3 shows under stronger conditions that ensure $\sigma \geq 0$, the optimal tax can be arbitrarily high for sufficiently high values of bias. One theoretical nuance that complicates these comparative statics is that the optimal commodity tax need not necessarily be a monotonic function of $\bar{\gamma}$. Because high levels of bias may also increase redistributive concerns, and therefore ρ, higher $\bar{\gamma}$ does not universally imply high t^G. The role of the regularity conditions in parts 2 and 3 is to ensure that for high commodity taxes and high bias, consumers do not become so inelastic to the commodity tax that the regressivity costs outweigh the bias correction benefits.

Part 4 shows that when u_1 is bounded – so that consumption of c_2 is zero for a high enough t – redistributive concerns can make the optimal corrective tax higher than the Pigovian benchmark. To obtain some intuition for this, note that when both $\bar{\gamma}$ and t are high, it can simply be better to ban the product completely, so as to reduce the financial burden on the low-income earners.

Finally, part 5, which follows immediately from (3.3), simply says that when bias is concentrated on the high income earners, redistributive concerns always decrease the size of the optimal commodity tax.

3.2.5 Comparative statics for "Virtue Goods"

We now consider implications for "virtue goods" that consumers may not consume enough of. This may include healthy foods or investments in energy-efficient durables and appliances (AT, AMT). One potentially important distinction between this case and the case of sin taxes is the relationship between σ and ρ. With regressive sin goods that are more likely to be consumed by low-income earners, both σ and ρ are likely to be positive. Thus even though the tax is regressive, the bias is also more concentrated on the low income earners,

which makes the policymaker put more weight on the internality. In contrast, with virtue goods that are more likely to be consumed by the high income earners, ρ will be negative, but σ may be negative as well because the higher income consumers also consume more of c_2. At the same time, the reason higher income consumer consume more of c_2 may be because they are less biased, which would push σ back toward being positive. Overall, the sign of σ is unclear for the case of virtue subsidies.

A simple condition that we will impose for some of the results is that there is a Laffer curve for taxing c_2. Letting $C_2(t^M)$ denote total consumption of c_2 as a function of t, we define:

Condition 3.1 (Laffer curve). *There exists a finite t that maximizes* $tC_2(t)$.

When Condition 3.1 holds, we will let t^M denote the revenue-maximizing tax rate.

Proposition 3.3. *Suppose that G is strictly concave, that* $\gamma(c_2, \theta) < 0$ *for all* c_2, θ, *and that* c_2 *is increasing in z. Holding constant U as well as the concentration weights* ω, *the following hold for an optimal tax policy* (t^G, T^G):

1. *There exists* $\gamma^\dagger < 0$ *such that* $t^G > 0$ *if* $\gamma(\theta) > \gamma^\dagger$ *for all* θ.

2. *Suppose* $\{|\zeta(z,t)|\}_{z,t}$ *is bounded from below and from above, that Condition 3.1 holds, and that* $c_2(t^M, \underline{\theta}) > 0$. *Then for each t′ there exists a* $\gamma^\dagger$ *such that* $t^G < 0$ *if* $\gamma(\theta) < \gamma^\dagger$ *for all* θ..

3. *In contrast, if* $\sigma(t) < 0$ *for all t then it always holds that* $t^G > t^{NR}$.

Proposition 3.3 is analogous to Proposition 3.2. Part 1 shows that when bias is concentrated on high earners, it may be optimal to tax a product that consumers under-consume, for sufficiently small bias. Part 2 shows that under some regularity conditions, it is optimal to have a positive subsidy, for sufficiently large bias. Finally, part 3 shows that when bias is concentrated on the high-income earners, the optimal subsidy with redistributive concerns is always smaller in magnitude than the Pigovian benchmark without redistributive concerns.

3.2.6 The key role of the price elasticity of demand

We now focus on the role that the price elasticity of demand plays in the magnitude of the optimal commodity tax. Solving for t^* in (3.2), it follows that

$$
t^G = \frac{\bar\gamma\left[1 + \frac{Cov_H[g(z),\omega(z)c_2(z)]}{C_2(t)}\right] - p\frac{Cov_H[g(z),c_2(z)]}{|\bar\zeta|C_2(t)}}{1 + \frac{Cov_H[g(z),c_2(z)]}{|\bar\zeta|C_2(t)}}
$$

$$
= \bar\gamma\overbrace{\frac{|\bar\zeta|C_2(t) + |\bar\zeta|Cov_H[g(z),\omega(z)c_2(z)]}{|\bar\zeta| + Cov_H[g(z),c_2(z)]}}^{\text{Corrective Benefits}} - \overbrace{\frac{pCov_H[g(z),c_2(z)]}{|\bar\zeta|C_2(t) + Cov_H[g(z),c_2(z)]}}^{\text{Regressivity Costs}} \quad (3.5)
$$

$$
= \bar\gamma|\bar\zeta|\overbrace{\frac{C_2(t) + \sigma}{|\bar\zeta|C_2(t) + \rho}}^{\text{Corrective Benefits}} - \overbrace{\frac{p\rho}{|\bar\zeta|C_2(t) + \rho}}^{\text{Regressivity Costs}} \quad (3.6)
$$

When ζ and γ do not vary with type, we have that

$$
t^G = \gamma|\zeta|\overbrace{\frac{C_2(t) + \rho}{|\zeta|C_2(t) + \rho}}^{\text{Corrective Benefits}} - \overbrace{\frac{p\rho}{|\zeta|C_2(t) + \rho}}^{\text{Regressivity Costs}} . \quad (3.7)
$$

The key insight in equation (3.5) is that while the corrective benefit term is proportional to the elasticity, the regressivity costs term is not. This suggests that the importance of correcting consumer bias, relative to regressivity costs of the commodity tax, is proportional to the price elasticity of demand. Intuitively, if consumers are not at all elastic to a regressive tax, then the tax only redistributes money from the poor to the rich, without correcting consumption of c_2. Thus in the extreme case in which $|\bar\zeta(t)| \approx 0$, it is clear that commodity tax should be used almost exclusively as a tool for redistribution, rather than as a tool for correcting consumption of c_2. More generally, formula (3.5) clarifies that outside of this extreme case, the role of consumer bias in shaping the optimal commodity tax is extremely sensitive to the elasticity.

We now formalize the intuitions obtained from (3.5). To formally discuss the role of the elasticity, we restrict to a family of utility functions U that keep certain statistics – such as baseline consumption when $t = 0$ – constant, and then discuss the implications of selecting

utility functions with higher or lower elasticities. The function $\hat{U}$ will be held constant for all comparative statics in this section.

Proposition 3.4 (Implications for sin taxes). *Fix a bias function γ with $\gamma(\theta, c_2) > 0$ for all θ, c_2. Let $\mathbb{U}$ denote a space of utility functions U such that the following are constant for all $U \in \mathbb{U}$: i) $c_2(\theta, 0)$, ii) $u(c_2(\theta, 0), \theta)$, iii) the relative elasticity function $e(\theta, t) = \zeta(\theta, t)/\bar{\zeta}(t)$. Suppose also that $c_2(\theta, t)$ is decreasing in θ for all $U \in \mathbb{U}$ and that $e(\theta, t)$ is bounded. Then for each strictly concave G and a scalar $t' > -p$ there exists a $k > 0$ such that $t^G < t'$ for any $U \in \mathbb{U}$ satisfying $|\bar{\zeta}(t)| < k$ for all z, t.*

Proposition 3.4 shows that for a family of utility functions U characterized by (i)-(iii), the optimal commodity tax can be made arbitrarily small for a utility functions with sufficiently small elasticities. Although we have not been able to prove a converse for high elasticities, our simulations similarly suggest that the optimal corrective tax is increasing in the magnitude of the elasticity, and surpasses the Pigovian benchmark for sufficiently high elasticities.

For regressive "virtue subsidies," we prove a similar analogue of Proposition 3.4. The proposition similarly establishes that for a given family of utility functions U, it may be optimal to tax, rather than subsidize the under-consumed good if the elasticity is sufficiently close to zero.

Proposition 3.5 (Implications for virtue subsidies). *Fix a bias function γ with $\gamma(\theta, c_2) < 0$ for all θ, c_2. Let $\mathbb{U}$ denote a space of utility functions U such that the following are constant for all $U \in \mathbb{U}$: i) $c_2(\theta, 0)$, ii) $u(c_2(\theta, 0), \theta)$, iii) the relative elasticity function $e(\theta, t) = \zeta(\theta, t)/\bar{\zeta}(t)$. Suppose also that $c_2(\theta, t)$ is increasing in θ for all $U \in \mathbb{U}$ and that $e(\theta, t)$ is bounded. Then for each strictly concave G there exists a $k > 0$ such that $t^G > 0$ for any $U \in \mathbb{U}$ satisfying $|\bar{\zeta}(t)| < k$ for all z, t.*

3.2.7 Comparison to externality taxation

To compare our results to externality taxes, suppose instead that consumers maximize U, but that their total utility is given by $V = U - X(C_2)$, where $X(C_2)$ is the externality from

total consumption of C_2. Then analogous to equation (3.1), we have that

$$dW = \int \left(-g(z)c_2(z,t) - g(z)X'(C_2)\frac{dC_2}{dt} + t\frac{dC_2}{dt} + C_2 \right) dH(z) \qquad (3.8)$$

from which it follows that

$$t^* = \underbrace{X'(C_2)}_{\text{Corrective Benefits}} + \underbrace{(p+t^*)\frac{\rho(t^*)}{\bar{\zeta}(t^*)C_2(t^*)}}_{\text{Regressivity Costs}} \qquad (3.9)$$

Note that contrary to (3.3) , the first component of t^* is simply the marginal damages from the externality-producing good. Whereas (3.3) shows that the weight given to the internality depends on whether consumption of c_2 is concentrated on the low-income or high-income portion of the population, (3.9) shows that the covariance between income and consumption of c_2 does not affect the weight given to the externality. Intuitively, this is because both low-income and high-income consumers have the same marginal impact on the externality damages that affect everyone in the population. The logic is different with internalities, because the internality generated by a low-income consumer impacts only that low-income consumer. And because internalities have a higher impact on the welfare of low-income consumers than on the welfare of high-income consumers, it matters whether it is the low-income consumers or the high-income consumer who are the most biased.

The concepts generalize when externality production is heterogeneous and has heterogeneous welfare effects. Suppose that a type θ individual contributes $x(c_2, \theta)$ to the externality, so that $X(C_2) = \int x(c_2, \theta)dF$, and derives disutility $y(\theta)X(C_2)$ from the externality due to c_2, for some some scalar $y(\theta)$. Then generalizing the Diamond (1973) formula, we have that

$$t^* = \underbrace{\bar{x}\left(1 + Cov[g(\theta), y(\theta)]\right)}_{\text{Corrective Benefits}} + \underbrace{(p+t^*)\frac{\rho(t^*)}{\bar{\zeta}(t^*)C_2(t^*)}}_{\text{Regressivity Costs}} \qquad (3.10)$$

where $\bar{x} = \frac{\int x_1(c_2,\theta)\frac{dc_2}{dt}dF}{\int \frac{dc_2}{dt}dF}$ is the average marginal contribution to the externality. The average marginal externality $\bar{x}$ is analogous to the average marginal bias $\bar{\gamma}$ in equation (3.3). Equation (3.10) appears more similar to (3.3), but with an important difference. When sensitivity to

128

externalities is allowed to be heterogeneous, then as with internalities, it matters whether those most sensitive affected are the poor or the rich. However, the difference remains that with internalities – but not with externalities – the optimal commodity tax depends on how income covaries with both consumption of c_2 and the price elasticity of demand for c_2. Intuitively, for internalities it matters whether low-income earners are also those who are the most responsive to the the internality tax, since their utility is affected only by their own behavior. With externalities, it does not matter whether low income earners are most responsive to the tax, since their utility is affected by aggregate behavior.

The analysis here clarifies that in the standard case in which $y(\theta)$ is homogeneous, redistributive concerns always make the optimal externality-corrective taxes lower than the Pigovian benchmark when the commodity tax is regressive. And more generally, greater redistributive concerns cannot make the policymaker more concerned about the externality – only more concerned about the regressivity of the commodity tax. In contrast, even when taxes are regressive, redistributive concerns *can* make a policymaker more concerned with the internality, and might even lead to a tax that is higher in magnitude than the average marginal bias.

3.2.8   The general case with income effects and non-separable preferences

In this section, we consider a more general setting in which an individual's preference are not necessarily weakly separable in c_1, c_2 and z. We allow general functions $\hat{U}(c_1, c_2, z, \theta), U(c_1, c_2, z, \theta)$, and $V(c_1, c_2, z, \theta)$. Our goals here are twofold: First, we examine the robustness of the conclusions from equation (3.5) about the interaction of corrective and redistributive motives. Second, although we do not derive closed-form solutions, we still implicitly characterize the optimal tax system in terms of measurable elasticities and a summary statistic of consumer bias. These formulas help clarify what key empirical parameters are needed to estimate the optimal policy.

For simplicity, we assume that $\hat{U}_1 / \hat{U}_3 = U_1 / U_3 = V_1 / V_3$, meaning that consumers correctly tradeoff labor effort and consumption of c_1. This allow us to focus our analysis on

incorrect tradeoffs between c_1 and c_2, without the additional complication that consumers may make suboptimal labor supply choices because of a general misoptimization with respect to labor and consumption tradeoffs.[6]

Let $\xi(z, \theta, t)$ denote a type θ's earned income elasticity with respect to the net of tax rate $1 - T'$. We follow Jacquet *et al.* (2013) in defining this elasticity locally at the status quo tax regime. Let $\xi^c(z)$ denote the compensated elasticity, and let $\eta = \xi - \xi^c$ denote the income effect.

Analogously, let $\zeta^i(z, \theta, t)$ denote the elasticity of c_i with respect to net of tax price. Here $\zeta^i(z)$ is the total elasticity: it captures how much a person currently at income z will reduce his consumption of c_2, taking into account the fact that some of this reduction may be due to an income effect that results from this person lowering his income by some amount dz. Let $\zeta^{i,c}$ denote the compensated elasticity, and let $\varepsilon^i = \zeta^i - \zeta^{i,c}$ denote the income effect. The income effect is given simply by $\varepsilon^i(z) = \frac{dc_i}{dz}(p + t)$. Finally, let χ denote a type θ's earned income elasticity with respect to the net of tax price $p + t$ of c_2.

Consider now increasing the commodity tax by a small amount dt. First, this has a mechanical revenue effect given by $M = \int c_2(\theta, t)dF$. Second, this causes individuals to lower their income, creating a fiscal externality of size $I = \int \chi(\theta)\frac{z(\theta)}{p+t}T'(z(\theta))dF$. Third, the tax decreases consumption of c_2. The revenue loss from this is given by $R = \int t\zeta(\theta)\frac{c_2(\theta,t)}{p+t}dF$.

Finally, set $\gamma(\theta) = U_2^\theta/U_1^\theta - V_2^\theta/V_1^\theta$ to be the difference in perceived versus experienced marginal rates of substitution between c_2 and c_1, where c_1 and c_2 are chosen to maximize U at income $z(\theta)$ and tax $T(z(\theta))$. Set $g(\theta) = V_1^\theta(c_1, c_2, z)/\lambda$ where c_1 and c_2 are chosen to maximize the consumer's decision utility U at income $z(\theta)$ and tax schedule T.[7]

Then because $U_2/U_1 = p + t$ at the consumer's perceived optimal choice of consumption, the welfare effect from substitution, and from the utility loss to consumers paying the tax

---

[6]See Farhi and Gabaix (2015) for a general discussion of labor supply misoptimization, and **?** for an application with present bias.

[7]Note that this definition is slightly different from conventional marginal social welfare weights, as it is not identical to the social utility of giving a marginal dollar to type θ, however it serves as a useful summary of redistributive concerns in this context.

(normalized by the cost of public funds) is given by

$$
\begin{aligned}
S &= \frac{1}{\lambda} \int \left[\left(-(p+t)V_1^\theta + V_2^\theta \right) \zeta(\theta) \frac{c_2}{p+t} - c_2 V_1^\theta \right] dF \\
&= \frac{1}{\lambda} \int V_1^\theta \left[\left(-(p+t) + V_2^\theta/V_1^\theta \right) \zeta(\theta) \frac{c_2}{p+t} - c_2 \right] dF \\
&= \int g(\theta) \left[\left(-U_2^\theta/U_1^\theta + V_2^\theta/V_1^\theta \right) \zeta(\theta) \frac{c_2}{p+t} - c_2 \right] \\
&= -\int g(\theta) c_2(z, \theta, t) \left[\frac{\gamma(\theta)\zeta(\theta)}{p+t} + 1 \right] dF
\end{aligned}
$$

Now at the optimum, $M + I + R + S = 0$, from which it follows that

$$
\begin{aligned}
t &= \frac{\int g(\theta) c_2(\theta, t) \left[\frac{\gamma(\theta)\zeta(\theta)}{p+t} + 1 \right] dF - \int c_2(\theta) dF - \int \chi(\theta) \frac{z}{p+t} T'(z) dF}{\int \frac{dc_2(\theta,t)}{dt} dF} \\
&= \frac{\int g(\theta) c_2(\theta) \gamma(\theta) \zeta(\theta)}{\int \zeta(\theta) c_2(\theta) dF} - (p+t) \frac{\int c_2(\theta)(1 - g(\theta)) dF + \int z(\theta) \chi(\theta) T' dF}{\bar{\zeta} C_2(t)} \\
&= \bar{g}\bar{\gamma} + \frac{Cov[g(\theta), \gamma(\theta) c_2(\theta) |\zeta(\theta)|]}{\int |\zeta(\theta)| c_2(\theta) dF} - (p+t) \frac{Cov[c_2(\theta), g(\theta)] + C_2(\bar{g} - 1) + \int z(\theta) \chi(\theta) T'(z(\theta)) dF}{|\bar{\zeta}| C_2(t)} \\
&= \bar{\gamma}(t) \left[\bar{g} + \frac{Cov[g(\theta), \omega(\theta) c_2(\theta)]}{C_2} \right] - (p+t) \frac{Cov[c_2(\theta), g(\theta)] + C_2(\bar{g} - 1) + \int z(\theta) \chi(\theta) T'(z(\theta)) dF}{|\bar{\zeta}| C_2(t)}
\end{aligned}
$$

where $\bar{\gamma} = \frac{\int \gamma(\theta) \frac{dc_2(\theta,t)}{dt} dF}{\int \frac{dc_2(\theta,t)}{dt} dF} = \frac{\int \gamma(\theta) c_2(\theta) \zeta(\theta) dF}{\int c_2(\theta) \zeta(\theta) dF}$ denotes the average marginal bias, $\bar{g} = \int g(\theta) dF$ denotes the population average of $g(\theta)$, and $\omega(\theta) = \frac{\gamma(\theta)|\zeta(\theta)|}{\bar{\gamma}|\bar{\zeta}|}$ denotes the how biased and how elastic a consumer is relative to the population. As before, we set $\sigma(t) = Cov[g(\theta, t), \omega(\theta, t) c_2(\theta, t)]$ to denote the extent to which the misperception corresponding to a marginal change in consumption is concentrated on the low income earners. And we set $\rho = Cov[c_2(\theta, t), g(\theta)]$ to denote the regressivity of the tax. The new term $FE(t) = \int z(\theta) \chi(\theta) T'(z(\theta)) dF$ is the fiscal externality from the change in labor supply caused by an increase in the tax t. Thus in terms of these two indices, we again have that

$$
t^* = \bar{\gamma} \left[\bar{g} + \frac{\sigma(t)}{C_2} \right] - (p+t) \left[\frac{\rho(t) + FE + \bar{c}_2(\bar{g} - 1)}{|\bar{\zeta}| C_2} \right]
$$

The key difference from our baseline formula is that now it is not only the regressivity

term ρ that matters, but also the fiscal externality FE from a change in the commodity tax t. Intuitively, the fiscal externality results when exogenous differences in income change consumption of c_2. If a consumer were to reduce his consumption of c_2 from c_2' to c_2'' when he decreases his earnings from z' to z'', then the total taxes he pays would decrease by $(c_2' - c_2'')t$. Thus when consumption of c_2 is rising in income, increasing the commodity tax t causes consumers to lower their labor supply.

To explore the connection to standard results on externalities, suppose consumption of c_2 produces a harmful externality, and let $X(C_2)$ denote the impact that the externality has on each consumer's utility. Then as before,

$$t^* = \frac{X'(C_2)}{\lambda} - (p+t)\left[\frac{\rho(t) + FE + \bar{c}_2(\bar{g}-1)}{|\bar{\zeta}|\bar{C}_2}\right]$$

Again, a key difference between externality taxes and internality taxes is the bias concentration term σ, which shows that concerns for correcting internalities are amplified when the low income earners consume more of c_2 and are the most elastic. The difference is particularly transparent in the extreme case in which types are one-dimensional and *decision utility U* is weakly separable, so that $\rho + FE + \bar{c}_2(\bar{g}-1) = 0$.[8] In this case, the optimal commodity tax is always set equal to the Pigovian benchmark $X'(C_2)$. In contrast, the optimal internality tax can be higher or lower than the Pigovian benchmark, depending on whether low income earners' consumption of c_2 is relatively more responsive to the commodity tax t.

3.3 Numerical analysis

We now explore the quantitative implications of the above results in a the context of a common application of sin taxes: cigarette consumption. Our approach to this topic will be

---

[8]This fact is a variation of the well-known Atkinson-Stiglitz Theorem. Note, however, it is not clear what is the "right" analog of the Atkinson-Stiglitz assumptions in our setting. Because there is a wedge between $\hat{U}$, U and V, it is not clear which of these utility functions "should" be homogeneous. In general, because all existing work that estimates bias at the individual level finds substantial heterogeneity in biases, the assumption of homogeneous U and $\hat{U}$ seems implausible in our setting.

a simplified one—rather than modeling a full dynamic model of cigarette consumption with addiction and present bias, as in Gruber and Kőszegi (2001, 2004), we adopt the approach of Saez (2002a) and O'Donoghue and Rabin (2006), in which consumption of the sin good imposes a separable (perhaps delayed) cost which is not internalized by the consumer.

Although the time separable model of consumption utility in this paper is perhaps most appropriately applied to unhealthy, nonaddictive products like sugary drinks or "junk food," we focus here on cigarettes for three reasons. First, there is extensive literature on smoking patterns and elasticities (e.g., and, crucially, the heterogeneity of those patterns across the income distribution. This literature is largely made possible by the extensive variation in cigarette taxes over time and across states – see, for example, **?** and **?** for aggregate elasticity estimates, and Gruber and Kőszegi (2004) and Goldin and Homonoff (2013) for estimates by income. One aim of this paper is to highlight the value of collecting such data for other sin goods, but rather than conjecture about such parameters for other goods, we prefer to draw evidence from a domain that has already received substantial empirical focus. Second, and related to the widespread variation in taxes, cigarettes are already widely recognized as a "sin good" with taxes that are driven by internalities. Indeed, the externalities from cigarette consumption are typically found to be much smaller than internalities, and much smaller than prevailing excise taxes. (See **?** for an extensive review of this literature.)

Finally, there is evidence that smoking preference is correlated with income earning ability, independent of income itself. As noted in the introduction, and discussed extensively in Saez (2002a), commodity taxes are useful instruments for redistribution if the variation in consumption across the income distribution is due to differences in preferences across abilities, rather than to income effects – otherwise the classic result from Atkinson and Stiglitz (1976) implies that all redistribution can be achieved through the nonlinear income tax. Most generally, if the negative relationship between smoking and income were driven by income effects, and not by preferences, cigarettes would need to be an inferior good – yet as shown below, rich smokers do not consume fewer cigarettes than poor smokers. More concretely, see **?** for evidence on the correlation between IQ scores and cigarette takeup

133

among adolescents, which remains negative even after controlling for parents' socioeconomic status.

We use the data set from Goldin and Homonoff (2013), drawn from the Behavioral Risk Factor Surveillance System telephone survey in years 1984–2000, to calibrate smoking rates and elasticities across the income distribution. Figure 3.1 displays the share of smokers and the number of cigarettes smoked per day among smokers, both as a function of income. Data for the year 2000, which we use to calibrate our model, is plotted in bold. These patterns suggest that the primary dimension of heterogeneity in smoking across the income distribution is on the extensive margin (whether an individual smokes at all) rather than the number of cigarettes consumed. This is also the primary source of variation in cigarette consumption over time—whereas the share of smokers declined by 37% from 1984 to 2000, the number of cigarettes consumed among smokers fell by less than 10%. The variation in smoking rates across incomes is substantial, and is consistent with more recent survey data reported by Gallup, displayed in Figure 3.2.
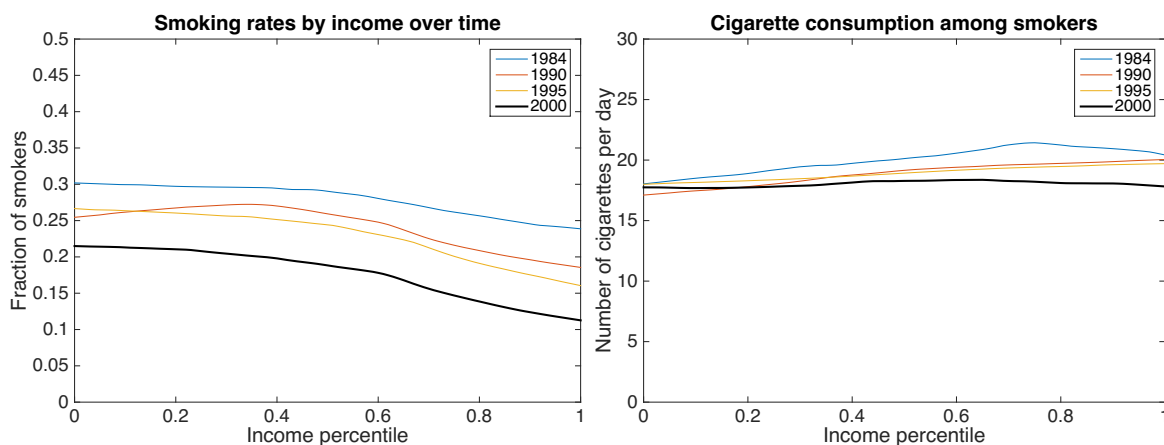
Figure 3.1: *Smoking patterns as a function of income across years. Data are from Goldin and Homonoff (2013), drawn from the Behavioral Risk Factor Surveillance System. These schedules are constructed using kernel regressions. The left panel regresses an indicator for whether an individual is a smoker (smokes at least one cigarette per day) on income percentile in each year. The right panel regresses the number of cigarettes smoked per day on income among smokers.*

Figure 3.2: *Smoking prevalence across income, according to Gallup.*

To fit the patterns in Figure 3.1, we adopt a slightly modified version of the general setup presented in Section 3.2. Specifically, we assume two types of consumers—those who have a taste for cigarettes ("smokers", denoted by an indicator variable $s = 1$), and those who do not ($s = 0$), with quasilinear demand for cigarettes among smokers. We let c_1 denote all consumption other than cigarettes and c_2 denote cigarettes (or the sin good of interest), both measured in dollars. As in the standard Mirrlees (1971b) model of redistributive income taxation, we further assume individuals have heterogeneous wages θ, and generate earnings z by supplying labor effort $\ell = z/\theta$. Individual decision utility thus takes the following form:

$$U(c_1, c_2, \ell, s) = \begin{cases} c_1 - v(\ell) & \text{if } s = 0 \\ c_1 + u(c_2) - v(\ell) & \text{if } s = 1. \end{cases} \tag{3.11}$$

Experienced utility differs from decision utility only for smokers, who experience an additional (uninternalized) cost γc_2. This cost is assumed to be linear so that the optimal sin tax absent redistributive considerations is simply γ. Thus experienced utility takes the form

$$V(c_1, c_2, \ell, s) = \begin{cases} c_1 - v(\ell) & \text{if } s = 0 \\ c_1 + u(c_2) - \gamma c_2 - v(\ell) & \text{if } s = 1. \end{cases} \tag{3.12}$$

135

In what follows, we use $v(\ell) = \frac{\ell^{1+1/e}}{1+1/e}$, where e is the constant compensated elasticity of labor supply, as in the Type I utility function from Saez (2001a), which we set equal to 0.33, the preferred value in Chetty (2012a). This specification avoids complications from income effects on the choice of labor supply, which are not of central concern for the results which follow. The distribution of wages is assumed to be continuous, denoted $F(\theta)$ with density $f(\theta)$, and is calibrated to match the empirical income distribution.[9]

We also use $u(c_2) = k \left(\frac{c_2^{1-1/\zeta}-1}{1-1/\zeta} \right)$, where ζ denotes the absolute value of the constant price elasticity of demand for cigarettes, with $u(c_2) = k \log(c_2)$ when $\zeta = 1$. The constant k controls the level of spending on cigarettes, which we set to generate \$1200 in spending on cigarettes per year among smokers, equal to approximately one pack per day, in the presence of a 34% tax (the mean cigarette excise tax in the US in 2000), consistent with Figure 3.1.

The planner's objective is to select a nonlinear income tax $T(z)$ and linear tax on cigarettes t, in order to maximize a weighted sum of experienced utility. Consistent with the literature on preference heterogeneity and redistribution, we assume that differences in preferences alone do not merit redistribution (see Fleurbaey and Maniquet (2006) and Lockwood and Weinzierl (2015)). Thus we write the marginal social welfare weights $g(\theta)$ as a decreasing function of ability θ, but not of smoking taste s, with $\int g(\theta) dF(\theta) = 1$. For numerical simplicity, we here assume marginal social welfare weights are a function of type directly, rather than being determined endogenously by the concave transformation G as in Section 3.2. As before, the optimum maximizes average weighted experienced utility, $g(\theta)V(c_1, c_2, z/\theta, s)$, where the choice variables c_1, c_2, and z are chosen optimally by each individual taking T and t as given, subject to a resource constraint that total income and commodity taxes sum to zero. (Imposing an exogenous government expenditure

---

[9]Specifically, we use the density of incomes as measured by the Current Population Survey for years 2003–2006, among single individuals and heads of households over 25 years old who earn at least 50% of household income. Due to the sparse coverage of CPS at the top of the income distribution, we follow Mankiw *et al.* (2009) and smooth paste a Pareto tail with parameter $\alpha = 2$ (consistent with Saez (2001a)) above the 97th percentile. Following Saez (2001a) and Saez (2002b), we back out the ability distribution so that the empirical income distribution would arise under a flat tax rate (equal to 30%). This avoids numerical complications from bunching at kinks in a piecewise linear tax function, which are not of central interest for our purposes.

requirement affects only the lump sum grant in this specification and can thus be ignored for our purposes.)

3.3.1 Baseline results

The optimal income tax function satisfies the standard first-order condition from Diamond (1998a) and Saez (2001a) (in the specification without income effects)—the schedule of marginal tax rates is plotted in Figure 3.3. To make our results comparable to other familiar results in the optimal tax literature, we set marginal social welfare weights $g(\theta)$ to equal $g(\theta) = c_1^{-\nu}$ among $s = 0$ consumers at the optimum, so that the parameter ν governs the redistributive taste of the planner, with $\nu = 1$ in our baseline specification. Under the quasilinear specification in (3.11) and (3.12), the the schedule of optimal marginal income tax rates does not depend on the share of smokers, nor on how smoking varies with income.
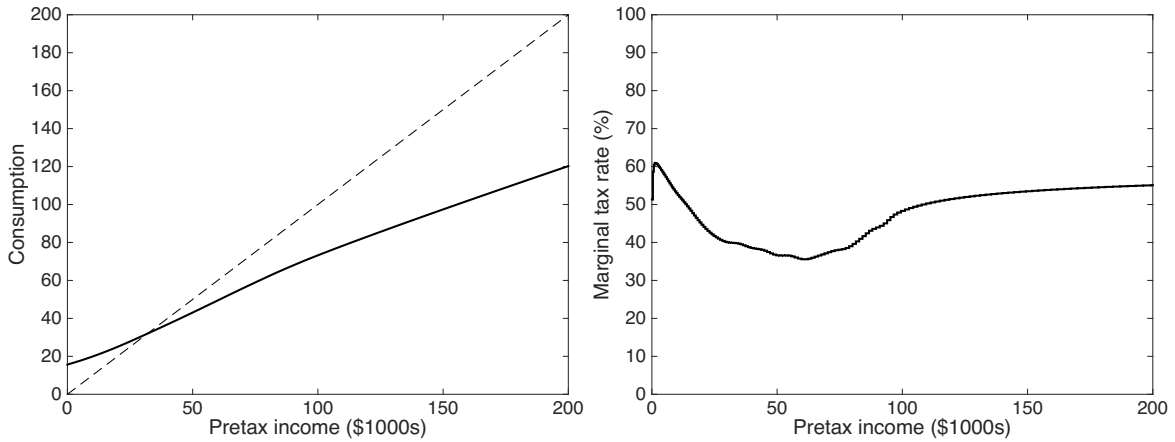
Figure 3.3: *The left panel displays the schedule of consumption (post-tax income) as a function of pre-tax earnings at the optimum (the dashed $45°$ line represents the case of no taxes for comparision). The right panel plots the schedule of optimal marginal income tax rates.*

In this setup, $\sigma(t) = \rho(t) = C_2(t)\left(\mathbb{E}\left[g|s = 1\right] - 1\right)$, and thus expression 3.3 takes a particularly simple form:

$$t = \gamma\mathbb{E}\left[g|s = 1\right] - \frac{1+t}{\zeta}\left(E\left[g|s = 1\right] - 1\right) \tag{3.13}$$

The optimal tax on c_2 is plotted as a function of the degree of bias γ for a range of demand

elasticities in Figure 3.4. The maximum value $\gamma = 1$ corresponds to the greatest degree of bias considered by O'Donoghue and Rabin (2006), in which the present bias parameter $\beta = 0.9$ and the future health cost of consuming a sin good is equal to ten times its pre-tax price. By construction, the optimal tax absent redistributive concerns is simply equal to γ, represented by the dashed $45°$ line through the origin. The line for $\zeta = 0.35$, corresponding to the preferred intensive margin elasticity estimate from Goldin and Homonoff (2013) (see Table 5 of that paper), is plotted in bold. Consistent with Proposition 3.2, part 1, the tax is negative (a subsidy) for sufficiently low levels of bias, and rises with bias. Conditional on a given degree of bias, the tax is lower when demand for c_2 is less elastic. When $\zeta = 0.05$, a subsidy is optimal for the entire range of bias considered. In contrast, for sufficiently high elasticity, and at high degrees of bias, the tax is higher than the non-redistributive benchmark. This illustrates Proposition 3.2, part 4: regressivity of sin taxes may raise the size of the optimal tax if elasticities are sufficiently high, although these simulations suggest such a result occurs only at elasticities significantly higher than those typically found in the literature on cigarette consumption.

Together, these results suggest that for modest to large degrees of bias, accounting for redistributive motives tends to reduce the optimal tax substantially, relative to the non-redistributive benchmark. For example, if smokers ignore about \$1 in future health costs for every \$1 worth of cigarettes consumed ($\gamma = 1$) then the optimal tax is 75%, rather than 100%, of the cost of a pack of cigarettes. Finally, the wide dispersion of the lines in Figure 3.4 emphasize the importance of accurately measuring the demand elasticity of sin goods – for example, a difference in elasticity of 0.35 versus 0.1 determines whether cigarettes should be subject to a substantial tax or a substantial *subsidy* for moderate degrees of bias.
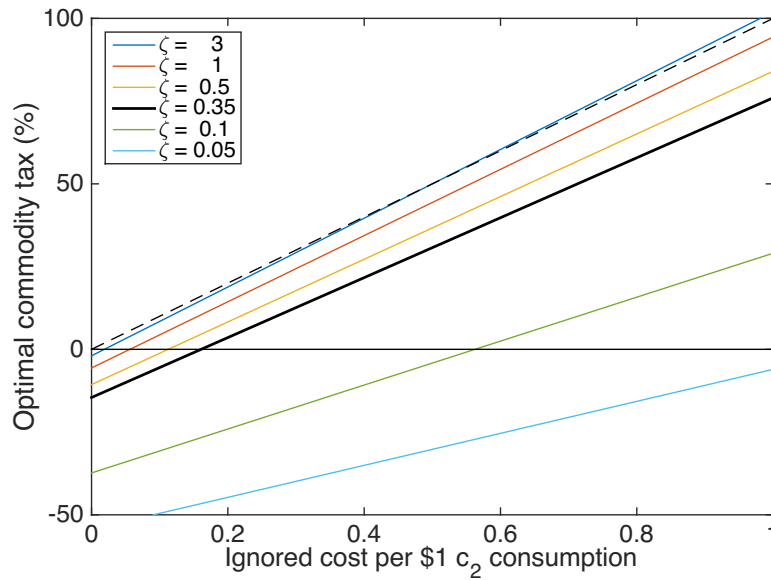
Figure 3.4: *Optimal linear tax on c_2 as a function of the ignored health cost γ, for a range of price elasticities of demand ζ.*

Figure 3.5 displays the optimal commodity tax on c_2 as a function of γ for a range of degrees of inequality aversion. The baseline case of $\nu = 1$ is plotted in bold—all are computed using the baseline elasticity of $\zeta = 0.35$. The schedule of income tax rates for each value of ν is plotted in the right panel. This figure highlights the key role of redistributive concerns in determining the optimal tax for goods which are preferred by the poor. Even with the relatively mild decline in smoking across the income distribution (Figure 3.1) a high aversion to inequality of $\nu = 4$ can reduce the optimal tax by between \$0.25 and \$0.50 for every \$1 of cigarette expenditures. On the other hand, if redistributive motives are mild ($\nu = 0.25$) the optimal tax is only slightly lower than the non-redistributive benchmark. Indeed, the implications of inequality aversion for the optimal commodity tax when $\gamma = 1$ are quantitatively similar to those for optimal marginal income tax rates – raising ν from 1 to 2 reduces the optimal cigarette tax by about 10 percentage points, and raises optimal marginal income tax rates by a comparable amount across the income distribution.

Finally, Figure 3.6 plots the gains in social welfare generated by implementing the optimal tax, and the naive Pigovian tax, relative to not having any tax on c_2. The optimal

commodity tax generates gains over the Pigovian tax equal to about $1.10 per person. Although this may not seem large, it can be quite substantial as a share of the potential gains from taxing cigarettes at all. As noted above, the average excise tax on cigarettes in 2000 was $0.34. If this were calibrated based on estimates that average marginal bias was $\bar{\gamma} = 0.34$, with no account of redistributive concerns, then Figure 3.6 indicates the (substantially lower) optimal cigarette tax would generate $1.27 per person in social welfare gains, rather than only $0.18 – that is, the welfare gains from the prevailing tax could be raised seven-fold by accounting for inequality aversion. Importantly, at low levels of bias (perhaps less relevant for cigarettes than for junk foods or other mildly unhealthy goods consumed by the poor) the Pigovian benchmark tax is actually harmful.
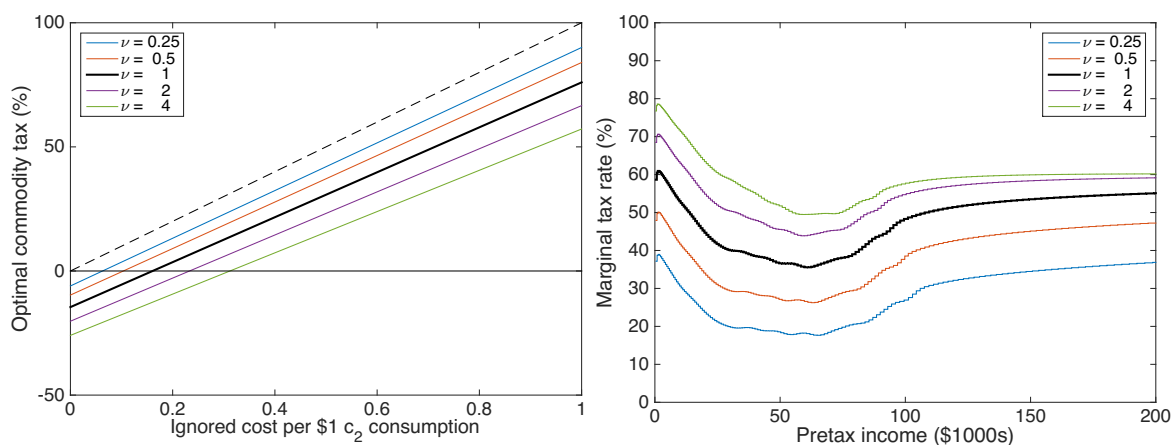
Figure 3.5: *Optimal linear tax on c_2 as a function of the ignored health cost γ, for a range of redistributive tastes ν, with $\zeta = 0.35$. Optimal marginal income tax rates in each case are shown in the right panel.*
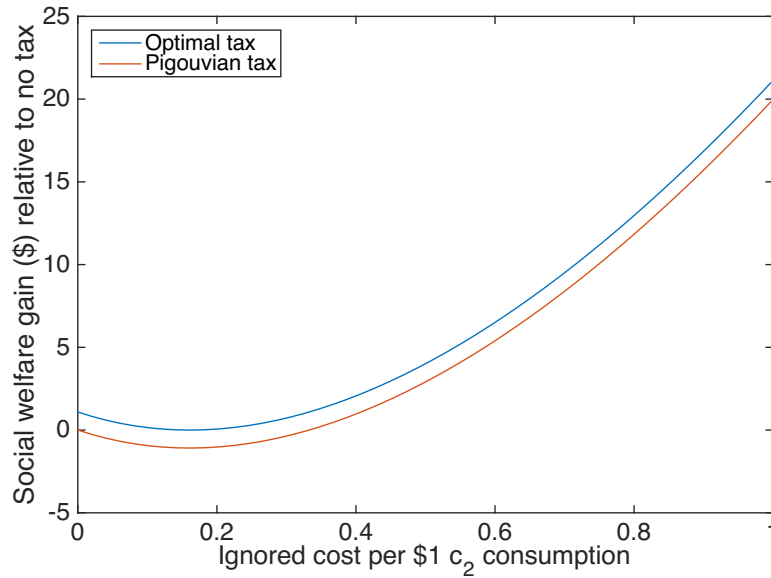
Figure 3.6: *Welfare gains (measured in dollars of public funds) from implementing the optimal commodity tax on c_2, and from implementing the suboptimal "Pigovian" corrective tax equal to γ, as a function of the bias γ.*

3.3.2 Heterogeneous demand elasticity

In the above analysis we assumed a constant price elasticity of demand for cigarettes of 0.35, consistent with the intensive margin elasticity estimated in Goldin and Homonoff (2013), and does not vary measurably with income. However other sources have estimated demand elasticities which decline in magnitude with income—for example, Gruber and Kőszegi (2004) use the Consumer Expenditure Survey to compute elasticities which fall (in absolute value) from 1.1 in the poorest quartile of consumers to 0.38 in the top quartile.[10] Although our goal is not to adjudicate between alternative elasticity computations, in light of the derivations in Section 3.2, we *are* interested in the implications of elasticities which covary with marginal social welfare weights.

Therefore we re-run the specification from the preceding section with the modification that ζ is set equal to 1.1 for the bottom quartile of earners, 0.70 for the second quartile,

---

[10]Gruber and Kőszegi (2004) report levels of cigarette expenditures (3.2% in the bottom quartile and 0.4% in the top) which are somewhat higher than those in the Goldin and Katz (2007) data, 2.5% and 0.2% respectively.

0.53 for the third quartile, and 0.39 for the fourth.[11] This is admittedly a somewhat ad-hoc method of incorporating elasticities which vary with incomes, as it assumes that variation is driven by variation in the utility function with θ. Yet it allows us to avoid the numerical complexities of income effects on c_2, which complicate the optimal income tax and which, in any case, appear inconsistent with the fairly constant level of cigarette consumption among smokers of all incomes.

The left panel of Figure 3.7 displays the optimal commodity tax on c_2 in this calibration, for a range of redistributive tastes (the optimal income tax schedules are as in Figure 3.5). The right panel displays the optimal tax t when the elasticity is constant at the average level from the left panel, $\zeta = 0.66$. This calibration generates optimal commodity tax rates that rise more quickly with the degree of bias γ, and we see levels of t that exceed the Pigovian benchmark (the dashed line) for $\gamma > 0.6$ even under our baseline redistributive tastes of $\nu = 1$. This stark result is driven in part by positive covariance between welfare weights and the demand elasticity, as emphasized by the importance of ω in 3.3. It is also driven by the fact that in this specification, a larger tax reduces the relationship between cigarette consumption and income – since low income consumers are more elasticity, a higher tax reduces their consumption by more, thereby reducing the regressivity cost of cigarette taxation through a reduction in ρ. However ρ remains positive even at the highest level of tax shown in Figure 3.7, indicating that the higher-than-Pigovian commodity taxes are driven by progressivity alone.[12]

---

[11]As before, we calibrate k so that annual cigarette expenditures is \$1200 across the income distribution when $t = 0.34$. This implies that when $t \neq 0.34$, cigarette consumption varies with income quartile.

[12]To be precise, ρ/C_2, which is not sensitive to the units on C_2, declines from 0.093 when $\gamma = 0$ to 0.023 when $\gamma = 1$ in the simulation with $\nu = 1$.
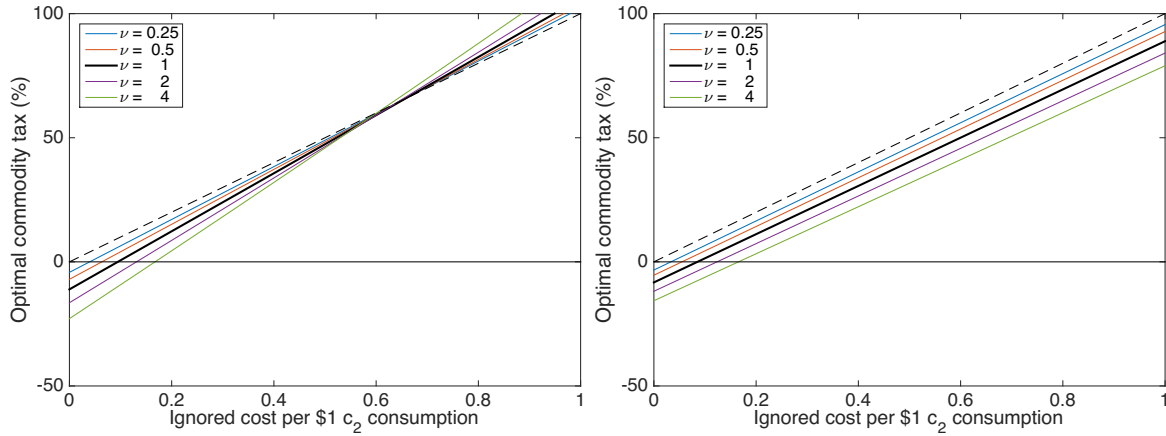
Figure 3.7: *The left panel displays the optimal linear tax on c_2 as a function of the ignored health cost γ, for a range of redistributive tastes v, with ζ declining in income, as in Gruber and Kőszegi (2004). The right panel shows the the optimal taxes arising when demand elasticities are constant, at the mean elasticity (-0.66) reported in Gruber and Kőszegi (2004).*

3.3.3 Heterogeneous bias

We also explore the quantitative implications of heterogenous bias γ across the ability distribution. Since we do not possess data on the degree of bias by income level, this analysis is necessarily hypothetical – however this can provide useful guidance on the importance of determining the extent to which such heterogeneity exists. These simulations, shown in the left panel of Figure 3.8, are constructed by assuming that bias decreases linearly across the income distribution, with $\gamma = 0$ for top earners. We continue to plot the mean level of bias on the horizontal axis, so for example the point 0.6 on the horizontal axis $\gamma = 1.2$ for the lowest income smokers and $\gamma = 0$ for the highest income smokers, with γ interpolated linearly across intervening percentiles. We show these results for a range of redistributive tastes v. The right panel of Figure 3.8 reproduces Figure 3.5 for comparison. The figure shows that heterogeneity in bias can be quite important for the optimal commodity tax, particularly at high levels bias, where the optimal tax is higher due to this heterogeneity. Intuitively, for a given average marginal bias $\bar{\gamma}$, the bias is more important to correct if it is concentrated among low income consumers. In the notation from Section 3.2, this heterogeneity raises σ, amplifying the corrective component of the

optimal tax. Importantly, both panels of Figure 3.5 are consistent with the same observed cigarette consumption behavior, highlighting the importance of identifying how behavioral biases vary with income.
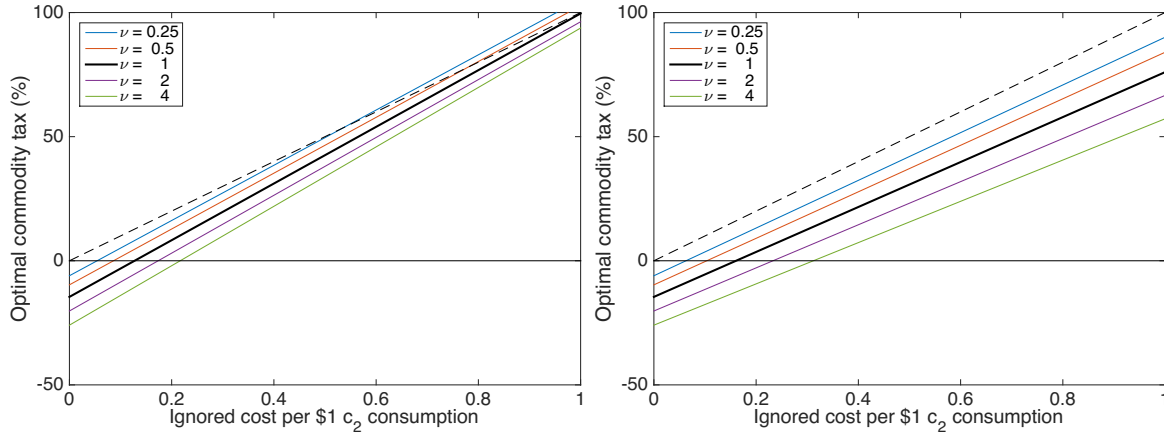
Figure 3.8: *The left panel displays the optimal linear tax on c_2 as a function of the ignored health cost γ when γ is heterogeneous, decreasing linearly across income percentiles (from twice the mean at the lowest income, to $\gamma = 0$ at the highest income) for a range of redistributive tastes ν. The right panel shows the the optimal taxes arising when γ is constant across incomes, reproducing Figure 3.5.*

3.4 Extensions

3.4.1 Tax salience

So far, we have assumed that the commodity tax is fully salient. However, recent empirical work suggests that taxes that are not included in posted prices may be ignored by consumers (Chetty *et al.*, 2009; Finkelstein, 2009; Goldin and Homonoff, 2013; **?**). An intuition informally suggested in the literature on tax salience is that taxes that are used for corrective purposes, such as reducing cigarette consumption, should be made maximally salient (**?**).

In this section, we thus consider a policymaker who can choose between either a fully salient commodity tax t that is included in the posted prices and a less salient commodity tax t_ϕ that is not included in posted prices and is paid by consumers at the register. We assume that the policymaker can choose either a salient or a non-salient commodity tax, but

cannot use a combination.[13] For the case of cigarette taxes, for example, the government can either include the tax in the price, or simply charge it at the register. Although state governments in the Unites States also use sales taxes, those are constrained to be the same for a wide variety of commodities and thus cannot target any one specific good in question. Although our analysis could be generalized to consider an optimal mix of both t and t_ϕ, we suspect that our starting point is a reasonable one because of political economy constraints – a politician may have trouble explaining to the public why he chose to break up an otherwise simple tax into shrouded and unshrouded subcomponents.

We begin with the case in which attention to the less salient tax is homogeneous. When the less salient tax is used, consumers choose c_2 to maximize $U = z - (p + \phi t_\phi)c_2 + u(c_2) - \psi(z/\theta)$, where $\phi \in [0,1]$ is consumers' attention to the less salient commodity tax t_ϕ. Because of quasilinearity, salience does not affect consumers' optimal choice of income z. Consumers' choice of z satisfies the first-order condition $\frac{1}{\theta}\psi'(z/\theta) = 1 - T'(z)$.

For simplicity, we will focus in this section on regressive sin taxes. We will assume that $\gamma(\theta, c_2) > 0$ for all θ, c_2 and that consumption of c_2 is decreasing in θ. Analogous results would hold for the case of regressive subsides for goods that people underconsume. Proposition 3.6 below characterizes conditions under which the policymaker will choose the more or less salient commodity tax.

Proposition 3.6. *Suppose attentiveness ϕ is homogeneous. Then:*

1. *A positive, less-salient commodity tax is never optimal. That is, the optimal policy cannot have* $t_\phi > 0$.

2. *A negative, fully salient commodity tax is never optimal. That is, the optimal policy cannot have* $t < 0$.

Proposition 3.6, part 1, says that if the policymaker sets a positive tax, then it always has to be the more salient tax. That is, the policymaker should not try to reduce consumption of

---

[13]See Goldin and Homonoff (2013) for a model (of otherwise optimizing consumers and no redistributive concerns) in which the policymaker can combine tax instruments of differing salience to raise revenue in the least distortionary way possible.

c_2 with the less salient policy instrument. Intuitively, this is because the less salient tax t_ϕ would have to be larger than t to achieve the same change in behavior. However, because taxes are regressive, it is not desirable to have a higher tax.

Conversely, part 2 of the proposition says that if the policymaker chooses to subsidize c_2, then he should do so with the less salient tax instrument. Intuitively, this is because a less salient subsidy can achieve the same redistributive properties as the more less salient subsidy, while having a smaller impact on behavior. A simple corollary of part 2 of the proposition above, combined with part 1 of Proposition 3.2, is that if consumer bias is sufficiently small, than the optimal policy must involve a subsidy in the form of the less salient tax instrument:

Corollary 3.1. *Under the assumptions of Proposition 3.2, part 1, there is $\gamma^\dagger > 0$ such that if $\gamma(\theta, c_2) < \gamma^\dagger$ then the planner uses the less salient commodity tax, setting $t_\phi < 0$.*

Heterogeneous attention

We now consider the case in which attentiveness to the tax is heterogeneous. We model salience as being a function of both ability θ, as well as income z. Both could plausibly affect attentiveness to taxes. On the one hand, higher ability types may be more cognitively skilled to correctly react to taxes. On the other hand, the more wealth a person has, the smaller his marginal utility from money, and thus the less important it is for this person to exert cognitive effort to pay attention to a not fully salient commodity tax.

When ϕ depends on z, consumers' attentiveness is then partly endogenous to the income tax, which the policymaker must take into account. But now while $\phi(\theta, z)$ is a function of both θ and z, note that because there is a bijection between θ and z, we can simply describe tax attentiveness via an indirect function $\hat{\phi}(z)$ that maps income z to tax attentiveness.[14]

While it is hard to separately identify how ϕ depends on both θ and z, the optimal commodity tax t_ϕ can still be expressed compactly in terms of $\hat{\phi}(z)$. To see this, note that

---

[14]Here, we continue assuming that salience does not affect period 0 income choices. These are still governed by the FOC $\frac{1}{\theta}\psi'(z/\theta) = 1 - T'(z)$.

salience bias operates much like overvaluation bias: at a commodity tax t_ϕ, a consumer earning z overestimates the marginal utility from consuming c_2 by $(1 - \hat{\phi}(z))t_\phi$ due to salience bias. Thus the total amount by which the consumer overestimates the marginal utility from c_2 is $\hat{\gamma}(z) = (1 - \hat{\phi}(z))t_\phi + \gamma(z)$. The optimal tax t_ϕ can then be expressed in terms of $\hat{\gamma}(z)$ by replacing $\gamma(z)$ in equation (3.2) with $\hat{\gamma}(z)$.

We now illustrate, by way of a simple two-type example, that with heterogeneous salience, it may be optimal to choose a positive, but non-salient sin tax. Suppose that there are just two ability types θ_L and θ_H, and let z_L and z_H denote income in the two-type model.

Proposition 3.7. *In the two-type model, let z_L^* and z_H^* denote the labor income choices that would result when the policymaker chooses an optimal policy (T, t) that restricts to only using the salient commodity tax. And suppose that $c_2 > 0$ for both types at this policy. Then using the salient commodity tax is suboptimal if $\hat{\phi}(z_L^*)$ is sufficiently close to 1, $\hat{\phi}(z_H^*)$ is sufficiently close to zero, and θ_H is sufficiently high.*

The intuition behind the Proposition follows in three steps. First, for θ_H sufficiently high, the social benefit of giving the high type more c_1 or c_2 approaches zero. This means that for θ_H sufficiently high, the benefits from correcting type H consumers' consumption of c_2 approach zero, and to a first-order, the policymaker wants to set taxes so as to maximize the revenue he can obtain from these type H consumers. But plainly, the less salient tax is a more effective way to raise revenue from type H consumers. At the same time, if $\hat{\phi}(z_L^*)$ is sufficiently high, then the less salient tax is only slightly less effective at changing type L's behavior than the more salient tax.

Numerical results for salience

For this analysis we return to the baseline (constant elasticity) specification from subsection 3.3.1 above. In Figure 3.9, we consider three different assumptions about salience, in addition to the baseline (plotted in bold). The left panels show the optimal commodity tax on c_2 for three cases with an average value of ϕ equal to 0.5. In the first, all consumers are equally inattentive. In the second, the bottom quartile of earners are fully attentive ($\phi = 1$), the top

quartile are completely inattentive ($\phi = 0$), and the second and third quartiles have $\phi = 0.75$ and $\phi = 0.25$, respectively, consistent with the finding in Goldin and Homonoff (2013) that low income consumers respond more to cigarette taxes which are not included in the posted price.[15] Finally, the last specification reverses this relationship, so that low income consumers are fully inattentive and high income consumers are fully attentive. Although extreme, this specification qualitatively corresponds to the finding in **?** that higher income consumers pay more attention to sales taxes on common household commodity items, at least in part because they are more financially sophisticated and are better at calculating taxes.

We display these results for two degrees of redistributive preference—the baseline $\nu = 1$ and a higher value $\nu = 4$. The right panels of Figure 3.9 shows the difference in welfare resulting from the use of the non-salient tax instrument relative to the baseline (fully salient) instrument. This difference is measured in dollars of public funds per person.

As is evident from the left panels, lower salience raises the size of the optimal subsidy when bias is low (indicated by the more negative intercept). Moreover, when the optimal policy is a subsidy, lower saliences makes nonsalient taxes superior instruments for raising welfare, as indicated by the positive welfare gains for low levels of bias in the right panels. These two results demonstrate Corollary 3.1.

Heterogeneous attention has large implications for the shape of the optimal tax, however. When low income earners are more attentive, the tax is higher than under uniform attention for all degrees of bias. For high bias, this reflects the higher public priority of correcting mistakes of poor consumers than of rich consumers. When high earners are more attentive, the effect is reversed: for high bias, the optimal tax is lower than under uniform attention, reflecting the lower priority of correcting rich consumers' mistakes, combined with the

---

[15]Specifically, Goldin and Homonoff (2013) estimates intensive margin elasticities among the top 75% of the income distribution equal to -0.31 for the excise tax (assumed fully salient) and 0.18 for the sales tax (possibly not salient, as sales taxes are typically not included in posted prices). Among the bottom quartile of consumers, they find elasticities of -0.3 (excise) and -0.59 (sales)—see Table 6 and footnotes 33 and 34 in that paper. We assume that the elasticity on the less salient instrument must lie between zero and the elasticity with respect to the fully salient (excise) tax, and thus we approximate their findings by assuming salience of 1 for the bottom quartile of earners and 0 for the rest.

desire not to levy heavy taxes on inattentive poor consumers. Finally, the bottom row (with $\nu = 4$) demonstrates Proposition 3.7 for high levels of bias: a nonsalient tax may be preferable to a salient one, even when the tax is positive and substantial, if attention is inversely correlated with income. Intuitively, when the tax is positive and large in the presence of substantial bias, a less salient tax instrument may be optimal as it allows the high tax to raise substantial revenues from rich consumers, which then raise the lump sum grant for redistributive benefit. Although the nonsalient instrument leads to greater consumption of c_2 among the rich, the resulting internalities have only small social welfare costs due to the low welfare weight on high income consumers.
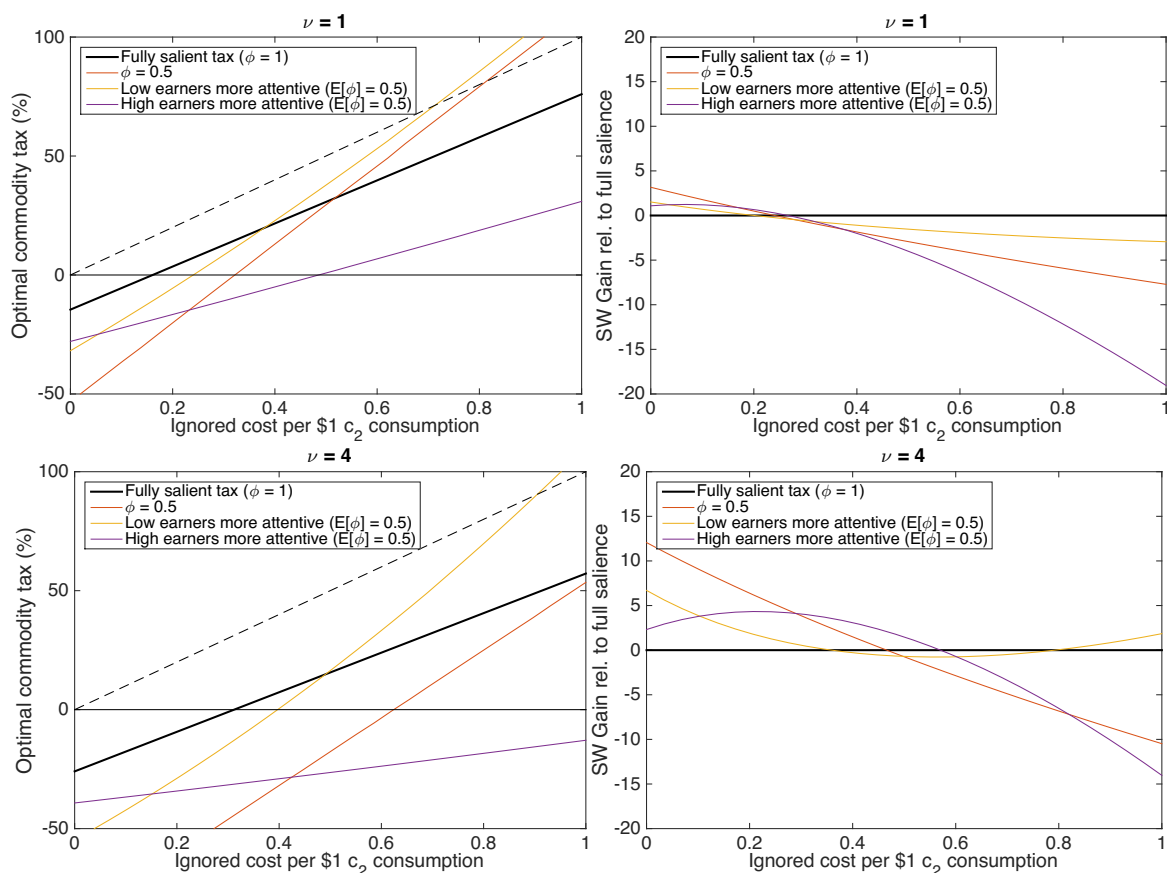
Figure 3.9: *The left panels display the optimal linear tax on c_2 under different assumptions about tax salience. The parameter ϕ represents the share of the tax that is perceived by consumers. The bold line represents the optimal tax under our baseline scenario, with $\phi = 1$ for all consumers. The next three lines plot the optimal tax when the tax is partially ignored. In each case, the average value of ϕ is 0.5. The first has $\phi = 0.5$ for all consumers. The second assumes low income consumers are more attentive, consistent with the results presented in Goldin and Homonoff (2013). The third assumes that attention rises with income. The right panel shows the social welfare gains from implementing the optimal tax in each case, relative to the optimal fully salient tax. (Thus the bold line is mechanically zero in the right panels.) The top row shows these results under our baseline assumption of inequality aversion—with $\nu = 1$, while the bottom shows results with higher inequality aversion of $\nu = 4$. All are computed assuming the baseline elasticity $\zeta = 0.35$.*

3.4.2 Non-financial instruments

We now consider the effects of "persuasion" – such as graphic imagery about the health costs of smoking – that might change consumers' consumption of c_2. As in Section 3.4.1, we continue restricting to the case in which low income earners have a higher preference for c_2,

and $\gamma > 0$.[16]

Suppose that decision utility in periods 0 and 1 is now given by $\hat{U} = c_1 + \hat{u}(c_2, \theta) - \psi(z/\theta) - sc_2$ and $U = c_1 + u(c_2, \theta) - \psi(z/\theta) - sc_2$, where s is the size of the persuasion campaign. The policy maker maximizes the weighted average of utility functions $V = c + m(c_2, \theta) - \psi(z/\theta) - \gamma(c_2, \theta)c_2 - nsc_2$, where $n \leq 1$ parametrizes the nudge aspect of the psychological tax. When $n = 0$, the policy is a pure nudge in the sense that it changes demand without affecting people's utility. When $n = 1$ the policy is a pure psychological tax.

The impact of increasing the tax t by some small amount dt is as before. Consider now the impact of increasing the psychological tax s by the same amount ds. Noting that $\frac{dc_2}{ds} = \frac{dc_2}{dt}$ by construction, the perturbation dt changes utility by $\int [nc_2(z) + s(1-n)]g(z)\frac{dc_2}{dt}dH(z)$. Second, this perturbation decreases revenue by $t\frac{dc_2}{dt}$. Assuming that the psychological tax costs the government κs to implement, the net welfare effect is

$$
\begin{aligned}
\frac{dW}{ds} &= \int \left[-nc_2(z)g(z) - \gamma(z)g(z)\frac{dc_2}{dt} + t\frac{dc_2}{dt} + s(1-n)g(z)\frac{dc_2}{dt} \right] dH(z) - \kappa \\
&= \int \left[-nc_2(z)g(z) - \gamma(z)g(z)\zeta(z)\frac{c_2(z)}{p+t} + (t + s(1-n))\zeta(z)\frac{c_2(z)}{p+t} \right] dH(z) - \kappa
\end{aligned}
$$

Thus

$$
\begin{aligned}
\frac{dW}{ds} - \frac{dW}{dt} &= \int \left[(1-n)g(z)c_2(z) - c_2(z) \right] dH(z) - \kappa \\
&= Cov_H \left[g(z), c_2(z) \right] + C_2 - n \int g(z)c_2(z) - C_2 - \kappa \qquad (3.14) \\
&= Cov_H \left[g(z), c_2(z) \right] - n \int g(z)c_2(z) - \kappa \qquad (3.15)
\end{aligned}
$$

Equation (3.15) shows that whether persuasion is more beneficial than taxation depends on three terms. The first term, $Cov_H \left[g(z), c_2(z) \right]$, is the regressivity of the tax. When the tax is not regressive, so that tax revenues are recycled to consumers in a way that does not

---

[16]See also **?** for a complementary analysis of the optimal combination of a tax and a nudge for a population of active and passive savers.

impede redistributive goals, persuasion cannot improve upon the tax. The more regressive the tax, however, the higher the relative benefits to persuasion, because persuasion does not a impose a relatively higher burden on the low income earners than on the high income earners. The second and third terms are simply the social cost of persuasion: the psychological cost that it imposes on consumers and the implementation cost that it imposes on the policymaker. The higher the social cost of persuasion, the lower its impact relative to the commodity tax. We formalize these insights in the proposition below:

Proposition 3.8. *Suppose that* $\gamma(\theta, c_2) > 0$ *for all* θ, c_2.

1. *Suppose that G is linear or that preferences for c_2 do not vary by type. Then the optimal tax system sets $t^G = \bar{\gamma}$ and $s^G = 0$. Psychological taxes are strictly suboptimal when there are no redistributive motives or when there is no preference heterogeneity.*

2. *Suppose that G is strictly concave and that c_2 is decreasing in z. Then the optimal policy sets $s^G > 0$ for low enough n and κ.*

3. *Under the additional conditions that $\gamma(\theta, c_2)$ is bounded and non-increasing in θ, and that $|G''|$ is bounded away from zero, the optimal policy sets $s^G > 0$ and $t^G < 0$ for low enough η and κ.*

Part 1 of the proposition states that when taxes are not regressive, the optimal policy mix should not rely on non-financial instruments. The intuition is that when the tax is not regressive, its revenues are recycled perfectly to consumers in a way that does not increase wealth-inequality, and thus it is a costless way to change behavior. In contrast, persuasion imposes both a psychological cost on consumers and an implementation cost on the policymaker, and thus it is always a costly way to change behavior.

Part 2, however, shows that because a regressive tax is no longer "costless," it may be optimal to use some persuasion when its total social cost is not too high. Part 3, in fact, shows that under some additional regularity assumptions, if persuasion is sufficiently cheap then it is best to correct behavior with persuasion, and then set a negative tax to redistribute wealth from high income earners to low income earners.

152

3.5 Conclusion

In this paper, we have analyzed optimal taxation in the presence of both consumer misoptimization and redistributive concerns. Hotly debated policies, such as cigarette taxes or energy efficiency subsidies, must address both consumer misoptimization and the redistributive goals of the government. The policy debates are often polarizing: some claim that regressive taxes are unfair and thus should be eliminated, while others argue for high taxes, focusing mostly on the corrective and revenue-raising aspects of the taxes. This paper provides a framework for tractably considering both motives. Theoretically and numerically, we show that both motives matter, and we characterize conditions under which each motive is likely to matter the most. Our simulation analysis shows that while the optimal tax is positive for many of the commonly made assumptions about the magnitude of bias, redistributive concerns significantly dampen the size of the optimal corrective tax.

Our framework also clarifies that corrective and redistributive motives do not simply oppose each other. Because an inequality averse government should care more about the poor leaving money on the table than the rich, redistributive motives can amplify corrective motives. We show that the extent to which these two motives reinforce each other can be expressed as an estimable covariance between a person's income and 1) his elasticity to the sin tax and 2) how biased he is relative to the average consumer. We show that such correlated heterogeneity can significantly change the magnitude of the optimal corrective tax and, for high levels of bias, can generate an optimal tax that is *higher* than what it would be in the absence of redistributive motives.

In addition to providing a tractable framework for studying corrective and redistributive motives jointly, our work thus underscores the importance of further empirical work on individual differences. Our generally-applicable, quantifiable formulas show that it is not only how biased people are "on average" that matters; it matters *who* is biased. Whether the mistake is being made by low income or high income consumers, whether the mistake is being made by those more or less elastic to the tax instrument, and whether the low income consumers are relatively more or less elastic to the tax instrument are critical questions for

generating robust policy recommendations.

References

ALLCOTT, H. and TAUBINSKY, D. (2013). The lightbulb paradox: evidence from two randomized experiments.

— and — (forthcoming). Evaluating behaviorally-motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*.

ANGRIST, J. D. and KRUEGER, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, **106** (4), 979–1014.

ARROW, K. J. and HAHN, F. H. (1971). *General Competitive Analysis*. Amsterdam: North-Holland.

ATKINSON, A. B. and STIGLITZ, J. (1976). Design of Tax Structure - Direct Versus Indirect Taxation. *Journal of Public Economics*, **6**, 55–75.

ATTARI, S., DEKAY, M., DAVIDSON, C. and DE BRUIN, W. B. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences*, **37**, 16054–16059.

AUGENBLICK, N., NIEDERLE, M. and SPRENGER, C. (2013). Working over time: Dynamic inconsistency in real effort tasks.

BAGWELL, K. (2007). The economic analysis of advertising. In M. Armstrong and R. H. Porter (eds.), *Handbook of Industrial Organization*, vol. 3, Amsterdam: North-Holland.

BAI, J., PHILIPPON, T. and SAVOV, A. (Forthcoming). Have financial markets become more informative? *Journal of Financial Economics*.

BAKIJA, J., COLE, A. and HEIM, B. T. (2012). Jobs and income growth of top earners and the causes of changing income inequality: Evidence from u.s. tax return data, manuscript, Department of Economics, Williams College.

BARRO, R. J. (1991). Economic growth in a cross-section of countries. *Quarterly Journal of Economics*, **106** (2), 407–443.

BAUMOL, W. J. (1990). Entrepreneurship: Productive, unproductive and destructive. *Journal of Political Economy*, **98** (5), 893–921.

BEAUDRY, P., BLACKORBY, C. and SZALAY, D. (2009). Taxes and employment subsidies in optimal redistribution programs. *The American Economic Review*, **99** (1), 216.

BECKER, G. S. and MURPHY, K. M. (1993). A simple theory of advertising as a good or a bad. *Quarterly Journal of Economics*, **108** (4), 941–964.

BERNHEIM, B. D. and RANGEL, A. (2007). Toward choice-theoretic foundations for behavioral welfare economics. *The American economic review*, pp. 464–470.

— and — (2009). Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, **124** (1), 51–104.

BERTRAND, M. and MULLAINATHAN, S. (2001). Are ceos rewarded for luck? the ones without principals are. *Quarterly Journal of Economics*, **116** (3), 901–932.

BESLEY, T. and COATE, S. (1992). Workfare versus welfare: Incentive arguments for work requirements in poverty-alleviation programs. *American Economic Review*, **82** (1), 249–261.

— and — (1995). The design of income maintenance programmes. *The Review of Economic Studies*, **62** (2), 187–221.

BLOOM, N., EIFERT, B., MAHAJAN, A., MCKENZIE, D. and ROBERTS, J. (2013). Does management matter? evidence from india. *Quarterly Journal of Economics*, **128** (1), 1–51.

BLUNDELL, R. and SHEPHARD, A. (2011). Employment, hours of work and the optimal taxation of low-income families. *The Review of Economic Studies*, p. rdr034.

BOLLINGER, B., LESLIE, P. and SORENSEN, A. (2011). Calorie posting in chain restaurants. *American Economic Journal: Economic Policy*, **3** (1), 91–128.

BOURGUIGNON, F. and SPADARO, A. (2012). Tax–benefit revealed social preferences. *The Journal of Economic Inequality*, **10** (1), 75–108.

BOYD-SWAN, C., HERBST, C. M., IFCHER, J. and ZARGHAMEE, H. (2013). The earned income tax credit, health, and happiness.

CARD, D. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, vol. 3, *30*, Amsterdam: Elsevier, pp. 1801–1863.

— and HYSLOP, D. R. (2005). Estimating the effects of a time-limited earnings subsidy for welfare-leavers. *Econometrica*, **73** (6), 1723–1770.

— and — (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, L. S., MEIER, S. and WANG, S. W. (2014). Poverty and economic decision making: evidence from changes in financial resources at payday. *Unpublished manuscript*.

CBO (2005). *Effective Marginal Tax Rates on Labor Income*. Tech. rep., Congressional Budget Office.

Center on Budget and Policy Priorities (2014). *Policymakers Often Overstate Marginal Tax Rates for Lower-Income Workers and Gloss Over Tough Trade-Offs in Reducing Them*. Tech. rep., http://www.cbpp.org/research/policymakers-often-overstate-marginal-tax-rates-for-lower-income-workers-and-gloss-over.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annu. Rev. Econ.*, **1** (1), 451–488.

— (2012a). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, **80** (3), 969–1018.

— (2012b). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, **80** (3), 969–1018.

— (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, **105** (5), 1–33.

—, Friedman, J. and Rockoff, J. E. (2014). Measuring the impact of teachers ii: Teacher value-added and student outcomes. *American Economic Review*, **104** (9), 2633–2679.

—, Friedman, J. N. and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *American Economic Review*, **103** (7), 2683–2721.

—, Looney, A. and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, **99** (4), 1145–1177.

Choné, P. and Laroque, G. (2010). Negative marginal tax rates and heterogeneity. *The American Economic Review*, pp. 2532–2547.

Colombi, R. (1990). A new model of income distributions: The pareto lognormal differentstribution. In C. Dagum and M. Zenga (eds.), *Income and Wealth Distribution, Inequality and Poverty*, Berlin: Springer, pp. 18–32.

Congressional Budget Office (2012). *Effective Marginal Tax Rates for Low- and Moderate-Income Workers*. Tech. rep., https://www.cbo.gov/publication/43709.

Cuff, K. (2000). Optimality of workfare with heterogeneous preferences. *Canadian Journal of Economics/Revue canadienne d'économique*, **33** (1), 149–174.

Dahl, G. B. and Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review*, **102** (5), 1927–1956.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, **47** (2), 315–372.

—, Lindner, A., Reizer, B. and Schmieder, J. F. (2015). Reference-dependent job search: Evidence from hungary. *Unpublished Working Paper*.

— and Paserman, M. D. (2005). Job search and impatience. *Journal of Labor Economics*, **23** (3).

157

Di Tella, R., MacCulloch, R. J. and Oswald, A. J. (2001). Preferences over inflation and unemployment: Evidence from surveys of happiness. *American Economic Review*, **91** (1), 335–341.

Diamond, P. (1980). Income taxation with fixed hours of work. *Journal of Public Economics*, **13** (1), 101–110.

Diamond, P. A. (1998a). Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review*, **88** (1), 83–95.

— (1998b). Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review*, **88** (1), 83–95.

Dixit, A. and Norman, V. (1978). Advertising and welfare. *Bell Journal of Economics*, **9** (1), 1–17.

Drenik, A. and Perez-Truglia, R. (2014). Sympathy for the diligent and the demand for workfare. *Available at SSRN 1599368*.

Fang, H. and Silverman, D. (2009). Time-inconsistency and welfare program participation: Evidence from the nlsy*. *International Economic Review*, **50** (4), 1043–1077.

Farhi, E. and Gabaix, X. (2015). Optimal taxation with behavioral agents.

Finkelstein, A. (2009). E-ztax: Tax salience and tax rates. *The Quarterly Journal of Economics*, **124** (3), 969–1010.

Fleurbaey, M. and Maniquet, F. (2006). Fair income tax. *The Review of Economic Studies*, **73** (1), 55–83.

French, K. R. (2008). The cost of active investing. *Journal of Finance*, **63** (4), 1537–1573.

Gabaix, X. and Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, **121** (2).

— and Landier, A. (2008). Why has ceo pay increased so much? *Quarterly Journal of Economics*, **123** (1), 49–100.

Gerritsen, A. (2015). Optimal taxation when people do not maximize well-being.

Goldin, C. and Katz, L. F. (2007). Long-run changes in the wage structure: Narrowing, widening, polarizing. *Brookings Papers on Economic Activity*, pp. 135–165.

—, —, Hausman, N. and Ward, B. (2013). *Harvard and Beyond Project*. Survey data project, Harvard University.

Goldin, J. and Homonoff, T. (2013). Smoke gets in your eyes: cigarette tax salience and regressivity. *American Economic Journal: Economic Policy*, **5** (1), 302–336.

Gruber, J. and Kőszegi, B. (2001). Is addiction "rational"? theory and evidence. *The Quarterly Journal of Economics*, **116** (4), 1261–1303.

— and Kőszegi, B. (2004). Tax incidence when individuals are time-inconsistent: the case of cigarette excise taxes. *Journal of Public Economics*, **88** (9), 1959–1987.

— and Saez, E. (2002). The elasticity of taxable income: evidence and implications. *Journal of public Economics*, **84** (1), 1–32.

Gul, F. and Pesendorfer, W. (2008). The case for mindless economics. In A. Caplin and A. Schotter (eds.), *The Foundations of Positive and Normative Economics: A Handbook*, New York: Oxford University Press, pp. 3–42.

Gurun, U. G., Matvos, G. and Seru, A. (Forthcoming). Advertising expensive mortgages. *Journal of Finance*.

Halpern-Meekin, S., Edin, K., Tach, L. and Sykes, J. (2015). *It's Not Like I'm Poor*. University of California Press.

Hastings, J. S., Neilson, C. A. and Zimmerman, S. D. (2013). Are some degrees worth more than others? evidence from college admission cutoffs in chile, working Paper no. 19241, NBER.

Hellwig, M. F. (2007). A contribution to the theory of optimal utilitarian income taxation. *Journal of Public Economics*, **91** (7), 1449–1477.

Hendren, N. (2014). The inequality deflator: Interpersonal comparisons without a social welfare function.

Herbst, C. M. (2013). Welfare reform and the subjective well-being of single mothers. *Journal of Population Economics*, **26** (1), 203–238.

Hicks, J. R. (1939). *Value and Capital*. Oxford: Clarendon.

Hirshleifer, J. (1971). The private and social value of information and the reward to inventive activity. *American Economic Review*, **61** (4), 561–574.

Ifcher, J. (2010). The happiness of single mothers after welfare reform. *The BE Journal of Economic Analysis & Policy*, **11** (1).

— and Zarghamee, H. (2014). The happiness of single mothers: Evidence from the general social survey. *Journal of Happiness Studies*, **15** (5), 1219–1238.

Jacquet, L. and Lehmann, E. (2014). Optimal nonlinear income taxation with multidimensional types: The case with heterogeneous behavioral responses.

—, — and Van der Linden, B. (2013). Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory*, **148** (5), 1770–1805.

Jaffe, A. B. (1989). Real effects of academic research. *American Economic Review*, **79** (5), 957–970.

Jones, C. I. and Williams, J. C. (1998). Measuring the social return to r & d. *Quarterly Journal of Economics*, **113** (4), 1119–1135.

Jones, D. (2010). Information, preferences, and public benefit participation: Experimental evidence from the advance eitc and 401 (k) savings. *American Economic Journal: Applied Economics*, pp. 147–163.

— and Mahajan, A. (2015). Time-inconsistency and savings: Experimental evidence from low-income tax filers. *Center for Financial Security Working Paper*.

Kanbur, R., Keen, M. and Tuomala, M. (1994). Optimal non-linear income taxation for the alleviation of income-poverty. *European Economic Review*, **38** (8), 1613–1632.

Kaplow, L. (2007). Optimal income transfers. *International Tax and Public Finance*, **14** (3), 295–325.

— and Shavell, S. (1992). Private versus socially optimal provision of ex ante legal advice. *Journal of Law, Economics, and Organization*, **8** (2), 306–320.

Kaur, S., Kremer, M. and Mullainathan, S. (2015). Self-control at work. *Forthcoming*.

Kirkebøen, L., Leuven, E. and Mogstad, M. (Forthcoming). Field of study, earnings, and self-selection. *Quarterly Journal of Economics*.

Kotlikoff, L. and Rapson, D. (2009). Does it pay, at the margin, to work and save? *Tax Policy and the Economy*, **21**.

Kroft, K., Kucko, K. J., Lehmann, E. and Schmieder, J. F. (2015). *Optimal Income Taxation with Unemployment and Wage Responses: A Sufficient Statistics Approach*. Tech. rep., National Bureau of Economic Research.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, **112** (2), 443–478.

—, Repetto, A. and Tobacman, J. (2007). *Estimating discount functions with consumption choices over the lifecycle*. Tech. rep., National Bureau of Economic Research.

Lockwood, B. B. and Taubinsky, D. (2015). Regressive sin taxes.

— and Weinzierl, M. (2015). De gustibus non est taxandum: Heterogeneity in preferences and optimal redistribution. *Journal of Public Economics*, **124** (0), 74–80.

— and — (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.

Malmendier, U. and Tate, G. (2009). Superstar ceos. *Quarterly Journal of Economics*, **124** (4), 1593–1638.

Mankiw, N. G. (2010). Spreading the wealth around: Reflections inspired by joe the plumber. *Eastern Economic Journal*, **36**, 285–298.

—, Weinzierl, M. and Yagan, D. (2009). Optimal taxation in theory and practice. *The Journal of Economic Perspectives*, pp. 147–174.

McGinn, K. L., Ruiz Castro, M. and Lingo, E. L. (2015). Mums the word! cross-national effects of maternal employment on gender inequalities at work and at home.

Meier, S. and Sprenger, C. (2015). Stability of time preferences. *Review of Economics and Statistics*.

Meyer, B. D. (2010). The effects of the earned income tax credit and recent reforms. In *Tax Policy and the Economy, Volume 24*, The University of Chicago Press, pp. 153–180.

— and Sullivan, J. X. (2008). Changes in the consumption, income, and well-being of single mother headed families. *The American Economic Review*, pp. 2221–2241.

Mirrlees, J. A. (1971a). An exploration in the theory of optimal taxation. *Review of Economic Studies*, **38** (2), 175–208.

— (1971b). An exploration in the theory of optimum income taxation. *The Review of Economic Studies*, pp. 175–208.

Moffitt, R. (2006). Welfare work requirements with paternalistic government preferences*. *The Economic Journal*, **116** (515), F441–F458.

Mullainathan, S., Schwartzstein, J. and Congdon, W. J. (2012). A reduced-form approach to behavioral public finance. *Economics*, **4**.

— and Shafir, E. (2013). *Scarcity: Why having too little means so much*. Macmillan.

Murphy, K. M., Shleifer, A. and Vishny, R. W. (1991). The allocation of talent: Implications for growth. *Quarterly Journal of Economics*, **106** (2), 503–530.

— and Topel, R. H. (2006). The value of health and longevity. *Journal of Political Economy*, **114** (5), 871–904.

O'Donoghue, T. and Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, **90** (10), 1825–1849.

Paserman, M. D. (2008). Job search and hyperbolic discounting: Structural estimation and policy evaluation*. *The Economic Journal*, **118** (531), 1418–1452.

Patterson, S. (2010). *The Quants*. New York: Random House.

Philippon, T. (2010). Financiers versus engineers: Should the financial sector be taxed or subsidized? *American Economic Journal: Microeconomics*, **2** (3), 158–182.

— and Reshef, A. (2012). Wages and human capital in the u.s. finance industry: 1909–2006. *Quarterly Journal of Economics*, **127** (4), 1551–1609.

Piketty, T. and Saez, E. (2003). Income inequality in the united states, 1913–1998. *Quarterly Journal of Economics*, **118** (1), 1–41.

—, — and Stantcheva, S. (2014). Optimal taxation of top labor incomes: A tale of three elasticities. *American Economic Journal: Economic Policy*, **6** (1), 230–271.

Posner, E. and Weyl, E. G. (2013). Benefit-cost analysis for financial regulation. *American Economic Review Papers and Proceedings*, **103** (3), 393–397.

Reyes, L., Rodríguez, J. and Urzúa, S. S. (2013). Heterogeneous economic returns to postsecondary degrees: Evidence from chile, working Paper no. 18817, NBER.

Rizzo, J. A. (1999). Advertising and competition in the ethical pharmaceutical industry: The case of antihypertensive drug. *Journal of Law and Economics*, **42** (1), 89–116.

Romich, J. L. and Weisner, T. (2000). How families view and use the eitc: Advance payment versus lump sum delivery. *National tax journal*, pp. 1245–1265.

Rothschild, C. and Scheuer, F. (2014). A theory of income taxation under multidimensional skill heterogeneity, manuscript, Department of Economics, Stanford University.

— and — (Forthcoming). Optimal taxation with rent-seeking. *Review of Economic Studies*.

Rothstein, J. (2010). Is the eitc as good as an nit? conditional cash transfers and tax incidence. *American Economic Journal: Economic Policy*, pp. 177–208.

Saez, E. (2001a). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies*, **68** (1), 205–229.

— (2001b). Using elasticities to derive optimal income tax rates. *The Review of Economic Studies*, **68** (1), 205–229.

— (2002a). The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics*, **83** (2), 217–230.

— (2002b). Optimal income transfer programs: intensive versus extensive labor supply responses. *The Quarterly Journal of Economics*, **117** (3), 1039–1073.

— (2002c). Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics*, **117** (3), 1039–1073.

—, Slemrod, J. and Giertz, S. H. (2012a). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, **50** (1), 3–50.

—, — and — (2012b). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature*, **50** (1), 3–50.

— and Stantcheva, S. (2016). Generalized social welfare weights for optimal tax theory. *American Economic Review*, **106** (1), 24–45.

Saxenian, A. (2006). *The New Argonauts*. Cambridge, MA: Harvard University Press.

Seade, J. (1982). On the sign of the optimum marginal income tax. *The Review of Economic Studies*, **49** (4), 637–643.

Seade, J. K. (1977). On the shape of optimal tax schedules. *Journal of public Economics*, **7** (2), 203–235.

SHAH, A. K., MULLAINATHAN, S. and SHAFIR, E. (2012). Some consequences of having too little. *Science*, **338** (6107), 682–685.

SHAH, M. and STEINBERG, B. M. (2015). Workfare and human capital investment: Evidence from india.

SPENCE, A. M. (1976). Product selection, fixed costs and monopolistic competition. *Review of Economic Studies*, **43** (2), 217–235.

SPINNEWIJN, J. (2014). Unemployed but optimistic: Optimal insurance design with biased beliefs. *Journal of the European Economic Association*.

STANTCHEVA, S. (2014). Optimal taxation and human capital policies over the life cycle.

STIGLITZ, J. E. (1982). Self-selection and pareto efficient taxation. *Journal of Public Economics*, **17** (2), 213–240.

VICKREY, W. (1945a). Measuring marginal utility by reactions to risk. *Econometrica: Journal of the Econometric Society*, pp. 319–333.

— (1945b). Measuring marginal utility by reactions to risk. *Econometrica*, **13** (4), 319–333.

WANE, W. (2001). The optimal income tax when poverty is a public 'bad'. *Journal of Public Economics*, **82** (2), 271–299.

WEINZIERL, M. (2014). The promise of poistive optimal taxation: Normative diversity and a role for equal sacrifice. *Journal of Public Economics*, **118**, 128–142.

WERNING, I. (2000). An elementary proof of positive optimal marginal taxes.

WILSON, R. B. (1993). *Nonlinear Pricing*. Oxford: Oxford University Press.

Appendix A

Appendix to Chapter 1

A.1 Proof of Proposition 1.1 using the Hamiltonian

Proof. Let $\ell(w) = z(w)/w$, and let $V(w)$ denote rescaled decision utility for ability w:

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

The individual's optimization implies

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

and so we have

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

Then experienced utility is equal to

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

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

$$\max \int_{w_0}^{w_1} \alpha(w)\left(V(w) + \left(\frac{1 - \beta(w)}{\beta(w)}\right)v(\ell(w))\right)f(w)dw$$

subject to

$$\int_{w_0}^{w_1} \left(V(w) + \frac{v(\ell(w))}{\beta(w)} \right) f(w)dw \leq \int_{w_0}^{w_1} w\ell(w)f(w)dw - E. \tag{A.1}$$

and

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

Letting λ denote the multiplier on the budget constraint in (A.1) $m(w)$ denote the multipliers on the constraint in (A.2), then the Hamiltonian for this problem is

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

The usual solution technique requires

$$m'(w) = -\frac{\partial \mathcal{H}}{\partial V} = (\lambda - \alpha(w)) f(w).$$

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

$$\left(-\alpha(w) \left(\frac{1-\beta(w)}{\beta(w)} \right) v'(\ell(w)) + \lambda \left(\frac{v'(\ell(w))}{\beta(w)} - w \right) \right) f(w)$$
$$= m(w) \left(\frac{v'(\ell(w)) + \ell(w)v''(\ell(w))}{\beta(w)w} + \frac{v'(\ell(w))\beta'(w)}{\beta(w)^2} \right). \tag{A.3}$$

Using the fact that $m(w_1) = 0$ (no distortion at the top) we have

$$m(w) = \int_{w_1}^{w} m'(w)dw = -\int_{w}^{w_1} m'(w)dw = \int_{w}^{w_1} (\alpha(w) - \lambda) f(w)dw.$$

Substituting into (A.3) and rearranging,

$$\frac{T'}{1-T'} =$$
$$\frac{1}{f(w)} \int_{w}^{w_1} (1 - g(w)) f(w)dw \left(\frac{1 + \ell(w)v''(\ell(w))/v'(\ell(w))}{w} + \frac{\beta'(w)}{\beta(w)} \right) - g(w) (1 - \beta(w)).$$

Then substituting in the expressions (from the text) for the elasticities of labor supply and present bias yields the expression in Proposition 1.1. □

165

A.2 Elasticity concepts

In this section I define formally the elasticities employed in Section 1.2.2. As in Jacquet and Lehmann (2014), these include "circularities" due to the curvature of the tax function, wherein an initial change causes a change in earnings, leading to a further change in the tax rate, etc. To this, it is useful to parameterize a local tax reform about earnings level z^* using r and τ, where r is a reduction in the level of taxes at z^* (identifying an income effect) and τ is a compensated change in the marginal tax rate at z^* (identifying a compensated elasticity):

$$\hat{T}(z; z^*, r, \tau) := T(z) - r + \tau(z - z^*).$$

Then let $\hat{z}(i; r, \tau)$ denote i's optimal choice of earnings as a function of the reform parameters r and τ, when the tax code is perturbed around i's status quo chosen level of earnings $z(i)$:

$$\hat{z}(i; r, \tau) := \arg\max_z \left\{ -v(z/w_i) - \chi_i \cdot 1\left\{z > 0\right\} + \beta_i u(z - \hat{T}(z; z(i), r, \tau)) \right\}.$$

The compensated elasticity is defined as the elasticity of income with respect to the marginal tax rate increase, τ, evaluated at the status quo:

$$\varepsilon(i) := \left(-\frac{d\hat{z}(i; 0, \tau)}{d\tau} \bigg|_{\tau=0} \right) \frac{1 - T'(z(i))}{z(i)}.$$

Computing this derivation using the separable form in Equation (1.1), we find that for any labor force participant,

$$\varepsilon(i) = \frac{1/z(i)}{\frac{v''(z(i)/w_i)/w_i}{v'(z(i)/w_i)} + \frac{T''(z(i))}{1 - T'(z(i))} - \frac{u''(z(i) - T(z(i)))}{u'(z(i) - T(z(i)))}(1 - T'(z(i)))}.$$

An implication of this result is that if labor supply disutility is isoelastic (i.e., $v(x) = ax^b$ for constants a and b) then the elasticity is constant conditional on income, as $\frac{v''(z(i)/w_i)/w_i}{v'(z(i)/w_i)} = \frac{b-1}{z(i)}$. That is, different combinations of ability and bias which give rise to the same chosen level of earnings also give rise to the same compensated elasticity of earnings.

The income effect is defined formally as

$$\eta(i) := \left(-\frac{d\hat{z}(i;r,0)}{dr}\bigg|_{r=0} \right) (1 - T'(z(i))).$$

Again computing the derivative using Equation (1.1), we have

$$\eta(i) = \frac{-u'((z(i) - T(z(i)))^{-1}}{\frac{v''(z(i)/w_i)/w_i}{v'(z(i)/w_i)} + \frac{T''(z(i))}{1-T'(z(i))} - \frac{u''(z(i)-T(z(i)))}{u'(z(i)-T(z(i)))}(1 - T'(z(i)))}(1 - T'(z(i))).$$

As with the compensated elasticity, this implies that if disutility of labor is isoelastic, then the income effect is the same for all individuals with a given level of earnings.

Since the participation elasticity is discrete, it is not defined as a function of individual type i, but rather as a function of income, representing the number of individuals who enter the labor force to earn z when the level of taxes at z is reduced slightly. Let $z^p(i)$ denote the optimal *participation earnings* of i—the earnings chosen if $z = 0$ is were removed from the budget set—under T, and $\hat{z}^p(i;r,\tau)$ the participation earnings under the reformed $\hat{T}$.[1] The participation elasticity quantifies the change in the share of labor force participants with a given participation earnings z^* in response to a small tax reduction at z^*. To write this formally, note that by construction $h(z^*)$ is equal to the measure of individuals with participation earnings $z^p(i) = z^*$ who actually participate:

$$h(z^*) = \int_{\mathcal{I}^p(z^*)} 1\left\{ \chi_i < -v\left(\frac{z^p(i)}{w_i}\right) + \beta_i\left(u\left(z^p(i) - T(z^p(i))\right) - u(-T(0))\right)\right\} d\mu(i)$$

Now define $\hat{h}(z^*;r,\tau)$ to be the measure of individuals with participation earnings $z^p(i) = z^*$ under the status quo tax $T(z)$ who participate under the reformed tax $\hat{T}(z;z^*,r,\tau)$:

$$\hat{h}(z^*;r,\tau) =$$
$$\int_{\mathcal{I}^p(z^*)} 1\left\{ \chi_i < -v\left(\frac{\hat{z}^p(i;r,\tau)}{w_i}\right) + \beta_i\left(u\left(\hat{z}^p(i;r,\tau) - \hat{T}(\hat{z}^p(i;r,\tau);z^*,r,\tau)\right) - u(-T(0))\right)\right\} d\mu(i).$$

Note that the set $I^p(z^*)$ is defined with respect to the status quo tax $T(z)$ and does not vary with the tax reform—this is because we are interested in the change in the share of labor

---

[1] That is, $\hat{z}^p(i;r,\tau) := \arg\max_z \{ -v(z/w_i) + \beta_i u(z - \hat{T}(z;z(i),r,\tau)) \}$.

force participants due solely to entry from outside the labor force, and not from the change in individuals earning z^* due to local adjustments in response to the reform.

Using this notation, the participation elasticity at z^* is defined as

$$\rho(z) := \left(\left. \frac{d\hat{h}(z; r, 0)}{dr} \right|_{r=0} \right) \frac{z - T(z) + T(0)}{h(z)}.$$

A.3 Derivation of Welfare Internality Terms

The change in welfare due to the welfare internality from the substitution effect (the local adjustment of earnings among individuals who face a higher marginal tax rate) is equal to

$$dS_W = d\tau\epsilon \left(\frac{1}{\lambda} \right) \int_{\mathcal{I}(z^*)} \alpha_i \left(\left. \frac{d\hat{z}(i; 0, \tau)}{d\tau} \right|_{\tau=0} \right) \left[-v'(z^*/w_i)/w_i + u'(z^* - T(z^*))(1 - T'(z^*)) \right] d\mu(i).$$

This is the change in experienced utility among individuals earning z^* whose earnings respond locally due to the rise in the marginal tax rate at z^*. The division by λ converts from utility (measured in the integral) into dollars, for summation with revenue effects. Individual optimization implies that $-v'(z^*/w_i)/w_i + \beta_i u'(z^* - T(z^*))(1 - T'(z^*)) = 0$; subtracting this from the term in the integral yields

$$\begin{aligned} dS_W &= d\tau\epsilon \int_{\mathcal{I}(z^*)} \left(\left. \frac{d\hat{z}(i; 0, \tau)}{d\tau} \right|_{\tau=0} \right) (1 - \beta_i) \left(\frac{\alpha_i u'(z^* - T(z^*))}{\lambda} \right) (1 - T'(z^*)) d\mu(i) \\ &= d\tau\epsilon \int_{\mathcal{I}(z^*)} \left(\left. \frac{d\hat{z}(i; 0, \tau)}{d\tau} \right|_{\tau=0} \right) (1 - \beta_i) \, g(i)(1 - T'(z^*)) d\mu(i) \\ &= d\tau\epsilon z^* \int_{\mathcal{I}(z^*)} -\varepsilon(i) \, (1 - \beta_i) \, g(i) d\mu(i). \end{aligned}$$

Rewriting the integral as an expectation yields the expression in the text. The derivation of dI_W is analogous.

The change in welfare due to the welfare internality on the extensive margin from a reduction in taxes dr for an income band of width $\epsilon \to 0$ around income level z^* is

$$dr\epsilon \left(\frac{1}{\lambda} \right) \int_{\mathcal{I}_{ext}(z^*)} \alpha_i \left[-v(z^*/w_i) - \chi_i + u(z^* - T(z^*)) - u(-T(0)) \right] d\mu(i).$$

By individual optimization, individuals on the extensive margin (i.e., those in $\mathcal{I}_{ext}(z^*)$)

satisfy $-v(z^*/w_i) - \chi_i + \beta_i u(z^* - T(z^*)) - \beta_i u(-T(0)) = 0$. Subtracting this from the term in brackets yields

$$d r \epsilon \left(\frac{1}{\lambda}\right) \int_{\mathcal{I}_{ext}(z^*)} \alpha_i (1 - \beta_i) \left[u(z^* - T(z^*)) - u(-T(0))\right] d\mu(i) =$$

$$d r \epsilon \left[z^* - T(z^*) + T(0)\right] \int_{\mathcal{I}_{ext}(z^*)} g_{ext}(i)(1 - \beta_i) d\mu(i),$$

employing the definition of the extensive margin welfare weight from the text. Rewriting the integral as a conditional expectation, and using the fact that $\int_{\mathcal{I}_{ext}(z^*)} d\mu(i) = \left. \frac{d\hat{h}(z;r,0)}{dr}\right|_{r=0}$, this expression can be written

$$d r \epsilon \rho(z^*) h(z^*) \mathbb{E}\left[g_{ext}(i)(1 - \beta_i) \big| i \in \mathcal{I}_{ext}(z^*)\right].$$

Under the perturbation in the text, this extensive margin internality effect occurs at all incomes above the point of the marginal tax reform. Integrating across those incomes yields the expression in the text.

A.4 Proof of Lemma 1.1

Proof. Suppose that for some χ' and χ'', $Y(\chi') = Y(\chi'') = 0$. Then by (1.10), $\beta \tilde{u}(Z(\chi') = \beta \tilde{u}(Z(\chi'')))$ (using the inequality in both directions), and therefore $Z(\chi') = Z(\chi'')$. Thus any individuals who do not work must receive the same compensation—call this value Z_0. Now suppose there is some type χ^* such that $Y(\chi^*) > 0$, and consider two types $\chi^\dagger$, $\chi^{\dagger\dagger}$, both less than χ^*. If any types do not work, then (1.10) requires $-v(Y(\chi^*)) - \chi^* + \beta \tilde{u}(Z(\chi^*)) \geq -v(0) + \beta \tilde{u}(Z_0)$. Since $\chi^\dagger < \chi^*$, it follows that $\chi^\dagger$ would rather mimic χ^* than mimic a non-worker, implying that $Y(\chi^\dagger) > 0$, and similarly $Y(\chi^{\dagger\dagger}) > 0$. Thus we have that if some χ^* works, all types $\chi < \chi^*$ also work. Let $\mu(\chi|\chi')$ denote the multiplier on the constraint (1.10), requiring that type χ' not prefer to mimic χ, and let λ denote the multiplier on (1.11). Then the derivative of the Lagrangian from (1.9)–(1.12) with respect to $Y(\chi^\dagger)$ is

$$\frac{\partial \mathcal{L}}{\partial Y(\chi^\dagger)} = -\tilde{v}'(Y(\chi^\dagger)) + \lambda + \int \mu(\chi^\dagger|\chi)\tilde{v}'(Y(\chi^\dagger))d\chi - \int \mu(\chi|\chi^\dagger)\tilde{v}'(Y(\chi^\dagger))d\chi = 0, \quad \text{(A.4)}$$

and

$$\frac{\partial \mathcal{L}}{\partial Y(\chi^{\dagger\dagger})} = -\tilde{v}'(Y(\chi^{\dagger\dagger})) + \lambda + \int \mu(\chi^{\dagger\dagger}|\chi)\tilde{v}'(Y(\chi^{\dagger\dagger}))d\chi - \int \mu(\chi|\chi^{\dagger\dagger})\tilde{v}'(Y(\chi^{\dagger\dagger}))d\chi = 0.$$

(A.5)

By (1.10), $-v(Y(\chi^{\dagger})) + \beta\tilde{u}(Z(\chi^{\dagger})) = -v(Y(\chi^{\dagger\dagger})) + \beta\tilde{u}(Z(\chi^{\dagger\dagger}))$. Therefore $\mu(\chi^{\dagger}|\chi') = \mu(\chi^{\dagger\dagger}|\chi')$ for all χ' (it is equally tempting for other types to mimic $\chi^{\dagger}$ and $\chi^{\dagger\dagger}$). Moreover, $\mu(\chi'|\chi^{\dagger}) = \mu(\chi'|\chi^{\dagger\dagger})$ for all χ' such that $Y(\chi') > 0$—it is equally tempting for $\chi^{\dagger}$ and $\chi^{\dagger\dagger}$ to mimic some other working types—and the assumption that $\chi^{\dagger\dagger\dagger}$ is greater than and $\chi^{\dagger}$ and $\chi^{\dagger\dagger}$ implies that $Y = 0$ is strictly more tempting for $\chi^{\dagger\dagger\dagger}$ than for $\chi^{\dagger}$ or $\chi^{\dagger\dagger}$. Therefore multipliers on the constraints preventing $\chi^{\dagger}$ and $\chi^{\dagger\dagger}$ from mimicking non-workers are all zero. Thus $\mu(\chi'|\chi^{\dagger}) = \mu(\chi'|\chi^{\dagger\dagger})$ for all χ'. Thus rearranging (A.4) and (A.5) yields

$$\tilde{v}'(Y(\chi^{\dagger})) = \frac{\lambda}{1 - \int \mu(\chi^{\dagger}|\chi)d\chi + \int \mu(\chi|\chi^{\dagger})d\chi} = \frac{\lambda}{1 - \int \mu(\chi^{\dagger\dagger}|\chi)d\chi + \int \mu(\chi|\chi^{\dagger\dagger})d\chi} = \tilde{v}'(Y(\chi^{\dagger\dagger})).$$

Thus if some χ^* works, then all types $\chi < \chi^*$ also work and have the same production and compensation bundle denoted Y^* and Z^*. Moreover, if some χ^{**} does not work, then all types $\chi > \chi^{**}$ do not work, so $Y(\chi) = 0$, and all receive the same compensation Z_0. Since the support of χ is connected, this implies there is some threshold χ^* below which all types work, and above which all types do not. It is easily verified that χ^* must therefore be indifferent between working and not, implying $-v(Y^*) - \chi^* + \beta\tilde{u}(Z^*) = \beta\tilde{u}(Z_0)$. Thus the optimal contract is characterized by the parameters Z_0, Z^*, Y^*, and χ^*, and the lemma is proved. □

A.5 Proof of Proposition 1.5

Proof. **Full commitment.** Let w^h solve (1.15) with equality when $Z_0 = z_0$:

$$-v(Y^*/w^h)/w^h - \overline{\chi} + \beta\tilde{u}(Z^*) = \beta\tilde{u}(z_0).$$

Then by letting $Y^* = Z^* - \kappa$, with $v'(Y^*/w^h)/w^h = \tilde{u}(Z^*)$, and $\chi^* = \overline{\chi}$, (1.13) is maximized subject to (1.14), and constraints (1.15) and (1.16) do not bind, demonstrating that this

constitutes a constrained optimum. Since $\frac{v'(Y^*/w^h)/w^h}{u'(Z^*-T(Z^*))(1-T'(Z^*))} = \frac{v'(Y^*/w^h)/w^h}{\tilde{u}'(Z^*)} = 1$, the misoptimization wedge is zero. The same trivially holds for all $w > w^h$.

No commitment. Substituting (1.16) into (1.15) yields a combined constraint

$$-\tilde{v}(Y^*) - \chi^* + \beta\tilde{u}(Z^*) \geq \beta\tilde{u}(z_0).$$

Moreover, using the fact that $Z^* \leq Y^*$, we can write

$$-v(Y^*/w) - \chi^* + \beta\tilde{u}(Y^*) \geq -\tilde{v}(Y^*) - \chi^* + \beta\tilde{u}(Z^*) \geq \beta\tilde{u}(z_0).$$

Letting w grow arbitrarily small, either $v(Y^*/w)$ grows arbitrarily large, or Y^* becomes smaller than z_0. In either case, the inequality above is violated, showing that for sufficiently low w, no contract with an interior χ^* satisfies (1.14)–(1.16). Therefore the only (possibly) feasible contract is $\chi^* = \underline{\chi}$ and $Z_0 = -\kappa$. If this contract violates (1.16), it is infeasible; if not, it is nevertheless dominated by the the outside option of flexibly setting labor supply in period 1, which provides a lower bound of utility equal to $Z_0 = 0$.

Limited commitment. Let λ denote the multiplier on (1.14), and since (1.15) binds at the optimum if $\chi^* < \overline{\chi}$ and the constraint is slack if $\chi^* = \overline{\chi}$, we can suppose that Z_0 satisfies (1.15) with equality in all cases, and we can use that equation to implicitly determine χ^* as a function of Y^*, Z^*, and Z_0. Let μ denote the multiplier on (1.16). Then the optimal contract satisfies

$$\frac{\partial\mathcal{L}}{\partial Y^*} = F(\chi^*)\left(-\tilde{v}'(Y^*) + \lambda\right) + \frac{\partial\mathcal{L}}{\partial\chi^*}\frac{\partial\chi^*}{\partial Y^*}$$

$$= F(\chi^*)\left(-\tilde{v}'(Y^*) + \lambda\right) - f(\chi^*)\left((1-\beta)\left(\tilde{u}(Z^*) - \tilde{u}(Z_0)\right) + \lambda\left(Y^* - Z^* + Z_0\right)\right)\tilde{v}'(Y^*),$$

implying

$$\tilde{v}'(Y^*) = \frac{\lambda}{1 + \frac{f(\chi^*)}{F(\chi^*)}\left((1-\beta)\left(\tilde{u}(Z^*) - \tilde{u}(Z_0)\right) + \lambda\left(Y^* - Z^* + Z_0\right)\right)}.$$

Similarly,

$$\frac{\partial \mathcal{L}}{\partial Z^*} = F(\chi^*)\left(\tilde{u}'(Z^*) - \lambda\right) + \frac{\partial \mathcal{L}}{\partial \chi^*}\frac{\partial \chi^*}{\partial Z^*}$$

$$= F(\chi^*)\left(\tilde{u}'(Z^*) - \lambda\right) + f(\chi^*)\left((1-\beta)\left(\tilde{u}(Z^*) - \tilde{u}(Z_0)\right) + \lambda\left(Y^* - Z^* + Z_0\right)\right)\beta\tilde{u}'(Z^*),$$

implying

$$\tilde{u}'(Z^*) = \frac{\lambda}{1 + \beta\frac{f(\chi^*)}{F(\chi^*)}\left((1-\beta)\left(\tilde{u}(Z^*) - \tilde{u}(Z_0)\right) + \lambda\left(Y^* - Z^* + Z_0\right)\right)}.$$

Thus at the constrained optimum,

$$\frac{\tilde{v}'(Y^*)}{\tilde{u}'(Z^*)} = \frac{1 + \beta A}{1 + A},$$

with $A = \frac{f(\chi^*)}{F(\chi^*)}\left((1-\beta)\left(\tilde{u}(Z^*) - \tilde{u}(Z_0)\right) + \lambda\left(Y^* - Z^* + Z_0\right)\right)$. Since $A > 0$, this ratio lies between 1 (when $A = 0$) and β (as $A \to \infty$), implying that the misoptimization wedge lies between 0 and $1 - \beta$. $\qquad\square$

A.6   Details of numerical example in Section 1.3

This example uses a quasilinear utility function:

$$U(c, y, w) = c - \frac{(y/w)^{(1+1/e)}}{1 + 1/e},$$

where e is the elasticity of labor supply. I set $e = 0.3$, and I assume a flat tax marginal tax rate of 0.3 with no lump sum grant. (The nature of the results below is not sensitive to these assumptions.) I assume a population with heterogenous ability, with w distributed uniformly on $[\underline{w}, \overline{w}]$, and I assume all individuals face the same distribution of fixed costs $F(\chi)$.

A simplifying assumption.   The nature of the boundary between the "limited commitment" and "no commitment" contract regions depends on whether it is possible to sustain a desirable contract in which some individuals quit. To fix ideas while retaining simplicity,

172

here I assume it is not. For specificity, let $z_0 = 0$ (so that limited liability prohibits firms from fining workers). Since the firm participation constraint (1.14) also binds with equality under any optimal contract, we can substitute it into the the limited liability constraint to get a combined constraint:

$$-v(Y^*/w) - \chi^* + \beta \tilde{u} \left(Y^* - \frac{\kappa}{F(\chi^*)} \right) \geq \beta u(0).$$

Convexity of this constraint with respect to χ^* requires that

$$1 - \beta \tilde{u}' \left(Y^* - \frac{\kappa}{F(\chi^*)} \right) \left(\frac{\kappa f(\chi^*)}{F(\chi^*)^2} \right) > 0.$$

If this inequality is violated for a given χ^*, then that χ^* cannot be sustained—intuitively, a slight reduction in χ^* raises the share who quit, reducing the utility of remaining workers, further raising the share who quit, and unraveling the contract. In the example to follow, I assume that $f(\bar{\chi})$ is sufficiently high that such unraveling occurs for all $\chi^* < \bar{\chi}$. This simplifies the contract, as the only relevant feature of the distribution $F(\chi)$ is the upper bound $\bar{\chi}$. Letting Y_0^* and Y_1^* denote the labor supply levels which satisfy self 0 and self 1's first-order conditions without any quits, the three contracting regions have a simple interpretation in this case. In the full commitment region, self 1 gets greater utility from Y_0^* than from quitting, even under the worst-case fixed cost shock $\bar{\chi}$. In the no commitment region, self 1 prefers to quit than supply even the preferred Y_1^* under the worst-cased fixed cost shock $\bar{\chi}$. In the middle limited commitment region, self 0 chooses a labor supply between Y_1^* and Y_0^* to achieve partial commitment, while partially appeasing self 1 to prevent a quit under the maximal fixed cost $\bar{\chi}$.

Homogeneous present bias. I begin with the very simplest case of homogeneous present bias, with $\beta = 0.7$. Figure A.1 displays the utility from various labor supply and consumption bundles for self 0 (top panel) and self 1 (bottom panel), as a function of ability. The solid green lines plot the utility each self would derive from the labor supply which solves self 0's first-order condition. The dashed red lines show each self's utility from self 1's (lower) preferred labor supply. The horizontal dotted line in the lower panel shows the utility from

quitting for an individual with maximal fixed costs $\chi = \overline{\chi}$. (I set $\overline{\chi} = 0.42$, slightly above self 1's utility without commitment at the lowest ability, in order to ensure all three contracting regions are represented.) The thin vertical lines divide the type space into the three contract regions: no commitment, limited commitment, and full commitment (left to right).

In the right-most (full commitment) region, even a self 1 with a maximal fixed cost shock $\chi = \overline{\chi}$ prefers to provide self 0's favored labor supply rather than quit—thus the first-best commitment can be sustained, individuals in this region realize utilities on the green line in equilibrium. In the center (limited commitment) region, the self 0's first-best labor supply would induce individuals with high fixed cost realizations to quit, inflating the fixed costs born by the remaining workers, and causing the contract to unravel. To prevent such an outcome, the labor supply commitment must be lowered to the point that a self 1 with $\chi = \overline{\chi}$ is just willing to remain employed, so that z satisfies

$$\beta Z^* - \kappa - \frac{(Y^*/w)^{(1+1/e)}}{1+1/e} = \overline{\chi}.$$

This modification lowers self 0's utility below the that which would be realized under full commitment, as depicted by the dot-dashed orange line in the upper figure, which plots utility as a function of ability for individuals in this region in equilibrium. Finally, in the left-most (no commitment) region, a self 1 with a sufficiently high realization of χ will prefer to quit rather than provide even his most preferred level of labor supply. That positive probability of quitting unravels any potential contract, and thus in this region individuals do not contract and instead supply their preferred labor supply in period 1 after observing their fixed cost (if they work at all).

In this setup, the labor supply elasticity is equal to e in the left and right regions and zero in the center region. The income effect is equal to zero in the left and right regions (due to quasilinear utility of consumption) and is positive in the center region. These transitions are discrete when β is assumed to be homogeneous, though a smooth U-shape and inverse U-shape emerges under two dimensional heterogeneity.

174
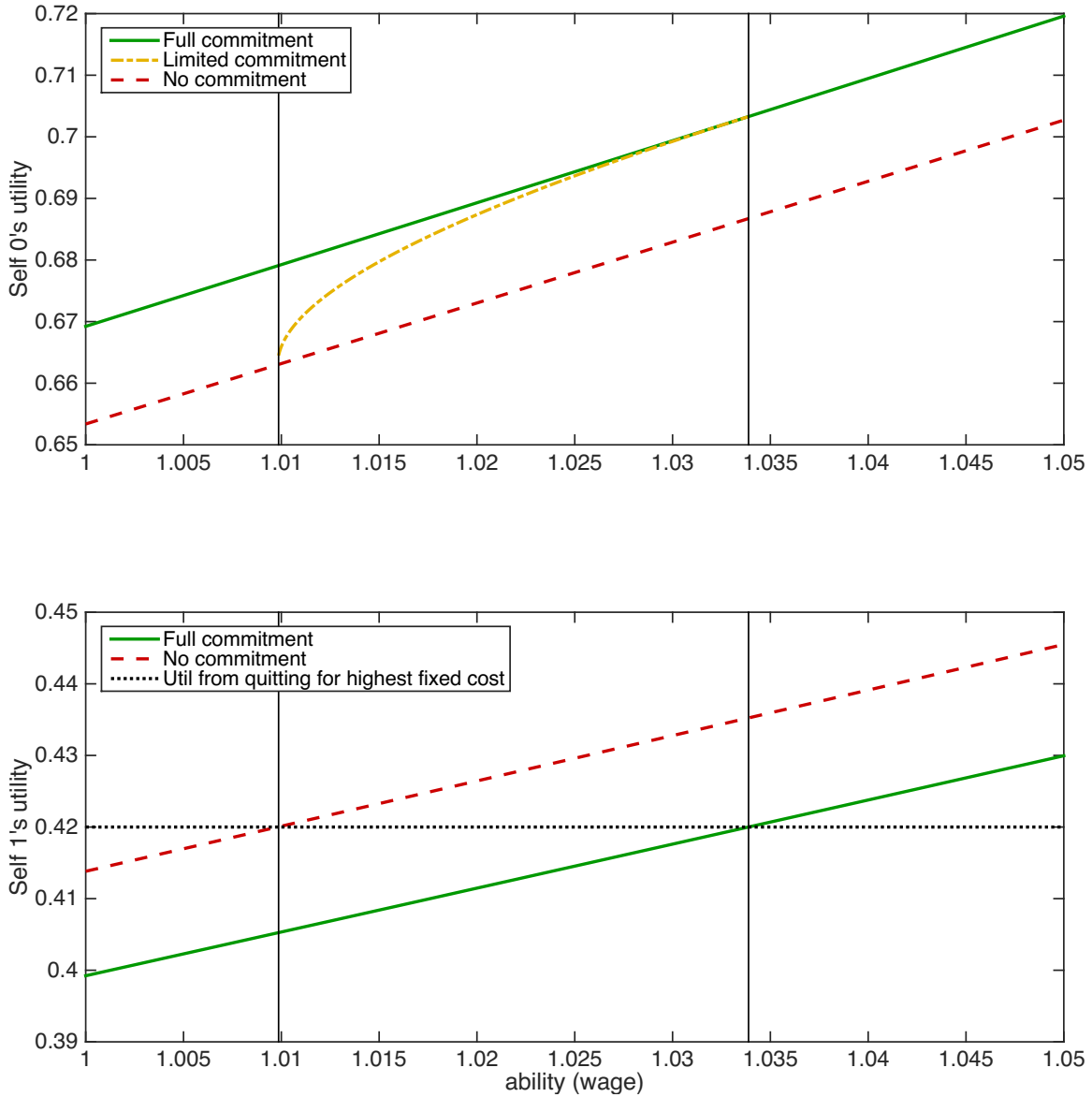
Figure A.1: *This figure plots the utility derived by the "long-term" self 0 (top panel) and the present biased self 1 under full commitment, no commitment, and limited commitment. Self 1's utility from the outside option of quitting, assuming the highest possible fixed cost draw $\overline{\chi}$, is shown by the dotted black line in the lower panel. Thin vertical lines divide the ability space into the three regions of employment contracts: full commitment (right), limited commitment (center), and no commitment (left).*

Heterogeneous present bias. I now extend the above example to a context with heterogeneous β. Suppose individuals are distributed with full support over $\left[\underline{\beta}, \overline{\beta}\right] \times [\underline{w}, \overline{w}]$. Figure A.2 displays the areas of the type space corresponding to various combinations of β and w for a range of abilities (for which the scaling is immaterial) and present bias.
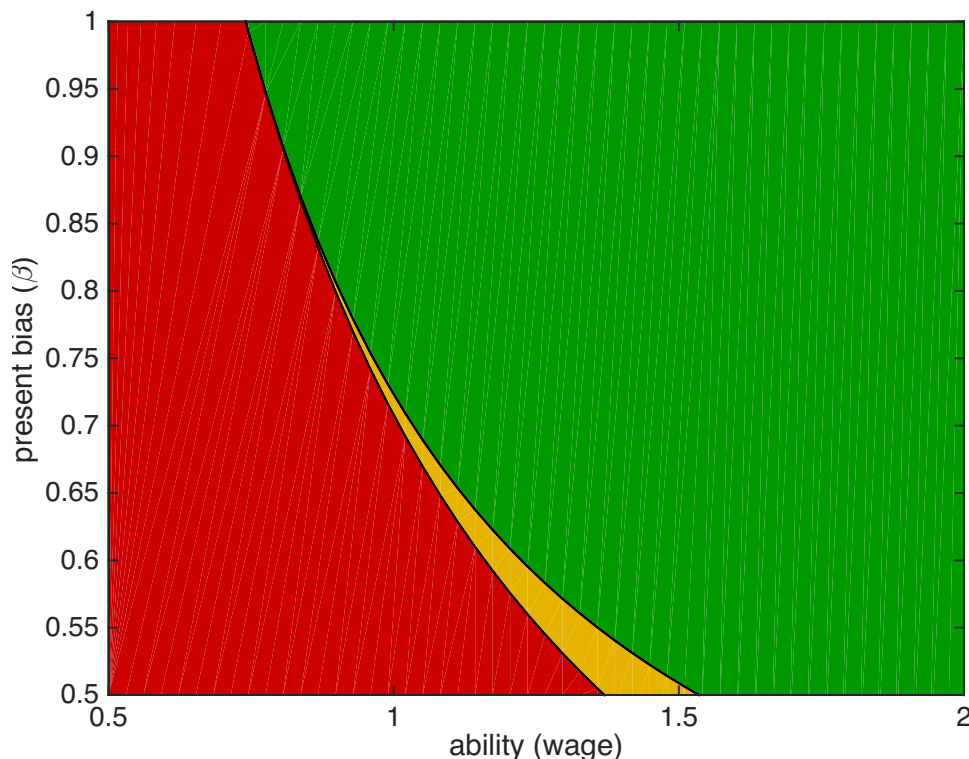
Figure A.2: *This figure plots the regions of the ability and present bias type space in which each contract type prevails. The upper right area represents full commitment; the middle area represents limited commitment; the lower left area represents no commitment.*

The region to the lower left represents the low ability, high bias types for whom commitment is unsustainable. The middle region represents the set of types for whom limited commitment is feasible, with the participation constraint binding for $\overline{\chi}$, and the upper right region represents the set of types for whom full commitment is sustainable.

Endogenous sorting on income. The results in Figures A.1–A.2 hold for any joint distribution over ability and present bias. However to simulate an income distribution, we must assume a distribution. For illustrative purposes, I explore the distribution of incomes

and contract behavior for a uniform joint distribution over the region in Figure A.2. The ability distribution can be extended upward without affecting the behavior of the income distribution above this range, for example by adding a nonparametric ability distribution calibrated to match the distribution of middle class incomes, with a Pareto tail at the top to represent the distribution of top incomes.

Figure A.3 displays the resulting income densities. As noted earlier, the density is continuous. The density of incomes is fairly uniform and dispersed in the full commitment region, as the sole dimension of heterogeneity determining income variation in this region is ability. (This also explains the positive income density at the upper bound of the distribution, which consists of all individuals on the right edge of Figure A.2, with $w = \overline{w}$.) In contrast, income density falls to zero at the bottom of the distribution, reflecting the variation of income with both w and β in this region—agents with earnings at the lower bound are represented by the point at the bottom left corner of the ability/bias space in Figure A.2, and thus they have measure zero.
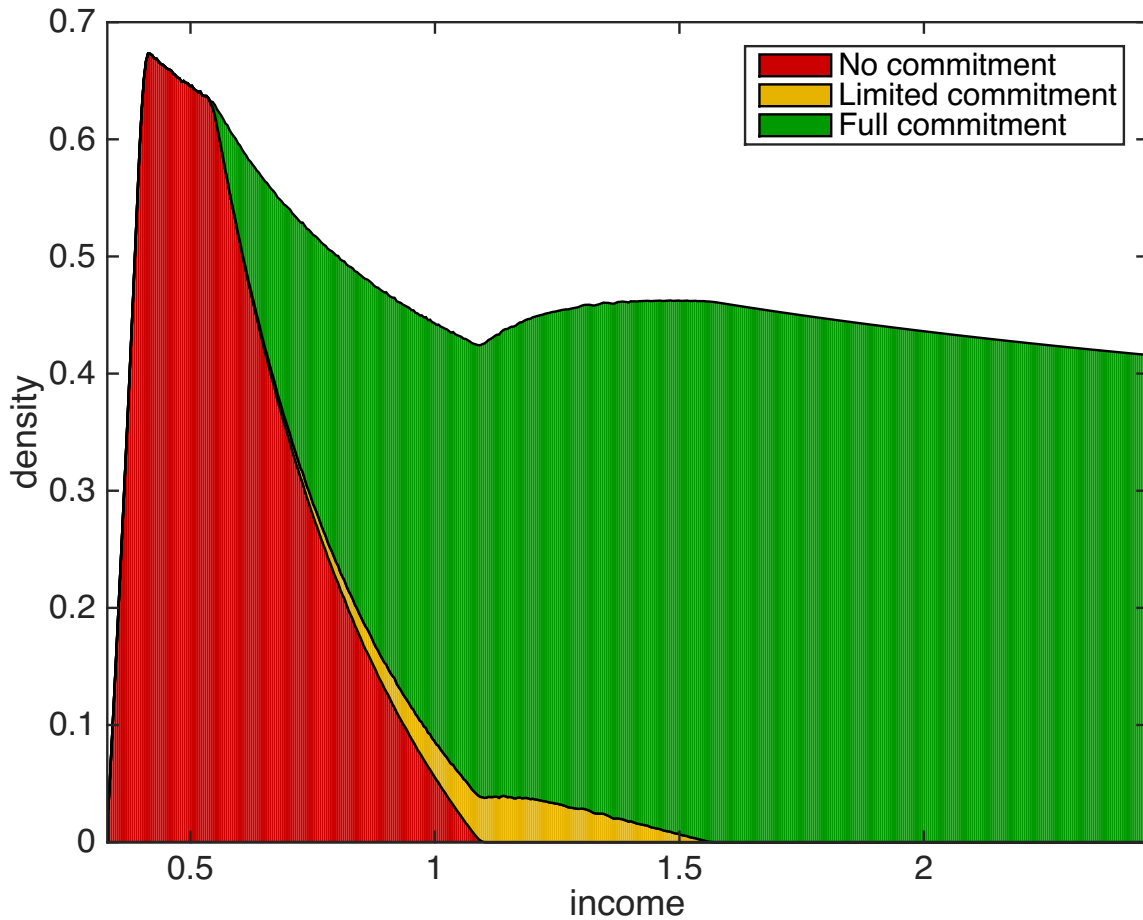
Figure A.3: *This figure displays the density of incomes resulting from a uniform distribution over the set of ability and bias in Figure A.2. The densities of individuals falling under each contract type are shaded as in Figure A.2, with full commitment dominating at high incomes, limited commitment bunched at middle incomes, and no commitment at low incomes.*

Endogenous sorting on income leads to greater concentration of uncommitted workers at low incomes than was apparent over the ability distribution—note that top incomes of those without commitment is below the median, though some such individuals have above median ability (see Figure A.2). This is a result of the fact that individuals with a given ability who have a sufficiently high β to sustain their first-best contract earn more than individuals who are unable to sustain a contract. Individuals with limited commitment have earnings which primarily overlap the lower tail of fully committed workers.

A.7 Review of studies estimating structural present bias

There is a substantial experimental literature—both in the lab and in the field—testing for present bias (see DellaVigna (2009) for a review).

Choice-based estimation. As discussed in the text, choice-based approach identifies a environment in which choices are trustworthy, and assumes that similar preferences persist in alternative environments. See Bernheim and Rangel (2007) and Bernheim and Rangel (2009) for a discussion of this theoretical framework, and Chetty, Looney and Kroft (2009) for an application to nonsalient taxes.

Augenblick, Niederle and Sprenger (2013) presents a lab experiment in which student participants face a fixed amount of effort to be performed within a given period. Absent access to a commitment device, students appear wait until late in the period to perform most of the effort—but a majority of students (59%) prefer a dominated commitment contract, which leads to smoother labor effort throughout the period. Commitment demand is greater among individuals who exhibit greater present bias. Individuals without commitment exhibit an apparent discount rate of about 11% per week. Thus if individuals were time consistent (with no discounting) beyond one week, this would suggest a misoptimization wedge of 0.89—this is the estimate reported in Table 1.1.

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" is larger than that which could be explained by a sensible time-consistent rate. Second, when given the option to choose a dominated commitment contract, in which pay is docked if an individual earns less than their pre-selected (and freely chosen) "goal amount." A substantial minority of individuals chose such contracts, and that share is larger among individuals who exhibit a more pronounced payday effect. Moreover, correlation between commitment demand and payday effect rises over time, consistent with

individuals learning about their self control problems (i.e., becoming more sophisticated) over time. Kaur *et al.* (2015) find that the payday effect is consistent with a daily discount factor of about 5%. This suggests a discount factor of about 0.7 at a one week horizon. Their results are inconsistent with a strict $\beta\,\delta$ model of quasihyperbolic discounting, as The results of commitment demand also shed light on the degree of sophistication and time inconsistency. The authors report that the demand for dominated commitment contracts suggest that "the difference in discounting of benefits between the self that chooses the contract and the one that works [is] at least 18%." If all discounting by the contract-selecting-self is rational, and if individuals are approximately time consistent beyond a one week horizon, this suggests $\beta = 1 - 0.18 = 0.82$, which is the value reported in Table 1.1. Among individuals with above-median payday effects, the difference is 64%, suggesting a misoptimization wedge among these most biased individuals of 0.64.

The remaining two studies in this category do not deal with tradeoffs over labor effort, however they are of particular interest for the population they study—both measure the preferences of EITC recipients in the US. Jones and Mahajan (2015) presents a field experiment in which EITC recipients have an opportunity to sign up for short-run savings accounts yielding high interest rates, with the option to make those accounts less liquid. 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. In both cases, subjects exhibit greater impatience at shorter horizons, suggestive of present bias. Meier and Sprenger (2015) perform a maximum likelihood estimation to calibrate the parameters of a quasi-hyperbolic discounting model. Figure 1.5 plots the nonparametric relationship (unreported in the paper) between the estimated β in each income quintile in their data, and average income within each quintile.[2]

The second source of evidence about present bias over labor supply comes from structural models, in which true utility is taken to be discounted exponentially. Laibson *et al.* (2007) use the method of simulated moments to perform a calibration using data on income, wealth,

---

[2]I am very grateful to Stephan Meier and Charlie Sprenger for sharing their data for this analysis.

and credit use. Fang and Silverman (2009) calibrate a quasi-hyperbolic model of model of welfare takeup and labor supply using data from the NLSY.

Structural estimation. Of particular interest, DellaVigna and Paserman (2005) analyzes the relationship between measured impatience and unemployment exits, and finds evidence for present bias over labor search effort, and Paserman (2008) extends the analysis to a calibrated structural model, finding evidence consistent with quasi-hyperbolic discounting behavior. The analysis is conducted separately for the bottom quartile, the middle half, and the top quartile of wage earners. The estimated β values are $\beta = 0.40$ in the bottom quartile, $\beta = 0.48$ or $\beta = 0.81$ in the middle half (depending on assumptions about the shape of the wage distribution) and $\beta = 0.89$ in the top quartile. I use the average of the two values for the middle half, $\beta = 0.65$, and I plot the three resulting values of β against average wage earnings in the bottom quartile, middle half, and top quartile of the 2010 wage distribution in Figure 1.5. (Results are nearly identical if I instead multiply the previous weekly wage for each group, as reported in Paserman (2008), by 52, and convert to 2010 dollars.)

Appendix B

Appendix to Chapter 2

B.1 Proofs and Derivations

B.1.1 Section 2.2

Formalization of No-Bunching Condition

Let w denote the total productivity of a worker. His optimal income choice is $y^* = \arg\max_y y - T(y) - \phi(y/w)$. This optimal y^* moves smoothly in response to perturbations in T as long as it strictly maximizes utility for all w. This property holds when the second-order condition is strictly satisfied when the first-order condition holds.

The first-order condition is $1 - T'(y) - \phi'(y/w)/w = 0$, and the second-order condition is $-T''(y) - \phi''(y/w)/w^2 < 0$. Because $\phi(h) = h^{1/1+\sigma}/(1/1+\sigma)$, $\phi''(h) = \phi'(h)/(\sigma h)$. Applying this equality, we find $\phi''(y/w)/w^2 = \phi'(y/w)/(\sigma y w) = (1 - T'(y))/(\sigma y)$, where the last equality used the first-order condition. The second-order condition thus simplifies to $-T''(y) - (1 - T'(y))/(\sigma y) < 0$, which reduces to the inequality in Assumption 2.1.

Proof of Lemma 2.1

Note $h^*(\theta)$ depends on $T(\cdot)$ only through $T'(\cdot)$. A worker prefers profession i over j if and only if

$$y_i^*(\theta) - T(y_i^*(\theta)) - \phi(h_i^*(\theta)) + \psi_i(\theta) > y_j^*(\theta) - T(y_j^*(\theta)) - \phi(h_j^*(\theta)) + \psi_j(\theta).$$

This equation depends on T through the intensive margin and through the difference $T(y_i^*(\theta)) - T(y_j^*(\theta))$, but from (2.2), this difference depends only on $T'(\cdot)$ and not on T_0. Therefore, $i^*(\theta)$ depends on $T'(\cdot)$ and not on T_0, so the equilibrium depends only on $T'(\cdot)$.

Let R^a and R^b be two revenue requirements, and let T^a and T^b be the respective optimal tax rates. Let $\mathcal{U}^a$ and $\mathcal{U}^b$ be the respective values of the government's objective function under T^a and T^b. Consider the tax schedule $T^a + R^b - R^a$ formed by adding $R^b - R^a$ to T_0^a but leaving $(T^a)'$ unchanged. This tax schedule raises R^b in revenue, and the value of the objective function under it is $\mathcal{U}^a + R^b - R^a$ because the equilibrium is the same as under T^a. By the optimality of T^b, $\mathcal{U}^a + R^b - R^a \leq \mathcal{U}^b$. We can make the same argument with a and b reversed to obtain $\mathcal{U}^b + R^a - R^b \leq \mathcal{U}^a$. It follows that $\mathcal{U}^a + R^b - R^a = \mathcal{U}^b$, so $T^a + R^b - R^a = T^b$ and $(T^a)' = (T^b)'$.

Calculation of Externality Ratios

We show the externality ratios solve the system of equations

$$e_j = \sum_{i=1}^{n} \frac{\partial \log E_i(Y_1^*, ..., Y_n^*)}{\partial Y_j} \left(a_i + \sum_{k=1}^{n} b_{i,k} e_k \right), \tag{B.1}$$

where a_i and $b_{i,k}$ are constants that depend on the equilibrium under consideration. Each a_i represents the direct effect of an increase in productivity in i on welfare. $b_{i,k}$ measure the changes to output in each k, which themselves cause externalities. All of these coefficients depend on both intensive and allocative margin labor-supply adjustments.

To derive these constants, consider the effect of increasing log productivity in i on hours, income, and utility. The first-order condition for each worker is $h_i^* = w_i^\sigma(1 - T'(y_i^*))^\sigma$, so $y_i^* = w_i^{1+\sigma}(1 - T'(y_i^*))^\sigma$. Differentiating this equation yields $dy_i^*/d\log w_i = (1+\sigma)(1 -$

183

$T'(y_i^*))y_i^*/(1-T'(y_i^*)+\sigma y_i^* T''(y_i^*))$. The change in the cost of effort is $\phi'(h_i^*)dh_i^*/d\log w_i = y_i^*(1-T'(y_i^*))d\log h_i^*/d\log w_i$. Solving for this derivative and substituting yields a total change in the effort cost of $\sigma y_i^*(1-T'(y_i^*))(1-T'(y_i^*)-y_i^* T''(y_i^*))/(1-T'(y_i^*)+\sigma y_i^* T''(y_i^*))$. Finally, the change in utility is simply $y_i^*(1-T'(y_i^*))$ from the envelope theorem.

The change in productivity induces switching on the allocative margin. Denote $\Theta_i = \{\theta \mid \{i\} \subset I^*(\theta)\}$ to be the set of workers in i (or indifferent) and denote $\partial\Theta_i = \{\theta \mid \{i\} \subsetneq I^*(\theta)\}$ to be the set of workers indifferent between i and another profession. For these latter type of workers, define $i'(\theta)$ to be a uniquely chosen element of $I^*(\theta)$ not equal to i. The productivity change induces a switch between i and $i'(\theta)$. Because the worker is indifferent to the post-tax utility of these professions, the change in the pre-tax utility is $T(y_i^*(\theta)) - T(y_{i'(\theta)}^*(\theta))$. The switch also causes an externality. Output rises in i by $y_i^*(\theta)$ and falls in $i'(\theta)$ by $y_{i'(\theta)}^*(\theta)$, leading to a change in social welfare of $e_i y_i^*(\theta) - e_{i'} y_{i'(\theta)}^*(\theta)$.

We can now calculate the constants. For ease of notation, we define f_i on Θ_i by $f_i(\theta) = y_i^*(\theta)(1-T'(y_i^*(\theta)))f(\theta)$. Then

$$a_i = \int_{\Theta_i} \frac{1+\sigma T'(y_i^*(\theta)) + \sigma y_i^*(\theta)T''(y_i^*(\theta))}{1-T'(y_i^*(\theta)) + \sigma y_i^*(\theta)T''(y_i^*(\theta))} f_i(\theta)d\theta + \int_{\partial\Theta_i}\left(T(y_i^*(\theta)) - T\left(y_{i'(\theta)}^*\right)\right)f_i(\theta)d\theta$$

$$b_{i,i} = \int_{\Theta_i} \frac{1+\sigma}{1-T'(y_i^*(\theta)) + \sigma y_i^*(\theta)T''(y_i^*(\theta))} f_i(\theta)d\theta + \int_{\partial\Theta_i} y_i^*(\theta)f_i(\theta)d\theta$$

$$b_{i,k} = -\int_{\Theta_i \cap \Theta_k} y_{i'(\theta)}^* f_i(\theta)d\theta,$$

where the last equation is defined for $k \neq i$.

Returning to (B.1), note it takes the form $e = a + Be$, where lowercase letters are n dimensional column vectors and the upper case B is an $n \times n$, not necessarily symmetric, matrix. This has solution $e = [I - B]^{-1}a$. Because B need not be symmetric, neither does $I - B$ need to be. Properties of $I - B$ likely play an important role in the existence, uniqueness, and stability in this model. A natural conjecture by analogy to classical general equilibrium theorem (Arrow and Hahn, 1971) is that a sufficient condition for at most a single equilibrium to exist, which is stable, is that $-[I - B]$ is globally stable (stable for every value of the vector Y) in the sense of Hicks (1939) that all the principal minors of

$I - B$ are positive. This condition, combined with some boundary conditions, likely ensures existence of such an equilibrium. This conjecture is consistent with our empirical findings that when externalities (and thus B) become too large, we cannot find an equilibrium, or multiple local steady states exist. Investigating these issues in general equilibrium theory at a general level with greater depth is beyond the scope of this paper, however.

B.1.2 Section 2.3

Optimal Top Tax in General Three-Profession Model

The following lemma gives the first-order condition τ^* must satisfy, which is the equation in Proposition 2.1 computed for the current example.

Lemma B.1. *In this example, the optimal top tax rate τ^* solves the equation*

$$0 = \frac{\sigma}{1 - \tau^*}(\tau^* + s_H e_H + s_L e_L) + \frac{2\beta \tilde{s}_H s_L (r^\alpha - 1)(1 - \tau^*)^\sigma}{\alpha(r+1)}(\tau^*(r-1) + re_H - e_L), \quad \text{(B.2)}$$

where $\tilde{s}_H$ is the share of skilled top earners that choose H, conditional on ability.

Proof. The equation follows from Proposition 2.1 in the limit of large y. The intensive part follows immediately. We show a constant limiting tax rate is optimal, which shows $T'' = 0$ at high income levels. For the allocative part, we must calculate $f_S(y)/f(y)\Delta_T(y)$ and $f_S(y)/f(y)\Delta_e(y)$ for large y.

We first solve for the profession shares for skilled workers. No externalities affect L or H, so the total productivity of a worker in either profession i is private productivity a_i. The solution to the optimization $\max_y y - T(y) - \phi(y/a_i(\theta))$ is $a_i(\theta)^{1+\sigma}(1-\tau)^{1+\sigma}/(1+\sigma)$. A skilled worker chooses H if and only if $\psi_H(\theta) - \psi_L(\theta) > -(1-\tau)^{1+\sigma}a_L(\theta)^{1+\sigma}(r-1)/(1+\sigma)$. The difference between two variables following Gumbel distributions with the same scale parameter is logistically distributed, so $\tilde{s}_L = F^{\mathcal{L}}(-2\beta(1-\tau)^{1+\sigma}(r-1)(1+\sigma)^{-1}(r+1)^{-1} - \Delta\overline{\psi})$, where $\Delta\overline{\psi} = \overline{\psi}_H - \overline{\psi}_L$ and $F^{\mathcal{L}}$ is the CDF of the standard logistic distribution. Note this share is independent of income.

Conditional on being skilled with productivity a_L, the CDF of ψ_i is $F_\psi(2\beta(1+r)^{-1}a_L^{-(1+\sigma)}\psi_i -$

$\overline{\psi}_i$). Therefore, the conditional CDF of $\Delta \psi = \psi_H - \psi_L$ equals $F^{\mathcal{L}}(2\beta(1+r)^{-1}a_L^{-(1+\sigma)}\Delta\psi - \Delta\overline{\psi})$. The PDF of $\Delta\psi$ equals $2\beta(1+r)^{-1}a_L^{-(1+\sigma)}f^{\mathcal{L}}$. A standard fact about the logistic distribution is that $f^{\mathcal{L}} = F^{\mathcal{L}}(1 - F^{\mathcal{L}})$. Therefore, the conditional measure of indifferent workers equals $2\beta(1+r)^{-1}a_L^{1+\sigma}\tilde{s}_H\tilde{s}_L$. Using the formula for income in the text, we simplify this expression to $2\beta(1+r)^{-1}(1-\tau)^{\sigma}y_L^{-1}\tilde{s}_H\tilde{s}_L$.

At income y, the share of workers in L is s_L. The measure of workers who are skilled and for whom $y_L^*(\theta) = y$ equals $s_L f(y)/\tilde{s}_L$. Therefore, the measure of such workers who are indifferent between H and L is $2\beta(1+r)^{-1}(1-\tau)^{\sigma}y^{-1}\tilde{s}_H s_L f(y)$. It follows that

$$\Delta_T(y) = \int_{y/r}^{y} \frac{\tau(r-1)y'}{y}\frac{2\beta(1-\tau)^{\sigma}\tilde{s}_H s_L f(y')}{(1+r)y'f(y)}dy' = \frac{\tau(r-1)2\beta(1-\tau)^{\sigma}\tilde{s}_H s_L (r^{\alpha}-1)}{\alpha(1+r)},$$

where we have used the fact that the distribution of income is Pareto with parameter α. This fact follows because the ability distributions are all Pareto with parameter $\alpha(1+\sigma)$ and log income equals $1+\sigma$ times log ability. Similarly,

$$\Delta_e(y) = \int_{y/r}^{y} \frac{(re_H - e_L)y'}{y}\frac{2\beta(1-\tau)^{\sigma}\tilde{s}_H s_L f(y')}{(1+r)y'f(y)}dy' = \frac{(re_H - e_L)2\beta(1-\tau)^{\sigma}\tilde{s}_H s_L (r^{\alpha}-1)}{\alpha(1+r)}.$$

Putting these equations together and factoring yields the desired result.

$\square$

As can be seen from Lemma B.1, the relative weight on τ_{int} scales with the labor-supply elasticity σ, whereas the relative weight on τ_{all} scales with the profession-switching sensitivity β. The larger β is, the more sensitive profession choices are to relative income and the greater the importance of the allocative margin in the optimal tax.

Note $\tau^{\star}$ is always in the interval between τ_{int} and τ_{all} because only in this interval will the two terms in (B.2) have opposite signs and thus only there can the equation be satisfied. Thus, $\tau^{\star}$ must be a convex combination of τ_{int} and τ_{all}, though no simple closed-form solution exists for the relevant weights.

B.1.3 Section 2.4

Identification of Externality Coefficients

First, we derive (2.6). Note $Y_i = H_i E_i(Y_1, ..., Y_n)$, where $H_i = \int_\Theta a_i(\theta) h_i(\theta) d\theta$. Given how it is defined, the partial derivative $\partial Y / \partial Y_j$ equals $(\partial Y / \partial H_j)/(\partial Y_j / \partial H_j)$, where these partial derivatives are calculated holding each H_i constant:

$$\frac{\partial Y_i}{\partial H_j} = 1_{i,j} E_j + H_i \sum_k \frac{\partial E_i}{\partial Y_k} \frac{\partial Y_k}{\partial H_j} = 1_{i,j} E_j + Y_i \sum_k \frac{\partial E_i / E_i}{\partial Y_k} \frac{\partial Y_k}{\partial H_j}.$$

Define the quasi-Jacobian matrix J by $J = \{(Y_i / E_i) \partial E_i / \partial Y_k\}_{i,k}$. Let $\partial \mathcal{Y} / \partial H_j$ be the column matrix whose ith entry equals $\partial Y_i / \partial H_j$. Then the above equation can be written in matrix form as

$$\frac{\partial \mathcal{Y}}{\partial H_j} = 1_j E_j + J \frac{\partial \mathcal{Y}}{\partial H_j},$$

where 1_j is the vector with a 1 in the jth spot and 0 otherwise. Therefore,

$$\frac{\partial \mathcal{Y}}{\partial H_j} = (I - J)^{-1} 1_j E_j \implies \frac{\partial Y}{\partial Y_j} = 1'(I - J)^{-1} 1_j,$$

where we have used the facts that $\partial Y_j / \partial H_j = E_j$ and $\partial Y / \partial Y_j = 1' \partial \mathcal{Y} / \partial Y_j$. Note that when externalities are absent, J is identically 0 so $\partial Y / \partial Y_j = 1$. Finally, directly taking the derivatives of E using our specification gives the equation for J in the text.

Proof of Lemma 2.2

We begin by proving statements made in the text before the lemma. Consider (2.7). Conditional on $i^*(\theta) = i$, the worker's maximization is $\max_{y_i} y_i - T(y_i) - \phi\left(y_i a_i^{-1} E_i(Y_0, ..., Y_n)^{-1}\right)$. The solution satisfies $y_i = (1 - T'(y_i))^\sigma (a_i E_i(Y_0, ..., Y_n))^{1+\sigma}$. By using $T = T_{2005}$ and solving for a_i, we immediately obtain (2.7).

Now we prove (2.9). The result of the maximization just described is $y_i - T(y_i) - (\sigma/(1 + \sigma))(1 - T'(y_i)) y_i$. Using this equation and (2.7), as well as the definition for relative utility in the text, we derive (2.9).

Next, we prove the distribution of a_{-i} conditional on a_i follows a Gaussian copula. By

187

definition, the $\Phi^{-1}(F_i^a(a_i))$ are jointly normal with mean 0 and covariance Σ. A standard result is that a multivariate normal conditioned on some of the variates is also multivariate normal, with mean and covariance given by formulas. Applying these formulas, we obtain that conditional on $\Phi^{-1}(F_i^a(a_i))$, the remaining $\Phi^{-1}(F_j^a(a_j))$ are multivariate normal with mean $\Phi^{-1}(F_i^a(a))\varrho$ and covariance $\Sigma_{-1} - \varrho'\varrho$.

We finally move on to the lemma itself. Consider a worker with productivity vector a. She chooses i to maximize $U_i^*(\theta)$, where θ restricted to productivity is a. This optimization is equivalent to maximizing $U_i^*(\theta) - \psi_i(\theta) + \psi_i(\theta) = n^{-1}\left(\sum_j a_j^{1+\sigma}\right)\widetilde{u}_i(a) + \psi_i(\theta)$, which is equivalent to maximizing $\widetilde{u}_i(a) + \psi_i(\theta)/\left(n^{-1}\sum_j a_j^{1+\sigma}\right)$. This latter term is distributed as $\beta^{-1}(\overline{\psi}_i + F_\psi)$, where F_ψ is a standard Gumbel distribution. If we let this Gumbel draw be $\widetilde{\psi}_i(\theta)$, the worker is choosing i to maximize $\beta\widetilde{u}_i(a) + \overline{\psi}_i + \widetilde{\psi}_i(\theta)$. A result from Gumbel distributions is that the probability that $A_i + B_i > A_j + B_j$ for all j when B_j are independent standard Gumbel distributions is $e^{A_i}/\sum_j e^{A_j}$. Applying this result, we conclude the share of workers with productivity a who choose i is

$$\Pr(i^*(\theta) = i \mid \theta|_a = a) = \frac{e^{\beta\widetilde{u}_i(a)+\overline{\psi}_i}}{\sum_j e^{\beta\widetilde{u}_j(a)+\overline{\psi}_j}}. \tag{B.3}$$

To prove (2.10), we compute in two different ways the share of all skilled workers such that the worker is in i at productivity a_i. First is the density of the i empirical productivity distribution $\widetilde{f}_i^a(a_i)$, times the share s_i of all workers in i, divided by the measure of skilled workers $1 - s_0$. This product gives the left side of (2.10). Alternatively, consider the probability that any worker would have productivity in i equal to a_i were she to choose i. This probability is $f_i^a(a_i)$. But only some of such workers choose i. To compute that conditional probability, we integrate over the conditional distribution of a_{-i}, using (B.3) as the probability of choosing i for each productivity profile. The result is the right side of (2.10).

B.2 Effort Specification of Externalities

Here we show that the baseline optimal taxes are not sensitive to whether externalities are specified as a function of output or effort. The effort specification is

$$\widetilde{E}_i(H_0, ..., H_n) = \prod_{j=1}^{n}(1 + \widetilde{\epsilon}_{i,j}H_j^{\gamma}),$$

where $H_j = \int_{\Theta} a_j(\theta)h_j(\theta)f(\theta)d\theta$. To match all the moments and distributions targeted in Section 2.4, it suffices to set $\widetilde{\epsilon}_{i,j} = \epsilon_{i,j}E_j(Y_0^*, ..., Y_n^*)^{\gamma}$ and use the same values for σ, β, ρ, and γ; here E_j is the externality function used in the paper and each argument Y_j^* is the empirical total income in j. This claim holds because $E(Y_0^*, ..., Y_n^*) = \widetilde{E}(H_0^*, ..., H_n^*)$ and because $\partial Y/\partial Y_j$ is the same under each specification at the estimation point. The first point is immediate from $Y_i = H_iE_i = H_i\widetilde{E}_i$. From Appendix B.1.3 ("Identification of Externality Coefficients"), the second point follows as long as $\partial E_i/\partial Y_j = \partial\widetilde{E}_i/\partial Y_j$. In the baseline specification, $\partial E_i/\partial Y_j = \gamma\epsilon_{i,j}(Y_j^*)^{\gamma-1}E_i(Y_0^*, ..., Y_n^*)[1 + \epsilon_{i,k}(Y_j^*)^{\gamma}]^{-1}$. In the effort specification, $\partial\widetilde{E}_i/\partial Y_j = \partial\widetilde{E}_i/\partial(\widetilde{E}_jH_j) = \widetilde{E}_j^{-1}\gamma\widetilde{\epsilon}_{i,j}(H_j^*)^{\gamma-1}\widetilde{E}_i[1 + \widetilde{\epsilon}_{i,j}(H_j^*)^{\gamma}]^{-1} = \gamma\epsilon_{i,j}(Y_j^*)^{\gamma-1}E_i(Y_0^*, ..., Y_n^*)[1 + \epsilon_{i,k}(Y_j^*)^{\gamma}]^{-1}$, where the last equality follows from the definition of $\widetilde{\epsilon}_{i,j}$. Thus the two partials coincide.

As a result, to compute the optimal tax under the effort specification, we keep the estimated parameters and distributions constant and calculate the optimal tax schedule when each worker's income is $y_i(\theta) = a_i(\theta)h_i(\theta)\widetilde{E}_i(H_0, ..., H_n)$. Appendix Table B.1 shows the results. Optimal rates and welfare are similar under this specification and the baseline.

B.3 Calibration Details

The Pareto parameter α is calibrated as follows. Using data from Bakija *et al.* (2012) described in Section 2.4, we use the fact that 16.97% of US income in 2005 went to those earning at least \$280,000, in order to calculate that the average income of such earners equals \$800,000. We calculate this average using the aggregate income and number of earners covered in the Bakija *et al.* (2012) data. In a Pareto distribution, the average value over a threshold

(within the support of the distribution) equals the value of that threshold times $\alpha/(\alpha - 1)$. Therefore, this fraction equals about 2.86 (the value we use does not involve intermediate rounding), leading to an α of 1.5.

We use engineering, research, and teaching to represent L, and finance and law to represent H. The 99th percentile incomes are the average of these statistics across these professions. Specifically, we use the average weighted by the share of each profession among earners in the top 1% of the total income distribution. These shares, given by Bakija *et al.* (2012), are 4.6% for engineering, 1.8% for research, 0.8% for teaching, 13.9% for finance, and 8.4% for law. The 99th percentile incomes are reported in Table 2.2.

Our simplified approximation to externality ratios is just the ratio of a profession's aggregate spillover to that profession's aggregate income. The reason this ratio is an approximation is because such externality estimates consider only income, whereas the true externality ratios should consider utility, which includes the cost of labor. Using the formulas for the externality ratios from the previous appendix, we calculate that in the case in which the tax is linear, using income externalities underestimates the true externality ratios by a factor of $1 + \sigma T'$. Because $\sigma = 0.24$ and the average tax rate in the United States is around 30%, we underestimate the externality ratios by only 7%, which is small enough to ignore for this illustrative example. Section 2.4 uses a more complex method that does not involve an approximation.

According to data we use from the BLS, 6.3% of the labor force in 2005 were in engineering, research, or teaching, and 1.3% were in finance or law. Therefore, we assume 7.6% of the total population is skilled and 17% of skilled workers choose H over L. They make this choice given r and the prevailing taxes in 2005, which we assume for simplicity are constant at a 30% marginal rate. The resulting preference parameters imply the share $\tilde{s}_H$ of highly skilled workers choosing H was 23.3%.[1] From $\tilde{s}_H$, we infer $\overline{\psi}_H - \overline{\psi}_L$, which allows us to recalculate the s_i as tax rates move around.[2]

---

[1] The measure of workers in H earning y equals $\tilde{s}_H/\tilde{s}_L$ times the measure of workers in L earning y/r. Thus, $s_H/s_L = r^\alpha \tilde{s}_H/\tilde{s}_L$. This equation allows us to obtain $\tilde{s}_H$ and $\tilde{s}_L$ from s_H and s_L, as the former two sum to 1.

[2] As shown in the proof of Lemma B.1, $\tilde{s}_L = F^{\mathcal{L}}(-2\beta(1-\tau)^{1+\sigma}(r-1)(1+\sigma)^{-1}(r+1)^{-1} - \overline{\psi}_H + \overline{\psi}_L)$, where

B.4 Estimation Details

B.4.1 Professional classifications

We map the IRS profession classifications in Bakija *et al.* (2012) to ours in the following manner. *Art* is "Arts, media, sports," *Engineering* is "Computer, math, engineering, technical (nonfinance)," *Finance* is "Financial professions, including management," *Law* is "Lawyers," *Management* is "Executive, non-finance, salaried" plus "Executive, non-finance, closely held business" plus "Manager, non-finance, salaried" plus "Manager, non-finance, closely held business," *Medicine* is "Medical," *Operations* is "Business operations (nonfinance)," *Real Estate* is "Real estate," *Research* is "Professors and scientists," and *Sales* is "Skilled sales (except finance or real estate)." Bakija *et al.* (2012) use a combined category "Government, teachers, social services." We apportion worker counts from this category to *Teaching* and *Other* using the ratio in the BLS data of teachers (SOCs below) to government workers (NAICS = 92). We subtract teachers in government from the count of government workers (this adjustment is de minimis). The remainder of *Other* is "Blue collar, miscellaneous service" plus "Unknown" plus "Farmers & ranchers" plus "Pilots" plus "Supervisor, non-finance, salaried" plus "Supervisor, non-finance, closely held business." We use Tables 2 and 3, "Percentage of primary taxpayers in the top one [resp. 0.1] percent of the distribution of income (excluding capital gains) that are in each profession."

In the BLS data, we aggregate SOCs into our professions using a classification similar to that in Bakija *et al.* (2012). This similarity justifies matching the BLS and IRS data. The exact list of SOCs we use to define each profession in the BLS is below:

Art: Art directors (27-1011), Craft artists (27-1012), Fine artists, including painters, sculptors, and illustrators (27-1013), Multi-media artists and animators (27-1014), Artists

---

$F^{\mathcal{L}}$ is the standard logistic distribution. We solve this equation using $\tilde{s}_L = 76.7\%$ and $\tau = 35\%$. To recalculate the s_i, pick an ability level and let s_U^* be the share of workers at that ability in either U or L who are in U. This share does not depend on ability or the tax rate at high ability levels. Then $s_U = s_U^* / (s_U^* + (1 - s_U^*)\tilde{s}_L + (1 - s_U^*)r^\alpha \tilde{s}_H)$. Using this equation, we calculate s_U^* at $\tau = 35\%$ and then use the updated $\tilde{s}_H$ and $\tilde{s}_L$ to update s_U at different tax levels, and then use $s_L = s_U(1 - s_U^*)\tilde{s}_L / s_U^*$ and $s_H = s_U(1 - s_U^*)r^\alpha \tilde{s}_H / s_U^*$.

and related workers, all other (27-1019), Commercial and industrial designers (27-1021), Fashion designers (27-1022), Floral designers (27-1023), Graphic designers (27-1024), Interior designers (27-1025), Merchandise displayers and window trimmers (27-1026), Set and exhibit designers (27-1027), Designers, all other (27-1029), Actors (27-2011), Producers and directors (27-2012), Athletes and sports competitors (27-2021), Dancers (27-2031), Choreographers (27-2032), Music directors and composers (27-2041), Musicians and singers (27-2042), Radio and television announcers (27-3011), Public address system and other announcers (27-3012), Broadcast news analysts (27-3021), Reporters and correspondents (27-3022), Public relations specialists (27-3031), Editors (27-3041), Technical writers (27-3042), Writers and authors (27-3043), Interpreters and translators (27-3091), Media and communication workers, all other (27-3099).

Engineering: Computer programmers (15-1021), Computer software engineers, applications (15-1031), Computer software engineers, systems software (15-1032), Aerospace engineers (17-2011), Agricultural engineers (17-2021), Biomedical engineers (17-2031), Chemical engineers (17-2041), Civil engineers (17-2051), Computer hardware engineers (17-2061), Electrical engineers (17-2071), Electronics engineers, except computer (17-2072), Environmental engineers (17-2081), Health and safety engineers, except mining safety engineers and inspectors (17-2111), Industrial engineers (17-2112), Marine engineers and naval architects (17-2121), Materials engineers (17-2131), Mechanical engineers (17-2141), Mining and geological engineers, including mining safety engineers (17-2151), Nuclear engineers (17-2161), Petroleum engineers (17-2171), Engineers, all other (17-2199).

Finance: Chief executives (11-1011) in Finance and Insurance (NAICS = 52), General and operations managers (11-1021) in Finance and Insurance (NAICS = 52), Financial managers (11-3031), Financial analysts (13-2051), Personal financial advisors (13-2052), Securities, commodities, and financial services sales agents (41-3031).

Law: Lawyers (23-1011), Administrative law judges, adjudicators, and hearing officers (23-1021), Arbitrators, mediators, and conciliators (23-1022), Judges, magistrate judges, and magistrates (23-1023).

Management: Chief executives (11-1011) outside Finance and Insurance (NAICS $\neq$ 52), General and operations managers (11-1021) outside Finance and Insurance (NAICS $\neq$ 52), Advertising and promotions managers (11-2011), Marketing managers (11-2021), Sales managers (11-2022), Public relations managers (11-2031), Administrative services managers (11-3011), Computer and information systems managers (11-3021), Compensation and benefits managers (11-3041), Training and development managers (11-3042), Human resources managers, all other (11-3049), Industrial production managers (11-3051), Purchasing managers (11-3061), Transportation, storage, and distribution managers (11-3071), Farm, ranch, and other agricultural managers (11-9011), Farmers and ranchers (11-9012), Construction managers (11-9021), Education administrators, preschool and child care center/program (11-9031), Education administrators, elementary and secondary school (11-9032), Education administrators, postsecondary (11-9033), Education administrators, all other (11-9039), Engineering managers (11-9041), Food service managers (11-9051), Funeral directors (11-9061), Gaming managers (11-9071), Lodging managers (11-9081), Medical and health services managers (11-9111), Natural sciences managers (11-9121), Social and community service managers (11-9151), Managers, all other (11-9199).

Medicine: Chiropractors (29-1011), Dentists, general (29-1021), Oral and maxillofacial surgeons (29-1022), Orthodontists (29-1023), Prosthodontists (29-1024), Dentists, all other specialists (29-1029), Anesthesiologists (29-1061), Family and general practitioners (29-1062), Internists, general (29-1063), Obstetricians and gynecologists (29-1064), Pediatricians, general (29-1065), Psychiatrists (29-1066), Surgeons (29-1067), Physicians and surgeons, all other (29-1069), Podiatrists (29-1081).

193

Operations: Agents and business managers of artists, performers, and athletes (13-1011), Purchasing agents and buyers, farm products (13-1021), Wholesale and retail buyers, except farm products (13-1022), Purchasing agents, except wholesale, retail, and farm products (13-1023), Claims adjusters, examiners, and investigators (13-1031), Insurance appraisers, auto damage (13-1032), Compliance officers, except agriculture, construction, health and safety, and transportation (13-1041), Cost estimators (13-1051), Emergency management specialists (13-1061), Employment, recruitment, and placement specialists (13-1071), Compensation, benefits, and job analysis specialists (13-1072), Training and development specialists (13-1073), Human resources, training, and labor relations specialists, all other (13-1079), Logisticians (13-1081), Management analysts (13-1111), Meeting and convention planners (13-1121), Business operations specialists, all other (13-1199).

Real Estate: Property, real estate, and community association managers (11-9141), Appraisers and assessors of real estate (13-2021), Real estate brokers (41-9021), Real estate sales agents (41-9022).

Research: Computer and information scientists, research (15-1011), Animal scientists (19-1011), Food scientists and technologists (19-1012), Soil and plant scientists (19-1013), Biochemists and biophysicists (19-1021), Microbiologists (19-1022), Zoologists and wildlife biologists (19-1023), Biological scientists, all other (19-1029), Conservation scientists (19-1031), Epidemiologists (19-1041), Medical scientists, except epidemiologists (19-1042), Life scientists, all other (19-1099), Astronomers (19-2011), Physicists (19-2012), Atmospheric and space scientists (19-2021), Chemists (19-2031), Materials scientists (19-2032), Environmental scientists and specialists, including health (19-2041), Geoscientists, except hydrologists and geographers (19-2042), Hydrologists (19-2043), Physical scientists, all other (19-2099), Economists (19-3011), Sociologists (19-3041), Urban and regional planners (19-3051), Anthropologists and archeologists (19-3091), Geographers (19-3092), Historians (19-3093), Political scientists (19-3094), Social scientists and related workers, all other (19-3099), Business teachers, postsecondary (25-1011), Computer science teachers, postsecondary (25-1021),

Mathematical science teachers, postsecondary (25-1022), Architecture teachers, postsecondary (25-1031), Engineering teachers, postsecondary (25-1032), Agricultural sciences teachers, postsecondary (25-1041), Biological science teachers, postsecondary (25-1042), Forestry and conservation science teachers, postsecondary (25-1043), Atmospheric, earth, marine, and space sciences teachers, postsecondary (25-1051), Chemistry teachers, postsecondary (25-1052), Environmental science teachers, postsecondary (25-1053), Physics teachers, postsecondary (25-1054), Anthropology and archeology teachers, postsecondary (25-1061), Area, ethnic, and cultural studies teachers, postsecondary (25-1062), Economics teachers, postsecondary (25-1063), Geography teachers, postsecondary (25-1064), Political science teachers, postsecondary (25-1065), Psychology teachers, postsecondary (25-1066), Sociology teachers, postsecondary (25-1067), Social sciences teachers, postsecondary, all other (25-1069), Health specialties teachers, postsecondary (25-1071), Nursing instructors and teachers, postsecondary (25-1072), Education teachers, postsecondary (25-1081), Library science teachers, postsecondary (25-1082), Criminal justice and law enforcement teachers, postsecondary (25-1111), Law teachers, postsecondary (25-1112), Social work teachers, postsecondary (25-1113), Art, drama, and music teachers, postsecondary (25-1121), Communications teachers, postsecondary (25-1122), English language and literature teachers, postsecondary (25-1123), Foreign language and literature teachers, postsecondary(25-1124), History teachers, postsecondary (25-1125), Philosophy and religion teachers, postsecondary (25-1126).

Sales: Advertising sales agents (41-3011), Insurance sales agents (41-3021), Sales representatives, services, all other (41-3099), Sales representatives, wholesale and manufacturing, technical and scientific products (41-4011), Sales representatives, wholesale and manufacturing, except technical and scientific products (41-4012), Sales engineers (41-9031), Sales and related workers, all other (41-9099).

Teaching: Preschool teachers, except special education (25-2011), Kindergarten teachers, except special education (25-2012), Elementary school teachers, except special education (25-2021), Middle school teachers, except special and vocational education (25-2022), Vocational

education teachers, middle school (25-2023), Secondary school teachers, except special and vocational education (25-2031), Vocational education teachers, secondary school (25-2032), Special education teachers, preschool, kindergarten, and elementary school (25-2041), Special education teachers, middle school (25-2042), Special education teachers, secondary school (25-2043).

Other: All SOCs not listed above.

B.4.2 Income by ability

Appendix Figure B.1 plots our estimated marginal distributions of pre-tax income for each quantile of the income distribution in our baseline estimation, where $\rho = 1$ so that individuals have a single dimension of ability. The figure therefore represents the pre-tax earnings from which an individual at the given quantile of the ability distribution could choose if she entered each of the professions.

The patterns are quite intuitive. At low levels of ability, stable professions, such as Engineering and Law, have the highest earnings, whereas "starving artists" are at the bottom. Toward the top end of the income distribution, finance, law and medicine are most lucrative, but even art does well given superstar effects. Teaching is at the bottom given the limited upside.

B.4.3 Calibration of profession-switching sensitivity

Let $s_{i,t}$ denote the share of the population in (other) finance at time t. We denote the share of workers flowing into finance by $s_{i,t}^f$. The stock $s_{i,t}$ and flow $s_{i,t}^f$ are related by the differential equation $\dot{s}_{i,t} = \delta(s_{i,t}^f - s_{i,t})$, where $\delta > 0$ is a replacement rate. Solving this equation from some reference time 0 yields

$$s_{i,t} = e^{-\delta t}s_{i,0} + \int_0^t \delta e^{-\delta(t-\tau)}s_{i,\tau}^f d\tau.$$

The share of the stock that is replaced in a year equals $1 - e^{-\delta}$, and we use this expression to calibrate δ. For instance, if $1/30$ of the stock is replaced annually because people work for 30 years, we choose δ such that $1/30 = 1 - e^{-\delta}$.

Some elasticity exists that expresses the flows into finance as a function of relative log wages. The specification we adopt is $s^f_{i,t} = b_0 + b_1 \log \widetilde{w}_{i,t}$, where $\widetilde{w}_{i,t}$ is the relative wage in finance. If relative wages are a random walk, a worker's best predictor of lifetime relative wages is the current value, and hence present relative wages alone guide labor flows. We want to compute b_1, which we will use as our moment to match.

To estimate b_1, we use Philippon and Reshef (2012)'s annual data on $\widetilde{w}_{i,t}$ and $s_{i,t}$. They show that in the period 1950-1980, these series were roughly flat. Both increased sharply after 1980, and the increases are close to linear in time. We therefore assume that in 1980, which we denote $t = 0$, finance employment was in equilibrium, so that $s^f_{i,0} = s_{i,0}$. It follows that $s^f_{i,t} = s_{i,0} + b_1(\log \widetilde{w}_{i,t} - \log \widetilde{w}_{i,0})$. We also assume that log relative wages increased linearly, so that $\log \widetilde{w}_{i,t} = \log \widetilde{w}_{i,0} + b_w t$. The data strongly bear out this linearity. When we regress $\log \widetilde{w}_{i,t}$ on time, the estimated coefficient for $b_w = 0.0478$ and has a t-stat of 22; the $R^2 = 95\%$. It follows that $s_{i,t} = s_{i,0} + b_1 b_w \int_0^t \delta e^{-\delta(t-\tau)} \tau d\tau$, so that

$$b_1 = \frac{\delta}{t\delta + e^{-\delta t} - 1} \frac{s_{i,t} - s_{i,0}}{b_w}.$$

The quantity $s_{i,t} - s_{i,0}$ equals the increase in the share of the population in finance between 1980 and 2005. Philippon and Reshef (2012) directly reports these numbers; the share increased from 0.35% to 0.87%, yielding $s_{i,t} - s_{i,0} = 0.52\%$. In the above formula, $t = 25$. Therefore, b_1 emerges as a function of our input for δ. Using the calibration method described earlier, we arrive at Appendix Table B.2.

B.4.4   Externality calculations

Engineering   We use Murphy *et al.* (1991)'s preferred estimates that are restricted to the 55 countries in which more than 10,000 students are in college. Their result is that a one percentage-point increase in the share of students studying engineering raises real per-capita

GDP growth by 0.054% points. According to the OECD, 10.7% of college students in the United States in 2005 studied engineering. The total externalities from engineering therefore amount to $(10.7)(0.054) = 0.6\%$ of GDP. In our context, we interpret the externalities from engineering as 0.6% of total income.[3]

Finance   We interpret the entirety of French (2008)'s estimates as negative externalities from finance. In 2005, he estimates these externalities at 0.63% of US stock market capitalization, or $90.7 billion. This externality amounts to -1.4% of the total income we calculate in Table 2.2, which is $6.3 trillion.

Law   Murphy *et al.* find a one percentage-point increase in the share of students in a country studying law lowers real per-capita GDP growth by 0.078% points. We again interpret this effect as a one-time change to the level of output. According to the OECD, 2.4% of students in the United States study law. Externalities from law therefore equal $-(2.4)(0.078) = -0.2\%$ of GDP.

Research   Jaffe's model allows university research to have a direct effect on commercial patents as well as an indirect effect through influencing industrial R&D. His preferred estimate is that the total elasticity of patents with respect to university research is 0.6. The direct elasticity of industrial R&D on patents is 0.94, and industrial R&D expenditures are six times larger than university research expenditures. Therefore, a dollar in university research is equivalent to $6(0.6)/0.94 = 3.83$ dollars spent on industrial R&D in terms of resulting patents. According to the National Science Foundation, $45 billion was spent on university R&D in 2005. Using the estimate from Jaffe, we conclude the total externality from this activity was $172 billion, which amounts to 2.7% of total income.

---

[3]The total income we measure in Table 2.2 is less than GDP because it excludes capital gains, transfers, and investment. We assume the engineering externality raises each component of GDP by the same proportion, so it raises the total income measure we focus on by 0.6%.

Sales   Informative and purely rational consumption theories of advertising imply advertising will tend to be undersupplied in most cases (Becker and Murphy, 1993), whereas persuasive theories suggest it will be oversupplied (Dixit and Norman, 1978). But although empirical efforts have sought to quantify the welfare effects of advertising in particular markets, such as pharmaceuticals (Rizzo, 1999) and subprime mortgages (Gurun *et al.*, Forthcoming), none attempts a comprehensive, profession-wide study, so we are hesitant to use these estimates.

Teaching   Card (1999) reviews the literature on the causal effect of education on earnings and finds results between a 0.05 and 0.15 log increase for each year of schooling. We use the midpoint of this interval, 0.1, which also equals the estimate of Angrist and Krueger (1991). The annual social product of teaching therefore equals $e^{0.1} - 1 = 10.5\%$ of total income. According to our estimates of profession-specific income distributions in Table 2.2, the private product of teachers equals 3.2% of total income. The total externalities of teachers amount to the difference, which is 7.3%.

Chetty *et al.* (2014) calculate that a one-standard-deviation increase in teacher quality raises eventual student earnings by 1.34%. They also calculate that the present value of future earnings for a middle-school student is $468,000 in 2005 dollars.[4] In our model, all variations in teacher pay come from differences in quality $a_i(\theta)$. The standard deviation of teacher pay in our data equals $27,000. The total pay equals $203 billion. According to the Digest of Education Statistics published by the National Center for Education Statistics, the number of students in all elementary and secondary schools in the United States in 2005 was 54 million.[5] Our data contain 4.2 million teachers. The total social product of teachers is

$$(\$203 \text{ billion}/\$27,000)(1.34\% * \$468,000)(54 \text{ million}/4.2 \text{ million}) = \$606 \text{ billion}.$$

---

[4]They use a 5% discount rate and assume earnings grow 2% annually.

[5]The NCES reports enrollments of 53.4 million in 2000 and 54.9 million in 2010, which we average. Data accessed at `http://nces.ed.gov/programs/digest/d13/tables/dt13_105.20.asp`.

This social product equals 9.6% of the $6.3 trillion of total income in the economy. This number is slightly less than the 10.5% figure we calculated using the social returns to education, but it is very close.

B.4.5 Alternative local optimum for marginal tax rates

As discussed in the text, a second local optimum exists for marginal tax rates under the baseline parameters in which welfare is slightly higher ($22 per person). We choose not to focus on it because it is likely an artifact of the way the brackets are constructed. It is present on only the smallest bracket (in log terms), and it disappears when we change the $150k-$200k bracket to $150k-$250k. Appendix Figure **??** reproduces Figure 2.2 for the secondary optimum and also for the case in which the bracket is enlarged.
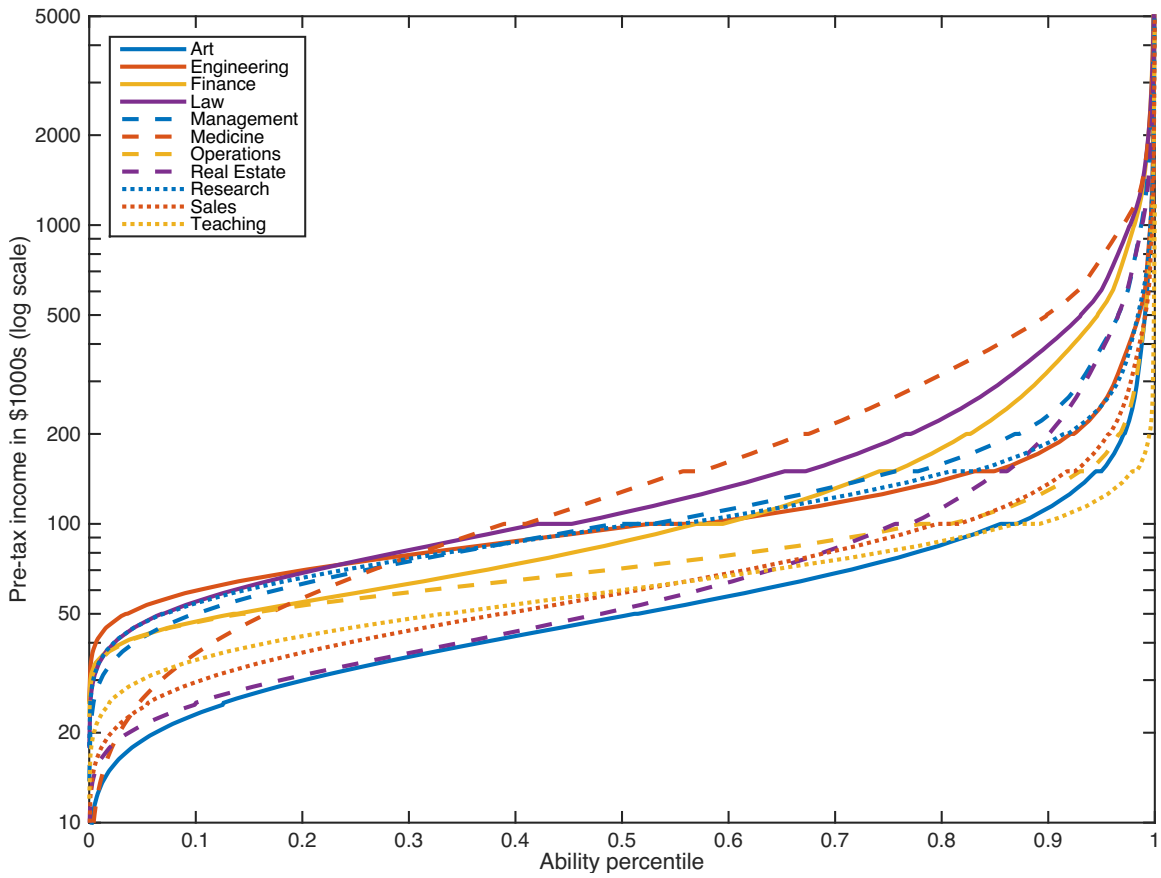
Figure B.1: *This figure plots, for each profession, the income earned by a worker in that profession whose ability a_i is at each percentile of the underlying distribution of ability across all workers in the economy. We compute the realized income at equilibrium under the optimal income tax given in Figure 2.2.*
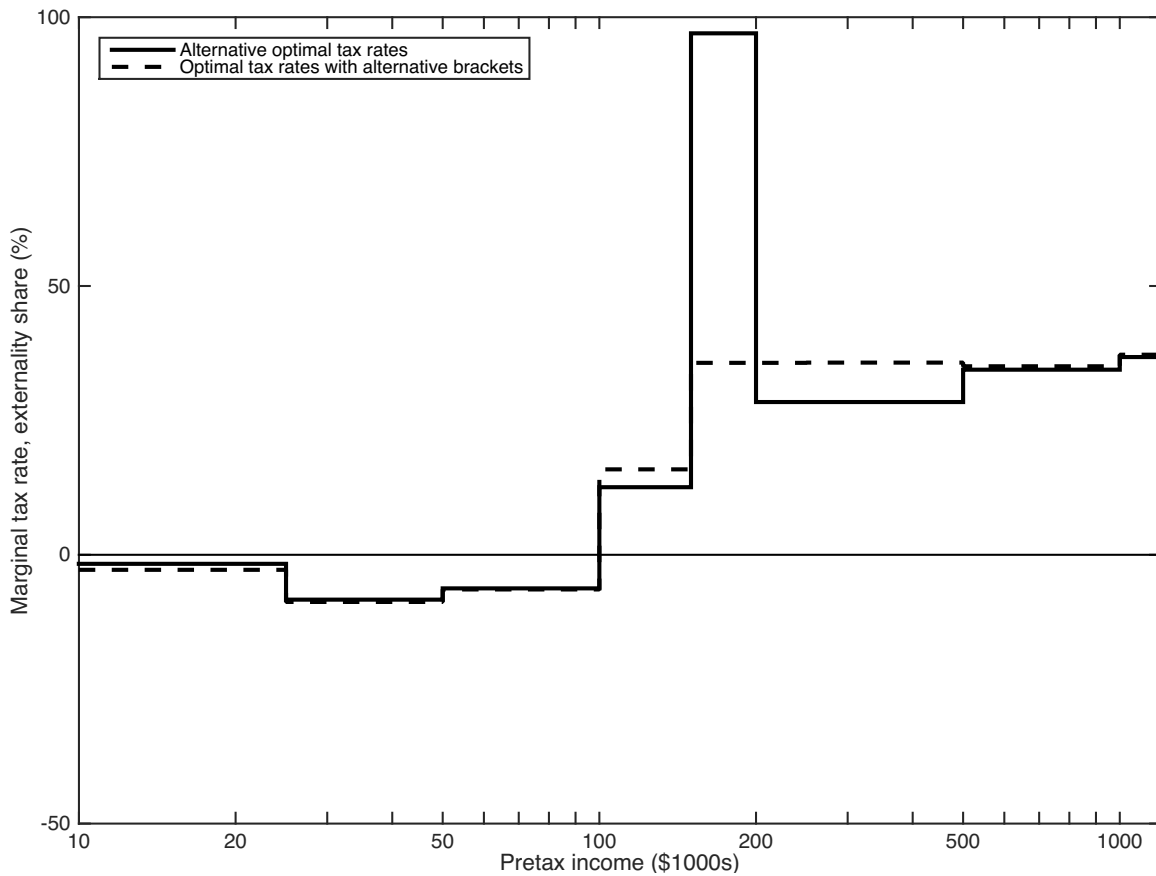
Figure B.2: *This figure displays an alternative local maximum for the marginal tax schedule using the same baseline assumptions and parameters behind Figure 2.2. The dashed line displays the only optimum we find when the $150k-$200k bracket is changed to $150k-$250k. While other local optima cannot be ruled out with certainty, we confirm that using the vector of optimal tax rates from the solid line as the seed value for the optimization when computing the dotted line does alter it.*

Table B.1: *Optimal Tax Rates and Welfare Gains Relative to* Laissez-Faire

	Tax Rate Bracket								
	$0-$25k	$25k-$50k	$50k-$100k	$100k-$150k	$150k-$200k	$200k-$500k	$500k-$1m	$1m+	Welfare Gain
Baseline	−**2.8%**	−**8.8%**	−**6.4%**	**16.0%**	**32.6%**	**37.2%**	**34.9%**	**37.2%**	**1.2%**
Effort specification	−2.3%	−7.3%	−5.4%	14.0%	28.5%	31.6%	29.2%	31.5%	1.2%

Table B.2: *Estimated Sensitivity of Flows into Finance with Respect to Relative Wages*

Assumed Tenure (Years)	Implied δ	$\widehat{b}_1$
0	∞	0.0044
1	0.63	0.0045
5	0.18	0.0054
10	0.095	0.0069
20	0.049	0.0101
30	0.033	0.0135
40	0.025	0.0170
50	0.020	0.0204
100	0.010	0.0378

Appendix C

Appendix to Chapter 3

C.1 Dynamic model with addiction

Here we show how our model can capture basic dynamic considerations. Suppose that consumers live for K periods $k = 1, \dots K$. In each period, a consumer chooses income z_k and pays a tax $T(z^k)$. After-tax income is thus $z^k - T(z^k)$. Note that here, we are assuming (realistically) that the income tax T cannot be age-dependent. Similarly, we are assuming that the commodity tax t cannot be age-dependent. The consumer chooses a consumption bundle (c_1^k, c_2^k) subject to the budget constraint $c_1^k + (p + t)c_2^k \leq z^k$. Note that by assumption, we abstract here from saving and borrowing, as well as time-dependent earnings ability.

The consumer's true, period k flow utility is given by $V(c_1, c_2, S, z, \theta)$, where S is the stock of c_2 consumption that evolves according to $S^{k+1} = (1 - d)(c_2^k + S^k)$.

Let $\mathcal{U}^k(c_1^k, c_2^k, S^k, z^k, \theta)$ be the utility function that the consumer maximizes in period k. Note that this taxes into account the consumer's dynamic considerations. The $\mathcal{U}$ could be the objective function maximized by either a sophisticated or naive present-biased consumer, for example.

The planner maximizes $\int G \left(\sum_{k=1}^{K} \delta^{k-1} V^k(c_1^k(\theta), c_2^k(\theta), S^k(\theta), z^k(\theta), \theta) \right) dF(\theta)$, where V^k is the period k flow utility and $\mu(\theta)$ is the welfare weight on type θ in period k. In each period k let $\mathcal{V}^k(c_1^k, c_2^k, S^k, z^k, \theta)$ denote the true utility that results from a choice of (c_1^k, c_2^k, z^k)

in period k, taking into account the how the consumer will choose in periods $k+1, \ldots K$. As before, set $\gamma(c_2, S, \theta) = \mathcal{V}_2^k / \mathcal{V}_1^k - \mathcal{U}_2^k / \mathcal{U}_1^k$.

Then calculations analogous to those in Section 3.2 show that

$$t = \bar{\gamma}(t) \left[\bar{g} + \frac{Cov[g(\vartheta), \omega(\vartheta)c_2(\vartheta)]}{C_2} \right] - (p+t) \frac{Cov[c_2(\vartheta), g(\vartheta)] + C_2(\bar{g} - 1) + \int z(\vartheta)\chi(\vartheta)T'(z(\vartheta))d\mathcal{F}}{|\bar{\zeta}|C_2}$$

where $\vartheta = (k, \theta)$ is the age-dependent type encoding both the consumer's age and intrinsic type θ, $\mathcal{F}$ is the distribution of ϑ, where $g(\vartheta) := G'\mathcal{V}_1^k$, and where $\bar{\gamma}$ is the average marginal bias with respect to age-dependent types. This shows that all of the core economic concepts from the static framework carry over to dynamic case.

C.2 Proofs of Propositions (sketches)

Lemma C.1. *Let $y(\theta)$ and $z(\theta)$ denote a type θ's post-tax and pre-tax earnings at the optimal tax system. Then $z(\theta) - z(\underline{\theta}) > y(\theta) - y(\underline{\theta}) > (\log(\theta) - \log(\underline{\theta}))\psi'(0)$.*

Proof. The first order condition that $y(\theta)$ must satisfy is $y'(\theta) = \frac{1}{\theta}\psi'(z(\theta)/\theta)$. Thus

$$
\begin{aligned}
y(\theta^*) - y(\underline{\theta}) &= \int \frac{1}{\theta}\psi'(z(\theta)/\theta) \\
&\geq \int \frac{1}{\theta}\psi'(0) = (\log(\theta^*) - \log(\underline{\theta})\psi'(0)
\end{aligned}
$$

$\square$

Proof of Proposition 3.2

Proof. Part 1: We have that

$$
\begin{aligned}
\frac{dW}{dt} &= \int \left(-g(z)c_2(z, t) - \gamma(z)g(z)\frac{dc_2(z, t)}{dt} + t\frac{dc_2(z, t)}{dt} + c_2(z, t) \right) dH(z) \\
&= \int \left(-g(z)c_2(z, t) + \gamma(z)g(z)\frac{|\zeta|c_2}{p+t} - t\frac{|\zeta|c_2}{p+t} + c_2(z, t) \right) dH(z) \qquad \text{(C.1)}
\end{aligned}
$$

By assumption, there is a finite A such that $\zeta(c_2, \theta) / \zeta(c_2, \theta) < A$ for all c_2, θ.

Thus by assumption, $\int g(z) |\zeta(z)| c_2(z) dH(z) \leq A \int |\zeta(z)| c_2(z) dH(z)$. Thus if $\gamma(z, c_2) \leq \gamma^\dagger$, then the optimal commodity tax must satisfy $t < A\gamma^\dagger$.

Now let $W(t, \gamma)$ denote the highest possible welfare, conditional on a bias function γ and a commodity tax t. Now by assumption, the derivative $W_t(0, 0) < 0$. Now by continuity, for any interval $[0, a]$ and any $\epsilon > 0$ there exists a $\gamma^\dagger$ such that $|W(t, \gamma) - W(t, 0)| < \epsilon$ for all $t \in [0, a]$ and all γ such that $\gamma(z, c_2) < \gamma^\dagger \; \forall z, c_2$. This implies that there exists a $\gamma^\dagger$ small enough such that $W_t(t, \gamma) < 0$ for all $t \in [0, A\gamma^\dagger]$ and all γ bounded above by $\gamma^\dagger$. But since the optimal commodity tax satisfies $t < A\gamma^\dagger$, it thus follows that $t^G < 0$ for $\gamma^\dagger$ small enough.

Part 2: Now if $|\bar{\zeta}(z, t)|$ and $\omega(z, t)$ are bounded away from zero, then there exists a high enough $\gamma^\dagger$ such that $|\zeta| c_2 (\gamma^\dagger / (p + t)) > 1$ for all z and $t \leq 0$. By (C.1), it thus follows that when $\gamma(c_2, \theta) \geq \gamma^\dagger$, we must have $\frac{dW}{dt} > 0$ for all $t \leq 0$.

Part 3: Now note that for $\gamma(\theta) \geq \gamma^\dagger > t'$, we have that

$$
\begin{aligned}
\frac{dW}{dt} &= \int \left(-g(z) c_2(z, t) + (\gamma(z) - t) g(z) \frac{|\zeta| c_2(z, t)}{p + t} + t g(z) \frac{|\zeta| c_2(z, t)}{p + t} - t \frac{|\zeta| c_2(z, t)}{p + t} + c_2(z, t) \right) dH(z) \\
&= \int \left(c_2(z, t) - g(z) c_2(z, t) + (\gamma(z) - t) g(z) \frac{|\zeta| c_2(z, t)}{p + t} \right) dH(z) + \frac{t}{p + t} Cov[g(z), |\zeta(z)| c_2] \quad \text{(C.2)}
\end{aligned}
$$

Under these same conditions, there exists a large enough $\gamma^\dagger$ such that $(\gamma^\dagger - t) \frac{|\zeta|}{p+t} > 1$ for all $t \leq t'$ and all z. Thus $\frac{dW}{dt} > 0$ for all $t \leq t'$ as long as $Cov[g(z), |\zeta(z)| c_2(z, t)] \geq 0$.

Part 4: Suppose, for the sake of contradiction, that $t \leq \gamma^\dagger$. Using the computations from part 3, we have that

$$
\frac{dW}{dt} = \frac{t|\bar{\zeta}|}{p + t} Cov[g(z), \omega(z) c_2(z, t)] - Cov[g(z), c_2(z, t)] + \int (\gamma^\dagger - t) g(z) \frac{|\zeta| c_2(z, t)}{p + t} dH(z).
$$
(C.3)

Now at $t = \gamma^\dagger$, we have that $\frac{dW}{dt} = \frac{t|\bar{\zeta}|}{p+t} Cov[g(z), \omega(z) c_2(z, t)] - Cov[g(z), c_2(z, t)]$. Note, however, that because $c_2(z, t) \to 0$, while $\frac{dc_2(\theta, t)}{dt}$ is bounded away from zero for $t < m_1(0, \underline{\theta})$,

it follows that $|\bar{\zeta}| \to \infty$ on the interval $(0, m_1(0, \underline{\theta}))$ as $t \to m_1(0, \underline{\theta})$. It thus follows that for $\gamma^\dagger$ sufficiently close to $m_1(0, \underline{\theta}), \frac{dW}{dt} > 0$ at $t = \gamma^\dagger$. By quasiconcavity, this implies that we cannot have $t^G \leq \gamma^\dagger$, a contradiction.

Part 5: Obvious.

$\square$

Proof of Proposition 3.3

Proof. Part 1. Analagous to the proof of part 1 of Proposition 3.2.

Part 2. Let $A > a > 0$ be such that $A > |\zeta(c_2, \theta)| > a$ for all c_2, θ. Now clearly, the optimal commodity tax cannot be greater than t^M. Then from (C.1), we have that for $t \in [t', t^M]$ and for $\gamma \leq \gamma^\dagger < 0$

$$
\begin{aligned}
\frac{dW}{dt} &< \int \left(\gamma^\dagger g(z) \frac{a c_2(t^M, \underline{\theta})}{p + t} - t' \frac{A c_2(z, t)}{p + t'} + c_2(z, t) \right) dH(z) \\
&< \frac{\gamma^\dagger c_2(t^M, \underline{\theta})}{p + t^M} - \frac{t'}{p + t'} A C_2(t') + C_2(t')
\end{aligned}
$$

Thus for $\gamma^\dagger$ sufficiently negative, it follows that $\frac{dW}{dt} < 0$ for all $t \in [t', t^M]$.

Part 3. Obvious.

$\square$

Proof of Proposition 3.4

Proof. Take $\bar{e}$ and $\underline{e}$ such that $e(\theta, t) \in (\underline{e}, \bar{e})$ for all θ, t. Take $\gamma_{max} = max_{\theta, t} \gamma(\theta, c_2(\theta, t))$. Then from (C.1) it follows that

$$
\frac{dW}{dt} < \int \left(\gamma_{max} g(z) |\bar{\zeta}(t)| \frac{\bar{e} c_2(z, t)}{p + t} - |\bar{\zeta}(t)| t \frac{\underline{e} c_2(z, t)}{p + t} \right)
$$

This shows that regardless of what value $|\bar{\zeta}(t)|$ takes on, we must have $t^G < \frac{\gamma_{max} \bar{e}}{\underline{e}}$. Let $\bar{t}^*$ denote this upper bound. Now let $g_0(z)$ denote the social marginal welfare weights when the commodity tax is set to $t = 0$ and the optimal income tax is set optimally. By conditions

(i) and (ii) of the proposition, the function g_0 is constant over all $U \in \mathbb{U}$. Now because $Cov[g_0(z), c_2(z,0)] < 0$ by assumption, there exists a low enough k such that

$$\int \left(-g_0(z)c_2(z,0) + |\bar{\zeta}(t)|\gamma(z)g_0(z)\frac{e(z,0)c_2(z,0)}{p+t} - t|\bar{\zeta}(t)|\frac{e(z,0)c_2(z,0)}{p+t} + c_2(z,0) \right) dH(z) < 0$$

if $|\bar{\zeta}(t)| < k$ for all $t \in [t', \bar{t}^*]$. At the same time, we have that $\max_{z, t \leq \bar{t}^*} \{|c_2(z,t) - c_2(z,0)|\} \to 0$ as $|\bar{\zeta}(t)| \to 0$. To see this, note that

$$c_2(z,t) - c_2(z,0) = \int \frac{|\zeta(z,t)|c_2(z,t)}{p+t} \leq c_2(z,0) \int \frac{|\zeta(z,t)|}{p+t} dH(z) \leq k\bar{e} \int \frac{1}{p+t} dH(z)$$

and thus $c_2(z,t) - c_2(z,0)$ converges uniformly to 0 as $k \to 0$. From this, it follows that for each $\epsilon > 0$ there exists a small enough k such that

$$\left| \int \left(-g_0(z)c_2(z,0) + |\bar{\zeta}(t)|\gamma(z)g_0(z)\frac{e(z,0)c_2(z,0)}{p+t} - t|\bar{\zeta}(t)|\frac{e(z,0)c_2(z,0)}{p+t} + c_2(z,0) \right) dH(z) \right|$$
$$\leq \left| \int \left(-g(z)c_2(z,t) + |\bar{\zeta}(t)|\gamma(z)g(z)\frac{e(z,t)c_2(z,t)}{p+t} - t|\bar{\zeta}(t)|\frac{e(z,t)c_2(z,t)}{p+t} + c_2(z,t) \right) dH(z) \right| + \epsilon$$

if $|\bar{\zeta}(t)| < k$ for all $t \in [t', \bar{t}^*]$. But now for ϵ small enough, it follows that

$$\int \left(-g(z)c_2(z,t) + |\bar{\zeta}(t)|\gamma(z)g(z)\frac{e(z,t)c_2(z,t)}{p+t} - t|\bar{\zeta}(t)|\frac{e(z,t)c_2(z,t)}{p+t} + c_2(z,t) \right) dH(z) < 0$$

for all $t \in [t', \bar{t}^*]$, from which the statement of the proposition follows.

$\square$

Proof of Proposition 3.5

Proof. The proof is identical to the proof of Proposition 3.4. First, for a bounded bias function there exists some $\underline{t}^*$ such that for any ζ, the optimal tax is $t^G \geq \underline{t}^*$. Second, almost identical reasoning shows that there exists a small enough k such that $\frac{dW}{dt} > 0$ if $|\bar{\zeta}(t)| < k$ for all $t \in [\underline{t}*, 0]$.

□

Proof of Proposition 3.6. Part 1. Suppose, for the sake of contradiction, that the optimal policy sets $t_\phi^* > 0$. Let T^* denote the corresponding income tax, let $c_2^*(z)$ denote consumption of $c_2$by a consumer earning z at this tax policy, and let C_2^* denote the total consumption of c_2 that occurs at this tax policy. Now consider instead a policy that sets a fully salient commodity tax $t = \phi t_\phi^*$, and sets an income tax T given by $T(z) = T^*(z) - (1 - \phi)t_\phi^* C_2^*$. By construction, a type θ consumer chooses the same level of both z and c_2 under these two tax policies. Thus the only difference is to each consumer's consumption of c_1. Compared to the first policy, the change in c_1 consumption of a consumer earning z is given by

$$\Delta(z) := \underbrace{[-\phi t_\phi^* c_2^*(z) - (-t_\phi^* c_2^*(z))]}_{\text{Change in tax paid for} c_2} - \underbrace{(1 - \phi)t_\phi^* C_2^*}_{\text{Change in income tax}} = (1 - \phi)t_\phi^*[c_2^*(z) - C_2^*]$$

Now by construction, $\int \Delta(z)dH(z) = 0$, but $\Delta(z)$ is also decreasing in z. Thus this change in policy is preferable because it distributes more resources to the lower income earners.

Part 2. Suppose, for the sake of contradiction, that the optimal policy sets $t^* < 0$. Let T^* denote the corresponding income tax, let $c_2^*(z)$ denote consumption of c_2 by a consumer earning z at this tax policy, and let C_2^*denote the total consumption of c_2 that occurs at this tax policy. Now consider instead a policy that sets the less salient commodity tax $t_\phi = t^*/\phi$, and sets an income tax T given by $T(z) = T^*(z) - (1 - 1/\phi)t^* C_2^*$. By construction, a type θ consumer chooses the same level of both z and c_2 under these two tax policies. Thus the only difference is to each consumer's consumption of c_1. Compared to the first policy, the change in c_1 consumption of a consumer earning z is given by

$$\Delta(z) := \underbrace{[-t^* c_2^*(z)/\phi - (-t^* c_2^*(z))]}_{\text{Change in tax paid for} c_2} - \underbrace{(1 - 1/\phi)t^* C_2^*}_{\text{Change in income tax}} = (1 - 1/\phi)t^*[c_2^*(z) - C_2^*]$$

Now by construction $\int \Delta(z)dH(z) = 0$, but $\Delta(z)$ is also decreasing in z. Thus this change

in policy is preferable because it distributes more resources to the lower income earners.

$\square$

Proof of Proposition 3.8

Proof. Part 1. This is clear from equation (3.15), which shows that $\frac{dW}{ds} - \frac{dW}{dt} < 0$ for any $s > 0$.

Part 2. Suppose, for the sake of contradiction, that the optimal policy sets $s^G = 0$ and sets a commodity tax $\hat{t}$. Now by assumption, $Cov[g(z), c_2(z)] > 0$ at $\hat{t}$. Thus by (3.15), $\frac{dW}{ds} - \frac{dW}{dt} > 0$ at this tax policy for low enough n and κ, from which it follows that $s^G = 0, t^G = \hat{t}$ cannot be optimal, thus generating a contradiction.

Part 3. As in the proof of Proposition 3.4, the boundedness assumption on γ implies that there is a $\bar{t}^*$ such that $t^G \le \bar{t}^*$. Now set $\hat{\sigma}(t) = Cov[g(z,t), c_2(z)]$, where $g(z,t)$ and $c_2(z,t)$ are defined as follows: given commodity tax t, let T_t denote the optimal income tax, and let $g(z,t)$ and $c_2(z,t)$ denote the corresponding social marginal welfare weights and consumption choices. As we will show below, $\hat{\sigma}(t)$ is bounded away from zero on the interval $[0, \bar{t}^*]$. Call the lower bound $\underline{\sigma}$. Next, because $\int g(z)c_2(z)dH = Cov[(g(z), c_2(z)] + C_2(z)$, we have that for all $t \in [0, \bar{t}^*]$,

$$
\begin{aligned}
\frac{dW}{ds} - \frac{dW}{dt} &= (1-\eta)\hat{\sigma}(t) - \eta C_2(z,t) - \kappa \\
&= (1-\eta)\hat{\sigma}(t) - \eta C_2(z,0) - \kappa
\end{aligned}
$$

is positive for η and κ small enough. Thus there exist small enough η and κ such that that $\frac{dW}{ds} - \frac{dW}{dt} > 0$ for all potential optimal taxes $t \in [0, \bar{t}^*]$.

We now show that $\hat{\sigma}(t)$ is bounded from below on the interval. To see this, notice that $c_2(z,t)$ is increasing in z and does not depend on the optimal income T. For a given t, $\hat{\sigma}$ is minimized when $g(z,t)$ is decreasing as little as possible in z. Now since $g(z,t) = G'(z - T(z) - (p+t)c_2(z,t) - \gamma(z,t))$, the assumptions in the proposition imply that

$$g_z(z,t) \leq (1 - T' - (p+t)\frac{d}{dz}(c_2 + \gamma))G''(z - T(z) - (p+t)c_2 - \gamma) \leq$$

$$- (p+t)\frac{d}{dz}(c_2)G''(z - T(z) - (p+t)c_2 - \gamma).$$

Now define $a < 0$ such that $G''(x) < a$ for all x. Next, take some $\theta^* > \underline{\theta}$, and let $b = \max_{z' \leq z(\theta^*,t)} \left\{ \frac{dc_2(z,t)}{dz}|_{z=z'} \right\}$. Note that b is well-defined because $\frac{dc_2(z,t)}{dz}|_{z=z'}$ is continuous in z' and t. It thus follows that $g_z(z,t) \leq -(p+t)ab$ when $\theta \leq \theta^*$ and $t \in [0, \bar{t}^*]$. We now have that

$$
\begin{aligned}
\hat{\sigma}(t) &\leq F(\theta^*)Cov[g(z(\theta),t), c_2(z(\theta),t)|\theta \leq \theta^*] \\
&\leq F(\theta^*)Cov[g(z(\underline{\theta}),t) - (p+t)abz(\theta), c_2(z(\underline{\theta})) + bc_2z(\theta)] \\
&= F(\theta^*)(p+t)|a|b^2 Var[z(\theta)|\theta \leq \theta^*]
\end{aligned}
$$

Now from lemma C.1 it follows that $Var[z(\theta)|\theta \leq \theta^*]$ is bounded away from zero for all $t \in [0, \bar{t}^*]$, from which it follows that $\hat{\sigma}(t)$ is bounded away from zero on the interval.

$\square$