Optimal Income Taxation:
Mirrlees Meets Ramsey[*]

Jonathan Heathcote[†]

Hitoshi Tsujiyama[‡]

Federal Reserve Bank of Minneapolis and CEPR

Goethe University Frankfurt

August 27, 2019

Abstract

What is the optimal shape of the income tax schedule? We address this question in an environment featuring distinct roles for public and private insurance. In our baseline calibration to the United States, optimal marginal tax rates increase in income and can be well approximated by a simple two-parameter function. The shape of the optimal schedule is sensitive to the amount of fiscal pressure to raise revenue that the government faces. As fiscal pressure increases, the optimal schedule becomes first flatter and then U-shaped, reconciling various findings in the literature.

Keywords: Optimal income taxation; Mirrlees taxation; Ramsey taxation; Tax progressivity; Flat tax; Private insurance; Social welfare functions

1 Introduction

We revisit a classic and important question in public finance: what structure of income taxation maximizes the social benefits of redistribution while minimizing the social harm associated with distorting the allocation of labor input? We focus on the Mirrleesian approach (Mirrlees 1971), which seeks to characterize the optimal tax system subject only to the constraint that taxes must be a function of individual earnings. Taxes cannot be explicitly conditioned on individual productivity or individual labor input because these are assumed to be unobserved by the tax authority. The Mirrleesian approach is attractive because it places no constraints on the shape of the tax schedule and because the implied allocations are constrained efficient.

Following this approach, Mirrlees (1971) found the optimal tax schedule to be close to linear in his numerical exercises, a finding mirrored more recently by Mankiw et al. (2009). In contrast, starting from the influential papers of Diamond (1998) and Saez (2001), most recent quantitative papers have argued that marginal tax rates should be U-shaped, with higher rates at low and high incomes compared with the middle of the income distribution.

We consider a model environment similar to the ones in these existing papers. Agents differ with respect to productivity, and the government chooses an income tax system to redistribute and finance exogenous government purchases. One innovation relative to most of the existing literature is that we allow for partial private insurance. In particular, we assume that idiosyncratic labor productivity has two orthogonal components: $\log(w) = \alpha + \varepsilon$. The first component α cannot be privately insured and is unobservable by the planner—the standard Mirrlees assumptions. The second component ε can be perfectly privately insured. For the purposes of providing concrete practical advice on tax system design, it is important to appropriately specify the relative roles of public and private insurance. When agents can insure more risks privately, the government has a smaller role, and the optimal tax schedule is less redistributive.

In our baseline model calibration, we use cross-sectional evidence on income and consumption inequality to discipline the relative magnitudes of uninsurable and insurable wage risk, and the shapes of the corresponding distributions. We then solve for the optimal allocation numerically. We find that the tax and transfer system chosen by a utilitarian planner features marginal tax rates that are increasing across the entire income distribution, a finding that contrasts with the existing literature.

The second contribution of the paper is to develop new intuition for what determines the shape of the optimal tax schedule, which we use to better understand the disparate results in the literature. We emphasize the idea that the shape of the optimal tax schedule is sensitive to the amount of fiscal pressure to raise revenue that the government faces. When fiscal pressure is relatively low—for example, because required government expenditure is low—the optimal tax schedule is upward sloping. When fiscal pressure is sufficiently increased, the optimal schedule becomes first flatter and then U-shaped, as in Saez (2001).

At a general level, the optimal tax schedule trades off equity and efficiency considerations. An increasing marginal rate profile is attractive from an equity standpoint, since a progressive marginal tax schedule redistributes the tax burden upward within the income distribution. A decreasing marginal rate profile is attractive from an efficiency standpoint, since a regressive profile lowers marginal tax rates on average and thus reduces distortions to households' labor supply choices.

To better understand the equity versus efficiency trade-off, we formalize measures of the marginal distributional gains and efficiency costs associated with changing marginal tax rates at different points along the income distribution. These gains and costs are equated at all income levels at the optimum.

Distributional gains from raising marginal rates are always high at the top of the income distribution, so the planner always sets high marginal rates there, tolerating the high associated efficiency costs. At the bottom of the income distribution, the size of distributional gains—and the optimal level for marginal rates—depends on fiscal pressure. When fiscal

2

pressure is low, distributional gains from raising marginal rates at low income levels are small because generous lump-sum transfers, mainly funded through taxing the rich, imply that consumption is relatively compressed. Thus, there is little to gain from taxing the moderately poor to increase transfers that help the very poorest. On the other hand, when the government needs to finance more purchases, lump-sum transfers are smaller and distributional gains at low income levels are larger. Larger distributional gains then incentivize the planner to choose higher marginal rates at low income levels, leading to a U-shaped profile.

We then show that introducing private insurance has similar effects to reducing required government expenditure, in terms of the impact on the optimal tax schedule. In particular, when private insurance is extensive, a combination of limited public transfers financed by high taxes on the rich coupled with private insurance in the background ensures relatively modest consumption inequality at low income levels, and thus there is no reason to impose high marginal tax rates on the poor. In contrast, when private insurance is absent, higher marginal tax rates at low income levels are optimal, and if the risk of very low uninsurable productivity realizations is sufficiently large, the optimal schedule becomes U-shaped.

We use our distributional gain/efficiency loss decomposition to revisit the conditions for a U-shaped optimal tax schedule discussed by Diamond (1998) in the context of a preference specification without income effects. Here we develop a new theoretical result on how raising required government purchases changes how distributional gains vary with income, a result that is consistent with the fiscal pressure intuition we use to interpret our numerical results.

Overall, our intuition about how fiscal pressure interacts with the standard equity-efficiency trade-off is a useful way to understand the shape of the optimal tax schedule. One reason it has not been developed to date is that it is not apparent in the functional equation (Diamond 1998 and Saez 2001) that is the usual starting point for interpreting the optimal tax schedule.

In the rest of the paper, we consider several important extensions to our baseline analysis. First, we consider alternatives to a utilitarian welfare criterion. We focus on a class of Pareto

weight functions in which the weight on an agent with uninsurable idiosyncratic productivity α is $\exp(-\theta\alpha)$. The parameter θ determines the planner's taste for redistribution, with $\theta > 0$ indicating greater than utilitarian concern for the poor. What value for θ is consistent with the extent of redistribution built into the actual U.S. tax and transfer system? To answer this question, we approximate the current tax system using the parametric tax and transfer scheme adopted in Benabou (2000) and Heathcote et al. (2017), where taxes net of transfers are given by the following function of income: $T(y) = y - \lambda y^{1-\tau}$. In this scheme, which we henceforth label "HSV," the parameter τ indexes the progressivity of the system. We develop a closed-form mapping between θ and the corresponding optimal choice for τ, a mapping that can be inverted to infer the taste for redistribution for the United States, θ^{US}, that rationalizes the observed degree of tax progressivity, τ^{US}. Given this "empirically motivated" social welfare function, we find that the optimal marginal tax schedule is again increasing, as in the utilitarian case.

Next, we compare the optimal Mirrleesian policy to the best possible policies when the tax and transfer system is restricted to simple parametric functional forms, à la Ramsey. We contrast two simple functional forms that are perhaps the most widely used in the literature: affine tax functions and the HSV tax scheme. These two schemes allow us to compare two alternative ways to redistribute income: the affine scheme allows for lump-sum transfers but imposes constant marginal tax rates, while the HSV scheme rules out transfers but allows for a progressive tax schedule. We find that the best policy in the HSV class is preferred to the best policy in the affine class, indicating that tax progressivity is more important than lump-sum transfers.

Finally, we explore Pareto-improving tax reforms. We consider the problem of a utilitarian planner who must ensure that tax reform leaves all households at least weakly better off. We find that such a planner would lower marginal rates at the top of the income distribution and raise them at the bottom, relative to our approximation of the current system. This reform leads to welfare gains in the tails of the income distribution, although the average

overall welfare gain is quite small. At the Pareto-improving optimum there is a range of values for productivity where Pareto-improving constraints bind. One interesting theoretical result is that within this range, allocations and taxes—and not just utility values—are identical to those under the status quo tax system.

Related Literature Seminal papers in the literature on taxation in the Mirrlees tradition include Mirrlees (1971), Diamond (1998), and Saez (2001). More recent work has focused on extending the approach to dynamic environments: Farhi and Werning (2013) and Golosov et al. (2016) are the most important examples. Golosov and Tsyvinski (2015) offer a survey of the key policy conclusions from this literature.

There are also many papers on tax design in the Ramsey (1927) tradition in economies with heterogeneity and incomplete private insurance markets. Recent examples include Conesa and Krueger (2006), who explore the Gouveia and Strauss (1994) functional form for the tax schedule, and Heathcote et al. (2017), who explore the HSV form developed by Feldstein (1969), Persson (1983), and Benabou (2000). Relative to those papers, the advantage of our non-parametric Mirrleesian approach is that we can characterize the entire shape of the optimal tax and transfer schedule. In particular, we can explore whether and when the optimal tax system exhibits lump-sum transfers or a non-monotone (e.g., U-shaped) profile for marginal tax rates; the HSV functional form allows for neither property.

Our interest in constructing a Pareto weight function that is consistent with observed tax progressivity is related to the inverse optimum taxation problem, which is to characterize the non-parametric profile for social welfare weights that precisely rationalizes a particular observed tax system; see Bourguignon and Spadaro (2012), Brendon (2013), and Heathcote and Tsujiyama (2017). The approach in this paper restricts the Pareto weight function to a one-parameter functional form that only allows for a simple tilt in planner preferences toward (or against) relatively high-productivity workers. Restricting the Pareto weight function to belong to a parametric class is analogous to restricting the tax function to a parametric class à la Ramsey) rather than solving for the fully optimal non-parametric Mirrlees schedule.

Werning (2007) describes how to test for Pareto efficiency of any given tax schedule, given an underlying skill distribution. Because our approximation to the current U.S. tax and transfer system violates several known properties of any optimal tax scheme, it is immediate that the associated allocations are not efficient. This motivates our extension to characterize a specific Pareto-improving reform.

Weinzierl (2014), Saez and Stantcheva (2016), and Hendren (2017) propose various interesting ways to generalize interpersonal comparisons that allow one to go beyond an assessment of Pareto efficiency, without insisting on a specific set of Pareto weights. For example, Saez and Stantcheva (2016) advocate the use of generalized social marginal welfare weights, which represent the value that society puts on providing an additional dollar of consumption to any given individual. In contrast, all our analyses specify fixed Pareto weights ex ante. One advantage is that we can evaluate non-marginal tax reforms, implying large differences in equilibrium allocations, in addition to local perturbations around a given tax system.

Chetty and Saez (2010) is one of the few papers to explore the interaction between public and private insurance in environments with private information. Section III of their paper explores a similar environment to ours, in which there are two components of productivity and differential roles for public versus private insurance with respect to the two components. Like us, they conclude that the government should focus on insuring the source of risk that cannot be insured privately. Relative to Chetty and Saez (2010), our contributions are twofold: (i) we consider optimal Mirrleesian tax policy in addition to affine tax systems, and (ii) our analysis is more quantitative in nature.

2 Environment

Labor Productivity   There is a unit mass of individuals. They differ only with respect to labor productivity w, which has two orthogonal idiosyncratic components: $\log w = \alpha + \varepsilon$. The first component $\alpha \in \mathcal{A} \subset \mathbb{R}$ represents shocks that cannot be insured privately. The second component $\varepsilon \in \mathcal{E} \subset \mathbb{R}$ represents shocks that can be privately observed and perfectly

privately insured. Neither α nor ε is observed by the tax authority. A natural motivation for the informational advantage of the private sector relative to the government with respect to ε shocks is that these are shocks that can be observed and pooled within a family (or other risk-sharing group), whereas the α shocks are shared by all members of the family but differ across families.[1] For the purposes of optimal tax design, the details of how private insurance is delivered do not matter as long as the set of risks that is privately insurable remains independent of the choice of tax system, which is our maintained assumption.

We let the vector (α, ε) denote an individual's type and F_α and F_ε denote the distributions for the two components. We assume F_α and F_ε are differentiable.

In the simplest description of the model environment, the world is static, and each agent draws α and ε only once. However, there is an isomorphic dynamic interpretation in which α represents fixed effects that are drawn before agents enter the economy, whereas ε captures a mix of predictable life-cycle productivity variation and life-cycle shocks against which agents can purchase insurance.[2]

Preferences Agents have identical preferences over consumption c and work effort h. The utility function takes the separable form

$$u(c, h) = \frac{c^{1-\gamma}}{1-\gamma} - \frac{h^{1+\sigma}}{1+\sigma},$$

where $\gamma > 0$ and $\sigma > 0$. The Frisch elasticity of labor supply is $1/\sigma$. We denote by $c(\alpha, \varepsilon)$ and $h(\alpha, \varepsilon)$ consumption and hours worked for an individual of type (α, ε).

Technology Aggregate output in the economy is aggregate effective labor supply. Output is divided between private consumption and a nonvalued publicly provided good G. The

---

[1] In Appendix A.1, we consider an alternative model for insurance in which there is no family and individual agents buy insurance against ε on decentralized financial markets.

[2] We discuss this interpretation further in Appendix A.2. Although explicit insurance against life-cycle shocks may not exist, households can almost perfectly smooth transitory shocks to income by borrowing and lending. A more challenging extension to the framework would be to allow for persistent shocks to the unobservable noninsurable component of productivity α. However, Heathcote et al. (2014) estimate that life-cycle uninsurable shocks account for only 17 percent of the observed cross-sectional variance of log wages.

resource constraint of the economy is thus

$$\iint c(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon) + G = \iint \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon). \tag{1}$$

Insurance We imagine insurance against ε shocks as occurring via a family planner who dictates hours worked and private within-family transfers for a continuum of agents who share a common uninsurable component α and whose insurable shocks ε are distributed according to F_ε.

Government The planner/tax authority observes only end-of-period family income, which we denote $y(\alpha)$ for a family of type α, where

$$y(\alpha) = \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\varepsilon(\varepsilon). \tag{2}$$

The tax authority does not directly observe α or ε, does not observe individual wages or hours worked, and does not observe the within-family transfers associated with within-family private insurance against ε.

Let $T(\cdot)$ denote the income tax schedule. Given that it observes income and taxes collected, the authority also effectively observes family consumption, since

$$\int c(\alpha, \varepsilon) dF_\varepsilon(\varepsilon) = y(\alpha) - T\left(y(\alpha)\right). \tag{3}$$

Family Head's Problem The timing of events is as follows. The family first draws a single $\alpha \in \mathcal{A}$. The family head then solves

$$\max_{\{c(\alpha,\varepsilon), h(\alpha,\varepsilon)\}_{\varepsilon \in \mathcal{E}}} \int \left[\frac{c(\alpha, \varepsilon)^{1-\gamma}}{1-\gamma} - \frac{h(\alpha, \varepsilon)^{1+\sigma}}{1+\sigma} \right] dF_\varepsilon(\varepsilon) \tag{4}$$

resource constraint of the economy is thus

$$\iint c(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon) + G = \iint \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon). \tag{1}$$

Insurance We imagine insurance against ε shocks as occurring via a family planner who dictates hours worked and private within-family transfers for a continuum of agents who share a common uninsurable component α and whose insurable shocks ε are distributed according to F_ε.

Government The planner/tax authority observes only end-of-period family income, which we denote $y(\alpha)$ for a family of type α, where

$$y(\alpha) = \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\varepsilon(\varepsilon). \tag{2}$$

The tax authority does not directly observe α or ε, does not observe individual wages or hours worked, and does not observe the within-family transfers associated with within-family private insurance against ε.

Let $T(\cdot)$ denote the income tax schedule. Given that it observes income and taxes collected, the authority also effectively observes family consumption, since

$$\int c(\alpha, \varepsilon) dF_\varepsilon(\varepsilon) = y(\alpha) - T\left(y(\alpha)\right). \tag{3}$$

Family Head's Problem The timing of events is as follows. The family first draws a single $\alpha \in \mathcal{A}$. The family head then solves

$$\max_{\{c(\alpha,\varepsilon), h(\alpha,\varepsilon)\}_{\varepsilon \in \mathcal{E}}} \int \left[\frac{c(\alpha, \varepsilon)^{1-\gamma}}{1-\gamma} - \frac{h(\alpha, \varepsilon)^{1+\sigma}}{1+\sigma} \right] dF_\varepsilon(\varepsilon) \tag{4}$$

subject to (2) and the family budget constraint (3).[3] The first-order conditions (FOCs) are

$$c(\alpha, \varepsilon) = c(\alpha) = y(\alpha) - T(y(\alpha)), \tag{5}$$

$$h(\alpha, \varepsilon)^\sigma = [y(\alpha) - T(y(\alpha))]^{-\gamma} \exp(\alpha + \varepsilon) [1 - T'(y(\alpha))]. \tag{6}$$

The first FOC indicates that the family head wants to equate consumption within the family. The second indicates that the family equates—for each member—the marginal disutility of labor supply to the marginal utility of consumption times individual productivity times one minus the marginal tax rate on family income. If the tax function satisfies

$$T''(y) > -\gamma \frac{[1 - T'(y)]^2}{y - T(y)} \tag{7}$$

for all feasible y, then the second derivative of family welfare with respect to hours for any type (α, ε) is strictly negative, and the first-order conditions (5) and (6) are sufficient for optimality.

Equilibrium Given the income tax schedule T, a *competitive equilibrium* for this economy is a set of decision rules $\{c, h\}$ such that

(i) The decision rules $\{c, h\}$ solve the family's maximization problem (4),

(ii) The resource feasibility constraint (1) is satisfied, and

(iii) The government budget constraint is satisfied: $\int T(y(\alpha))dF_\alpha(\alpha) = G$.

3 Planner's Problems

The planner maximizes social welfare given Pareto weights $W(\alpha)$ that may vary with α.[4]

---

[3]In Appendix A.3, we show that allowing the planner to observe and tax income (after within-family transfers) at the individual level would not change the solution to the family head's problem. Thus, there would be no advantage to taxing at the individual rather than the family level.

[4]We assume symmetric weights with respect to ε to focus on the government's role in providing public insurance against privately uninsurable differences in α. In addition, we will show that constrained efficient allocations cannot be conditioned on ε.

3.1 Mirrlees Problem: Constrained Efficient Allocations

In the Mirrlees formulation of the program that determines constrained efficient allocations, we envision the Mirrlees planner interacting with family heads for each α type. Thus, each family is effectively a single agent from the perspective of the planner. The planner chooses both aggregate family consumption $c(\alpha)$ and income $y(\alpha)$ as functions of the family type α. The Mirrleesian planner's problem includes incentive constraints that guarantee that for each and every type α, a family of that type weakly prefers to deliver to the planner the value for income $y(\alpha)$ the planner intends for that type, thereby receiving $c(\alpha)$, rather than delivering any alternative level of income.

The timing within the period is as follows. Families first decide on a reporting strategy $\hat{\alpha} : \mathcal{A} \to \mathcal{A}$. Each family draws $\alpha \in \mathcal{A}$ and makes a report $\tilde{\alpha} = \hat{\alpha}(\alpha) \in \mathcal{A}$ to the planner. In a second stage, given the values for $c(\tilde{\alpha})$ and $y(\tilde{\alpha})$, the family head decides how to allocate consumption and labor supply across family members.

Family Problem   As a first step toward characterizing efficient allocations, we start with the second stage. Taking as given a report $\tilde{\alpha} = \hat{\alpha}(\alpha)$ and a draw α, the family head solves

$$U(\alpha, \tilde{\alpha}) \equiv \max_{\{c(\alpha,\tilde{\alpha},\varepsilon),h(\alpha,\tilde{\alpha},\varepsilon)\}_{\varepsilon \in \mathcal{E}}} \int \left[\frac{c(\alpha, \tilde{\alpha}, \varepsilon)^{1-\gamma}}{1-\gamma} - \frac{h(\alpha, \tilde{\alpha}, \varepsilon)^{1+\sigma}}{1+\sigma} \right] dF_\varepsilon(\varepsilon), \qquad (8)$$

$$\text{subject to} \quad \int c(\alpha, \tilde{\alpha}, \varepsilon) dF_\varepsilon(\varepsilon) = c(\tilde{\alpha}),$$

$$\int \exp(\alpha + \varepsilon) h(\alpha, \tilde{\alpha}, \varepsilon) dF_\varepsilon(\varepsilon) = y(\tilde{\alpha}).$$

Solving this problem gives the following indirect utility function:

$$U(\alpha, \tilde{\alpha}) = \frac{c(\tilde{\alpha})^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\tilde{\alpha})}{\exp(\alpha)} \right)^{1+\sigma}, \text{ where } \Omega = \left(\int \exp(\varepsilon)^{\frac{1+\sigma}{\sigma}} dF_\varepsilon(\varepsilon) \right)^{-\sigma}. \qquad (9)$$

First-Stage Planner's Problem The planner maximizes social welfare, evaluated according to $W(\alpha)$, subject to the resource constraint and to incentive constraints:

$$\max_{\{c(\alpha),y(\alpha)\}_{\alpha \in \mathcal{A}}} \int W(\alpha)U(\alpha,\alpha)dF_\alpha(\alpha), \tag{10}$$

$$\text{subject to} \quad \int c(\alpha)dF_\alpha(\alpha) + G = \int y(\alpha)dF_\alpha(\alpha), \tag{11}$$

$$U(\alpha,\alpha) \geq U(\alpha,\tilde{\alpha}) \qquad \text{for all } \alpha \text{ and } \tilde{\alpha}. \tag{12}$$

Note that ε does not appear anywhere in this problem (the distribution F_ε is buried in the constant Ω). The problem is therefore identical to a standard static Mirrlees type problem, where the planner faces a distribution of agents with heterogeneous unobserved productivity α.[5] We will solve this problem numerically.

Decentralization with Income Taxes Instead of thinking of the planner as offering agents a menu of alternative pairs for income and consumption, we can instead conceptualize the planner offering a mapping from any possible value for family income to family consumption. Such a schedule can be decentralized via a tax schedule on family income y of the form $T(y)$ that defines how rapidly consumption grows with income.[6]

Substituting the first-order condition with respect to hours (6) into the second constraint in problem (8) and letting $c^*(\alpha)$ and $y^*(\alpha)$ denote the values for family consumption and income that solve the Mirrlees problem (10), we can recover how optimal marginal tax rates vary with income:

$$1 - T'(y^*(\alpha)) = \frac{\Omega}{c^*(\alpha)^{-\gamma}\exp(\alpha)}\left(\frac{y^*(\alpha)}{\exp(\alpha)}\right)^\sigma. \tag{13}$$

3.2 Ramsey Problem

We use the label "Ramsey planner" to describe a planner who chooses the optimal tax function in a given parametric class $\mathcal{T}$. For the class of affine functions, $\mathcal{T} = \{T : \mathbb{R}_+ \to$

---

[5]Note that the weight on hours in the agents' utility function is now Ω rather than 1.
[6]Note that some values for income might not feature in the menu offered by the Mirrlees planner. Those values will not be chosen in the income tax decentralization if income at those values is heavily taxed.

$\mathbb{R}|T(y) = \tau_0 + \tau_1 y$ for $y \in \mathbb{R}_+$, $\tau_0 \in \mathbb{R}$, $\tau_1 \in \mathbb{R}\}$. For the HSV class, $\mathcal{T} = \{T : \mathbb{R}_+ \to \mathbb{R}|T(y) = y - \lambda y^{1-\tau}$ for $y \in \mathbb{R}_+$, $\lambda \in \mathbb{R}_+$, $\tau \in [-1, 1]\}$.

The Ramsey problem is to maximize social welfare by choosing a tax schedule in $\mathcal{T}$ subject to allocations being a competitive equilibrium:

$$\max_{T \in \mathcal{T}} \int W(\alpha) \int u(c(\alpha, \varepsilon), h(\alpha, \varepsilon)) dF_\varepsilon(\varepsilon) dF_\alpha(\alpha) \qquad (14)$$

subject to (1) and to $c(\alpha, \varepsilon)$ and $h(\alpha, \varepsilon)$ being solutions to the family problem (4).[7]

3.3 Decomposing the Trade-offs in Setting Tax Rates

We now describe a decomposition of the welfare effects of changing marginal tax rates at different points along the income distribution, which we will use later to develop intuition about the shape of the optimal marginal tax schedule. This decomposition is similar to the expressions developed by Diamond (1998) and Saez (2001).[8] One advantage of our decomposition, which we will later exploit, is that it can be used to evaluate the welfare effects of tax reform starting from any tax system, even if it is non-optimal.[9]

Consider the effect of increasing the marginal tax rate at some income level $\hat{y}$, so as to collect one dollar more from everyone with income above $\hat{y}$. Assume that all extra revenue generated is used to increase lump-sum transfers. Consider, first, the hypothetical welfare

---

[7]Note that in the affine tax function case, condition (7) is satisfied because

$$T''(y) + \gamma \frac{[1 - T'(y)]^2}{y - T(y)} = \gamma \frac{(1 - \tau_1)^2}{y - T(y)} > 0.$$

In the HSV tax function case, condition (7) becomes

$$T''(y) + \gamma \frac{[1 - T'(y)]^2}{y - T(y)} = \lambda y^{(-\tau-1)} (1 - \tau) [\tau + \gamma (1 - \tau)] > 0.$$

This is satisfied for any progressive tax, $\tau \in [0, 1)$, because $\tau + \gamma (1 - \tau) > 0$. It is also satisfied for any regressive tax, $\tau < 0$, if $\gamma \geq 1$, because $\gamma \geq 1 > \frac{-\tau}{1-\tau}$. Therefore, for all relevant parameterizations, condition (7) is also satisfied for this class of tax functions.

[8]In Appendix B.2, we also derive the standard Diamond-Saez formula for our economy.

[9]Equation (19) in Saez (2001) includes the multiplier on the government budget constraint at the optimum (see his footnote 14), but there is no such multiplier in our equation. Thus, our expressions can be used away from the optimum, where this multiplier is not well defined.

gain that this reform would deliver if there was no behavioral response. The revenue collected, and thus the increase in lump-sum transfers, would be $1 - F_y(\hat{y})$, where F_y denotes the distribution of income. The value of an extra dollar of lump-sum transfers to the planner is the Pareto-weighted average marginal utility of consumption, $\chi \equiv \int_0^\infty W_y(y) u_c(y) dF_y(y)$ where $W_y(y(\alpha)) \equiv W(\alpha)$ and $u_c(\cdot)$ is the marginal utility of consumption. The welfare cost to the planner from raising a dollar from all households earning above $\hat{y}$ is the (weighted) average marginal utility of that set of households. Thus, the *distributional welfare gain* from the reform is $[1 - F_y(\hat{y})] \chi - \int_{\hat{y}}^\infty W_y(y) u_c(y) dF_y(y)$. It is convenient to measure this gain in units of consumption per dollar of revenue collected. Thus, we define

$$D(\hat{y}) \equiv 1 - \frac{\int_{\hat{y}}^\infty W_y(y) u_c(y) dF_y(y)}{[1 - F_y(\hat{y})] \chi}. \tag{15}$$

The cost of this tax reform is that it will reduce labor supply and thus tax revenue. We define the *efficiency cost* of the reform due to behavioral responses to be the revenue that would be collected from increasing the marginal tax rate at $\hat{y}$ absent a behavioral response (i.e., $1 - F_y(\hat{y})$,), minus the actual extra transfers that can be funded in equilibrium, which we denote $\Delta Tr(\hat{y})$. Again, we express this measure per unit of hypothetical revenue. Thus,

$$E(\hat{y}) = 1 - \frac{\Delta Tr(\hat{y})}{1 - F_y(\hat{y})}.$$

This efficiency cost measure can be interpreted as the fraction of hypothetical revenue that leaks away because of behavioral responses: if $\Delta Tr(\hat{y}) = 1 - F_y(\hat{y})$, there is no leakage and $E(\hat{y}) = 0$, whereas if $\Delta Tr(\hat{y}) = 0$, there is 100 percent leakage and $E(\hat{y}) = 1$.[10]

---

[10]This efficiency cost can be written as

$$E(\hat{y}) = \frac{-I(0)}{1 - I(0)} - \frac{1}{1 - F_y(\hat{y})} \frac{S(\hat{y}) - I(\hat{y})}{1 - I(0)},$$

where $S(y) < 0$ denotes the revenue loss from households at income level y working less because of a substitution effect, and $I(y) < 0$ denotes the loss in revenue from all individuals with income above y working less via a wealth effect because they receive an extra dollar of unearned income. See Appendix B.1 for the derivation.

If the tax system is optimal, the distributional gain from our hypothetical tax reform exactly equals the efficiency cost at every income level, and the equation $D(\hat{y}) = E(\hat{y})$ is the standard Diamond-Saez formula.

4 Calibration

Preferences   We assume preferences are logarithmic in consumption:

$$u(c, h) = \log c - \frac{h^{1+\sigma}}{1 + \sigma}.$$

This balanced growth specification is the same one adopted by Heathcote et al. (2017). We choose $\sigma = 2$ so that the Frisch elasticity $(1/\sigma)$ is 0.5. This value is consistent with the microeconomic evidence (see, e.g., Keane 2011) and is very close to the value estimated by Heathcote et al. (2014). The compensated (Hicks) elasticity of hours with respect to the marginal net-of-tax wage is approximately equal to $1/(1 + \sigma)$ (see Keane 2011, eq. 11) which, given $\sigma = 2$, is equal to 1/3. Again, this value is consistent with empirical estimates: Keane reports an average estimate across 22 studies of 0.31. Given our model for taxation, the elasticity of average income with respect to one minus the average income-weighted marginal tax rate is also equal to $1/(1 + \sigma)$.[11]   According to Saez et al. (2012), the best available estimates for the long-run version of this elasticity range from 0.12 to 0.40.

Tax and Transfer System   The class of tax functions that we label "HSV" was perhaps first used by Feldstein (1969) and introduced into dynamic heterogeneous agent models by Persson (1983) and Benabou (2000).

Heathcote et al. (2017) begin by noting that the HSV tax function implies a linear relationship between $\log(y)$ and $\log(y - T(y))$, with a slope equal to $1 - \tau$. Thus, given micro data on household income before taxes and transfers and income net of taxes and transfers, it is straightforward to estimate τ by ordinary least squares. Using micro data

---

[11]The average income-weighted marginal tax rate is $1 - (1 - g)(1 - \tau)$, where g is the ratio of government purchases to output (see Heathcote et al. 2017, eq. 4).

from the Panel Study of Income Dynamics (PSID) for working-age households over the period 2000 to 2006, Heathcote et al. (2017) estimate $\tau = 0.181$.

The remaining fiscal policy parameter λ is set such that government purchases G is equal to 18.8 percent of model GDP, which was the ratio of government purchases to output in the United States in 2005. When we evaluate alternative tax policies, we always hold fixed G at its baseline value.

Wage Distribution and Insurance   Our strategy for calibrating the model distribution of wages and the relative importance of uninsurable versus insurable shocks is as follows. First, we assume that log wages are drawn from an exponentially modified Gaussian (EMG) distribution. Second, we parameterize the overall wage variance and the fraction of this variance that is privately uninsurable to replicate the observed cross-sectional variances of log earnings and log consumption, exploiting the standard result that uninsurable wage risk will show up in consumption, whereas insurable shocks will not.[12]

We assume that the insurable component of productivity is normally distributed, $\varepsilon \sim N(-\sigma_\varepsilon^2/2, \sigma_\varepsilon^2)$, and that the uninsurable component follows an EMG distribution: $\alpha = \alpha_N + \alpha_E$, where $\alpha_N \sim N(\mu_\alpha, \sigma_\alpha^2)$ and $\alpha_E \sim Exp(\lambda_\alpha)$ so that $\alpha \sim EMG(\mu_\alpha, \sigma_\alpha^2, \lambda_\alpha)$. It follows that the log wage, $\log w = \alpha + \varepsilon$, is itself EMG (the sum of the two normally distributed random variables α_N and ε is normal), so the *level* wage distribution is Pareto lognormal.

Given our baseline HSV tax system, the equilibrium distributions for log earnings and

---

[12]One measurement issue we need to address is that some of the observed cross-sectional inequality in earnings and consumption reflects systematic variation by age, but there is no notion of age in our static model. To guide our calibration choices here, Appendix A.2 lays out a simple life-cycle overlapping-generations model with both predictable life-cycle variation in wages and idiosyncratic insurable life-cycle shocks. We show that our benchmark static model is isomorphic to this extended model, as long as the static model is calibrated to replicate total cross-sectional dispersion in wages, earnings, and consumption, with both predictable wage changes and life-cycle shocks captured in the insurable component of wages.

log consumption are also EMG with[13]

$$\text{Var}\left(\log y\right) = \left(\frac{1+\sigma}{\sigma}\right)^2 \sigma_\varepsilon^2 + \sigma_\alpha^2 + \frac{1}{\lambda_\alpha^2}, \tag{16}$$

$$\text{Var}\left(\log c\right) = (1-\tau)^2\sigma_\alpha^2 + \frac{(1-\tau)^2}{\lambda_\alpha^2}. \tag{17}$$

Our calibration strategy is to first use an empirical distribution for log earnings to estimate the normal variance $\sigma_y^2 = \left(\frac{1+\sigma}{\sigma}\right)^2 \sigma_\varepsilon^2 + \sigma_\alpha^2$ and the tail parameter λ_α. Given our external estimates for σ and τ and this estimate for λ_α, we then use an estimate for the variance of log consumption to infer σ_α^2 from eq. (17). Finally, given σ_α^2 and λ_α, the variance of log earnings (16) residually exactly identifies σ_ε^2.

As Mankiw et al. (2009) emphasize, it is difficult to sharply estimate the shape of the productivity distribution given typical household surveys, such as the Current Population Survey (CPS), in part because high-income households tend to be underrepresented in these samples. We therefore turn to the Survey of Consumer Finances (SCF), which uses data from the Internal Revenue Service (IRS) Statistics of Income program to ensure that wealthy households are appropriately represented. We estimate λ_α and σ_y^2 by maximum likelihood, searching for the values of the three parameters in the EMG distribution that maximize the likelihood of drawing the observed 2007 distribution of log labor income.[14] The resulting

---

[13]Equilibrium allocations for hours, individual earnings, and consumption are given by

$$h(\varepsilon) = (1-\tau)^{\frac{1}{1+\sigma}} \left\{\mathbb{E}\left[\exp(\varepsilon)^{\frac{1+\sigma}{\sigma}}\right]\right\}^{\frac{-1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\varepsilon\right),$$

$$y(\alpha,\varepsilon) = (1-\tau)^{\frac{1}{1+\sigma}} \left\{\mathbb{E}\left[\exp(\varepsilon)^{\frac{1+\sigma}{\sigma}}\right]\right\}^{\frac{-1}{1+\sigma}} \exp(\alpha)\exp\left(\frac{1+\sigma}{\sigma}\varepsilon\right),$$

$$c(\alpha) = \lambda(1-\tau)^{\frac{1-\tau}{1+\sigma}} \left\{\mathbb{E}\left[\exp(\varepsilon)^{\frac{1+\sigma}{\sigma}}\right]\right\}^{\frac{\sigma(1-\tau)}{1+\sigma}} \exp((1-\tau)\alpha).$$

Note that hours worked are independent of the uninsurable shock α—preferences have the balanced growth property—whereas the elasticity of hours to the insurable shock ε is exactly the Frisch elasticity. The elasticities of log earnings (log productivity plus log hours) to uninsurable and insurable shocks are therefore 1 and $1+\frac{1}{\sigma}$, respectively. Consumption does not respond to insurable shocks, and the elasticity of consumption to uninsurable shocks is $1-\tau$.

[14]The empirical distribution for labor income in 2007 is constructed as follows. We define labor income as wage income plus two-thirds of income from business, sole proprietorship, and farm. We then restrict our sample to households with at least one member aged 25-60 and with household labor income of at least $10,000 (mean household labor income is $77,325).
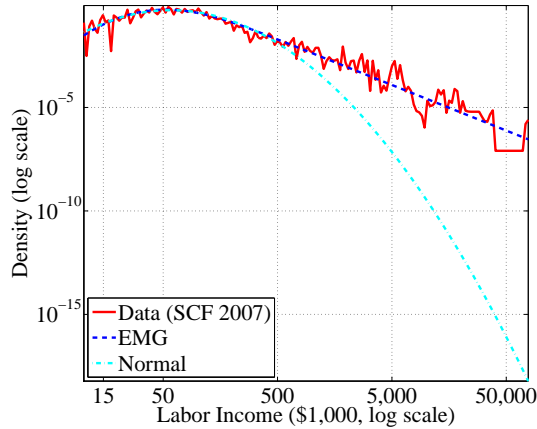
Figure 1: Fit of EMG distribution. The figure plots the empirical earnings density from the SCF against the estimated EMG distribution and against a normal distribution.

estimates are $\lambda_\alpha = 2.2$ and $\sigma_y^2 = 0.412$, implying a total variance for log earnings of 0.618.[15]

Figure 1 plots the empirical density against the estimated EMG distribution and a normal distribution with the same mean and variance. The density is plotted on a log scale to magnify the tails. It is clear that the heavier right tail that the additional parameter in the EMG specification introduces delivers an excellent fit, substantially improving on the normal specification.

We require an estimate of the cross-sectional variance of log consumption to calibrate v_α. Using the Consumer Expenditure Survey, Heathcote et al. (2010, figure 13) report a variance of 0.332 in 2006.[16] However, Heathcote et al. (2014, table 3) estimate that 29.6 percent of the variance of measured consumption reflects measurement error, implying a true variance of 0.234.[17] Given $\lambda_\alpha = 2.2$, the model replicates this variance when $\sigma_\alpha^2 = 0.142$. Finally, using eq. (16) to residually infer σ_ε^2 gives $\sigma_\varepsilon^2 = 0.120$. In Section 5.1, we will explore how changing the relative magnitudes of insurable and uninsurable wage risk changes the optimal tax schedule.

---

[15]Bootstrapped 95 percent confidence intervals for the point estimates for λ_α and σ_y^2 are [1.86,2.56] and [0.303,0.501], respectively.

[16]Other estimates in the literature are consistent with this estimate. Meyer and Sullivan (2017, figures 6 and 7) report 90/50 and 50/10 percentile ratios in the mid-2000s that are both close to 2. The same ratios are also close to 2 in Heathcote et al. (2010, figure 13). Attanasio and Pistaferri (2014, figure 1) report a standard deviation of log consumption in the PSID of around 0.6, implying a variance of 0.36.

[17]They estimate that none of the measured variance of earnings reflects measurement error.

Table 1: Model Productivity Distribution and Offered Wage Distribution in Low and Pistaferri (2015)

Percentile Ratios	Model	Low and Pistaferri
P5/P1	1.46	1.48
P10/P5	1.23	1.20
P25/P10	1.42	1.40

Note: Px denotes the xth percentile.

Given all these values, the total model variance for log wages is $\sigma_\varepsilon^2 + \sigma_\alpha^2 + \lambda_\alpha^{-2} = 0.469$. For comparison, Heathcote et al. (2010, figure 5) report a similar log wage variance for men of 0.499 in the CPS in 2005. We have documented that our assumptions on the wage distribution deliver an extremely close approximation to the top of the earnings distribution, as reflected in the SCF. It is also important to assess whether we accurately capture the distribution of labor productivity at the bottom. A well-known challenge here is that some low productivity workers choose not to work, and thus their productivity cannot be directly observed. Low and Pistaferri (2015) estimate a rich structural model of participation in which workers face disability risk and can apply for disability insurance. Table 1 compares statistics for the left tail of our calibrated productivity distribution to corresponding statistics from the distribution of *latent offered* wages from their estimated model.[18] Reassuringly, the two sets of statistics are very similar.

Our calibration is designed to replicate the empirical variance of log consumption, but it is also important to ask whether it implies a realistic *shape* for the consumption distribution. Because we have attributed the heavy right tail in the log wage distribution to the uninsurable component of wages, the model implies a heavy right tail in the distribution for consumption. Toda and Walsh (2015) estimate that the distribution of household consumption does in fact have fat tails, and they estimate an average right tail Pareto parameter of 3.38. The value for λ_c implied by our estimates is similar at 2.69, providing empirical support for our assumption that the exponential component of log wages is uninsurable.

---

[18]We thank Low and Pistaferri for sharing their estimates.

Discretization   In solving the Mirrlees problem to characterize efficient allocations, the incentive constraints only apply to the uninsurable component of the wage α, and the distribution for ε appears only in the constant Ω. Thus, there is no need to approximate the distribution for ε, and we therefore assume these shocks are drawn from a continuous unbounded normal distribution with mean $-\sigma_\varepsilon^2/2$ and variance σ_ε^2.

We take a discrete approximation to the continuous EMG distribution for α that we have discussed thus far. We construct a grid of I evenly spaced values $\{\alpha_1, \cdots, \alpha_I\}$ with corresponding probabilities $\{\pi_1, \cdots, \pi_I\}$ as follows. We make the endpoints of the grid, α_1 and α_I, sufficiently extreme that only a tiny fraction of individuals lie outside these bounds in the true continuous distribution. In particular, we set α_1 such that $\exp(\alpha_1)/\sum_i (\pi_i \exp(\alpha_i)) = 0.05$ and set α_I such that $\exp(\alpha_I)/\sum_i (\pi_i \exp(\alpha_i)) = 74$, which corresponds to household labor income at the 99.99th percentile of the SCF labor income distribution (\$6.17 million).[19] We read corresponding probabilities π_i directly from the continuous EMG distribution, rescaling to ensure that (i) $\sum_i \pi_i = 1$, (ii) $\sum_i \pi_i \exp(\alpha_i) = 1$, and (iii) the variance of (discretized) α is equal to $\sigma_\alpha^2 + \lambda_\alpha^{-2}$. For our baseline set of numerical results, we set $I = 10,000$. The resulting model distribution for α is plotted in panel A of figure 2. The distribution appears continuous, even though it is not, because our discretization is very fine. In Appendix F.6, we report how the results change when we increase or reduce I.

5   Quantitative Analysis

We explore the structure of the optimal tax and transfer system, given the model specification described above.[20] We focus initially on a utilitarian social welfare function.

---

[19] Assuming $2,000$ household hours worked, the average hourly wage is \$41.56, so 5 percent of the average corresponds to \$2.08, which is less than half the federal minimum wage in 2007 (\$5.85). Reducing α_1 further would not materially affect any of our results, since given the parameters for the EMG distribution, the probability of drawing $\alpha < \log(0.05)$ is vanishingly small.

[20] In Appendix C.1, we explain how we numerically solve the Mirrlees optimal tax problem.

5.1 Increasing versus U-Shaped Marginal Rates

The extensive literature exploring the Mirrlees optimal taxation problem has established that the shape of the optimal tax schedule is sensitive to all elements of the environment, including the shape of the skill distribution, the form of the utility function, the planner's taste for redistribution, and the government revenue requirement (see, for example, Tuomala 1990). However, starting from the influential papers of Diamond (1998) and Saez (2001), most quantitative applications of the theory to the United States have found a U-shaped profile for optimal marginal tax rates.[21]

Figure 2 plots the marginal and average tax schedules (panels A and B) that decentralize the constrained efficient allocation against the baseline HSV approximation to the current U.S. tax and transfer system (HSV^US).[22] In contrast to Diamond and Saez, optimal marginal tax rates are always increasing in income (except at the very top).[23] The marginal rate starts at 5.5 percent for the least productive households, is fairly flat (between 30 and 40 percent) up to half of average productivity, and rises rapidly to peak at 66.9 percent at 15 times average productivity.[24] Panel B indicates that the optimal schedule is considerably more redistributive than our approximation to the current U.S. system.

Panel C plots the distributional gain $D(\alpha)$ (equivalently, the efficiency cost $E(\alpha)$) from changes to marginal tax rates starting from the optimal policy (see Section 3.3 for the definitions of these measures).[25] Panel D plots the distributional gain/efficiency cost multiplied by the hazard ratio $[1 - F_\alpha(\alpha)]/f_\alpha(\alpha)$. Note that this plot qualitatively resembles the optimal marginal tax schedule in panel A. In fact, the two series would be exactly proportional under

---

[21]See also Diamond and Saez (2011) and Golosov et al. (2016).

[22]The profiles for marginal and average tax rates look very similar plotted against log household income rather than against log household productivity. Figure A1 in Appendix D plots the marginal tax rate against the *level* of income.

[23]Like us, Tuomala (2010) finds an increasing marginal rate schedule to be optimal. However, his results hinge on assuming a utility function that is quadratic in consumption with a bliss point.

[24]Because our discrete distribution for α is bounded, the Mirrleesian marginal tax rate drops to zero at the very top. However, marginal tax rates only dip very close to the upper bound for α, the choice for which is somewhat arbitrary.

[25]In Appendix B.1, we describe how we derive $D(\alpha)$ and $E(\alpha)$ from $D(y)$ and $E(y)$.
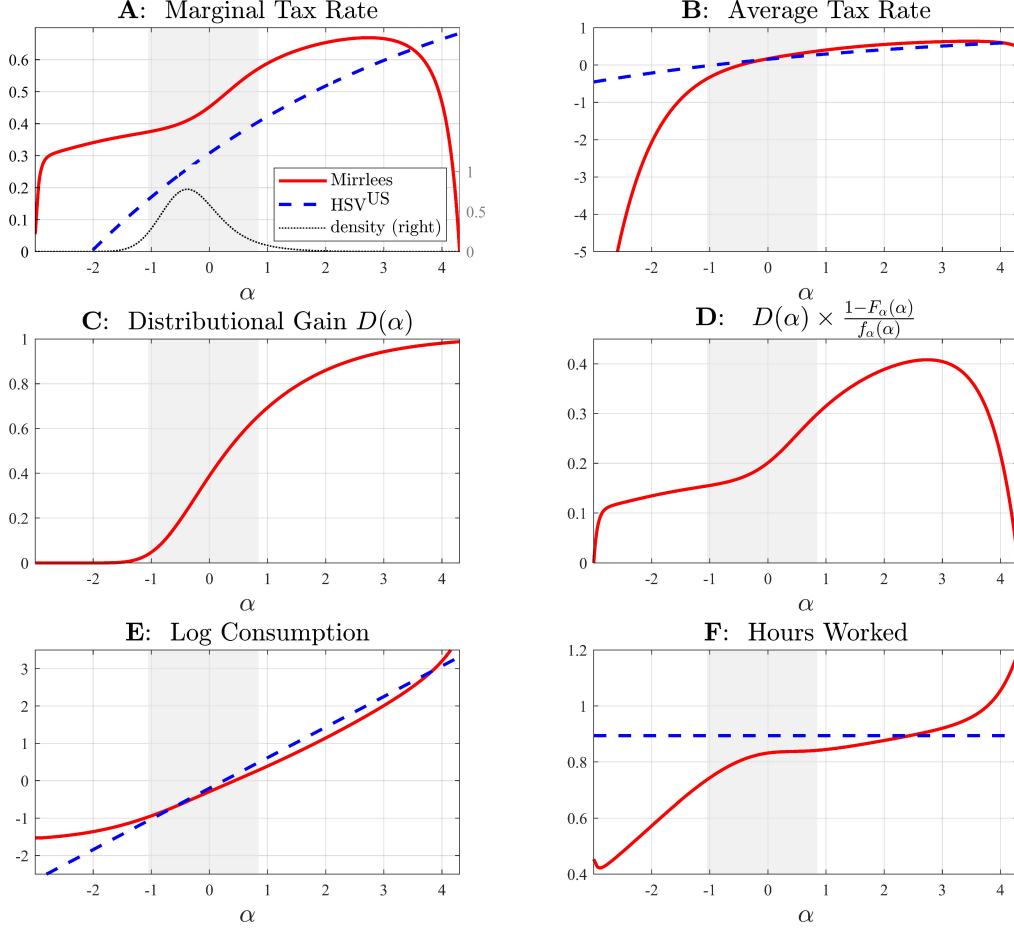
Figure 2: Optimal Tax Policy. Panels A and B plot the optimal Mirrleesian tax schedules against the baseline HSV approximation to the U.S. tax and transfer system and the productivity density. The area between the 5th and 95th percentiles is shaded gray. Panels C and D plot the distributional gain (equivalently, the efficiency cost) under the optimal policy. Panels E and F plot decision rules for consumption and hours worked (for an agent with average ε).

a preference specification without income effects (see Section 5.2). This resemblance indicates that if we can understand what drives distributional gains, we can better understand optimal taxation.

In our baseline calibration, distributional gains are large at high income levels because these households enjoy much higher consumption than the poor (see panel E). Thus, the utilitarian planner wants high marginal tax rates on the rich to finance lump-sum transfers. Because these distributional gains are so large, the planner tolerates equally large efficiency costs. For example, at $\alpha = \ln(3)$ (i.e., at three times average productivity), the efficiency

cost is 0.71, indicating that 71 percent of every hypothetical marginal dollar tax in revenue leaks away via behavioral responses. As we increase α further, the efficiency cost rises further toward one. This reflects the well-known result that a planner with a concern for equity will seek to maximize redistribution down from the very richest households.[26] It is over the rest of the income distribution that the shape of the optimal marginal tax schedule is less well understood and where our results disagree with Diamond (1998) and Saez (2001).

In our calibrated model, the tax revenue generated by soaking the rich funds sufficiently generous lump-sum transfers that consumption inequality across the bottom half of the productivity distribution is quite low (see panel E of figure 2). Thus, the distributional gains from raising marginal rates at low income levels are very small, implying that the planner does not want to set high (and highly distortionary) marginal tax rates in this part of the income distribution. For example, at $\alpha = \ln(1/3)$, the efficiency cost is only 0.03, indicating that the planner chooses not to raise the marginal rate here even though only 3 percent of each marginal tax dollar would leak away. Because distributional gains are very small at low income levels, the optimal marginal tax schedule is upward sloping. At the very bottom of the income distribution, bunching is optimal in our economy, implying sharply rising marginal tax rates at very low income levels.[27]

An upward-sloping profile for marginal tax rates is desirable because it pushes the tax burden upward within the income distribution, allowing the planner to redistribute from the richest agents toward everyone else. Thus, equity considerations will generally dictate an upward-sloping marginal tax schedule. However, as we will shortly see, efficiency considerations can also play an important role. From an efficiency standpoint, a downward-sloping

---

[26]Assuming an unbounded Pareto distribution for earnings, the well-known formula for the rate that maximizes revenue collection from the most productive households (eq. 9 in Saez 2001) is $\bar{T}' = \left[1 + \bar{\zeta}^u + \bar{\zeta}^c(\lambda_y^* - 1)\right]^{-1} = \frac{1+\sigma}{\sigma+\lambda_a}$, where $\bar{T}'$, $\bar{\zeta}^u$ and $\bar{\zeta}^c$ are limiting values of the marginal tax rate and uncompensated and compensated labor supply elasticities, and where λ_y^* is the Pareto parameter defining the right tail of the optimal earnings distribution. Given our utility function, $\bar{\zeta}^u = 0$, $\bar{\zeta}^c = (1+\sigma)^{-1}$, and $\lambda_y^* = \lambda_a$, which delivers the second equality. Note that this expression is independent of the value for government purchases. Evaluated at our calibrated values for σ and λ_a, the above equation implies that there is nothing to be gained from raising marginal rates above 71 percent at the top.

[27]We discuss this bunching in more detail in Appendix D.

profile for marginal rates is preferred because such a profile implies that agents face relatively low marginal tax rates—implying modest distortions—but relatively high average tax rates—translating into high revenue.

Under our baseline model calibration, the fact that the optimal marginal tax schedule is upward sloping indicates that equity concerns dominate. We will now show that alternative model parameterizations in which the planner faces greater fiscal pressure can change the trade-off and thus the shape of the optimal tax schedule. The main message will be that a downward-sloping or U-shaped optimal marginal tax schedule is optimal when there are large distributional gains from imposing high marginal tax rates at low income levels. Such gains can arise when (i) the government must deliver high government consumption, which crowds out lump-sum transfers, or (ii) when there is high uninsurable productivity dispersion at low income levels. In Section 5.3 we show that a U-shaped optimal profile also emerges when the planner puts very high welfare weight on low-productivity households.

Increasing Government Purchases   Panel A of figure 3 plots the optimal marginal tax schedules when we increase G from the baseline value (18.8 percent of output under the HSV^{US} policy) to higher levels (40 percent and 70 percent). Panel B plots the corresponding distributional gain/efficiency cost functions, in each case relative to the baseline.[28]

Raising required expenditure leads the government to raise marginal tax rates across the productivity distribution and by much more at low productivity levels. The result is that the schedule eventually becomes generally U-shaped.[29] This new pattern of marginal rates is optimal because a higher required expenditure squeezes lump-sum transfers, which in turn amplifies the gains from redistributing downward even from relatively unproductive agents (panel B). This leads the planner to impose high marginal tax rates at relatively low productivity levels, thereby sacrificing redistribution to the middle class in order to focus on

---

[28]To avoid the visual distraction of the zero-top-tax-rate property, we have truncated the visible range for productivity α at the 99.95^{th} percentile of the baseline model distribution for α in this figure and in subsequent similar ones.

[29]We add the caveats that the marginal rate is still increasing in the very low productivity interval where bunching occurs and still declines to zero at the very top.
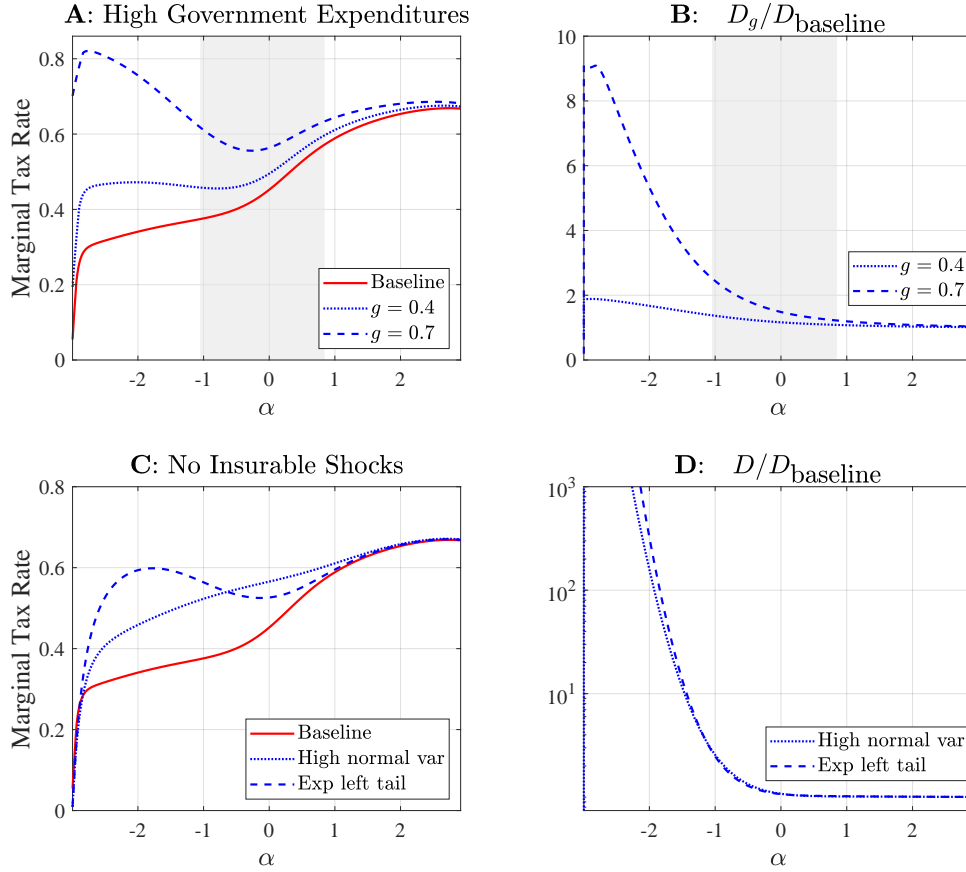
Figure 3: Increasing Fiscal Pressure. Panels A and B plot the optimal marginal tax schedules and the corresponding distributional gains for different values of government purchases. For example, the line labeled $g = 0.4$ corresponds to a value for G equal to 40 percent of output under the HSV^{US} policy. Likewise, panels C and D are the economies with no insurable shocks and higher uninsurable risk. The dotted lines indicate when the uninsurable risk has a high normal variance, and the dashed lines indicate when the uninsurable risk has a thick left exponential tail.

the very poorest.[30] Note that the distributional gains from raising tax rates also rise at high income levels, but by much less because raising rates further here will not generate much additional revenue. A complementary way to frame the intuition for the effect of raising the revenue requirement is that increasing fiscal pressure on the planner leads it to prioritize a more efficient tax system (i.e., flatter/declining marginal tax rates) over a more redistributive one.

---

[30]Slemrod et al. (1994) explored the sensitivity of optimal policy with respect to the government's revenue requirement in a two-tax-bracket economy. They found that the optimal marginal tax rate in the bottom bracket is more sensitive to the revenue requirement than the rate in the top bracket. However, they consistently found decreasing marginal rates to be optimal, in contrast to our baseline calibration results.

Of course, the level of G is not the only parameter determining the shape of the optimal tax schedule. The shape of the productivity distribution also plays an important role. In particular, efficiency costs from taxation are proportional to the productivity density, and thus the government wants to keep marginal rates relatively low where the heaviest population mass is located. This plays a role in generating a U-shaped tax schedule for high values for G. In particular, there is always some convexity in the middle of the optimal marginal tax schedule, which depresses rates around the mode of the α distribution (in the range $\alpha = -0.5$ to $\alpha = 0$). This convexity appears as something resembling an upward step in the marginal tax schedule when G is low and as a U-shape when G is high.[31]

One might think that the traditional Diamond-Saez implicit formula for optimal marginal tax rates could be used to build intuition for how the level of G affects the shape of the optimal tax schedule. That equation for our economy is

$$\frac{T'(y(\alpha))}{1 - T'(y(\alpha))} = (1 + \sigma) \frac{1 - F_\alpha(\alpha)}{f_\alpha(\alpha)} \int_\alpha^\infty \left[1 - \frac{W(s) \cdot C}{c(s)}\right] \frac{c(s)}{c(\alpha)} \frac{dF_\alpha(s)}{1 - F_\alpha(\alpha)},$$

where C denotes aggregate (and average) consumption.[32] Note, however, that this expression is formally identical for all values for G! The value for G does affect the right-hand side of the formula via the endogenous consumption allocation, but the consumption allocation varies with G only because the optimal tax schedule itself varies with G. Thus, the Diamond-Saez equation provides only limited intuition for the link between fiscal pressure and optimal taxation.

Increasing Uninsurable Risk   Panel C of figure 3 plots the optimal marginal schedules when we assume there are no insurable shocks and increase uninsurable risk to leave the total model variance of earnings unchanged. We run two versions of this experiment: one in which the extra uninsurable risk is assumed normal and another in which it translates

---

[31]We have verified that if G is increased sufficiently, the optimal tax schedule eventually becomes monotonically declining.

[32]See Appendix B.2 for the derivation and a longer discussion.

into a heavy left tail for the distribution.[33] The first case gives a flatter marginal tax profile than the baseline, but the schedule is still increasing. The second case gives a U-shaped profile. Thus, increasing uninsurable risk can rationalize a U-shaped marginal tax profile, especially when the risk of very low productivity is large. This is exactly when there is significant inequality in the bottom half of the productivity distribution and thus when there are large distributional gains from taxing the moderately poor to increase lump-sum transfers benefiting the very poor (panel D). Note that, as in the experiment in which we increase required expenditure, eliminating insurance has little impact on optimal marginal rates near the top of the income distribution.[34]

Comparison to Saez (2001) The previous two experiments help to explain the difference between the increasing optimal marginal tax schedule under our baseline calibration and the examples in Saez (2001) that find U-shaped marginal rate schedules. Relative to Saez, we impose a smaller value for government purchases, and optimal transfers are smaller in our model, in part because we allow for private insurance.[35] In Saez's calibration reported in column (3) of his table 2, optimal transfers are 31 percent of GDP, and government purchases are 25 percent of GDP.[36] Thus, the required government tax take is 56 percent of GDP. In our baseline parameterization, the corresponding number is 42 percent (transfers are 21.5 percent of GDP, and purchases are 20.3 percent).

---

[33]We generalize the baseline EMG distribution for α to a normal-Laplace distribution: $\alpha = \alpha_N + \alpha_{E_1} - \alpha_{E_2}$ where $\alpha_N \sim N(\hat{\mu}_\alpha, \hat{\sigma}_\alpha^2)$, $\alpha_{E_1} \sim Exp(\lambda_\alpha)$ and $\alpha_{E_2} \sim Exp(\hat{\lambda}_\alpha)$. This second exponential component allows for a heavier than normal left tail in the log productivity distribution. The baseline calibration is nested as $\hat{\sigma}_\alpha^2 = \sigma_\alpha^2$ and $\hat{\lambda}_\alpha = \infty$. We retain our estimate for the right tail parameter $\lambda_\alpha = 2.2$ and consider two alternative values for $(\hat{\sigma}_\alpha^2, \hat{\lambda}_\alpha^{-2})$, where in each case we ensure that the model replicates the total observed empirical variance for log earnings. In the first, the extra uninsurable risk is normal: $\hat{\sigma}_\alpha^2 = \sigma_\alpha^2 + ((1+\sigma)/\sigma)^2\sigma_\varepsilon^2, \hat{\lambda}_\alpha = \infty$. In the second, the increase translates into a thicker left exponential tail: $\hat{\sigma}_\alpha^2 = \sigma_\alpha^2, \hat{\lambda}_\alpha^{-2} = ((1+\sigma)/\sigma)^2\sigma_\varepsilon^2$.

[34]Kuziemko et al. (2015) find that educating people about the extent of inequality in the United States does not significantly change their views about optimal top marginal rates.

[35]Golosov et al. (2016) and Mankiw et al. (2009) also find U-shaped marginal rates. Both papers abstract from private insurance. The Golosov et al. (2016) calibration implies that most households have very low productivity, while Mankiw et al. (2009) assume that 5 percent of households have zero productivity. Together these assumptions translate into strong fiscal pressure to finance large lump-sum transfers, which in turn translates into very high and U-shaped marginal rates.

[36]In this calibration, Saez assumes a utilitarian welfare criterion, a utility function with income effects, and a compensated elasticity of 0.5.

If we change our calibration to deliver a similar average tax rate to Saez, we also get a U-shaped profile for marginal rates. For example, in the two economies that give U-shaped optimal tax schedules in figure 3 (dashed blue lines), government purchases plus transfers are 60 percent (panel A) and 55 percent (panel C) of output.

In 2015, total U.S. government spending including public consumption, gross investment, transfer payments and interest on debt was 33.5 percent of GDP.[37] This total is smaller than the value in our baseline model and much smaller than the value in Saez's economy. Such a modest level of revenue can be raised via an upward-sloping marginal tax schedule—which is preferable from a distributional standpoint—without generating large efficiency costs.[38]

5.2 Preferences without Income Effects à la Diamond (1998)

In this section, we specialize to the case of preferences that have no income effects:

$$\log \left(c - \frac{h^{1+\sigma}}{1+\sigma} \right). \tag{18}$$

This assumption allows us to do two useful things. First, we can compare our quantitative results directly with the well-known theoretical results in Diamond (1998).[39] Second, this specification simplifies the expressions for efficiency costs, which allows us to develop a partial theoretical characterization of the comparative statics of distributional gains with respect to government purchases G, as a complement to the numerical exploration in figure 3. This analysis will reinforce the point that the shape of the optimal tax schedule is closely tied to the shape of the distributional gain function.

---

[37]National income and product accounts, table 3.1.

[38]In Section 5.3, we will explore alternative social welfare functions, which can rationalize the fact that the actual U.S. tax and transfer system is less redistributive that our utilitarian-optimal policy.

[39]To facilitate comparison to Diamond (1998), we abstract from insurable risk when considering this preference specification. Our economy is then identical to the case considered by Diamond when the $G(.)$ function in his eq. (1) is logarithmic.

Given the utility function (18), optimal tax rates must satisfy

$$\underbrace{\int_{\alpha}^{\infty} \left\{ 1 - \frac{u_c(s)}{\chi} \right\} dF_\alpha(s)}_{\tilde{D}(\alpha) \equiv [1 - F_\alpha(\alpha)]D(\alpha)} = \underbrace{\frac{1}{1+\sigma} \frac{T'(\alpha)}{1 - T'(\alpha)} f_\alpha(\alpha)}_{\tilde{E}(\alpha) \equiv [1 - F_\alpha(\alpha)]E(\alpha)} \quad \text{for all } \alpha, \tag{19}$$

where χ is the average marginal utility of consumption in the population (see Section 3.3) and $\tilde{D}(\alpha) \equiv [1 - F_\alpha(\alpha)]D(\alpha)$ and $\tilde{E}(\alpha) \equiv [1 - F_\alpha(\alpha)]E(\alpha)$ are total distributional gains and efficiency costs (recall $D(\alpha)$ and $E(\alpha)$ are per dollar of revenue raised). This is exactly equation (9) in Diamond (1998). Note that efficiency costs are increasing in the labor supply elasticity, in the marginal tax rate, and in the productivity density.

Because the marginal utility of consumption $u_c(\alpha)$ is decreasing in α under the optimal policy, there exists a productivity value α^* such that $u_c(\alpha^*) = \chi$ and thus $\tilde{D}(\alpha)$ is maximized. Note that α^* is endogenous: it depends on the shape of the marginal utility profile, which in turn depends on the tax system. Let α_m denote the mode of the distribution for α. Diamond (1998) notes that if $\alpha^* < \alpha_m$ under the optimal tax policy, then there must be a range of values for productivity $\alpha \in [\alpha^*, \alpha_m]$ in which optimal marginal tax rates are declining. The logic is simply that the optimality condition can be rearranged as

$$\frac{T'(\alpha)}{1 - T'(\alpha)} = (1 + \sigma) \frac{\tilde{D}(\alpha)}{f_\alpha(\alpha)},$$

and for $\alpha \in [\alpha^*, \alpha_m]$, $\tilde{D}(\alpha)$ is declining while $f_\alpha(\alpha)$ is increasing. But note that assuming $\alpha^* < \alpha_m$ is *equivalent* to assuming that there is a downward-sloping portion to the optimal marginal tax schedule. In discussing this condition, Diamond (1998) writes that "[t]his seems like the more interesting case, assuming that the mode of skills is near the median and the government would like to redistribute toward a fraction of the labor force well below one-half" (p. 87). This is a conjecture that the optimal system has a downward-sloping portion.

Panel A of figure 4 plots the total distributional gain term $\tilde{D}(\alpha)$ under the optimal policy
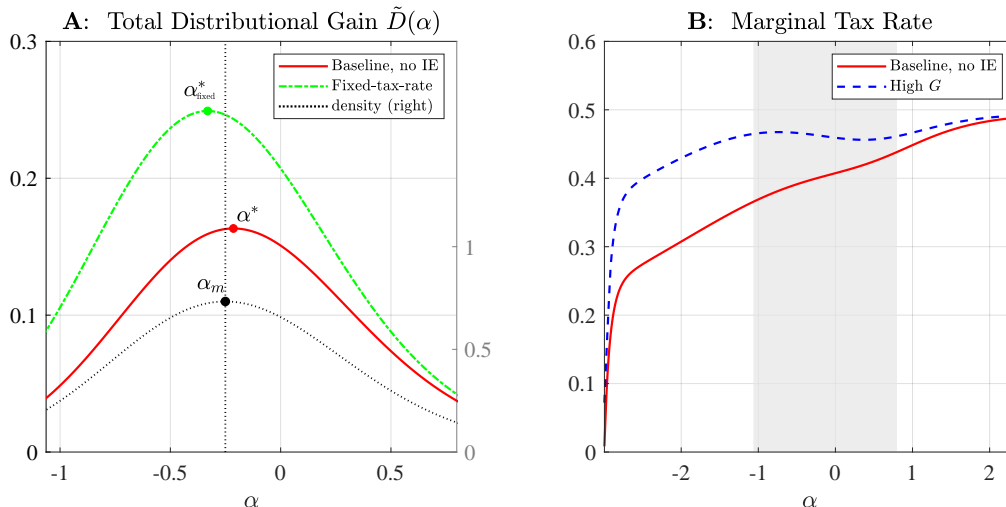
Figure 4: Preferences without Income Effects. Panel A plots distributional gains and the productivity density (between the 5th and 95th percentiles). The peak of each curve is indicated by a dot. The solid red line is the baseline case. The dotted green line corresponds to the interim case in which G is higher but marginal tax rates are unchanged. Panel B plots the optimal marginal tax schedules for the baseline (solid red) and high G (dashed blue) cases.

and the density $f_\alpha(\alpha)$ for a parameterization similar to the one described in Section 4.[40] Note that the distributional gain term peaks *after* $f_\alpha(\alpha)$(i.e., $\alpha^* > \alpha_m$), contrary to Diamond's assumption. The optimal marginal tax schedule plotted in panel B is everywhere increasing, as in our baseline calibration (see figure 2), and contrary to Diamond's intuition.

Consider now an increase in G from its baseline (low) value that is financed by reducing lump-sum transfers with an unchanged marginal tax rate schedule.

Proposition 1 *Given a utility function of the form (18), a reduction in lump-sum transfers (i) has no effect on efficiency costs, $\tilde{E}(\alpha)$, (ii) increases distributional gains $\tilde{D}(\alpha)$ for all finite α, and (iii) reduces the value α^* at which $\tilde{D}(\alpha)$ is maximized.*

Proof. See Appendix E.1. ∎

Result (i) is trivial: lump-sum transfers do not affect labor supply given the preferences in eq. (18). Result (ii) is intuitive and reflects the fact that with lower lump-sum transfers, there

---

[40]Here we assume $\sigma = 2$. The distribution for α is EMG with variance $\sigma_\alpha^2 = 0.218$ and tail parameter $\lambda_\alpha = 3.03$. Government purchases G are such that they would account for 18.8 percent of output under the tax system estimated in Section 4. Given these choices and under that tax system, the distribution for labor earnings would be identical to the EMG distribution estimated in Section 4.

is more inequality in consumption and in the marginal utility of consumption. The intuition behind result (iii) is that reducing lump-sum transfers hurts the poor disproportionately, in the sense that marginal utility becomes a more convex function of productivity. Thus, distributional gains increase relatively more at low income levels.

Panel A of figure 4 illustrates Proposition 1 with a numerical example. The dotted green line plots distributional gains when G is increased but the marginal tax schedule is unchanged relative to the (initially optimal) baseline.[41] Let α^*_{fixed} denote the distributional gain maximizing value for α in this case.

The shift in the distributional gain function can be used to interpret the change in the optimal tax schedule plotted in panel B. First, combining results (i) and (ii), it cannot be optimal to finance an increase in G solely by reducing lump-sum transfers: distributional gains (dotted green line) would then exceed efficiency costs (solid red line) at all productivity levels. This explains why optimal marginal tax rates increase across the distribution. Second, thanks to result (iii), increasing G shifts the argmax of the $\tilde{D}(\alpha)$ function to the left, holding marginal rates fixed: $\alpha^*_{\text{fixed}} < \alpha^*$. In fact, in this example, the argmax shifts from above the mode for productivity to below the mode: $\alpha^*_{\text{fixed}} < \alpha_m < \alpha^*$. This change in the shape of the $\tilde{D}(\alpha)$ function implies that the welfare gains from raising marginal tax rates (i.e., $D(\alpha) - E(\alpha)$) are larger below α_m than above α_m, which in turn accounts for why the planner raises marginal tax rates by more below α_m than above α_m. This explains why the new optimal tax schedule is flatter (panel B).[42]

5.3 Alternative Social Preferences

To this point, we have explored optimal policy assuming the planner is utilitarian, the most common assumption in the literature. We now consider alternative Pareto weight functions. Our key findings here are two. First, for a Pareto weight function that rationalizes the

---

[41]This higher value is 40 percent of output under the baseline HSV tax system.

[42]Let α^*_{G+} denote the corresponding distribution gain maximizing value for α. In this example, it turns out that $\alpha^*_{G+} < \alpha_m$. Thus, Diamond's condition for the optimal marginal tax schedule to have a downward-sloping portion is satisfied, which accounts for the U-shape of the optimal profile in panel B.

amount of redistribution embedded in the actual U.S. tax and transfer system, the optimal marginal tax schedule is again increasing, as it is under the utilitarian objective. Second, if the planner has a sufficiently strong taste for redistribution, the optimal schedule becomes first U-shaped and then decreasing.

We assume that the Pareto weight function takes the form

$$W(\alpha; \theta) = \frac{\exp(-\theta\alpha)}{\int \exp(-\theta\alpha)dF_\alpha(\alpha)} \quad \text{for } \alpha \in \mathcal{A}. \tag{20}$$

Here the parameter θ controls the planner's taste for redistribution. With a negative (positive) θ, the planner puts relatively high weight on more (less) productive agents.

This one-parameter specification nests several classic social preference specifications in the literature. The case $\theta = 0$ corresponds to the baseline utilitarian case, with equal Pareto weights on all agents. The case $\theta = -1$ corresponds to a laissez-faire planner, with planner weights inversely proportional to equilibrium marginal utility absent redistributive taxation.[43] The case $\theta \to \infty$ corresponds to the maximal desire for redistribution. We label this the Rawlsian case because in our environment, a planner with this objective function will seek to maximize the minimum level of welfare in the economy.[44]

Empirically Motivated Pareto Weight Function We are especially interested in the value for θ that rationalizes the extent of redistribution embedded in the actual U.S. tax and transfer system. Consider a Ramsey problem of the form (14) where the planner uses a Pareto weight function of the form (20) and is restricted to choosing a tax-transfer policy within the HSV class. The planner has to respect the government budget constraint and therefore effectively has a single choice variable, τ. Let $\hat{\tau}(\theta)$ denote the welfare-maximizing choice for τ given a Pareto weight function indexed by θ, and let τ^{US} denote the estimated

---

[43]A government with this objective function and the ability to apply α-specific lump-sum taxes would choose consumption proportional to productivity, $c(\alpha) \propto \exp(\alpha)$, and hours worked independent of α.

[44]With elastic labor supply and unobservable shocks, the rankings of productivity and welfare will always be aligned. So, maximizing minimum welfare is equivalent to maximizing welfare for the least productive household.

degree of progressivity for the actual U.S. tax and transfer system. We define an *empirically motivated Pareto weight function* $W(\alpha; \theta^{\text{US}})$ as the special case of the function defined in eq. (20) in which the taste for redistribution θ^{US} satisfies $\hat{\tau}(\theta^{\text{US}}) = \tau^{\text{US}}$.[45]

The Pareto weight function $W(\alpha; \theta^{\text{US}})$ is appealing for two related reasons. First, it offers a positive theory of the observed tax system: given θ^{US} a Ramsey planner restricted to the HSV functional form would choose exactly the observed degree of tax progressivity τ^{US}. Second, given $\theta = \theta^{\text{US}}$, any tax system that delivers higher welfare than the HSV function with $\tau = \tau^{\text{US}}$ must do so by redistributing in a cleverer way; by virtue of how θ^{US} is defined, simply increasing or reducing τ within the HSV class cannot be welfare improving. In this sense, the case $\theta = \theta^{\text{US}}$ isolates the efficiency gains from replacing the HSV parametric function with the optimal non-parametric schedule.

A Closed-Form Link between Tax Progressivity and Taste for Redistribution   A closed-form expression for social welfare can be derived in our economy. The first-order condition with respect to τ then offers a closed-form mapping between τ and θ.

Proposition 2 *The social preference parameter θ^{US} consistent with the observed choice for progressivity τ^{US} is a solution to the following quadratic equation:*

$$\sigma_\alpha^2 \theta^{\text{US}} - \frac{1}{\lambda_\alpha + \theta^{\text{US}}} = -\sigma_\alpha^2 (1 - \tau^{\text{US}}) - \frac{1}{\lambda_\alpha - 1 + \tau^{\text{US}}} + \frac{1}{1+\sigma} \left[\frac{1}{(1 - g^{\text{US}})(1 - \tau^{\text{US}})} - 1 \right], \quad (21)$$

where g^{US} is the observed ratio of government purchases to output.[46]

Proof. See Appendix E.2. ∎

Equation (21) is novel and useful. Given observed choices for g^{US} and τ^{US}, and estimates for the uninsurable productivity distribution parameters σ_α^2 and λ_α and for the labor elas-

---

[45]This approach to estimating a Pareto weight function can be generalized to apply to alternative tax function specifications. In particular, for any representation of the actual tax and transfer scheme T, one can always compute the value for θ that maximizes the social welfare associated with $W(\alpha; \theta)$, given the equilibrium allocations corresponding to T.

[46]The relevant root of this quadratic equation can be deduced by comparison with the special case in which $\lambda_\alpha \to \infty$, in which case one can explicitly solve for θ^{US} in closed form.

ticity parameter σ, we can immediately infer θ^{US}.[47] Given our baseline parameter values and $g^{\text{US}} = 0.188$, the implied empirically motivated taste for redistribution is $\theta^{\text{US}} = -0.517$. Thus, the fact that the current U.S. tax and transfer system is only modestly redistributive points to a weaker than utilitarian taste for redistribution.

Our finding of a negative θ may be interpreted in two ways. One is that the U.S. political system appropriately aggregates Americans' preferences, so we should use these weights to evaluate social welfare. Consistent with this idea, Weinzierl (2017) reports survey support for the idea that there should be a link between taxes paid and government benefits received, and that respondents who emphasize that principle are not enthusiastic about using the tax system to reduce inequality. An alternative interpretation is that the political system has been captured by the elites and that a utilitarian (or Rawlsian) objective would better reflect the preferences of "average" Americans. Gilens and Page (2014) find that the preferences of affluent citizens have a much greater impact on policy outcomes than the preferences of those in the middle of the income distribution. The probabilistic voting model (see Persson and Tabellini 2000) is one model that can account for this pattern.[48]

Given $\theta = \theta^{\text{US}}$, we compute the welfare gain of switching from our HSV approximation to the current tax system to the optimal Mirrleesian policy. We find a tiny welfare gain of 0.05 percent of consumption, indicating that the current tax system is close to efficient.
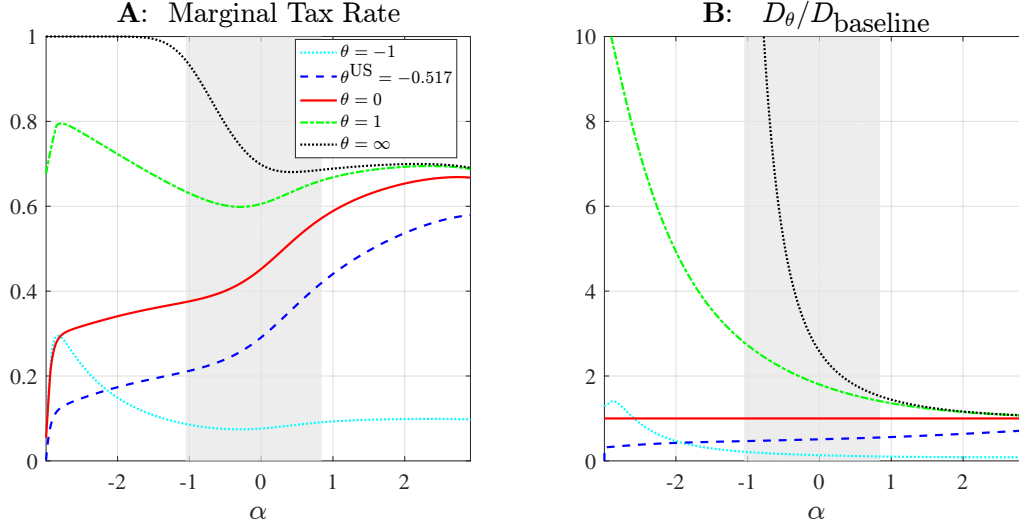
Optimal Taxation under Alternative Social Preferences Panel A of figure 5 plots optimal Mirrleesian marginal tax rate profiles for $\theta \in \left\{ -1, \theta^{\text{US}}, 0, 1, \infty \right\}$.[49] Panel B plots the corresponding distributional gain functions, in each case relative to the utilitarian baseline.

---

[47]For the purpose of inferring θ^{US}, we can treat g^{US} as exogenous.

[48]Here, two candidates for political office (who care only about getting elected) offer platforms that appeal to voters with different preferences over tax policy and over some orthogonal characteristic of the candidates. If the amount of preference dispersion over this orthogonal characteristic is systematically declining in labor productivity, then by tilting their tax platforms in a less progressive direction, candidates can expect to attract more marginal voters than they lose. Thus, in equilibrium, both candidates offer tax policies that maximize social welfare under a Pareto weight function that puts more weight on more productive (and more tax-sensitive) households.

[49]When we compute the Rawlsian case, we simply maximize welfare for the lowest α type in the economy, subject to the usual feasibility and incentive constraints. A numerical value for θ is not required for this program.

Figure 5: Alternative Social Preferences. Panel A plots optimal marginal tax schedules corresponding to Pareto weight functions with the following values for the taste for redistribution parameter: $\theta = -1$ (laissez-faire), $\theta = -0.517$ (empirically motivated), $\theta = 0$ (utilitarian), $\theta = 1$ (more redistributive), and $\theta = \infty$ (Rawlsian). Panel B plots distributional gain functions for the same set of values for θ, each relative to the baseline utilitarian objective ($\theta = 0$).

The first key message from panel A is that the optimal marginal tax schedule is increasing under the empirically motivated Pareto weight function, as it is for the utilitarian case considered previously.

Considering Pareto weight functions with a stronger than utilitarian taste for redistribution, the optimal marginal tax schedule shifts upward. In addition, the optimal schedule changes from upward sloping to U-shaped. These changes are qualitatively similar to those in the earlier experiments in which we confronted a utilitarian planner with larger expenditure requirements or with more uninsurable productivity dispersion. The reason is that a stronger taste for redistribution implies larger distributional gains from raising marginal tax rates at low income levels (panel B). Thus, the planner is willing to tolerate the larger efficiency costs associated with higher marginal tax rates in order to increase lump-sum transfers that benefit the very poorest.[50]

Note that in the laissez-faire case ($\theta = -1$), the optimal marginal tax schedule is quite

---

[50]Welfare gains from tax reform here are very large, reaching 662 percent of consumption in the Rawlsian case (see Appendix F.3). They are so large because the least productive households rely almost entirely on transfers for consumption, and the Rawlsian planner maximizes transfers.

different, with low and generally declining marginal rates. In this case, the planner does not perceive large distributional gains from downward redistribution, and focusses instead on efficiency. From an efficiency standpoint, a generally declining marginal tax rate is desirable because it implies a low average marginal rate.

5.4 Summary of Findings

We take away several related messages from this analysis. First and foremost, the U-shaped profile for marginal rates emphasized by Saez (2001) is not a general feature of an optimal tax system: our model with private insurance calibrated to the United States indicates upward-sloping marginal tax rates. Second, the commonly held notion that marginal rates should be high at the bottom in order to rapidly tax away transfers intended only for the very poor is misleading. In particular, reducing fiscal pressure on the government (e.g., by reducing G) both *increases* lump-sum transfers and *reduces* marginal tax rates on low incomes. Third, if an optimizing government needs to increase net tax revenue (e.g., to finance a war), it should do so primarily by raising marginal tax rates at the bottom of the productivity distribution rather than at the top.[51] Finally, a useful way to build intuition about optimal tax design is to conceptualize the optimal tax schedule as trading off distributional gains versus efficiency costs, where this trade-off is mediated by the amount of fiscal pressure the planner faces: higher fiscal pressure leads the planner to prioritize efficiency, and a U-shaped or downward-sloping marginal tax schedule.

6 Further Explorations

We extend our exploration in two different directions. First, we compare optimal non-parametric "Mirrleesian" policies to the best that can be achieved when the tax and transfer scheduled is restricted to simple functional forms. Next, we explore optimal tax reform when the planner faces the additional constraint that no households can be left worse off relative

---

[51]In contrast, the "equal sacrifice" principle (see, e.g., Scheve and Stasavage 2016) would dictate increasing tax progressivity during wartime, based on the idea that the rich should sacrifice more through taxes if the poor are asked to do the actual fighting.

to the current tax system.[52]

6.1 Mirrlees versus Ramsey Taxation

We compute the best tax and transfer systems in two parametric classes: (i) the HSV class and (ii) the affine class. Assuming a utilitarian objective, we compare allocations and welfare in each of those cases with their counterparts under the fully optimal Mirrleesian policy and under our baseline HSV approximation to the current U.S. tax and transfer system.

Table 2 presents outcomes for each tax function. The outcomes reported, relative to the baseline (HSV$^{\text{US}}$), are (i) the change in welfare, ω (percent), (ii) the change in aggregate output, ΔY (percent), (iii) the average income-weighted marginal tax rate, $\overline{T'}$, and (iv) the size of the net transfer (transfers net of taxes) received by the lowest α type household, relative to average income, Tr/Y.[53]

Moving from the baseline policy HSV$^{\text{US}}$ to the optimal Mirrleesian one, as noted previously, translates into a much more redistributive tax system, with a higher average marginal tax rate and larger lump-sum net transfers. This comes at the cost of a 7.3 percent decline in output relative to the baseline. However, the additional redistribution translates into an overall welfare gain of 2.07 percent.

When we restrict the new fiscal policy to the parametric HSV class, we find an increase from 0.181 to 0.331 in the progressivity parameter τ. This reform generates a welfare gain equivalent to giving all households 1.65 percent more consumption, which is 80 percent of the gain under the best-possible Mirrleesian policy. The best policy in the affine class does less well, delivering only 66 percent of the welfare gains from the optimal Mirrlees reform.

---

[52]In Appendix F, we conduct extensive additional sensitivity analyses.

[53]We define the welfare gain of moving from policy T to $\hat{T}$ as the percentage increase in consumption for all agents under policy T that leaves the planner indifferent between T and $\hat{T}$. Given logarithmic utility in consumption, this gain, which we denote $\omega(T, \hat{T})$, is given by $1 + \omega(T, \hat{T}) = \exp\left[V(\hat{T}, \theta) - V(T, \theta)\right]$, where $V(T, \theta)$ denotes the planner's realized value under a policy T given a taste for redistribution θ :

$$V(T, \theta) = \int W(\alpha; \theta) \int \left[\log c(\alpha, \varepsilon; T) - \frac{h(\alpha, \varepsilon; T)^{1+\sigma}}{1 + \sigma}\right] dF_\varepsilon(\varepsilon) dF_\alpha(\alpha).$$

For the welfare numbers in table 2, the baseline policy T is HSV$^{\text{US}}$ and $\theta = 0$.
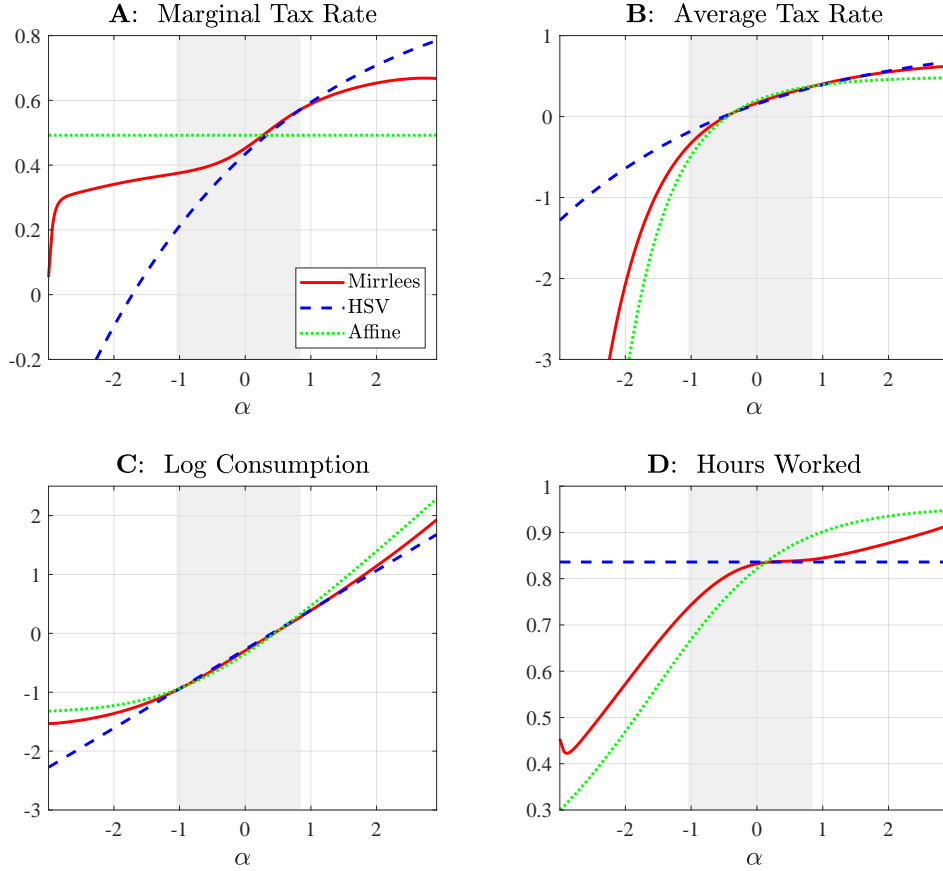
Table 2: Mirrlees versus Ramsey Taxation

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV$^{\mathrm{US}}$	$\lambda : 0.840$	$\tau : 0.181$	–	–	0.335	0.023
HSV	$\lambda : 0.817$	$\tau : 0.331$	1.65	-6.53	0.466	0.064
Affine	$\tau_0 : -0.259$	$\tau_1 : 0.492$	1.36	-7.31	0.477	0.281
Mirrlees			2.07	-7.32	0.491	0.215

This indicates that for welfare, it is more important that marginal tax rates increase with income—which the HSV functional form accommodates but which the affine scheme rules out—than that the government provides universal lump-sum transfers—which only the affine scheme admits.

Figure 6 plots marginal and average tax schedules (panels A and B) and decision rules for consumption and hours (panels C and D) for each best-in-class tax and transfer scheme. Over most of the shaded area, covering 90 percent of the population, allocations under the HSV policy are very similar to those in the constrained efficient Mirrlees case. Allocations are similar because the HSV marginal and average tax schedules are broadly similar to those under the optimal policy. As we have already emphasized, the profile for marginal tax rates that decentralizes the constrained efficient allocation is increasing in productivity, and the optimal HSV schedule mirrors this. Because marginal rates are too high at the top under the HSV scheme, very productive agents work too little. At the same time, because transfers are too small, very unproductive agents work too much. However, the mass of agents in these tails is small.

Panel A of figure 6 offers a straightforward visualization of why an affine tax schedule is welfare inferior to the HSV form. Because any affine tax function features a constant marginal rate, an affine scheme cannot replicate the increasing optimal marginal tax schedule. Under the best affine scheme, low-wage households face marginal rates that are too high and work too little relative to the constrained efficient allocation. At the same time, because marginal

Figure 6: Mirrlees versus Ramsey Taxation. The figure contrasts allocations under the HSV tax system, the affine system, and the Mirrlees system. Panels A and B plot marginal and average tax schedules, and panels C and D plot decision rules for consumption and hours worked (for an agent with average ε).

tax rates are too low at high income levels, high-productivity workers consume too much.[54]

6.2 Pareto-Improving Tax Reforms

Exploring Pareto-improving tax reforms is of interest for two reasons. First, one would expect Pareto-improving reforms to be easier to implement in practice compared with reforms that create winners and losers. Second, as figure 5 illustrates, the welfare-maximizing policy under the traditional Mirrlees approach is highly sensitive to the taste for redistribution embedded in the planner's objective function, and people might disagree about how much emphasis the planner should put on reducing inequality. Insisting that any tax reform be

---

[54]Note that an affine tax scheme might be appealing for reasons that our theoretical framework does not capture, such as being easy to communicate and administer.

Pareto improving makes the choice of planner weights less critical.

To characterize Pareto-improving reforms, we adapt the Mirrlees problem (10) by adding a set of additional constraints of the form $U(\alpha, \alpha) \geq U^{\text{US}}(\alpha)$ for all $\alpha \in \mathcal{A}$, where $U^{\text{US}}(\alpha)$ denotes expected utility for a household of type α under our HSV approximation to the current U.S. tax and transfer system.[55]

Figure 7 plots tax rates and decision rules under three different tax systems: the optimal Mirrlees scheme, our approximation to the current system, and the scheme that is optimal subject to also being weakly Pareto improving. For both the Mirrlees and Pareto improving cases, we assume the planner has a utilitarian objective.

Because the Mirrleesian planner chooses a more redistributive tax scheme than the current system (panel A), relatively productive households are worse off, and thus the Mirrlees reform is not Pareto improving (panel B). In fact, Mirrleesian tax reform leaves 44.5 percent of households worse off.

Consider now the optimal Pareto-improving reform. The Pareto-improving constraints bind for households in the middle of the productivity distribution. In this region, where the majority of households are located, allocations and tax rates are identical to those under the current tax system. We formalize this result in the following proposition.

Proposition 3 *Let T^{US} be the current tax system, and let T^{PI} be the optimal Pareto-improving system. If the Pareto-improving constraints bind in an open interval $\Gamma \subset \mathcal{A}$, i.e., $U(\alpha; T^{US}) = U(\alpha; T^{PI})$ for all $\alpha \in \Gamma$, and if T^{US} and T^{PI} are differentiable on Γ, then allocations and tax rates under T^{PI} are identical to those under T^{US} for all $\alpha \in \Gamma$.*

Proof. See Appendix E.3. ∎

The Pareto-improving reform leaves most households indifferent relative to the baseline tax system. However, households in both tails of the productivity distribution are strictly better off. In the right tail, marginal tax rates are lower than under the baseline HSV tax

---

[55]Adding these Pareto-improving constraints is challenging computationally because the pattern of which subset of constraints is binding at the optimum is unknown ex ante. We describe our computational approach in Appendix C.2.
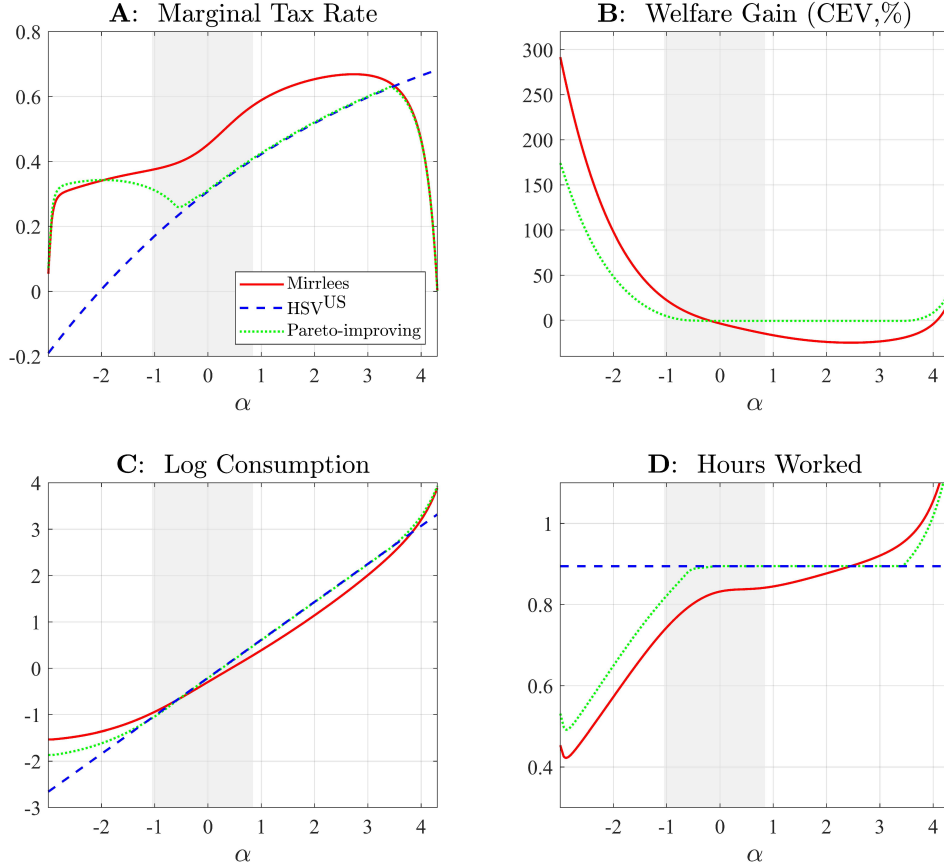
Figure 7: Pareto-improving Tax Reforms. The figure plots marginal tax schedule and welfare gains (CEV) (panels A and B) and decision rules for consumption and hours worked (panels C and D) under the Mirrlees system, the approximation to the current system, and the system that is optimal subject to being weakly Pareto improving The plot for hours worked is for an agent with average ε.

system and decline to zero at the upper bound for productivity, an established property of any Pareto-efficient system. These lower tax rates leave the very rich better off and also increase revenue that can be redistributed to the poor. In the left tail of the productivity distribution, marginal tax rates under the Pareto-improving reform are higher than under the HSV system and are everywhere strictly positive. Again, this change generates additional tax revenue that can be used to increase lump-sum transfers.

However, the welfare gains from Pareto-improving tax reform turn out to be small. The Pareto-improving reform generates a gain equivalent to giving all households 0.41 percent more consumption, compared to a 2.07 percent gain in the same economy when the planner

is not required to leave all households weakly better off.

Note that insisting that tax reforms be Pareto improving and endowing the planner with a weak taste for redistribution (Section 5.3) are two different ways to reduce the welfare gains from making the tax system more redistributive. In both cases, we find small welfare gains from tax reform and optimal systems that resemble the current one. We conclude that the majority of the welfare gains in the utilitarian baseline Mirrlees experiment reflect gains from redistributing the tax burden toward higher-income households rather than gains from making the system more efficient.

7   Conclusions

We now highlight five lessons from our analysis that should be useful for future work that aims to provide quantitative advice on tax and transfer design.

First, incorporating private insurance has an important quantitative impact on the shape of the optimal tax and transfer schedule.

Second, for interpreting the shape of the optimal tax schedule, it is useful to consider how much pressure the planner faces to raise revenue. When fiscal pressure is low, the optimal marginal tax schedule will be an upward-sloping function of income. As fiscal pressure is progressively increased, the optimal schedule becomes first flatter, then U-shaped in income, and ultimately downward sloping.

Third, the specification of the planner's objective function has an enormous impact on policy prescriptions. We have proposed a functional form for Pareto weights indexed by a single taste for redistribution parameter and have argued that a natural baseline for this parameter is the value that rationalizes the progressivity embedded in the current tax and transfer system.

Fourth, the optimal profile for marginal tax rates may be well approximated by the simple two-parameter power function used by Benabou (2000) and Heathcote et al. (2017).

Fifth, Pareto-improving tax reforms may imply that most households face no changes in

average or marginal tax rates.

Our model environment could be enriched along several dimensions. First, labor supply is the only decision margin distorted by taxes. Although this has been the focus of the optimal tax literature, skill investment and entrepreneurial activity are additional margins that are likely sensitive to the tax system. Second, our model features no uninsurable life-cycle shocks to productivity: modeling such shocks would allow the Mirrlees planner to increase welfare by making taxes history dependent. The associated welfare gains may be modest, however, given that privately uninsurable life-cycle shocks are small relative to permanent productivity differences.

References

ATTANASIO, O. AND L. PISTAFERRI (2014): "Consumption Inequality over the Last Half Century: Some Evidence Using the New PSID Consumption Measure," *American Economic Review*, 104, 122–126.

BENABOU, R. (2000): "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review*, 90(1), 96–129.

BOURGUIGNON, F. AND A. SPADARO (2012): "Tax-Benefit Revealed Social Preferences," *Journal of Economic Inequality*, 10, 75–108.

BRENDON, C. (2013): "Efficiency, Equity, and Optimal Income Taxation," mimeo, European University Institute.

CHETTY, R. AND E. SAEZ (2010): "Optimal Taxation and Social Insurance with Endogenous Private Insurance," *American Economic Journal: Economic Policy*, 2, 85–114.

CONESA, J. C. AND D. KRUEGER (2006): "On the Optimal Progressivity of the Income Tax Code," *Journal of Monetary Economics*, 53, 1425–1450.

DIAMOND, P. AND E. SAEZ (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165–190.

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

FARHI, E. AND I. WERNING (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80, 596–635.

FELDSTEIN, M. S. (1969): "The Effects on Taxation on Risk Taking," *Journal of Political Economy*, 77, 755–764.

GILENS, M. AND B. I. PAGE (2014): "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens," *Perspectives on Politics*, 12, 564–581.

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2016): "Redistribution and Social Insurance," *American Economic Review*, 106, 359–386.

GOLOSOV, M. AND A. TSYVINSKI (2015): "Policy Implications of Dynamic Public Finance," *Annual Review of Economics*, 7, 147–171.

GOUVEIA, M. AND R. P. STRAUSS (1994): "Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis," *National Tax Journal*, 47, 317–339.

HEATHCOTE, J., F. PERRI, AND G. L. VIOLANTE (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 13, 15–51.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104, 2075–2126.

——— (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132, 1693–1754.

HEATHCOTE, J. AND H. TSUJIYAMA (2017): "Inverse Optimal Taxation in Closed Form," mimeo.

HENDREN, N. (2017): "Efficient Welfare Weights," Working Paper 20351, National Bureau of Economic Research.

KEANE, M. P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49, 961–1075.

KUZIEMKO, I., M. NORTON, E. SAEZ, AND S. STANTCHEVA (2015): "How Elastic Are Preferences for Redistribution: Evidence from Randomized Survey Experiments," *American Economic Review*, 105, 1478–1508.

LOW, H. AND L. PISTAFERRI (2015): "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off," *American Economic Review*, 105, 2986–3029.

MANKIW, N. G., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23, 147–174.

MEYER, B. D. AND J. X. SULLIVAN (2017): "Consumption and Income Inequality in the U.S. Since the 1960s," Working Paper 23655, National Bureau of Economic Research.

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

PERSSON, M. (1983): "The Distribution of Abilities and the Progressive Income Tax," *Journal of Public Economics*, 22, 73–88.

PERSSON, T. AND G. TABELLINI (2000): *Political Economics: Explaining Economic Policy*, Cambridge, MA: MIT Press.

RAMSEY, F. P. (1927): "A Contribution to the Theory of Taxation," *Economic Journal*, 37, 47–61.

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

SAEZ, E., J. SLEMROD, AND S. H. GIERTZ (2012): "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature*, 50, 3–50.

SAEZ, E. AND S. STANTCHEVA (2016): "Generalized Social Welfare Weights for Optimal Tax Theory," *American Economic Review*, 106, 24–45.

SCHEVE, K. AND D. STASAVAGE (2016): *Taxing the Rich: A History of Fiscal Fairness in the United States and Europe*, Princeton, NJ: Princeton University Press.

SLEMROD, J., S. YITZHAKI, J. MAYSHAR, AND M. LUNDHOLM (1994): "The optimal two-bracket linear income tax," *Journal of Public Economics*, 53, 269–290.

TODA, A. A. AND K. WALSH (2015): "The Double Power Law in Consumption and Implications for Testing Euler Equations," *Journal of Political Economy*, 123, 1177–1200.

TUOMALA, M. (1990): *Optimal Income Tax and Redistribution*, Oxford: Clarendon Press.

——— (2010): "On Optimal Non-linear Income Taxation: Numerical Results Revisited," *International Tax and Public Finance*, 17, 259–270.

WEINZIERL, M. (2014): "The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice," *Journal of Public Economics*, 118, 128–142.

WEINZIERL, M. C. (2017): "Popular Acceptance of Inequality Due to Innate Brute Luck and Support for Classical Benefit-based Taxation," *Journal of Public Economics*, 155, 54–63.

WERNING, I. (2007): "Pareto Efficient Income Taxation," mimeo.

Optimal Income Taxation: Mirrlees Meets Ramsey
Online Appendix

Jonathan Heathcote and Hitoshi Tsujiyama

Contents

A   Alternative Specification

A.1   Insurance via Family versus Insurance via Financial Markets

We show that, with one caveat, all the analysis of the paper remains unchanged if we consider an alternative model of insurance against ε shocks. In particular, we put aside the model of the family and suppose instead that each agent is autonomous, buys private insurance in decentralized financial markets against ε shocks, and is taxed at the individual level.

Decentralized Economy   Suppose agents first observe their idiosyncratic uninsurable component α and then trade in insurance markets to purchase private insurance at actuarially fair prices against ε. The budget constraint for an agent with α is now given by

$$\int B(\alpha, \varepsilon)Q(\varepsilon)d\varepsilon = 0, \tag{A1}$$

where $B(\alpha, \varepsilon)$ denotes the quantity (positive or negative) of insurance claims purchased that pay a unit of consumption if and only if the draw for the insurable shock is $\varepsilon \in \mathcal{E}$ and where $Q(E)$ is the price of a bundle of claims that pay one unit of consumption if and only if $\varepsilon \in E \subset \mathcal{E}$ for any Borel set E in $\mathcal{E}$. In equilibrium, these insurance prices must be actuarially fair, which implies $Q(E) = \int_E dF(\varepsilon)$.

In this decentralization, taxation occurs at the individual level and applies to earnings plus insurance payments. Thus, the individual's budget constraints are

$$c(\alpha, \varepsilon) = y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon)) \qquad \text{for all } \varepsilon, \tag{A2}$$

where individual income before taxes and transfers is given by

$$y(\alpha, \varepsilon) = \exp(\alpha + \varepsilon)h(\alpha, \varepsilon) + B(\alpha, \varepsilon) \qquad \text{for all } \varepsilon. \tag{A3}$$

The individual agent's problem is then to choose $c(\alpha, \cdot)$, $h(\alpha, \cdot)$, and $B(\alpha, \cdot)$ to maximize expected utility (4) subject to eqs. (A1), (A2), and (A3). The equilibrium definition in this case is similar to that for the specification in which insurance takes place within the family.

It is straightforward to establish that the FOCs for this problem are exactly the same as those for the family model of insurance with taxation at the individual level. Thus, given the same tax function T, allocations with the two models of insurance are the same. Part of the reason for this result is that each family is small relative to the entire economy and takes the tax function as parametric. Moreover, taxes on income after private insurance / family transfers do not crowd out risk sharing with respect to ε shocks.

Planner's Problem   Now consider the Mirrlees planner's problem in the environment with decentralized insurance against ε shocks. We first establish that if the planner is restricted to only ask agents to report α, the solution is the same as the one described previously for the family model. We then speculate about what might change if the planner can also ask agents to report ε.

Suppose that the planner asks individuals to report α before they draw ε. Then, given their true type α and a report $\tilde{\alpha}$ and associated contract $(c(\tilde{\alpha}), y(\tilde{\alpha}))$, agents shop for insurance. Consider the agent's problem at this stage:

$$\max_{\{h(\alpha, \tilde{\alpha}, \varepsilon), B(\alpha, \tilde{\alpha}, \varepsilon)\}} \int \left\{ \frac{c(\tilde{\alpha})^{1-\gamma}}{1-\gamma} - \frac{h(\alpha, \tilde{\alpha}, \varepsilon)^{1+\sigma}}{1+\sigma} \right\} dF_\varepsilon(\varepsilon),$$

subject to

$$\int B(\alpha, \tilde{\alpha}, \varepsilon) Q(\varepsilon) d\varepsilon = 0,$$

$$\exp(\alpha + \varepsilon) h(\alpha, \tilde{\alpha}, \varepsilon) + B(\alpha, \tilde{\alpha}, \varepsilon) = y(\tilde{\alpha}).$$

Substituting the second constraint into the first, and assuming actuarially fair insurance prices, we have

$$\int \left[y(\tilde{\alpha}) - \exp(\alpha + \varepsilon) h(\alpha, \tilde{\alpha}, \varepsilon) \right] dF_\varepsilon(\varepsilon) = 0.$$

The first-order condition for hours is

$$h(\alpha, \tilde{\alpha}, \varepsilon)^\sigma = \mu(\alpha, \tilde{\alpha}) \exp(\alpha + \varepsilon),$$

where the budget constraint can be used to solve out for the multiplier $\mu(\alpha, \tilde{\alpha})$:

$$h(\alpha, \tilde{\alpha}, \varepsilon) = \frac{y(\tilde{\alpha})}{\exp(\alpha)} \frac{\exp(\varepsilon)^{\frac{1}{\sigma}}}{\int \exp(\varepsilon)^{\frac{1+\sigma}{\sigma}} dF_\varepsilon(\varepsilon)}.$$

Now note that this expression is exactly the same as the one for the family planner decentralization (the first-order condition with respect to hours from problem (8)). Moreover, in both cases $c(\alpha, \varepsilon) = c(\tilde{\alpha})$. It follows that for any values for $(\alpha, \tilde{\alpha})$, expected utility for the agent in this decentralization with private insurance markets is identical to welfare for the family head in the decentralization with insurance within the family. Thus, the set of allocations that are incentive compatible when the social planner interacts with the family head are the same as those that are incentive compatible when the planner interacts agent by agent. It follows that the solution to the social planner's problem is the same under both models of ε insurance. Similarly, the income tax schedule that decentralizes the Mirrlees solution is also the same under both models of ε insurance, and marginal tax rates are given

in both cases by eq. (13). Note that marginal tax rates do not vary with ε under either insurance model because income (including insurance payouts/family transfers) does not vary with ε.[1]

Finally, note that if insurance against ε is achieved via decentralized financial markets, the planner could conceivably ask agents to report ε after the ε shock is drawn and offer allocations for consumption $c(\tilde{\alpha}, \tilde{\varepsilon})$ and income $y(\tilde{\alpha}, \tilde{\varepsilon})$ indexed to reports of both α and ε. With decentralized insurance, the planner might be able to offer contracts that separate agents with different values for ε (recall that under the family model for insurance, this was not possible). One might think there would be no possible welfare gain to doing so, since private insurance already appears to deliver an efficient allocation of hours and consumption within any group of agents sharing the same α. However, it is possible that by inducing agents to sacrifice perfect insurance with respect to ε, the planner can potentially loosen incentive constraints and thereby provide better insurance with respect to α.[2] We plan to explore this issue in future work. For now, we simply focus on the problem in which the planner offers contracts contingent only on α, which is the natural benchmark under our baseline interpretation that the family is the source of insurance against shocks to ε.

A.2 Dynamic Model with Life-Cycle Shocks

Consider the following overlapping-generations economy. Individuals live for two periods, $t = 1$ and $t = 2$. There is no discounting ($\beta = 1$) and individuals can borrow and lend freely at a gross interest rate $R = 1$. We assume log utility for consumption as in our quantitative section.

At labor market entry, individuals draw a "permanent wage" $\alpha \sim F_\alpha$. Wages grow over the life cycle at gross rate ρ. The wage at age t is

$$w_t(\alpha, \varepsilon_t) = \frac{2\rho^{t-1}}{1+\rho} \exp(\alpha) \exp(\varepsilon_t),$$

where ε_t denotes an *i.i.d.* insurable shock, drawn from F_ε anew at each age. The average wage (averaging by t, by α, and by ε) is equal to one, by construction.

---

[1]It is clear that the Mirrlees solution could equivalently be decentralized using consumption taxes. In that case we would get

$$1 + T'(c^*(\alpha)) = \frac{c^*(\alpha)^{-\gamma} \exp(\alpha)^{1+\sigma} \left(\int \exp(\varepsilon)^{\frac{1+\sigma}{\sigma}} dF_\varepsilon(\varepsilon) \right)^\sigma}{y^*(\alpha)^\sigma}.$$

[2]When we introduce publicly observable (but privately uninsurable) differences in productivity, we see that constrained efficient allocations typically have the property that agents with the same unobservable component α but different observable components of productivity κ are allocated different consumption (see Section F.5.2).

The individual's log wage is

$$\log w_t(\alpha, \varepsilon_t) = \log\left(\frac{2\rho^{t-1}}{1+\rho}\right) + \alpha + \varepsilon_t,$$

and thus cross-sectional dispersion in wages has three uncorrelated components related to age t, permanent income α, and insurable shocks ε.

We assume a progressive tax on consumption, so that c units of consumption requires $\left(\frac{c}{\lambda}\right)^{\frac{1}{1-\tau}}$ units of income.[3]

Absent borrowing constraints, the individual problem can be written with a single lifetime budget constraint. In particular, an individual with given values for α and ρ chooses $\{c_1(\alpha, \varepsilon_1), c_2(\alpha, \varepsilon_1, \varepsilon_2).h_1(\alpha, \varepsilon_1), h_2(\alpha, \varepsilon_1, \varepsilon_2)\}$ to solve

$$\max \iint \left\{\log c_1(\alpha, \varepsilon_1) - \frac{h_1(\alpha, \varepsilon_1)^{1+\sigma}}{1+\sigma} + \log c_2(\alpha, \varepsilon_1, \varepsilon_2) - \frac{h_2(\alpha, \varepsilon_1, \varepsilon_2)^{1+\sigma}}{1+\sigma}\right\} dF_\varepsilon(\varepsilon_1) dF_\varepsilon(\varepsilon_2)$$

subject to

$$\int Q(\varepsilon_1)\lambda^{\frac{-1}{1-\tau}}c_1(\alpha, \varepsilon_1)^{\frac{1}{1-\tau}}d\varepsilon_1 + \iint Q(\varepsilon_1, \varepsilon_2)\lambda^{\frac{-1}{1-\tau}}c_2(\alpha, \varepsilon_1, \varepsilon_2)^{\frac{1}{1-\tau}}d\varepsilon_1 d\varepsilon_2$$

$$\leq \int Q(\varepsilon_1)\frac{2}{1+\rho}\exp(\alpha)\exp(\varepsilon_1)h_1(\alpha, \varepsilon_1)d\varepsilon + \iint Q(\varepsilon_1, \varepsilon_2)\frac{2\rho}{1+\rho}\exp(\alpha)\exp(\varepsilon_2)h_2(\alpha, \varepsilon_1, \varepsilon_2)d\varepsilon_1 d\varepsilon_2,$$

where $Q(\varepsilon_1)$ and $Q(\varepsilon_1, \varepsilon_2)$ are the prices of insurance contracts that deliver consumption in the corresponding idiosyncratic states. In equilibrium, these prices must be actuarially fair, implying, for example, $Q(E) = \int_E dF_\varepsilon(\varepsilon)$ for any Borel set E in $\mathcal{E}$.

Let ζ denote the multiplier on the budget constraint. Given fairly priced insurance, the FOCs for consumption choices are

$$\frac{1}{c_1(\alpha, \varepsilon_1)} = \zeta\lambda^{\frac{-1}{1-\tau}}\frac{1}{1-\tau}c_1(\alpha, \varepsilon_1)^{\frac{1}{1-\tau}-1},$$

$$\frac{1}{c_2(\alpha, \varepsilon_1, \varepsilon_2)} = \zeta\lambda^{\frac{-1}{1-\tau}}\frac{1}{1-\tau}c_2(\alpha, \varepsilon_1, \varepsilon_2)^{\frac{1}{1-\tau}-1}.$$

which immediately imply

$$c_1(\alpha, \varepsilon_1) = c_2(\alpha, \varepsilon_1, \varepsilon_2) = c(\alpha),$$

$$\zeta = (1-\tau)\lambda^{\frac{1}{1-\tau}}c(\alpha)^{\frac{-1}{1-\tau}}.$$

---

[3]This tax system can equivalently be described as an HSV-style progressive tax on income where savings are tax deductible (see Heathcote et al. 2019).

The FOCs for hours are

$$h_1(\alpha, \varepsilon_1) = \left[(1-\tau)\lambda^{\frac{1}{1-\tau}}c(\alpha)^{\frac{-1}{1-\tau}}\right]^{\frac{1}{\sigma}}\left(\frac{2}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\alpha\right)\exp\left(\frac{1}{\sigma}\varepsilon_1\right),$$

$$h_2(\alpha, \varepsilon_1, \varepsilon_2) = \left[(1-\tau)\lambda^{\frac{1}{1-\tau}}c(\alpha)^{\frac{-1}{1-\tau}}\right]^{\frac{1}{\sigma}}\left(\frac{2\rho}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\alpha\right)\exp\left(\frac{1}{\sigma}\varepsilon_2\right).$$

Finally, we can solve for $c(\alpha)$ from the budget constraint:

$$2\lambda^{\frac{-1}{1-\tau}}c(\alpha)^{\frac{1}{1-\tau}} = \left[(1-\tau)\lambda^{\frac{1}{1-\tau}}c(\alpha)^{\frac{-1}{1-\tau}}\right]^{\frac{1}{\sigma}}\exp(\alpha)^{\frac{1+\sigma}{\sigma}}\left(\frac{2}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}\left\{\mathbb{E}\left[\exp(\varepsilon_1)^{\frac{1+\sigma}{\sigma}}\right]+\rho^{\frac{1+\sigma}{\sigma}}\mathbb{E}\left[\exp(\varepsilon_2)^{\frac{1+\sigma}{\sigma}}\right]\right\}$$

$$\Rightarrow c(\alpha) = \lambda(1-\tau)^{\frac{1-\tau}{1+\sigma}}X^{\frac{\sigma(1-\tau)}{1+\sigma}}\exp(\alpha)^{1-\tau},$$

$$\text{where } X = \mathbb{E}\left[\exp(\varepsilon)^{\frac{1+\sigma}{\sigma}}\right]\frac{1}{2}\sum_{t=1}^{2}\left(\frac{2\rho^{t-1}}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}.$$

Substituting the expression for $c(\alpha)$ into the decision rules for hours gives

$$h_1(\alpha, \varepsilon_1) = h_1(\varepsilon_1)$$

$$= (1-\tau)^{\frac{1}{\sigma}}\lambda^{\frac{1}{\sigma(1-\tau)}}\left[\lambda(1-\tau)^{\frac{1-\tau}{1+\sigma}}X^{\frac{\sigma(1-\tau)}{1+\sigma}}\right]^{\frac{-1}{\sigma(1-\tau)}}\left(\frac{2}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\varepsilon_1\right)$$

$$= (1-\tau)^{\frac{1}{1+\sigma}}X^{\frac{-1}{1+\sigma}}\left(\frac{2}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\varepsilon_1\right),$$

$$h_2(\alpha, \varepsilon_2) = h_2(\varepsilon_2)$$

$$= (1-\tau)^{\frac{1}{1+\sigma}}X^{\frac{-1}{1+\sigma}}\left(\frac{2\rho}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\varepsilon_2\right).$$

Earnings are given by

$$y_t(\alpha, \varepsilon_t) = \frac{2\rho^{t-1}}{1+\rho}\exp(\alpha)\exp(\varepsilon_t)(1-\tau)^{\frac{1}{1+\sigma}}X^{\frac{-1}{1+\sigma}}\left(\frac{2\rho^{t-1}}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\varepsilon_t\right)$$

$$= (1-\tau)^{\frac{1}{1+\sigma}}X^{\frac{-1}{1+\sigma}}\exp(\alpha)\left(\frac{2\rho^{t-1}}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}\exp\left(\frac{1+\sigma}{\sigma}\varepsilon_t\right).$$

Substituting the expressions for $c(\alpha)$, $h_1(\alpha, \varepsilon_1)$ and $h_2(\alpha, \varepsilon_2)$ into the expression for

expected lifetime utility, conditional on α, gives

$$
\begin{aligned}
U(\alpha) &= 2\log\left(\lambda\,(1-\tau)^{\frac{1-\tau}{1+\sigma}}X^{\frac{\sigma(1-\tau)}{1+\sigma}}\exp(\alpha)^{1-\tau}\right) \\
&\quad -\frac{(1-\tau)X^{-1}}{1+\sigma}\left\{\left(\frac{2}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}\mathbb{E}\left[\exp\left(\frac{1+\sigma}{\sigma}\varepsilon_1\right)\right]+\left(\frac{2\rho}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}\mathbb{E}\left[\exp\left(\frac{1+\sigma}{\sigma}\varepsilon_2\right)\right]\right\} \\
&= 2\log\left(\lambda\,(1-\tau)^{\frac{1-\tau}{1+\sigma}}X^{\frac{\sigma(1-\tau)}{1+\sigma}}\exp(\alpha)^{1-\tau}\right)-2\left(\frac{1-\tau}{1+\sigma}\right).
\end{aligned}
$$

Equivalence to Static Model We now compare these allocations to those from the static model studied in the paper. Let tildes index policy parameters and idiosyncratic shocks in the static model. In the static model we have[4]

$$
\begin{aligned}
w(\tilde{\alpha},\tilde{\varepsilon}) &= \exp(\tilde{\alpha})\exp(\tilde{\varepsilon}), \\
c(\tilde{\alpha}) &= \tilde{\lambda}\,(1-\tilde{\tau})^{\frac{1-\tilde{\tau}}{1+\sigma}}\left\{\mathbb{E}\left[\exp(\tilde{\varepsilon})^{\frac{1+\sigma}{\sigma}}\right]\right\}^{\frac{\sigma(1-\tilde{\tau})}{1+\sigma}}\exp(\tilde{\alpha})^{1-\tilde{\tau}}, \\
h(\tilde{\varepsilon}) &= (1-\tilde{\tau})^{\frac{1}{1+\sigma}}\left\{\mathbb{E}\left[\exp(\tilde{\varepsilon})^{\frac{1+\sigma}{\sigma}}\right]\right\}^{\frac{-1}{1+\sigma}}\exp\left(\frac{1}{\sigma}\tilde{\varepsilon}\right), \\
y(\tilde{\alpha},\tilde{\varepsilon}) &= (1-\tilde{\tau})^{\frac{1}{1+\sigma}}\left\{\mathbb{E}\left[\exp(\tilde{\varepsilon})^{\frac{1+\sigma}{\sigma}}\right]\right\}^{\frac{-1}{1+\sigma}}\exp(\tilde{\alpha})\exp\left(\frac{1+\sigma}{\sigma}\tilde{\varepsilon}\right),
\end{aligned}
$$

and expected utility, conditional on $\tilde{\alpha}$, is therefore

$$
U(\tilde{\alpha}) = \log\left(\tilde{\lambda}\,(1-\tilde{\tau})^{\frac{1-\tilde{\tau}}{1+\sigma}}\left\{\mathbb{E}\left[\exp(\tilde{\varepsilon})^{\frac{1+\sigma}{\sigma}}\right]\right\}^{\frac{\sigma(1-\tilde{\tau})}{1+\sigma}}\exp(\tilde{\alpha})^{1-\tilde{\tau}}\right)-\left(\frac{1-\tilde{\tau}}{1+\sigma}\right).
$$

Comparing expressions across the two economies, it is immediate that allocations and welfare are identical in the life-cycle and static economies as long as:

(i) $\tilde{\lambda}=\lambda$ and $\tilde{\tau}=\tau$,

(ii) $\tilde{\alpha}=\alpha$, and

(iii) $\tilde{\varepsilon}=\log\left(\frac{2\rho^{t-1}}{1+\rho}\right)+\varepsilon$, where t and ε are independent random variables such that $t\in\{1,2\}$ with equal probability, and $\varepsilon\sim F_{\varepsilon}$. In particular, given these distributional assumptions,

$$
\mathbb{E}\left[\exp(\tilde{\varepsilon})^{\frac{1+\sigma}{\sigma}}\right]=\mathbb{E}\left[\left(\frac{2\rho^{t-1}}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}\exp(\varepsilon)^{\frac{1+\sigma}{\sigma}}\right]=\mathbb{E}\left[\exp(\varepsilon)^{\frac{1+\sigma}{\sigma}}\right]\frac{1}{2}\sum_{t=1}^{2}\left(\frac{2\rho^{(t-1)}}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}=X.
$$

Extensions This life-cycle model could be extended in various ways. First, increasing the number of periods from two to any number N is trivial. Second, it is also immediate

---

[4]See, for example, Appendix A in Heathcote et al. (2014).

that one can introduce heterogeneity in expected wage growth over the life-cycle. For example, nothing would change if the age profile for some people was $\left\{\frac{2\rho}{1+\rho}, \frac{2}{1+\rho}\right\}$ rather than $\left\{\frac{2}{1+\rho}, \frac{2\rho}{1+\rho}\right\}$. Third, if we alternatively ruled out inter-temporal borrowing and lending, then the predictable age component of wages would effectively become uninsurable. The life-cycle and static models would still be isomorphic, however.

A.3 Individual- versus Family-Level Taxation

Our baseline model specification assumes that the planner only observes—and thus can only tax—total family income. However, taxing income at the individual level would have no impact on allocations. We now prove that if the tax function for individual income satisfies condition (7), then equilibrium consumption and income are independent of ε, as in the version when taxes apply to total family income.

Proposition 4 *If the tax schedule satisfies condition (7), then the solution to the family head's problem is the same irrespective of whether taxes apply at the family level or the individual level.*

Proof. We will show that given condition (7), the FOCs for the family head with individual-level taxation are identical to those with family-level taxation, namely, eqs. (5) and (6).

If income is taxed at the individual level, the family head's problem becomes

$$\max_{\{h(\alpha,\varepsilon), y(\alpha,\varepsilon)\}} \int \left\{ \frac{[y(\alpha,\varepsilon) - T(y(\alpha,\varepsilon))]^{1-\gamma}}{1-\gamma} - \frac{h(\alpha,\varepsilon)^{1+\sigma}}{1+\sigma} \right\} dF_\varepsilon(\varepsilon)$$

subject to

$$\int y(\alpha,\varepsilon) dF_\varepsilon(\varepsilon) = \int \exp(\alpha+\varepsilon) h(\alpha,\varepsilon) dF_\varepsilon(\varepsilon),$$

where $y(\alpha,\varepsilon)$ denotes pre-tax income allocated to an individual of type ε.

The FOCs are

$$[y(\alpha,\varepsilon) - T(y(\alpha,\varepsilon))]^{-\gamma} [1 - T'(y(\alpha,\varepsilon))] = \mu(\alpha), \tag{A4}$$

$$h(\alpha,\varepsilon)^\sigma = \mu(\alpha) \exp(\alpha+\varepsilon), \tag{A5}$$

where $\mu(\alpha)$ is the multiplier on the family budget constraint.

If the tax schedule satisfies condition (7) (the condition that guarantees first-order conditions are sufficient for optimality) then we can show that optimal consumption and income are independent of ε, as in the version when taxes apply to total family income.

In particular, differentiate both sides of FOC (A4) with respect to ε. The right-hand side is independent of the insurable shock ε, and hence its derivative with respect to ε is zero.

The derivative of the left-hand side of this equation with respect to ε is, by the chain rule,

$$\frac{\partial}{\partial \varepsilon} \left\{ [y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon))]^{-\gamma} [1 - T'(y(\alpha, \varepsilon))] \right\}$$

$$= \left\{ -\gamma (y - T(y))^{-1} [1 - T'(y)]^2 - T''(y) \right\} (y - T(y))^{-\gamma} \frac{\partial y(\alpha, \varepsilon)}{\partial \varepsilon}.$$

The first term is nonzero by condition (7), which immediately implies that $\frac{\partial y}{\partial \varepsilon} = 0$. Therefore, pre-tax income is independent of ε, and hence consumption is also independent of ε. Thus, the FOCs (A4) and (A5) combine to deliver exactly the original intratemporal FOC with family-level taxation, namely, eq. (6). $\mathcal{Q.E.D.}$

B   Decomposition of Welfare Effects of A Tax Reform

B.1   Distributional Gain and Efficiency Cost

Derivation of Efficiency Cost   In Section 3.3, we define the efficiency cost of increasing the marginal tax rate at income level $\hat{y}$ as

$$E(\hat{y}) = 1 - \frac{\Delta Tr(\hat{y})}{1 - F_y(\hat{y})},$$

where $\Delta Tr(\hat{y})$ denotes the extra transfers that can be funded by this tax reform in equilibrium.

To solve for $\Delta Tr(\hat{y})$ we need to work through how increasing the marginal tax rate at income level $\hat{y}$ changes behavior. First, it induces households at income level $\hat{y}$ to work less, due to a substitution effect, resulting in a loss of revenue, $S(\hat{y}) < 0$. This effect is given by

$$S(\hat{y}) = -e^c(\hat{y}) \frac{\hat{y} T'(\hat{y})}{1 - T'(\hat{y}) + e^c(\hat{y}) \hat{y} T''(\hat{y})} f_y(\hat{y}),$$

where $e^c(\hat{y}) > 0$ is the compensated (Hicksian) labor supply elasticity. Second, households earning more than $\hat{y}$ work more due to the income effect of paying one dollar more tax. This direct income effect is given by

$$-I(\hat{y}) = -\int_{\hat{y}}^{\infty} \eta(y) \frac{T'(y)}{1 - T'(y) + e^c(y) y T''(y)} dF_y(y),$$

where $\eta(y) < 0$ is the elasticity of earnings with respect to a change in unearned income. The second derivative of the tax schedule appears in $S(\hat{y})$ and $I(\hat{y})$ because changing hours implies a change in the household's marginal tax rate, which indirectly affects hours via the substitution effect (see Saez 2001).

Let $X(\hat{y})$ denote the change in government revenues associated with the direct mechanical increase in the marginal tax rate, and the substitution and income effects just described:

$$X(\hat{y}) = [1 - F_y(\hat{y})] + S(\hat{y}) - I(\hat{y}). \tag{A6}$$

The amount of extra lump-sum transfers $\Delta Tr(\hat{y})$ that can be financed in equilibrium is not quite $X(\hat{y})$ because increasing transfers itself reduces labor supply via additional income effects. The equilibrium increase in transfers can be computed from the government budget constraint:

$$\Delta Tr(\hat{y}) = X(\hat{y}) + \Delta Tr(\hat{y}) \times I(0), \tag{A7}$$

where $I(0) < 0$ denotes the income effect of increasing lump-sum transfers to all households in the economy.

Substituting eq. (A6) into eq. (A7), we have

$$\Delta Tr(\hat{y}) = \frac{[1 - F_y(\hat{y})] + S(\hat{y}) - I(\hat{y})}{1 - I(0)},$$

and thus

$$E(\hat{y}) = 1 - \frac{\Delta T(\hat{y})}{1 - F_y(\hat{y})} = \frac{-I(0)}{1 - I(0)} - \frac{1}{1 - F_y(\hat{y})}\frac{S(\hat{y}) - I(\hat{y})}{1 - I(0)}.$$

Distributional Gain and Efficiency Cost as A Function of Productivity   In Section 3.3, we define the distributional gain as

$$D(\hat{y}) \equiv 1 - \frac{\int_{\hat{y}}^{\infty} W_y(y)u_c(y)dF_y(y)}{[1 - F_y(\hat{y})]\chi}.$$

Notice that

$$F_y(y(\alpha)) = F_\alpha(\alpha),$$

and also

$$f_y(y)\frac{dy}{d\alpha} = f_\alpha(\alpha).$$

Therefore, the distributional gain of increasing the marginal tax rate at productivity level $\hat{\alpha}$ is simply

$$D(\hat{\alpha}) \equiv 1 - \frac{\int_{\hat{\alpha}}^{\infty} W(\alpha)u_c(\alpha)dF_\alpha(\alpha)}{[1 - F_\alpha(\hat{\alpha})]\chi},$$

where χ is given by

$$\chi = \int_0^{\infty} W(\alpha)u_c(\alpha)dF_\alpha(\alpha).$$

Similarly, we have

$$I(\hat{\alpha}) = \int_{\hat{\alpha}}^{\infty} \eta(\alpha) \frac{T'(y(\alpha))}{1 - T'(y(\alpha)) + e^c(\alpha)y(\alpha)T''(y(\alpha))} dF_\alpha(\alpha).$$

Also, given our baseline utility function, applying Lemma 1 in Saez (2001), we have

$$
\begin{aligned}
S(\hat{\alpha}) &= -e^c(\hat{\alpha}) \frac{T'(y(\hat{\alpha}))}{1 - T'(y(\hat{\alpha})) + e^c(\hat{\alpha})y(\hat{\alpha})T''(y(\hat{\alpha}))} y(\hat{\alpha})f_y(y(\hat{\alpha})) \\
&= -e^c(\hat{\alpha}) \frac{T'(y(\hat{\alpha}))}{1 - T'(y(\hat{\alpha})) + e^c(\hat{\alpha})y(\hat{\alpha})T''(y(\hat{\alpha}))} f_\alpha(\hat{\alpha}) \frac{1 - T'(y(\hat{\alpha})) + e^c(\hat{\alpha})y(\hat{\alpha})T''(y(\hat{\alpha}))}{[1 + e^u(\hat{\alpha})][1 - T'(y(\hat{\alpha}))]} \\
&= -e^c(\hat{\alpha}) \frac{T'(y(\hat{\alpha}))}{[1 + e^u(\hat{\alpha})][1 - T'(y(\hat{\alpha}))]} f_\alpha(\hat{\alpha}) \\
&= -\frac{1}{1 + \sigma} \frac{T'(y(\hat{\alpha}))}{1 - T'(y(\hat{\alpha}))} f_\alpha(\hat{\alpha}),
\end{aligned}
$$

where e^u is the uncompensated (Marshallian) labor supply elasticity.

Thus, the efficiency cost of increasing the marginal tax rate at productivity level $\hat{\alpha}$ is

$$E(\hat{\alpha}) = \frac{-I(-\infty)}{1 - I(-\infty)} - \frac{1}{1 - F_\alpha(\hat{\alpha})} \frac{S(\hat{\alpha}) - I(\hat{\alpha})}{1 - I(-\infty)}.$$

In Section 5.2, when the utility function is given by eq. (18) and thus $I(\alpha) = 0$, this expression simplifies to

$$E(\hat{\alpha}) = \frac{1}{1 + \sigma} \frac{T'(y(\hat{\alpha}))}{1 - T'(y(\hat{\alpha}))} \frac{f_\alpha(\hat{\alpha})}{1 - F_\alpha(\hat{\alpha})}.$$

B.2 Diamond-Saez Formula

We describe how the fiscal pressure intuition described in Section 5.1 meshes with the Diamond-Saez formula. We first derive the Diamond-Saez formula for our economy. We then use a modified version of the Diamond-Saez formula to discuss the factors that determine the shape of the optimal marginal tax schedule.

Derivation: Diamond-Saez Formula Reproducing the Mirrlees planner's problem from eqs. (10-12), we have

$$
\begin{cases}
\displaystyle \max_{\{c(\alpha),y(\alpha)\}} \quad \int W(\alpha) \left[\frac{c(\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\alpha)}{\exp(\alpha)} \right)^{1+\sigma} \right] dF_\alpha(\alpha) \\[3mm]
\text{s.t.} \quad \frac{c(\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\alpha)}{\exp(\alpha)} \right)^{1+\sigma} \geq \frac{c(\tilde\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\tilde\alpha)}{\exp(\alpha)} \right)^{1+\sigma} \quad \text{for all } \alpha \text{ and } \tilde\alpha, \\[3mm]
\int [y(\alpha) - c(\alpha)] \, dF_\alpha(\alpha) - G \geq 0.
\end{cases}
$$

The IC constraints state

$$
U(\alpha) \equiv \frac{c(\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\alpha)}{\exp(\alpha)} \right)^{1+\sigma} = \max_{\tilde\alpha} \frac{c(\tilde\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\tilde\alpha)}{\exp(\alpha)} \right)^{1+\sigma}.
$$

Using the envelope condition:

$$
c(\alpha)^{-\gamma} c'(\alpha) - \frac{\Omega}{\exp\left[(1+\sigma)\,\alpha \right]} y(\alpha)^\sigma y'(\alpha) = 0,
$$

we get

$$
U'(\alpha) = \frac{\Omega}{\exp\left[(1+\sigma)\,\alpha \right]} y(\alpha)^{1+\sigma}.
$$

Thus, we can reformulate the planner's problem as follows:

$$
\begin{cases}
\displaystyle \max_{\{U(\alpha),y(\alpha)\}} \quad \int W(\alpha)U(\alpha) dF_\alpha(\alpha) \\[3mm]
\text{s.t.} \quad U'(\alpha) = \frac{\Omega}{\exp[(1+\sigma)\alpha]} y(\alpha)^{1+\sigma} \quad \text{for all } \alpha, \\[3mm]
\int [y(\alpha) - c(\alpha; U, y)] \, dF_\alpha(\alpha) - G \geq 0,
\end{cases}
$$

where $c(\alpha; U, y)$ is determined by $U(\alpha) = \frac{c(\alpha)^{1-\gamma}}{1-\gamma} - \frac{\Omega}{1+\sigma} \left(\frac{y(\alpha)}{\exp(\alpha)} \right)^{1+\sigma}$. Denoting by $\mu(\alpha)$ and ζ the corresponding multipliers, we then set up a Hamiltonian with U as the state and y as the control:

$$
\mathcal{H} \equiv \{W(\alpha)U(\alpha) + \zeta\,[y(\alpha) - c(\alpha; U, y) - G]\}\, f_\alpha(\alpha) + \mu(\alpha) \frac{\Omega}{\exp\left[(1+\sigma)\,\alpha \right]} y(\alpha)^{1+\sigma},
$$

where f_α is the derivative of F_α. By optimal control, the following equations must hold

$$
\begin{cases}
0 = \zeta \left[1 - c(\alpha)^\gamma \Omega \exp(-(1+\sigma)\alpha) y(\alpha)^\sigma \right] f_\alpha(\alpha) + \mu(\alpha) \frac{\Omega(1+\sigma)}{\exp[(1+\sigma)\alpha]} y(\alpha)^\sigma, \\[3mm]
-\mu'(\alpha) = [W(\alpha) - c(\alpha)^\gamma \zeta]\, f_\alpha(\alpha), \\[3mm]
\mu(0) = \mu(\infty) = 0.
\end{cases} \tag{A8}
$$

Integrating the second equation over α and using $\mu(\infty) = 0$, we solve for the costate:

$$\mu(\alpha) = \int_\alpha^\infty [W(s) - c(s)^\gamma \zeta] \, dF_\alpha(s).$$

Using $\mu(0) = 0$, we also get the expression for ζ:

$$\zeta = \frac{\int W(s) dF_\alpha(s)}{\int c(s)^\gamma dF_\alpha(s)} = \frac{1}{\int c(s)^\gamma dF_\alpha(s)}.$$

We now consider the decentralization via income taxes (see Section 3.1). Using the FOC (13), the first equation in (A8) can be written as

$$0 = \zeta T'(y(\alpha)) f_\alpha(\alpha) + \mu(\alpha) [1 - T'(y(\alpha))] c(\alpha)^{-\gamma} (1 + \sigma),$$

where T' is the marginal tax rate. Rearranging terms, we obtain

$$\frac{T'(y(\alpha))}{1 - T'(y(\alpha))} = (1 + \sigma) \frac{1 - F_\alpha(\alpha)}{f_\alpha(\alpha)} \int_\alpha^\infty \left[1 - \frac{W(s) c(s)^{-\gamma}}{\zeta} \right] \frac{c(\alpha)^{-\gamma}}{c(s)^{-\gamma}} \frac{dF_\alpha(s)}{1 - F_\alpha(\alpha)},$$

$$\text{where} \quad \zeta = \frac{1}{\int c(s)^\gamma dF_\alpha(s)}.$$

Imposing logarithmic preferences in consumption, we finally get the Diamond-Saez formula for our economy:

$$\frac{T'(y(\alpha))}{1 - T'(y(\alpha))} = (1 + \sigma) \frac{1 - F_\alpha(\alpha)}{f_\alpha(\alpha)} \int_\alpha^\infty \left[1 - \frac{W(s) \cdot C}{c(s)} \right] \frac{c(s)}{c(\alpha)} \frac{dF_\alpha(s)}{1 - F_\alpha(\alpha)}, \qquad \text{(A9)}$$

where C denotes aggregate (and average) consumption.

Discussion After some straightforward algebra, eq. (A9) can be rewritten as

$$\frac{T'(y(\alpha))}{1 - T'(y(\alpha))} = A(\alpha) \times B(\alpha), \qquad \text{(A10)}$$

$$\text{where} \quad A(\alpha) = (1 + \sigma) \times \frac{1 - F_\alpha(\alpha)}{f_\alpha(\alpha)},$$

$$B(\alpha) = F_a(\alpha) \times \frac{\mathbb{E}[c(\tilde\alpha)|_{\tilde\alpha \geq \alpha}] - \mathbb{E}[c(\tilde\alpha)|_{\tilde\alpha < \alpha}]}{c(\alpha)}.$$

The two terms labelled $A(\alpha)$ and $B(\alpha)$ (as in Saez 2001) can be used to discuss the factors that determine the shape of the optimal marginal tax schedule. In the following we interpret these terms, taking the exercise varying government expenditure levels as an example. See

12

Section 5.1 for more detail.

The first component of the $A(\alpha)$ term, $(1 + \sigma)$, indicates that the more elastic is labor supply, the lower are optimal marginal tax rates, all else equal. The second component of the $A(\alpha)$ term is the ratio of fraction of households more productive than α relative to the density at α. Marginal rates should be high in regions of the productivity distribution where this ratio is high, so that there are lots of more productive agents who will pay extra taxes, but relatively few whose labor supply will be directly distorted by higher rates at the margin. While the components of the $A(\alpha)$ term are easy to interpret, since they involve only structural primitives of the model, they cannot explain the differential marginal tax profiles corresponding to different values for G, since the $A(\alpha)$ term is independent of G.

Instead the way changes in G show up in the right-hand side of the Diamond-Saez formula is in the $B(\alpha)$ term, which indicates a relationship between optimal marginal tax rates and the shape of the consumption distribution. In particular, this term indicates that marginal rates should be low when the particular measure of consumption inequality defined by $B(\alpha)$ is low. When G is low, this measure of consumption inequality is relatively low at low productivity values—because generous lump-sum transfers offer a decent consumption floor—which is consistent with low marginal tax rates at low income levels. Conversely, when G is high and optimal transfers are smaller, there is more consumption inequality at the bottom of the productivity distribution (a higher $B(\alpha)$) which is consistent, via eq. (A10), with higher optimal marginal tax rates.

While this discussion illustrates that the Diamond-Saez equation (A10) and Panel A in Figure 3 are mutually consistent, it does not quite get to the bottom of why the optimal consumption allocation looks the way it does. In particular, the $B(\alpha)$ term, which is the critical factor for interpreting the optimal tax schedule, involves the distribution of consumption, which is obviously endogenous to the tax system. The only reason that the consumption distribution—and thus the $B(\alpha)$ term—varies with G is because the optimal tax schedule itself varies with G. We thus conclude that while the Diamond-Saez formula is useful, it offers limited intuition about the fundamental drivers of the shape of the optimal tax schedule.

C   Computational Method

C.1   Mirrlees Optimal Taxation

We briefly describe how we compute the optimal allocation in the baseline economy. We solve the Mirrlees planner's problem (10) for our discretized economy numerically. We first note that the local downward and local upward incentive compatibility constraints are necessary

Table A1: Deviation from Diamond-Saez formula

Prod. Percentile	0.01	0.1	1	25	50	75	99	99.9	99.99
$A(\alpha)B(\alpha) - \frac{T'(y(a))}{1-T'(y(\alpha))}$	0.0006	0.0006	0.0006	0.0007	0.0008	0.0010	0.0019	0.0023	0.0018

and sufficient for the global incentive compatibility constraints (12) to be satisfied:

$$U(\alpha_i, \alpha_i) \geq U(\alpha_i, \alpha_{i-1}) \qquad \text{for all } i = 2, \cdots, I$$
$$U(\alpha_{i-1}, \alpha_{i-1}) \geq U(\alpha_{i-1}, \alpha_i) \qquad \text{for all } i = 2, \cdots, I.$$

We then solve for the allocation exactly at each grid point. Specifically, we use forward iteration (forward from α_1 to α_I) to search for an allocation that satisfies all the first-order conditions, the incentive constraints above, and the resource constraint (11). Finally, we confirm that before-tax income is nondecreasing in wages, concluding that the resulting allocation is optimal given that our utility function exhibits the single-crossing property. Note that we never assume that the upward incentive constraints are slack, because their slackness is not guaranteed for any economy with $I > 2$. In our baseline economy, some upward incentive constraints are indeed binding at the bottom of the α distribution, which results in bunching.

This computational method contrasts with the typical approach in the literature that looks for approximate marginal tax rate schedules that satisfy the Diamond-Saez formula (the social planner's first-order condition), which implicitly defines the optimal tax schedule (see, e.g., the appendix to Mankiw et al. 2009). Since we do not iterate back and forth between candidate tax schedules and agents' best responses to those schedules, our method is much faster, especially when the grid is very fine.

Table A1 shows that our numerical solution satisfies the Diamond-Saez formula (eq. A10) almost exactly, even though (i) we have assumed a discrete distribution for α, while the formula assumes a continuous distribution, and (ii) we have not used the formula directly for computation.

C.2 Pareto-Improving Tax Reforms

The first-stage planner's problem with a set of Pareto-improving constraints is given by

$$\max_{\{c(\alpha),y(\alpha)\}} \quad \sum_i \pi_i W(\alpha_i) U(\alpha_i, \alpha_i), \tag{A11}$$

$$\text{subject to} \quad \sum_i \pi_i c(\alpha_i) + G = \sum_i \pi_i y(\alpha_i),$$

$$U(\alpha_i, \alpha_i) \geq U(\alpha_i, \alpha_j) \qquad \text{for all } i, j \tag{A12}$$

$$U(\alpha_i, \alpha_i) \geq U^{\text{US}}(\alpha_i) \qquad \text{for all } i. \tag{A13}$$

Notice that the left hand sides of eqs. (A12) and (A13) are the same and thus some constraints will be slack. Solving this problem is thus challenging computationally because the pattern of which subset of constraints is binding at the optimum is unknown *ex ante*.

To solve this problem, first denote

$$\bar{U}_i = U^{\text{US}}(\alpha_i).$$

We then consider the following planner's problem that replaces the Pareto-improving constraints (A13) with a penalty function in the objective function:

$$\max_{\{c(\alpha),y(\alpha)\}} \quad \sum_i \pi_i \left[W(\alpha_i) U(\alpha_i, \alpha_i) - \gamma \min \left\{ U(\alpha_i, \alpha_i) - \bar{U}_i, 0 \right\}^2 \right], \tag{A14}$$

$$\text{subject to} \quad \sum_i \pi_i c(\alpha_i) + G = \sum_i \pi_i y(\alpha_i),$$

$$U(\alpha_i, \alpha_i) \geq U(\alpha_i, \alpha_j) \qquad \text{for all } i, j$$

where $\gamma > 0$ controls the magnitude of penalty. Since the objective function is still differentiable, we can solve this alternative problem by applying the computational method described in Appendix C.1. However, the resulting allocation will not satisfy the original Pareto-improving constraints (A13); it is optimal for the planner to pay the penalty for a range of $i's$ where the Pareto-improving constraints are tight.

In the next step, given the solution to the problem (A14), we update $\{\bar{U}_i\}$ by adding a sufficiently small value $\epsilon > 0$ to $\bar{U}_j$ for the grid point j where the Pareto-improving constraint is most violated:

$$j = \arg\min_i U(\alpha_i, \alpha_i) - \bar{U}_i .$$

With the updated $\{\bar{U}_i\}$, we solve problem (A14) again. Because the penalty at j is larger than before, the planner will provide more utility for the agent with α_j while respecting
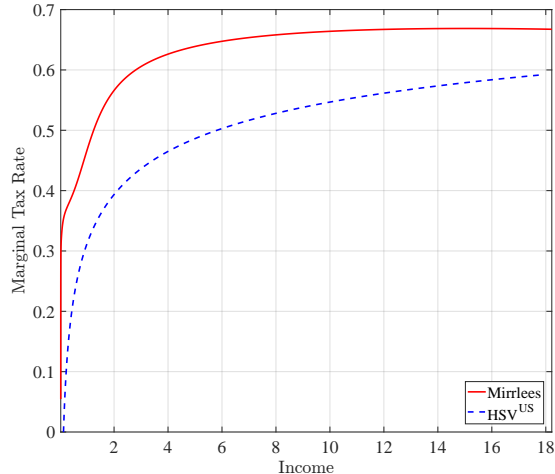
Figure A1: Optimal Tax Policy against Income. The figure plots the optimal Mirrleesian tax schedules against income and the baseline HSV approximation to the U.S. tax and transfer system.

all the incentive compatibility constraints. Therefore, the resulting allocation violates the original Pareto-improving constraints less severely.

We keep updating $\{\bar{U}_i\}$ and solving the problem (A14) until the resulting allocation satisfies all the Karush-Kuhn-Tucker conditions of the original planner's problem (A11).

D Baseline Optimal Tax Policy

Optimal Taxes as a Function of Income   Panel A of Figure 2 shows that the marginal tax schedule is generally increasing in productivity α. However, this does not necessarily mean that the marginal tax schedule is also increasing in income y, because y is a function of α and thus the marginal tax schedule depends on how y changes with α.

Figure A1 plots the optimal Mirrleesian tax schedules against income y. Reassuringly, the plot shows that the optimal marginal tax schedule is increasing in income, just as it is increasing in α (Panel A of Figure 2.)

Optimal Taxes at the Bottom   The optimal allocation features bunching at the bottom of the productivity distribution: all household types in a set $\{\alpha_1, \cdots, \alpha_k\}$ receive the same income and consumption allocation, labelled $(\underline{y}, \underline{c})$ in Panel A of Figure A2.[5] The red solid line to the right of this point traces optimal allocations $(y(\alpha_i), c(\alpha_i))$ for $i \geq k$. The indifference curves that go through the point $(\underline{y}, \underline{c})$ for the least and most productive households in the bunched set are labelled IC_1 and IC_k. To decentralize the optimal allocation, consumption $c(y)$ must be sufficiently small for $y < \underline{y}$ (and thus implied net taxes $T(y) = y - c(y)$

---

[5]This implies that hours are decreasing in α, while the marginal tax rate is strictly positive (see Ebert 1992) and increasing in α (see eq. 13). See Figure 2.
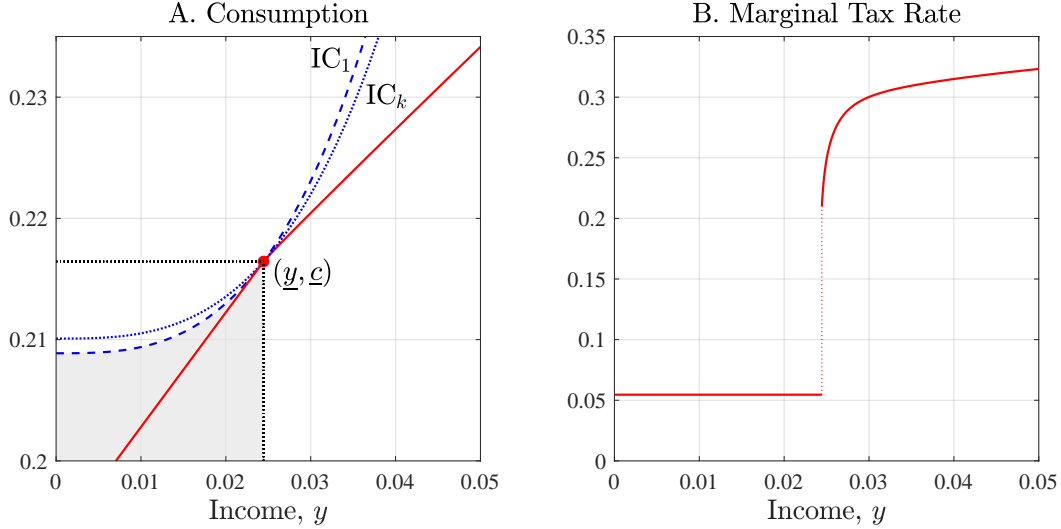
Figure A2: Allocations and Tax Rates for Low Income Households. Panel A plots household consumption against household income at the bottom of the income distribution (red solid line) and the indifference curves (blue dashed lines) for the least and most productive households in the bunched set, labelled IC_1 and IC_k. Panel B plots the marginal tax rate under the assumption that the marginal tax rate below the income level of the bunched set is given by the upper bound of the rate that implements the optimal allocation.

sufficiently large) so that no households will choose to deliver such low income. There are many possible consumption schedules that ensure this: the set of such schedules is shaded grey in the figure, where the upper bound of the set is given by $IC_1(y)$.

Restricting attention to continuous tax functions, it is immediate that as income approaches $\underline{y}$ from below, $1 - T'(y)$ (the slope of the budget line) must be greater than the slope of $IC_1(y)$ at $y = \underline{y}$. This translates into an upper bound on the marginal tax rate under any optimal tax scheme of 5.5%. In Panel B, we assume that this marginal tax rate applies for all $y < \underline{y}$ (the corresponding allocation is traced by the red solid line to the left of $(\underline{y}, \underline{c})$ in Panel A). As income approaches $\underline{y}$ from above, $1 - T'(y)$ must be smaller than the slope of $IC_k(y)$ at $y = \underline{y}$ in order to dissuade type k from delivering more income. This translates into a lower bound on the marginal tax rate of 21.0%. Thus, a jump in the marginal tax rate at $\underline{y}$ is a necessary property of any tax function that implements the optimal allocation.

It has long been recognized that tax systems featuring discrete steps in marginal tax rates will induce bunching at the income levels where rates jump. In this economy, bunching is optimal, and a jump in the marginal rate is required to deliver bunching.

E Proofs of Propositions

E.1 Proof of Proposition 1

Let $\Delta > 0$ denote the decline in lump-sum transfers.

Proof of Part (i) From eq. (19), $E(\alpha)$ is independent of Δ. ∎

Proof of Part (ii) Given a utility function of the form (18), define the marginal utility conditional on reduced lump-sum transfers:

$$u_c(\alpha, \Delta) = \frac{1}{c(\alpha) - \frac{h(\alpha)^{1+\sigma}}{1+\sigma} - \Delta},$$

and the average marginal utility:

$$\chi(\Delta) = \mathbb{E}\left[u_c(\alpha, \Delta)\right].$$

The gain from redistribution can be written as

$$\tilde{D}(\alpha, \Delta) = \int_\alpha^\infty \left[1 - \frac{u_c(s, \Delta)}{\chi(\Delta)}\right] dF_\alpha(s).$$

Note that the initial allocation (i.e., before reducing lump-sum transfers) corresponds to $u_c(\alpha, 0)$, $\chi(0)$ and $\tilde{D}(\alpha, 0)$.

We want to show that $\tilde{D}(\alpha, \Delta) > \tilde{D}(\alpha, 0)$ for any finite α and $\Delta > 0$. Note $\tilde{D}(-\infty, \Delta) = \tilde{D}(-\infty, 0) = 0$ and $\tilde{D}(\infty, \Delta) = \tilde{D}(\infty, 0) = 0$ by definition, and

$$\tilde{D}(\alpha, \Delta) - \tilde{D}(\alpha, 0) = \int_\alpha^\infty \left(\frac{u_c(s, 0)}{\chi(0)} - \frac{u_c(s, \Delta)}{\chi(\Delta)}\right) dF_\alpha(s).$$

It then suffices to show that there exists $\tilde{\alpha}$ such that

$$\frac{u_c(\alpha, 0)}{\chi(0)} < \frac{u_c(\alpha, \Delta)}{\chi(\Delta)} \text{ for all } \alpha \in (-\infty, \tilde{\alpha}), \text{ and} \tag{A15}$$

$$\frac{u_c(\alpha, 0)}{\chi(0)} > \frac{u_c(\alpha, \Delta)}{\chi(\Delta)} \text{ for all } \alpha \in (\tilde{\alpha}, \infty).$$

We show that $\tilde{\alpha}$ is given by the value for α such that

$$\frac{u_c(\tilde{\alpha}, 0)}{\chi(0)} = \frac{u_c(\tilde{\alpha}, \Delta)}{\chi(\Delta)}.$$

Note that there must be at least one such value, because the two sides of this equation are

two different continuous functions with the same mean when integrated over the distribution for α (1 in both cases).

We can prove the first case in (A15) because for $\alpha \in (-\infty, \tilde{\alpha})$,

$$\frac{\chi(0)}{\chi(\Delta)} = \frac{u_c(\tilde{\alpha}, 0)}{u_c(\tilde{\alpha}, \Delta)} = \frac{c(\tilde{\alpha}) - \frac{h(\tilde{\alpha})^{1+\sigma}}{1+\sigma} - \Delta}{c(\tilde{\alpha}) - \frac{h(\tilde{\alpha})^{1+\sigma}}{1+\sigma}} > \frac{c(\alpha) - \frac{h(\alpha)^{1+\sigma}}{1+\sigma} - \Delta}{c(\alpha) - \frac{h(\alpha)^{1+\sigma}}{1+\sigma}} = \frac{u_c(\alpha, 0)}{u_c(\alpha, \Delta)},$$

where the inequality comes from the fact that $c(\alpha) - \frac{h(\alpha)^{1+\sigma}}{1+\sigma}$ is positive and increasing in α. We can also prove the second case in (A15) analogously. ∎

Proof of Part (iii)   We want to show that $\alpha^*_{\text{fixed}} < \alpha^*$, where α^*_{fixed} is the distributional gain maximizing value of α when lump-sum transfers are reduced, but the marginal tax schedule is unchanged:

$$u_c(\alpha^*_{\text{fixed}}, \Delta) = \chi(\Delta) = \mathbb{E}[u_c(\alpha, \Delta)].$$

Notice that α^*_{fixed} can be regarded as the "certainty equivalent" given that $u_c(\cdot, \Delta)$ is a convex function. Therefore, following Pratt (1964), $\alpha^*_{\text{fixed}} < \alpha^*$ is equivalent to the statement that $u_c(\cdot, 0)$ is a concave transformation of $u_c(\cdot, \Delta)$. It thus suffices to show that there exists an increasing concave function $\psi(\cdot)$ such that $u_c(\alpha, 0) = \psi(u_c(\alpha, \Delta))$ for all α.

Now define $\psi(x)$ by

$$\psi(x) = \frac{1}{\frac{1}{x} + \Delta}.$$

It is then straightforward to show that $u_c(\alpha, 0) = \psi(u_c(\alpha, \Delta))$ for all α, and ψ is increasing and concave:

$$\frac{d\psi(x)}{dx} = \frac{1}{x^2 \left(\frac{1}{x} + \Delta\right)^2} = \frac{1}{(1 + \Delta x)^2} > 0,$$

$$\frac{d^2\psi(x)}{dx^2} = \frac{-2\Delta}{(1 + \Delta x)^3} = \frac{-2\Delta}{x}\psi(x)\frac{d\psi(x)}{dx} < 0,$$

because $x > 0$ and $\psi(x) > 0$. ∎ $\mathcal{Q.E.D.}$

E.2   Proof of Proposition 2

Given the HSV tax function, decision rules as a function of τ are as follows:

$$c(\alpha; \lambda, \tau) = \lambda(1-\tau)^{\frac{1-\tau}{1+\sigma}} \exp[(1-\tau)\alpha] \exp\left(\frac{1-\tau}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right), \tag{A16}$$

$$h(\varepsilon; \tau) = (1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{-1}{\sigma^2}\frac{\sigma_\varepsilon^2}{2}\right) \exp\left(\frac{\varepsilon}{\sigma}\right). \tag{A17}$$

19

Plugging these into the resource constraint (1), we get

$$\lambda(\tau) = \frac{(1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right) - G}{(1-\tau)^{\frac{1-\tau}{1+\sigma}} \exp\left(\frac{1-\tau}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right) \int \exp\left[(1-\tau)\alpha\right] dF_\alpha(\alpha)}.$$

We substitute these expressions into the planner's objective function in order to get an unconstrained optimization problem with one choice variable, τ. Specifically, the planner's objective function is

$$\int W(\alpha) \left[\log\left(c(\alpha;\tau)\right) - \int \frac{h(\varepsilon;\tau)^{1+\sigma}}{1+\sigma} dF_\varepsilon(\varepsilon)\right] dF_\alpha(\alpha),$$

and government expenditure is given by

$$G = g \iint \exp(\alpha + \varepsilon) h(\varepsilon;\tau) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon).$$

Substituting eqs. (A16) and (A17) into these, the optimization problem can be rewritten as

$$\max_\tau \quad (1-\tau)\int \alpha \cdot W(\alpha) dF_\alpha(\alpha) - \log(\int \exp\left[(1-\tau)\alpha\right] dF_\alpha(\alpha)) + \log\left[(1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right) - G\right] - \frac{1-\tau}{1+\sigma}$$

where

$$G = g(1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right). \tag{A18}$$

Note that the *level* of the government expenditure G is fixed when the planner is solving the problem, and hence it is not a function of τ.

Given the Pareto weight function (20), the optimization problem becomes[6]

$$\max_\tau \quad \frac{(1-\tau)}{\frac{\lambda_\alpha}{\lambda_\alpha+\theta} \exp\left[-\mu_\alpha\theta+\frac{\sigma_\alpha^2\theta^2}{2}\right]} \int \alpha \exp(-\theta\alpha) dF_\alpha(\alpha) - \log\left(\frac{\lambda_\alpha}{\lambda_\alpha-1+\tau}\right) - \mu_\alpha(1-\tau) - \frac{\sigma_\alpha^2(1-\tau)^2}{2}$$
$$+ \log\left[(1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right) - G\right] - \frac{1-\tau}{1+\sigma}. \tag{A19}$$

Assume this problem is well-defined; that is, $\int \alpha \exp(-\theta\alpha) dF_\alpha < \infty$. We want to further simplify this term.

---

[6]The moment-generating function for the EMG distribution, $EMG(\mu_\alpha, \sigma_\alpha^2, \lambda_\alpha)$, for $t \in \mathbb{R}$ is given by

$$\int_\alpha \exp\left(\alpha t\right) dF_\alpha = \frac{\lambda_\alpha}{\lambda_\alpha - t} \exp\left[\mu_\alpha t + \frac{\sigma_\alpha^2 t^2}{2}\right].$$

Define $V(\alpha, \theta) \equiv \exp(-\theta\alpha) f_\alpha(\alpha)$, where f_α is the derivative of F_α. We then have

$$\frac{\partial V(\alpha, \theta)}{\partial \theta} = -\alpha \exp(-\theta\alpha) f_\alpha(\alpha).$$

Lemma 5 *Assume the support of θ is compact, $[\underline{\theta}, \bar{\theta}]$. Then the integral and the derivative of V are interchangeable; that is,*

$$\int \frac{\partial}{\partial \theta} V(\alpha, \theta) d\alpha = \frac{\partial}{\partial \theta} \int V(\alpha, \theta) d\alpha.$$

Proof. It suffices to show that (i) $V : \mathbb{R} \times [\underline{\theta}, \bar{\theta}] \to \mathbb{R}$ is continuous and $\frac{\partial V}{\partial \theta}$ is well-defined and continuous in $\mathbb{R} \times [\underline{\theta}, \bar{\theta}]$, (ii) $\int V(\alpha, \theta) d\alpha$ is uniformly convergent, and (iii) $\int \frac{\partial}{\partial \theta} V(\alpha, \theta) d\alpha$ is uniformly convergent.

(i) is obvious since f_α is continuous.

To prove (ii), we rely on the Weierstrass M-test for uniform convergence. That is, if there exists $\hat{V} : \mathbb{R} \to \mathbb{R}$ such that $\hat{V}(\alpha) \geq |V(\alpha, \theta)|$ for all θ and $\hat{V}$ has an improper integral on $\mathbb{R}$, then $\int V(\alpha, \theta) d\alpha$ converges uniformly. Now define $\hat{V}(\alpha) \equiv \sup_{\theta \in [\underline{\theta}, \bar{\theta}]} |V(\alpha, \theta)|$. Then $\hat{V}(\alpha) \geq |V(\alpha, \theta)|$ by construction. Also $\hat{V}$ has an improper integral on $\mathbb{R}$ because

$$
\begin{aligned}
\int_{-\infty}^{\infty} \hat{V}(\alpha) d\alpha &= \int_{-\infty}^{0} V(\alpha, \bar{\theta}) d\alpha + \int_{0}^{\infty} V(\alpha, \underline{\theta}) d\alpha \\
&\leq \int_{-\infty}^{\infty} V(\alpha, \bar{\theta}) d\alpha + \int_{-\infty}^{\infty} V(\alpha, \underline{\theta}) d\alpha \\
&= \frac{\lambda_\alpha}{\lambda_\alpha + \bar{\theta}} \exp\left[-\mu_\alpha \bar{\theta} + \frac{\sigma_\alpha^2 \bar{\theta}^2}{2}\right] + \frac{\lambda_\alpha}{\lambda_\alpha + \underline{\theta}} \exp\left[-\mu_\alpha \underline{\theta} + \frac{\sigma_\alpha^2 \underline{\theta}^2}{2}\right] < \infty,
\end{aligned}
$$

where the first inequality comes from $V(\alpha, \theta) \geq 0$ for any α and $\theta \in [\underline{\theta}, \bar{\theta}]$. Thus, $\int V(\alpha, \theta) d\alpha$ is uniformly convergent.

We apply a similar logic to prove (iii) and find $\tilde{V} : \mathbb{R} \to \mathbb{R}$ such that $\tilde{V}(\alpha) \geq \left|\frac{\partial V(\alpha, \theta)}{\partial \theta}\right|$ for all θ and $\tilde{V}$ has an improper integral on $\mathbb{R}$. Specifically, define $\tilde{V}(\alpha) \equiv \sup_{\theta \in [\underline{\theta}, \bar{\theta}]} \left|\frac{\partial V(\alpha, \theta)}{\partial \theta}\right|$. Then $\tilde{V}(\alpha) \geq \left|\frac{\partial V(\alpha, \theta)}{\partial \theta}\right|$ by construction and $\tilde{V}$ has an improper integral on $\mathbb{R}$, because the original problem is assumed to be well-defined, and hence $\int \alpha \exp(-\theta\alpha) dF_\alpha < \infty$ for any $\theta \in [\underline{\theta}, \bar{\theta}]$.

∎

Applying this lemma, we get

$$
\begin{aligned}
\int \alpha \exp(-\theta\alpha) dF_\alpha(\alpha) &= -\frac{\partial}{\partial\theta} \int \exp(-\theta\alpha) dF_\alpha(\alpha) \\
&= -\frac{\partial}{\partial\theta} \left\{ \frac{\lambda_\alpha}{\lambda_\alpha + \theta} \exp\left[-\mu_\alpha\theta + \frac{\sigma_\alpha^2\theta^2}{2} \right] \right\} \\
&= \frac{\lambda_\alpha}{\lambda_\alpha + \theta} \exp\left[-\mu_\alpha\theta + \frac{\sigma_\alpha^2\theta^2}{2} \right] \left(\frac{1}{\lambda_\alpha + \theta} + \mu_\alpha - \sigma_\alpha^2\theta \right).
\end{aligned}
$$

Substituting this expression into eq. (A19), the optimization problem becomes

$$
\max_\tau \ (1-\tau)\left(\frac{1}{\lambda_\alpha+\theta} - \sigma_\alpha^2\theta - \frac{1}{1+\sigma} \right) + \log\left(\lambda_\alpha - 1 + \tau \right) - \frac{\sigma_\alpha^2(1-\tau)^2}{2} + \log\left[(1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{1}{\sigma}\frac{\sigma_\varepsilon^2}{2} \right) - G \right].
$$

The first-order condition with respect to τ is

$$
0 = -\frac{1}{\lambda_\alpha+\theta} + \sigma_\alpha^2\theta + \frac{1}{1+\sigma} + \frac{1}{\lambda_\alpha - 1 + \tau} + \sigma_\alpha^2(1-\tau) - \frac{\left[1 - \frac{G}{\exp\left(\frac{1}{\sigma}\frac{\sigma_\varepsilon^2}{2} \right)(1-\tau)^{\frac{1}{1+\sigma}}} \right]^{-1}}{(1-\tau)(1+\sigma)}. \quad \text{(A20)}
$$

Substituting eq. (A18) into this, we have

$$
\sigma_\alpha^2\theta - \frac{1}{\lambda_\alpha+\theta} = -\sigma_\alpha^2(1-\tau) - \frac{1}{\lambda_\alpha - 1 + \tau} + \frac{1}{1+\sigma}\left[\frac{1}{(1-g)(1-\tau)} - 1 \right].
$$

Therefore, the planner's weight θ must solve eq. (21). $\quad\quad\quad Q.\mathcal{E}.\mathcal{D}.$

E.3 Proof of Proposition 3

Consider general separable preferences:

$$
u(c) - v(h).
$$

For a generic tax function T, the household budget constraint is given by $c = y - T(y)$.

Let the function $U(\alpha; T)$ denote equilibrium utility for a household with productivity α facing a tax schedule T and let $c(\alpha; T)$ and $y(\alpha; T)$ denote the associated equilibrium allocations. We have

$$
U(\alpha; T) = u(y(\alpha; T) - T(y(\alpha; T))) - v\left(\frac{y(\alpha; T)}{\exp(\alpha)} \right).
$$

Lemma 6 *Consider two tax schedules T_1 and T_2. Suppose at some productivity level α, $U(\alpha; T_1) = U(\alpha; T_2)$, $U'(\alpha; T_1) = U'(\alpha; T_2)$ and T_1 and T_2 are differentiable. Then $c(\alpha; T_1) =$*

$c(\alpha;T_2)$, $y(\alpha;T_1) = y(\alpha;T_2)$, $T_1(y(\alpha;T_1)) = T_2(y(\alpha;T_2))$, and $T_1'(y(\alpha;T_1)) = T_2'(y(\alpha;T_2))$.

Proof. For a tax function T, we have

$$
\begin{aligned}
U'(\alpha;T) &= u'\left(y\left(\alpha;T\right) - T\left(y\left(\alpha;T\right)\right)\right) \frac{\partial y\left(\alpha;T\right)}{\partial \alpha}\left[1 - T'\left(y\left(\alpha;T\right)\right)\right] \\
&\quad - v'\left(\frac{y\left(\alpha;T\right)}{\exp(\alpha)}\right)\left[\frac{1}{\exp(\alpha)}\frac{\partial y\left(\alpha;T\right)}{\partial \alpha} - \frac{y\left(\alpha;T\right)}{\exp(2\alpha)}\right] \\
&= v'\left(\frac{y\left(\alpha;T\right)}{\exp(\alpha)}\right)\frac{y\left(\alpha;T\right)}{\exp(2\alpha)},
\end{aligned}
$$

where the last line comes from substituting in the household first-order condition.

It follows that if $U'(\alpha;T_1) = U'(\alpha;T_2)$, then $y(\alpha;T_1) = y(\alpha;T_2)$. Therefore, if it is also the case that $U(\alpha;T_1) = U(\alpha;T_2)$, then $c(\alpha;T_1) = c(\alpha;T_2)$. From the budget constraint and the FOC of the household it then follows that $T_1(y(\alpha;T_1)) = T_2(y(\alpha;T_2))$, and $T_1'(y(\alpha;T_1)) = T_2'(y(\alpha;T_2))$. ∎

If the Pareto-improving constraints bind in an open interval $\Gamma \subset \mathcal{A}$, then $U(\alpha;T^{\mathrm{US}}) = U(\alpha;T^{\mathrm{PI}})$ and $U'(\alpha;T^{\mathrm{US}}) = U'(\alpha;T^{\mathrm{PI}})$ for all $\alpha \in \Gamma$. Therefore, applying the lemma, if T^{US} and T^{PI} are differentiable on Γ, then we have $c(\alpha;T^{\mathrm{US}}) = c(\alpha;T^{\mathrm{PI}})$, $y(\alpha;T^{\mathrm{US}}) = y(\alpha;T^{\mathrm{PI}})$, $T^{\mathrm{US}}(y(\alpha;T^{\mathrm{US}})) = T^{\mathrm{PI}}(y(\alpha;T^{\mathrm{PI}}))$, and $T^{\mathrm{US}\prime}(y(\alpha;T^{\mathrm{US}})) = T^{\mathrm{PI}\prime}(y(\alpha;T^{\mathrm{PI}}))$. $\qquad\qquad \mathcal{Q.E.D.}$

F   Sensitivity Analysis

We explore the sensitivity of the optimal Mirrleesian tax schedule to the extent of private insurance, the shape of the productivity distribution, and the planner's taste for redistribution. We also explore the performance of our simple parametric functional forms for taxation, and the robustness of our result that the best policy in the HSV class is preferred to the best affine policy.[7] We then consider richer parametric tax structures: higher order polynomials that give the Ramsey planner more flexibility, and tax systems that condition on observables like age and education. Finally, we show how the coarseness of the discrete grid on productivity is quantitatively important for the shape of the optimal tax schedule.

F.1   Extent of Private Insurance

We show that the result that the best policy in the HSV class is preferred to the best affine policy hinges on the existence of private insurance. Table A2 shows how allocations and tax schedules change when we rule out private insurance by setting $\sigma_\varepsilon^2 = 0$ and increasing

---

[7]We have also conducted a sensitivity analysis with respect to preference parameters: the risk aversion coefficient, γ, and the labor supply elasticity parameter, σ. Holding fixed all other structural parameters, including the taste for redistribution θ, a higher value for risk aversion and a lower labor supply elasticity (higher σ) both translate into greater optimal redistribution.

Table A2: Optimal Tax and Transfer System with No Insurable Shocks

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV$^{\text{US}}$	$\lambda : 0.851$	$\tau : 0.181$	–	–	0.335	0.024
HSV*	$\lambda : 0.802$	$\tau : 0.422$	5.62	−10.97	0.544	0.103
Affine	$\tau_0 : -0.306$	$\tau_1 : 0.581$	8.35	−11.36	0.581	0.360
Mirrlees			8.63	−11.57	0.586	0.325

the variance of α_N, the normally distributed uninsurable component, so as to leave the total variance of log wages unchanged. All other parameter values are set to their values in the baseline calibration.

Since the dispersion of uninsurable shocks is now larger than in the baseline calibration, there would now be more poverty, absent public redistribution. Thus, second-best policy now features larger lump-sum transfers to provide a firmer consumption floor (32.5 percent of GDP rather than 21.5 percent) which in turn necessitates higher marginal tax rates: the utilitarian-optimal income-weighted marginal tax rate is 58.6 percent compared to 49.1 percent in the baseline model. The maximal welfare gains from tax reform are more than four times as large as in the baseline model and are associated with an output decline of 11.6 percent.

The finding we want to emphasize is that the best affine tax system is now preferred to the best policy in the HSV class. We conclude that to accurately characterize the qualitative nature of optimal taxation it is essential to explicitly account for the existence of private insurance.

F.2 Shape of the Wage Distribution

We consider (counterfactually) eliminating the Pareto right tail in the productivity distribution, by assuming that α is Normally distributed. We adjust the variance σ_α^2 so that the total variance of α is identical to the baseline case. Relative to the baseline, the distribution for the uninsurable component of wages has a much thinner right tail and a heavier left tail. We hold fixed all other parameter values and set $\theta = 0$. Figure A3 plots the optimal tax schedule in this log-normal case (red dashed line) relative to the baseline Pareto log-normal economy.

We find that the optimal nonlinear income tax is mildly decreasing in income, rather than increasing in income. This is broadly consistent with Mirrlees (1971) original finding of roughly constant optimal marginal rates (Mirrlees also assumed a log-normal productivity
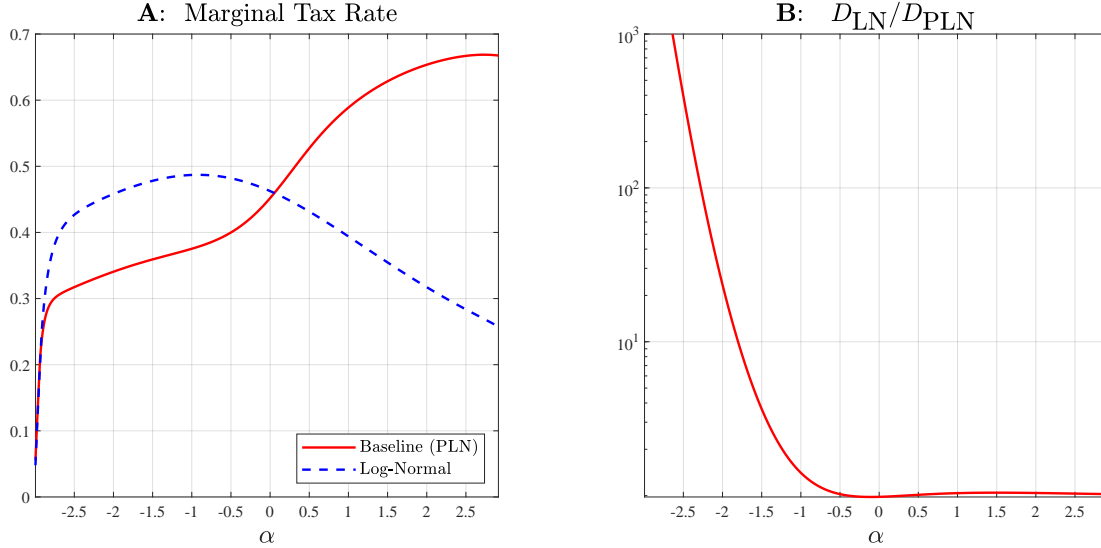
Figure A3: Log-Normal versus Pareto Log-Normal Wage Distribution. The left panel plots the ratios of the complementary CDF for household income relative to the income-weighted density. The right panel plots the profiles of optimal marginal tax rates. The plots are truncated at eight times average income.

Table A3: Optimal Tax and Transfer System with Log-Normal Wage Distribution

Tax System	Tax Parameters		Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV$^{\text{US}}$	$\lambda : 0.828$	$\tau : 0.181$	–	–	0.335	0.022
HSV*	$\lambda : 0.813$	$\tau : 0.287$	0.64	-4.50	0.427	0.048
Affine	$\tau_0 : -0.231$	$\tau_1 : 0.452$	1.96	-5.34	0.452	0.245
Mirrlees			2.06	-5.05	0.444	0.255

distribution).

The impact of the shape of the productivity distribution on the shape of the optimal tax schedule is easy to understand. Eliminating the heavy right tail in the productivity distribution reduces the distributional gains from high marginal tax rates on the rich, thus moderating optimal marginal rates at the top. Reduced revenue from soaking the rich increases the distributional gains from raising marginal rates at lower income levels. At the same time, the existence of more very low income households increases the planner's desire to provide transfers, further amplifying distributional gains at the bottom. The net result is that optimal marginal tax rates are now mildly declining in income.

Table A3 compares the optimal Mirrlees policy to the best-in-class HSV and affine tax schemes given a log-normal productivity distribution. A key result is that the best affine

Table A4: Alternative Social Preferences

Social Preferences		Mirrlees Allocations			Welfare Gain ω (%)			
	θ	$\overline{T'}$	Tr/Y	ΔY	Mirrlees	HSV*	Affine	$\frac{\omega(\text{HSV}^{\text{US}},\text{HSV}^*)}{\omega(\text{HSV}^{\text{US}},\text{Mirrlees})}$
Laissez-Faire	-1	0.082	-0.081	10.57	3.71	3.54	3.70	95%
Emp. Motivated	-0.517	0.334	0.068	0.03	0.05	–	-0.53	0%
Utilitarian	0	0.491	0.215	-7.32	2.07	1.65	1.36	80%
Rawlsian	∞	0.711	0.540	-21.98	661.6	329.3	606.2	50%

tax and transfer system dominates the best system in the HSV class and very closely approximates the second-best allocation. Thus, assuming a log-normal distribution for wages resurrects the original conclusion of Mirrlees (1971), namely, that the optimal nonlinear income tax is approximately linear. We conclude, like Saez (2001), that it is essential to carefully model the empirical productivity distribution for the purposes of providing quantitative guidance on the design of the tax and transfer system.

F.3 Alternative Social Preferences

Table A4 shows summary statistics for the optimal Mirrlees policies under the different Pareto weight functions discussed in Section 5.3.[8] Clearly, a stronger taste for redistribution translates into higher average marginal tax rates, larger lump-sum transfers, and larger output declines relative to the baseline tax policy (HSV with $\tau = 0.181$). Moving from the laissez-faire to the Rawlsian objective, the average income-weighted marginal tax rate rises from 8.2 percent to 71.1 percent.

The choice of Pareto weight function also has a huge impact on the potential welfare gains from policy reform. If we measure welfare gains using a Rawlsian welfare function as our baseline, we would conclude that tax reform could raise welfare by 662 percent. Given the empirically motivated Pareto weight function, in contrast, the maximum welfare gain from tax reform is only 0.05 percent, indicating that our HSV approximation to the U.S. tax system is close to efficient. The welfare gain here is very small because consumption and hours allocations under the current HSV schedule are very similar across most of the distribution for α to those chosen by the Mirrlees planner with taste for redistribution θ^{US} (see Figure A4). Allocations are more different at the extremes of the distribution, but the population density in those ranges is very small.[9]

---

[8]Public consumption G is fixed exogenously, and is thus invariant to θ.

[9]The maximum welfare gain from tax reform is small even though the HSV schedule violates some established theoretical properties of optimal tax schedules. In particular, it violates the prescriptions that marginal rates should be everywhere non-negative, and that the rate should be zero at the upper bound of
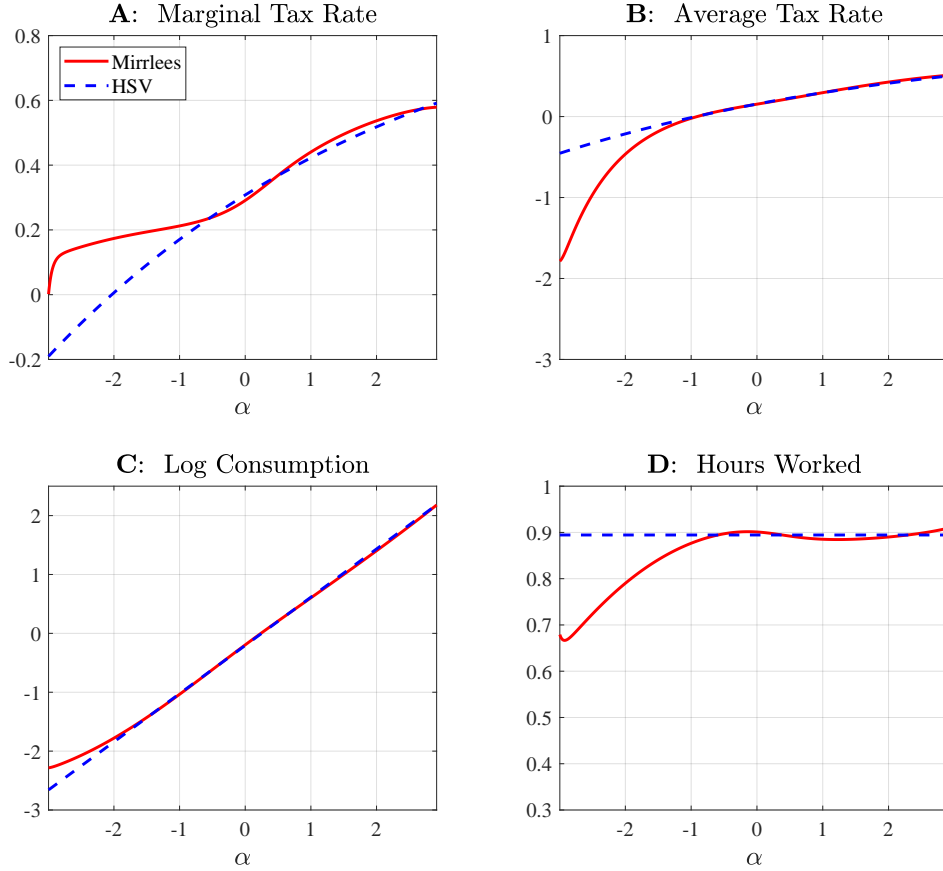
Figure A4: HSV versus Mirrlees tax functions with $\theta = \theta^{\mathrm{US}}$. The figure contrasts tax rates and allocations under the current HSV tax system to those under the Mirrlees policy using our empirically motivated Pareto weight function.

Table A4 indicates that assuming an empirically motivated Pareto weight function does not change our finding from the utilitarian case that the best-in-class HSV function is preferred to the best affine policy. In fact, given $\theta = \theta^{\mathrm{US}}$ moving from the current HSV system to the best possible affine tax scheme reduces welfare by 0.53 percent, in contrast to a 0.05 percent welfare gain under the best HSV system.

Figure A5 offers another perspective on the properties of optimal allocations at the bottom end of the income distribution. Here we plot the level of household consumption against the level of household income: net transfers is the difference between the two. We truncate the plot at 30 percent of average income to highlight how the different tax systems treat the poor. The red solid line traces out the budget set associated with constrained efficient allocations. The line stops at the red dot, which corresponds to the level of household income that the planner asks the least productive household to produce, $y^*(\alpha_1)$. As reported

the productivity distribution.
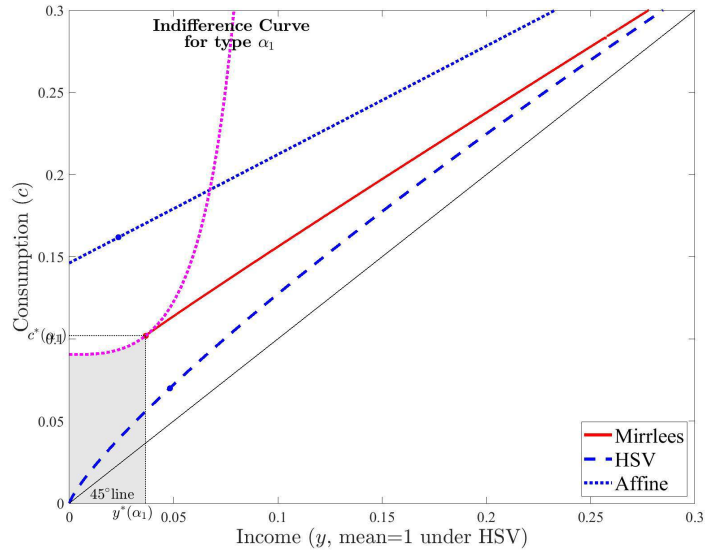
Figure A5: Allocations for low income households. The figure plots household consumption against household income at the bottom of the income distribution under the Mirrlees (red solid), HSV (blue dashed), and best-in-class affine (blue dotted) tax systems. Each tax scheme is best-in-class given the empirically motivated Pareto weight function.

in Table A4, this household receives a small net transfer. What does the Mirrlees tax schedule look like for lower income levels? An upper bound on net transfers is given by the indifference curve for the α_1 type that is tangent to the Mirrlees budget set at the point $(y^*(\alpha_1), c^*(\alpha_1))$. Any consumption schedule (and associated net tax schedule) that lies everywhere below this indifference curve will decentralize the Mirrlees solution; the set of possible such schedules is shaded light grey in the figure.

Figure A5 also plots the best income tax schedules in the affine and HSV classes. It is clear from the plot that the HSV schedule is closer than the affine one to the optimal Mirrlees schedule. The affine schedule implies net transfers that are much too generous at the bottom of the distribution, because the affine planner can only redistribute via transfers. In contrast, the Mirrlees planner prefers to redistribute primarily via an increasing marginal tax schedule. Transfers to the least productive households are small under the optimal Mirrlees policy in part because the planner puts relatively low weight on the least productive households, and in part because the fact that a portion of wage dispersion is privately insurable reduces the need for public insurance.

Figure A6 plots welfare gains under alternative tax systems, for a range of values for θ. The red solid line is the welfare gain associated with moving from the current HSV tax system to the optimal Mirrlees scheme, and the blue dashed and dotted lines are the gains moving from to the best-in-class HSV and affine schemes.

28
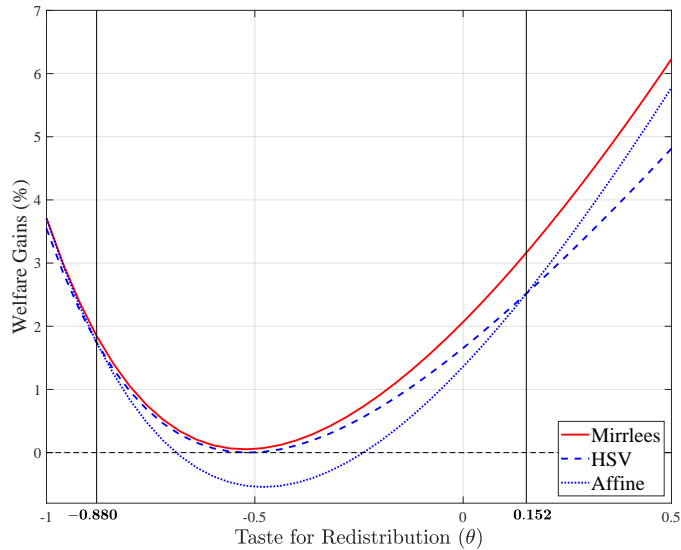
Figure A6: Maximum welfare gains from tax reform. The figure plots the maximum possible gains from tax reform for a range of values for the taste for redistribution parameter θ. Three lines are plotted, corresponding to the best policies in the unrestricted Mirrlees class (red solid), the HSV class (blue dashed), and the affine class (blue dotted).

The first message from Figure A6 is that for most intermediate values for θ, the red solid and blue dashed lines are not far apart, indicating that the lion's share of potential welfare gains from tax reform can be achieved by adjusting progressivity while retaining the HSV functional form. For example, the sizable welfare gains from tax reform that are possible under the utilitarian objective ($\theta = 0$) almost entirely reflect the fact that a utilitarian planner wants a more redistributive tax system—and do not signal that the current system redistributes in a very inefficient way.

Second, the optimal HSV scheme outperforms the optimal affine scheme for a wide range of intermediate values for θ between -0.880 and 0.152.

Third, when the taste for redistribution is either sufficiently weak or sufficiently strong, an affine scheme is preferred. For example, the laissez-faire planner prefers an affine tax because he wants to use lump-sum taxes to raise revenue; this planner chooses negative transfers. The Rawlsian planner prefers an affine tax because he values a high consumption floor for the least productive agents. However, as we argued earlier, it is difficult to reconcile the tax and transfer system currently in place in the United States with either a very low or a very high taste for redistribution.

F.4   More Elastic Labor Supply

We considered a case in which $\sigma = 0.5$, so that the labor supply is more elastic. This raises the efficiency cost of taxation, leading the planner to generally reduce marginal tax rates. The revenue-maximizing marginal tax rate on the rich is 56 percent, compared to 71 percent in the baseline parameterization. However, lower tax rates mean less revenue and lower lump-sum transfers. As in our other experiments, lower transfers mean greater distributional gains at low income levels. These larger distributional gains from raising marginal rates offset greater efficiency costs, with the net result that optimal marginal tax rates on the poor are little changed relative to the baseline calibration even as optimal rates on the rich fall sharply.

F.5   Richer Tax Structures

We explore richer tax structures. First, we consider polynomial tax functions that add quadratic and cubic terms to the affine functional form. Next, we consider an economy in which there is a third component of idiosyncratic productivity κ that is privately uninsurable but observed by the planner. We find that the potential welfare gains from indexing taxes to κ are as large as 6.2 percent of consumption.[10]

F.5.1   Polynomial Tax Functions

In the baseline model we have learned that for welfare it is more important that marginal tax rates increase with income—which the affine scheme rules out—than that the government provides universal lump-sum transfers. Relative to the affine case, we now ask how much better the Ramsey planner can do if we introduce quadratic and cubic terms in the net tax function, thereby allowing marginal tax rates to increase with income. We will find that as we give the Ramsey planner access to increasingly flexible tax functions, outcomes and welfare converge to the Mirrlees solution, which is reassuring.[11]

Let $T_n(y)$ denote an n-th order polynomial tax function: $T_n(y) = \tau_0 + \tau_1 y + \cdots + \tau_n y^n$. We assume that the marginal tax rate becomes constant above an income threshold $\bar{y}$ equal to 10 times average income in the baseline HSV-tax economy. We focus on the cases $n = 2$ and $n = 3$ (i.e., quadratic and cubic tax functions).

With polynomial tax systems, the households' first-order conditions are not sufficient in general. However, it is possible to prove that marginal utility is decreasing in income at sufficiently high income levels. Hence, for a given tax system, equilibrium allocations can be

---

[10]One interpretation of our previous analysis is that the κ component has always been present, but we have up to now imposed a restriction on tax functions such that net taxes must be independent of κ.

[11]By the Weierstrass Approximation theorem, a sufficiently high order polynomial tax function could approximate the Mirrlees solution to any desired accuracy.

Table A5: Polynomial Tax Functions

Tax System and Tax Parameters					Outcomes			
					ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
Baseline HSV								
			$\lambda : 0.840$	$\tau : 0.181$	$-$	$-$	0.335	0.023
Polynomial								
	τ_0	τ_1	τ_2	τ_3				
Proportional	$-$	0.176	$-$	$-$	-5.70	6.88	0.176	-0.009
Affine	-0.259	0.492	$-$	$-$	1.36	-7.31	0.492	0.281
Quadratic	-0.236	0.439	0.015	$-$	1.80	-7.33	0.492	0.256
Cubic	-0.213	0.371	0.049	-0.002	1.99	-7.32	0.491	0.230
Mirrlees					2.07	-7.32	0.491	0.215

found by evaluating all roots of the household first-order necessary conditions in the range $[0, y]$ with y sufficiently large. For both the quadratic and cubic polynomial cases we search for the optimal tax function coefficients using the Nelder-Mead simplex method. We check that the social welfare maximizing policy is independent of the initial set of tax parameters used to start the search process. Table A5 presents outcomes for the best policies in the quadratic and cubic classes.

With the quadratic function, marginal tax rates are increasing in income ($\tau_2 > 0$)—the key property of the optimal tax schedule. Relative to the affine case, the linear coefficient τ_1 is reduced, and lump-sum transfers τ_0 are also smaller. Thus, the planner relies more heavily on increasing marginal tax rates, rather than lump-sum transfers, as the primary tool for redistribution. Under the cubic system, the linear coefficient and lump-sum transfers are reduced still further, whereas the quadratic coefficient τ_2 is larger, so that marginal tax rates now rise more rapidly at low income levels. The cubic coefficient τ_3 is negative, which ensures that marginal rates flatten out before income reaches the threshold $\bar{y}$. Because marginal and average tax rates under the best cubic policy are generally very similar to those implied by the Mirrlees solution, allocations are close to being constrained efficient. Thus, moving to the best cubic policy generates about 97 percent of the maximum potential welfare gains from tax reform.

Figure A7 is the analogue to Figure 6 for the cubic tax function. The bottom panels show that allocations under the cubic policy are generally close to the constrained efficient Mirrlees solution. In particular, for intermediate values for productivity (where the vast majority of households are concentrated), marginal and average tax rates are very similar
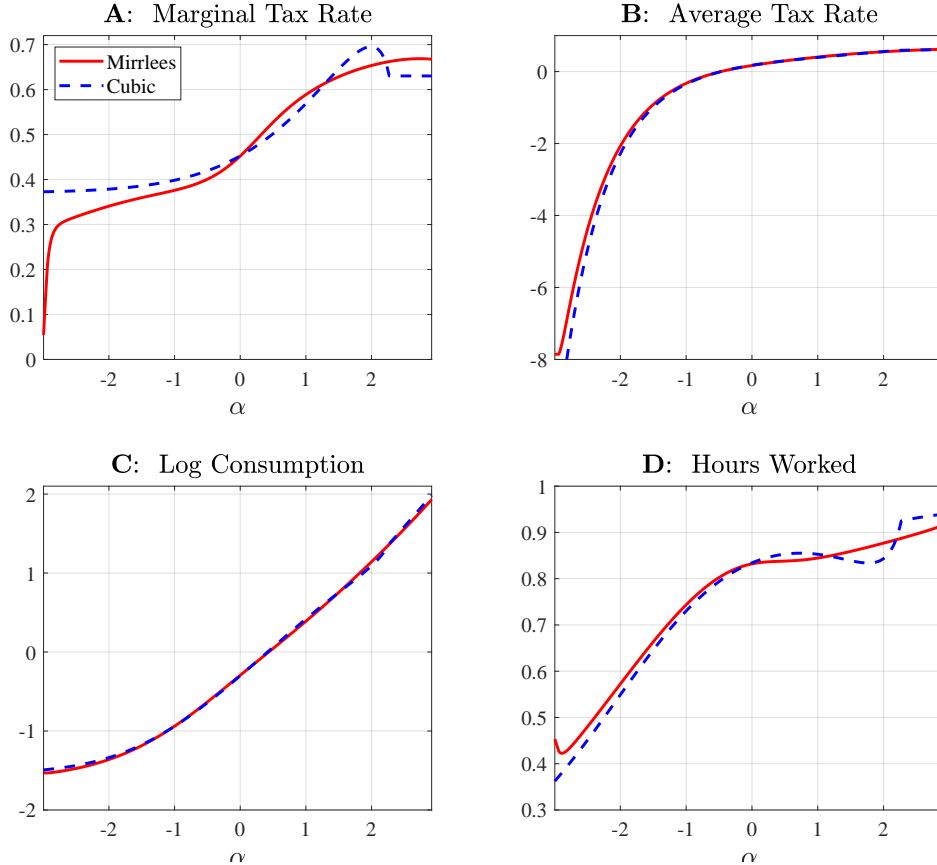
Figure A7: Cubic Tax Function. The figure contrasts allocations under the best-in-class cubic and Mirrlees tax systems. The top panels plot marginal and average tax schedules, and the bottom panels plot decision rules for consumption and hours worked. The plot for hours worked is for an agent with average ε.

to those implied by the Mirrlees solution. This explains why the cubic system comes very close, in welfare terms, to the Mirrlees solution.

F.5.2 Type-Contingent Taxes

In the baseline model, idiosyncratic productivity was divided into a privately uninsurable component α and a privately insurable component ε. Now we introduce a third component κ which is privately uninsurable but observed by the planner. This component is designed to capture differences in wages related to observable characteristics such as gender, age, and education. We assume that κ is drawn before family insurance comes into play and therefore cannot be insured privately.

We set the variance of this observable fixed effect σ_κ^2 equal to the variance of wage dispersion that can be accounted for by standard observables in a Mincer regression. Heathcote et al. (2010) estimate the variance of cross-sectional wage dispersion attributable to observables to be $\sigma_\kappa^2 = 0.108$. For the sake of simplicity, we assume a two-point equal-weight

distribution for κ. This gives $\exp(\kappa_{\text{High}})/\exp(\kappa_{\text{Low}}) = 1.93$.

The total variance of the privately uninsurable component of wages is unchanged relative to the baseline model, but we now attribute part of this variance to κ. The three parameters $\mu_\alpha, \sigma_\alpha^2$, and λ_α characterizing the EMG distribution for α are therefore recalibrated so that (i) the variance of (discretized) α is equal to that in the baseline model minus σ_κ^2, (ii) $\sum_i \pi_i \exp(\alpha_i) = 1$, and (iii) the value of the shape parameter $\sigma_\alpha \lambda_\alpha$ is the same as that in the baseline model (i.e., 0.829).[12]

When the planner can observe a component of productivity, the optimal tax system explicitly indexes taxes to that component (see, e.g., Weinzierl 2011). In the extreme case in which productivity is entirely observable, so that $\log w = \kappa$, the optimal system simply imposes a κ-specific lump-sum tax for each different value for κ. More generally, each different κ type faces a type-specific income tax schedule $T(y; \kappa)$.

Table A6 describes optimal type-contingent tax functions and the associated outcomes. The subscripts H and L correspond to tax schedule parameters for the κ_{High} and κ_{Low} types, respectively. We find that if the planner can condition taxes on the observable component of labor productivity, it can generate large welfare gains relative to the current tax system, which does not discriminate by type. The maximum (Mirrlees) welfare gain is now 6.18 percent of consumption, compared with 2.07 percent in the baseline analysis. This large welfare gain arises in part because the average effective marginal tax rate drops to 42 percent, which translates into a smaller output loss. Recall that if productivity were entirely observable, the planner could implement the first best, with a zero marginal rate for all households.

By implementing type-contingent tax systems, the Ramsey planner achieves welfare gains that nearly match those under the Mirrlees planner. Under an affine system, the high κ type faces a double whammy, paying higher marginal tax rates than the low type ($\tau_1^H > \tau_1^L$) and paying lump-sum taxes rather than receiving transfers ($\tau_0^H > 0 > \tau_0^L$). Higher marginal rates are an effective way for the planner to redistribute from the high to the low type (recall that κ enters the level wage multiplicatively), whereas the wealth effect associated with lump-sum taxes ensures that high κ households still work relatively hard.

One important caveat to this analysis is that we have treated all the variation in κ as exogenous and have therefore ignored potential feedback from the tax system to the distribution for κ. However, an education-dependent tax system would likely affect agents' educational decisions (see, e.g., Heathcote et al. 2017). In particular, relatively high taxation of high κ households would discourage education investment. Thus, we regard our 6.18 percent welfare gain as an upper bound on the feasible welfare gains from tagging.

---

[12]The shape parameter controls the relative importance of the normal and exponential distribution components.

Table A6: Type-Contingent Taxes

Tax System			Outcomes			
			ω (%)	ΔY (%)	$\overline{T'}$	Tr/Y
HSV$^{\text{US}}$	$\lambda : 0.834$	$\tau : 0.181$	–	–	0.335	0.019 / 0.025
HSV*	$\lambda^L : 1.067$ / $\lambda^H : 0.596$	$\tau^L : 0.481$ / $\tau^H : 0.075$	5.85	−2.09	0.417	0.148 / −0.019
Affine	$\tau_0^L : -0.402$ / $\tau_0^H : -0.034$	$\tau_1^L : 0.346$ / $\tau_1^H : 0.453$	5.79	−1.84	0.422	0.422 / 0.011
Mirrlees			6.18	−1.84	0.419	0.370 / 0.009

Table A7: Grid points

# of grid points I	ω (%, relative to HSV)		
	Affine	Mirrlees	First Best
10	1.32	20.38	46.37
100	1.36	4.41	44.81
1,000	1.36	2.28	44.72
10,000	1.36	2.07	44.72
100,000	1.36	2.05	44.72

F.6 Coarser Grids

In the baseline model, we set the number of grid points for α to $I = 10,000$. This is a very large number relative to grid sizes typically used the literature. However, assuming the true distribution for α is continuous, a very fine grid is required to accurately approximate the second-best allocation.

To make this point, in Table A7 we report welfare gains from tax reform (relative to the HSV baseline) as we increase the number of grid points from $I = 10$ to $I = 100,000$. Reassuringly, the number of grid points does not affect the results for the affine case or for the first best. However, as the number of grid points decreases, welfare gains for the Mirrlees planner increase substantially. For $I = 10$ these gains are 20.4 percent of consumption, compared with 2.07 percent with $I = 10,000$.

The intuition behind this result is that with a coarse grid, ensuring truthful reporting becomes easier for the Mirrlees planner. Consider a grid of size I. A common result is that only local downward incentive constraints bind at the solution to the planner's problem. Now

suppose that we remove every other point from the grid, leaving all else unchanged. At the original conjectured solution, none of the incentive constraints are now binding. Thus, the planner can adjust allocations to compress the distribution of consumption or to strengthen the correlation between productivity and hours worked.

In our static model, using a very fine grid is not too costly from a computational standpoint. In dynamic Mirrleesian settings, however, the presence of additional state variables typically necessitates a very coarse discretization of types, often using fewer than 50 points. The findings in Table A7 cast some doubt on the quantitative robustness of such analyses if the true underlying productivity distribution is continuous.

References

EBERT, U. (1992): "A reexamination of the optimal nonlinear income tax," *Journal of Public Economics*, 49, 47–73.

HEATHCOTE, J., F. PERRI, AND G. L. VIOLANTE (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 13, 15–51.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104, 2075–2126.

——— (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132, 1693–1754.

——— (2019): "Optimal Progressivity with Age-Dependent Taxation," mimeo.

MANKIW, N. G., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23, 147–174.

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

PRATT, J. W. (1964): "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 122–136.

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

WEINZIERL, M. (2011): "The Surprising Power of Age-Dependent Taxes," *Review of Economic Studies*, 78, 1490–1518.