Optimal Income Taxation: Mirrlees Meets Ramsey*

Jonathan Heathcote[†]

Federal Reserve Bank of Minneapolis and CEPR

Hitoshi Tsujiyama[‡]

Goethe University Frankfurt

May 2021

Abstract

We explore the optimal shape of the income tax and transfer schedule in an environment with distinct roles for public and private insurance. In a calibration to the U.S. we find that the optimal system features marginal tax rates that increase in income. When we increase pressure on the government to raise revenue, the optimal marginal tax schedule becomes first flatter and then U-shaped, reconciling various findings in the literature. A power function parametric tax schedule outperforms an affine one, indicating that tax progressivity is more important than lump-sum transfers. We also explore various social welfare objectives and Pareto-improving reforms.

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

---

*We thank Job Boerma, V.V. Chari, Mikhail Golosov, Nezih Guner, Christian Hellwig, Martin Hellwig, James Peck, Richard Rogerson, Florian Scheuer, Ctirad Slavik, Kjetil Storesletten, Aleh Tsyvinski, Gianluca Violante, Yuichiro Waki, Matthew Weinzierl, Tomoaki Yamada, and three anonymous referees for helpful comments. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

[†]Email: heathcote@minneapolisfed.org Address: Federal Reserve Bank of Minneapolis, Research Department, 90 Hennepin Ave, Minneapolis, MN 55401.

[‡]Email: hitoshi.tsujiyama@hof.uni-frankfurt.de Address: Goethe University Frankfurt, Department of Money and Macroeconomics, Theodor-W.-Adorno-Platz 3, 60629 Frankfurt, Germany.

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 features lower tax rates and smaller lump-sum transfers.

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. This pattern is robust to a range of alternative values for preference parameters.

We 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 amount of fiscal pressure that the government faces to raise revenue plays a key role in shaping the optimal tax schedule. When fiscal pressure is relatively low—for example, because required government expenditure is low—the optimal tax schedule is upward sloping. 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.

When fiscal pressure is sufficiently increased, the optimal schedule becomes first flatter and then U-shaped, as in Saez (2001). Flattening the marginal rate profile when fiscal pressure is high is desirable because this delivers high average tax rates (and thus high revenue) at the same time as keeping marginal tax rates relatively low.

Under our baseline calibration, total government spending (purchases plus transfers) is 41.8 percent of GDP under the optimal policy featuring increasing marginal rates. The corresponding figure for the U.S. in 2015 was 33.5 percent. Saez (2001) calibrations are associated with much higher spending levels, of around 56 percent of GDP. We conclude that the reason he finds a U-shaped optimal profile while we do not is that the planner in his calibration faces much greater pressure to raise revenue.

To better understand the equity versus efficiency trade-off in tax design, we formalize measures of the marginal distributional gains and efficiency costs associated with changing marginal tax rates. These gains and costs are equated at all income levels at the optimum. While both efficiency costs and distributional gains shape the optimal tax schedule, the more constructive understanding comes from focusing on distributional gains, since this is where the endogenous terms that respond to fiscal pressure appear.

Distributional gains from raising marginal rates are always high at the top of the income distribution, since taxes at the top redistribute away from the richest households. Thus the planner always sets high marginal rates at the top, 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 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 the consumption distribution is relatively compressed. Thus, the planner does not want to impose high marginal tax rates on the moderately poor to further increase lump-sum 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 flatter or decreasing marginal tax profile from low to middle incomes.

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.

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) and Brendon (2013). 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 (2020) 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]

[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. Note that in an explicit life-cycle framework one could consider the joint design of taxation and pension systems (see, e.g., Shourideh and Troshkin 2017, Choné and Laroque 2018, or Ndiaye 2020).

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 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

$$\begin{aligned} c(\alpha,\varepsilon) &= c(\alpha) = y(\alpha) - T(y(\alpha)), & (5)\\ h(\alpha,\varepsilon)^{\sigma} &= [y(\alpha) - T(y(\alpha))]^{-\gamma} \exp(\alpha+\varepsilon)\left[1 - T'(y(\alpha))\right]. & (6) \end{aligned}$$

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

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

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

[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.

(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]

3.1 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.

[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.

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

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

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

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

Solving this problem gives the following indirect utility function:

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

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

[5] Note that the weight on hours in the agents' utility function is now Ω rather than 1.

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:

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

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, $\mathcal{T} = \{T : \mathbb{R}_+ \to \mathbb{R} | T(y) = \tau_0 + \tau_1 y$ 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:

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

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

[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.

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

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

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

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

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.

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 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 hypothetical 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

[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.

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]

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 Our baseline calibration assumes 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)

[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.

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 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 average ratio of government purchases to output in the United States over the 2000-2006 period.[12] 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

[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).

[12] National Income and Product Accounts, Table 1.1.10. Heathcote et al. (2017) use the same value.

will show up in consumption, whereas insurable shocks will not.[13]

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 log consumption are also EMG with[14]

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

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

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

[13] 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.

[14] 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 $1-\tau$.

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.[15] The resulting estimates are $\lambda_\alpha = 2.2$ and $\sigma_y^2 = 0.412$, implying a total variance for log earnings of 0.618.[16]

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 the variance of α. Using the Consumer Expenditure Survey, Heathcote et al. (2010, figure 13) report a variance of 0.332 in 2006.[17] 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.[18] 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 5.2, we will explore

[15] 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.

[16] 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.

[17] 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.

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

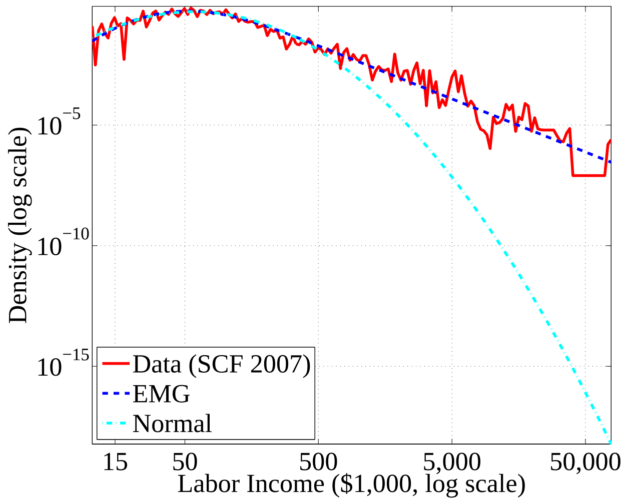

Figure 1: Fit of EMG distribution. The figure plots the empirical earnings density from the SCF against the estimated EMG distribution and against a normal distribution.

how changing the relative magnitudes of insurable and uninsurable wage risk changes the optimal tax schedule. 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.[19] 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.

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

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

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.

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.

Units Given our baseline model calibration, the model implies values for average earnings and average hours worked which we denote Y and H respectively. For the purpose of comparing model to data it is convenient to rescale model units. We will target average annual earnings in our 2007 SCF household sample, which is $\bar{Y} = \$77,326$. The SCF does not collect information on hours worked, so for an empirical target for average household hours we turn to the CPS. When using the same selection for age of household head (25–60) as in the SCF, average annual household hours worked in 2007 in the CPS are $\bar{H} = 3,075$.[20] When plotting model allocations we scale model earnings, consumption and taxes by a factor $\bar{Y}/Y$, and wages by $\bar{w} \equiv (\bar{Y}/Y)/(\bar{H}/H)$. In Appendix C, we provide the theoretical justification for rescaling model variables in this fashion.

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

[20] Using the same income definition (wage and salary income plus two-thirds of self-employment income) as in the SCF, average income in the CPS is $\$67,753$. It is thus somewhat lower than in the SCF. This likely reflects under-representation of high income households in the CPS.

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).[21] 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 Heathcote and Tsujiyama (2021), we show that a very fine grid is required to accurately solve the Mirrlees problem.

5 Quantitative Analysis

We explore the structure of the optimal tax and transfer system, given the model specification described above.[22] We assume that the Pareto weight function takes the form

$$W(\alpha;\theta) = \frac{\exp(-\theta\alpha)}{\int \exp(-\theta\alpha) dF_\alpha(\alpha)} \quad \text{for } \alpha \in \mathcal{A}. \tag{18}$$

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 households. We

[21]Given $Y/H = \exp(\frac{\sigma_\varepsilon^2}{\sigma}) = 1.06$, the average hourly wage is $\bar{w} = (\bar{Y}/Y)/(\bar{H}/H) = \$23.68$, so 5 percent of the average corresponds to \$1.18, which is less than a quarter of 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.

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

focus initially on a utilitarian social welfare function, $\theta = 0$, with equal Pareto weights on all households.

5.1 Increasing Optimal Marginal Tax 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.[23]

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}).[24] In contrast to Diamond and Saez, optimal marginal tax rates are always increasing in income (except at the very top).[25] The marginal rate starts at 5.5 percent for the least productive households, is fairly flat (between 30 and 40 percent) for households making up to around \$15 per hour, and rises rapidly to peak at 66.9 percent for those making around \$350 per hour.[26] Panel A indicates that the optimal schedule imposes much higher marginal tax rates than our approximation to the current U.S. system: the average income-weighted marginal tax rate is 49.1 percent, compared to 33.5 percent under the current policy (see table 2).

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 defi-

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

[24]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 E plots the marginal tax rate against the *level* of income.

[25]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.

[26]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.

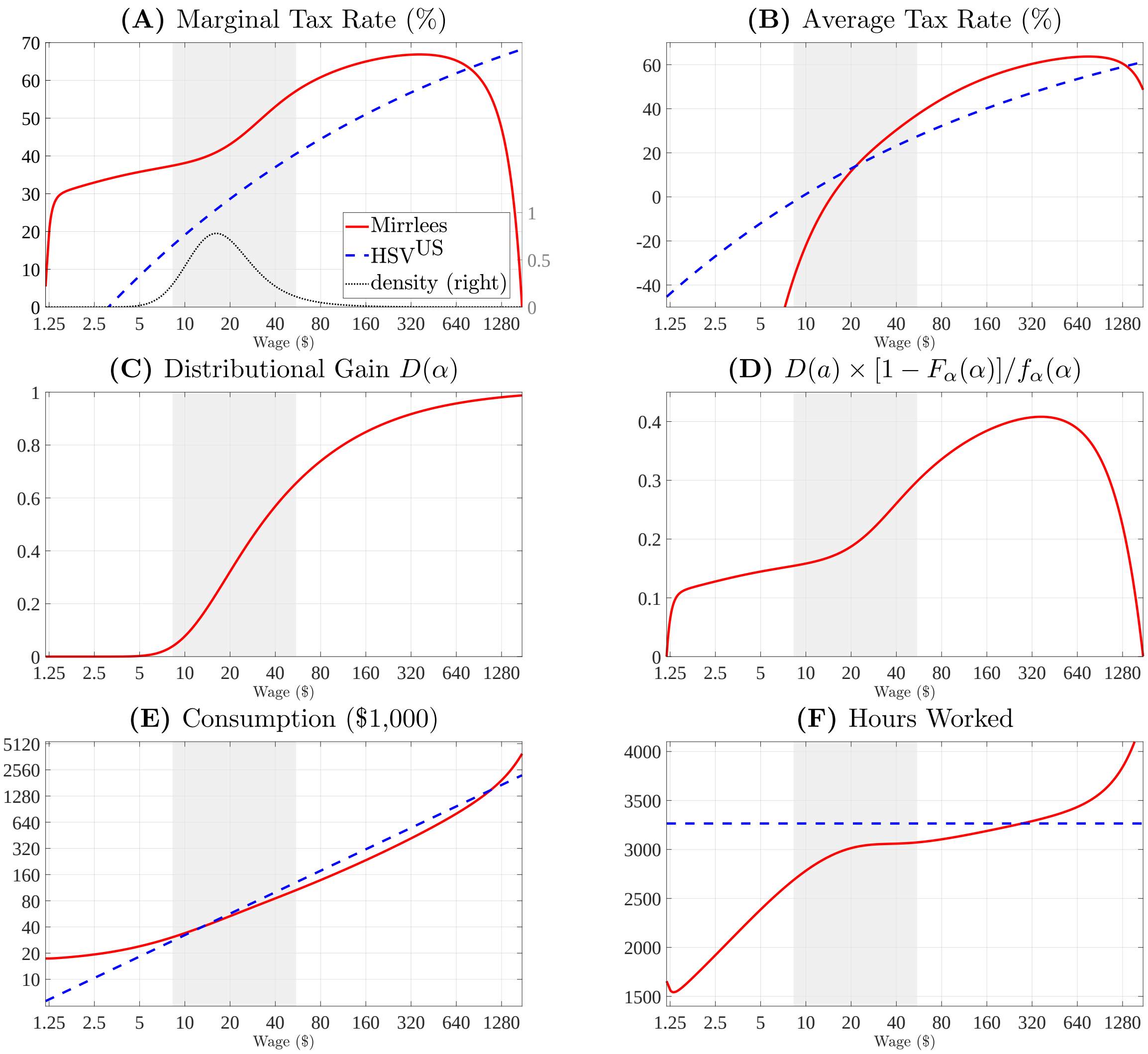

Figure 2: Optimal Tax Policy. Panels A and B plot the optimal Mirrleesian tax schedules against the baseline HSV approximation to the U.S. tax and transfer system and the productivity density. Panels C and D plot the distributional gain (equivalently, the efficiency cost) under the optimal policy. Panels E and F plot decision rules for consumption and hours worked. The x axis for each plot shows the household average wage, $\bar{w}\exp(\alpha)$. Hours are defined as household earnings divided by the household average wage. The area between the 5th and 95th percentiles is shaded gray.

nitions of these measures).[27] Panel D plots the distributional gain/efficiency cost multiplied by the Mills 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 a preference specification without income effects (see Section 5.3). This resemblance indicates that if we can understand what drives distributional gains, we can better understand

[27]In Appendix B.1, we describe how we derive $D(\alpha)$ and $E(\alpha)$ from $D(y)$ and $E(y)$.

optimal taxation.[28]

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 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 or a wage of \$71 per hour), the efficiency cost is 0.71, indicating that 71 percent of every hypothetical marginal dollar in tax 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.[29] 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 (\$15,400 or 21.5 percent of average income) that consumption inequality across the bottom half of the productivity distribution is quite low (see panel E). 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 $\alpha = \ln(1/3)$ (\$7.89 per hour) 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

[28] In Section 5.3, we will further develop the argument that understanding how distributional gains vary with productivity is the key to understanding the shape of the optimal tax schedule.

[29] 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 $\bar{T}' = \left[1 + \bar{\zeta}^u + \bar{\zeta}^c(\lambda_y^* - 1)\right]^{-1} = \frac{\sigma+\gamma}{\sigma+\lambda_y^*}$, 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, $\bar{\zeta}^u = \frac{1-\gamma}{\sigma+\gamma}$, $\bar{\zeta}^c = (\sigma+\gamma)^{-1}$, and $\lambda_y^* = \lambda_\alpha \frac{\sigma+\gamma}{1+\sigma}$. In the baseline case where $\gamma = 1$, we obtain $\bar{T}' = \frac{1+\sigma}{\sigma+\lambda_\alpha}$. Note that this expression is independent of the value for government purchases. Evaluated at our calibrated values for σ and λ_α, the above equation implies that there is nothing to be gained from raising marginal rates above 71.4 percent at the top.

Table 2: Key Statistics

Model	Outcomes					
	$\overline{T'}$	Tr	$\frac{Tr}{Y}$	$\frac{Tr+G}{Y}$	ω	ΔY
HSV$^{\text{US}}$	33.5	$1,753$	2.3	21.1	—	—
Baseline	49.1	$15,400$	21.5	41.8	2.07	-7.32
High Risk Aversion: $\gamma = 2$	59.8	$22,722$	32.1	52.9	5.12	-9.63
High Labor Elasticity: $\sigma = 1$	42.6	$12,310$	17.4	37.4	0.87	-5.85
High Government Expenditure: $g = 0.4$	52.9	$5,633$	7.3	47.8	0.19	-1.07
No Insurable Shocks: High Normal Var.	58.6	$21,586$	32.5	53.8	8.63	-11.57

Note: $\overline{T'}$ is the average income-weighted marginal tax rate in percent. Tr is transfers defined as consumption minus income for the lowest earning household in 2007 dollars. $\frac{Tr}{Y}$ is transfers as a percentage of average income. $\frac{Tr+G}{Y}$ is total government spending, measured as transfers plus government purchases, as a percentage of average income. ω is the welfare gain of moving from the current tax system T to the optimal one $\hat{T}$, defined as the percentage increase in consumption for all agents under policy T that leaves the planner indifferent between T and $\hat{T}$. ΔY is the associated percentage change in aggregate output.

distributional gains are very small at low income levels, optimal marginal tax rates are much lower at low income levels than at high income levels. At the very bottom of the income distribution, bunching is optimal in our economy, implying marginal tax rates that rise quickly with wages.[30]

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. From an efficiency standpoint, in contrast, a downward-sloping 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.

Table 2 reports some properties of taxes and transfers under our approximation to the

[30]Plotted against income, marginal tax rates jump upwards at the income bunching value. We discuss this bunching in more detail in Appendix E.

current US tax and transfer system (HSV^{US}), under the optimal policy given the baseline parameterization, and under optimal policy for various alternative parameterizations.

Robustness We now show that our finding of an upward-sloping optimal marginal tax schedule is robust to a range of alternative values for the risk aversion coefficient, γ, and the curvature parameter over labor supply, σ.[31]

We start by considering $\gamma = 2$ and $\gamma = 5$, and compare the optimal tax schedule in these cases to the one under our baseline logarithmic specification ($\gamma = 1$). Panel A of figure 3 plots optimal marginal tax rates.[32] With higher risk aversion, the planner chooses uniformly higher marginal tax rates. The optimal marginal tax schedule becomes flatter as γ is increased, but remains generally upward-sloping. There are two forces behind these higher tax rates. First, with more curvature in utility, the planner sees larger gains from redistribution, pointing to higher tax rates to fund larger transfers. Second, the larger is γ, the stronger are income effects in labor supply choices. Higher tax rates therefore depress labor supply by less, especially toward the top of the productivity distribution, because the depressing effect of taxes on consumption is associated with a stronger positive income effect on labor supply.

Thus, as risk aversion increases, the optimal tax and transfer system becomes much more redistributive, with the average income-weighted marginal tax rate rising from 49.1 percent when $\gamma = 1$ to 59.8 percent when $\gamma = 2$, and net transfers for the least productive households rising to $\$22,722$ (see table 2).

Next we turn to the elasticity of labor supply. Panel B of figure 3 describes the optimal marginal tax rates when $\sigma = 1$ and $\sigma = 4$, implying Frisch elasticities of 1 and 0.25, respectively, in addition to the baseline case in which $\sigma = 2$. When labor supply is more

[31] We leave all other parameters unchanged, besides the value for public consumption G that must be financed. We adjust G each time we change γ or σ so that the ratio of government purchases to output in the economy with HSV taxation remains identical to the value in the data. In Appendix G we discuss these exercises in more detail and conduct extensive additional sensitivity analyses.

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

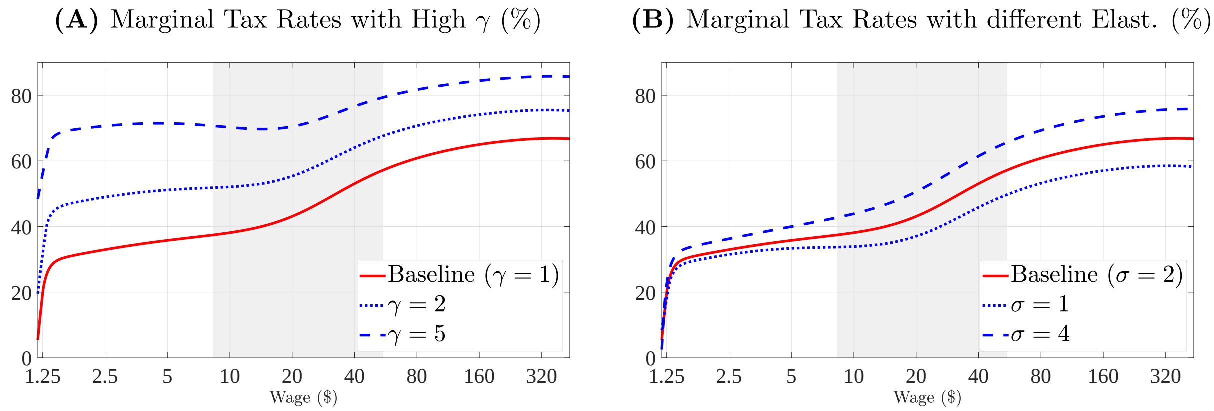

Figure 3: Sensitivity. The figure plots the optimal Mirrleesian tax schedules with higher risk aversion (panel A) and with a higher/lower labor supply elasticity (panel B).

(less) elastic, the efficiency cost of taxation is larger (smaller), and optimal marginal tax rates are reduced (increased). With $\sigma = 1$, the optimal policy is closer to our approximation to the current one, and the welfare gains from the optimal reform are smaller.

5.2 High Fiscal Pressure and U-Shaped Optimal Taxes

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 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 are many very low productivity households. Neither of these scenarios applies to our baseline calibration to the United States, but these experiments are useful for better understanding what shapes the optimal tax schedule. In Section 5.4 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 4 plots the optimal marginal tax schedules when we increase G from the baseline value (18.8 percent of output under the HSV$^{\text{US}}$ 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.

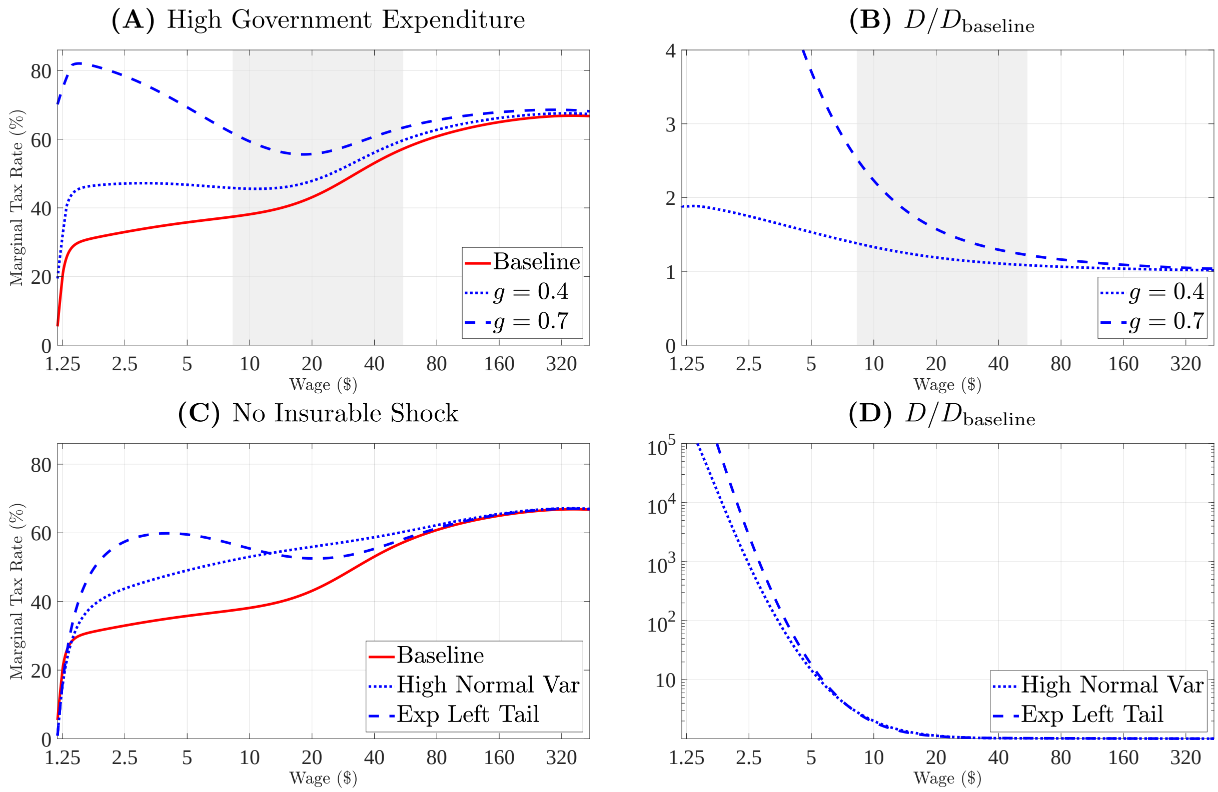

Figure 4: Increasing Fiscal Pressure. Panels A and B plot the optimal marginal tax schedules and the corresponding distributional gains for different values of government purchases. For example, the line labeled $g = 0.4$ corresponds to a value for G equal to 40 percent of output under the HSV$^{\text{US}}$ 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.

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.[33] This new pattern of marginal rates is optimal because a higher required expenditure squeezes lump-sum transfers (see table 2), 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 the very poorest.[34] Distributional gains near the top of the income

[33]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.

[34]Slemrod et al. (1994) explored the sensitivity of optimal policy with respect to the government's revenue

distribution are always close to one at the optimum, the maximum possible value, because asking the rich to pay higher taxes always has a small negative effect on social welfare. Thus distributional gains at the top do not rise much when fiscal pressure is increased. 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.

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 wage distribution. 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.[35]

The Diamond-Saez implicit formula for optimal marginal tax rates only limited intuition for the link between fiscal pressure and optimal taxation. That formula 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.[36] The problem is 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.

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.

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

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

Increasing Uninsurable Risk The only parameters we have not yet explored are those describing the distribution for idiosyncratic labor productivity, and the extent to which this dispersion can be insured privately.

Panel C of figure 4 plots the optimal marginal schedules when we assume there are no insurable shocks—a lower bound for the extent of private insurance—and increase uninsurable risk to leave the total model variance of earnings unchanged.

We run two versions of this experiment.[37] First, we assume that the extra uninsurable risk is normally distributed. This case gives a flatter optimal marginal tax profile than the baseline, but even in this extreme case without private insurance, the optimal marginal tax profile remains upward-sloping. Thus our finding of an upward-sloping profile is robust to any plausible variation in the extent of private insurance, as long as we retain the assumption that wages follow a Pareto lognormal distribution.[38]

Second, we introduce a heavy left tail in the wage distribution so that more individuals have very low (and privately uninsurable) productivity.[39] This second case gives a U-shaped optimal marginal tax profile. This specification implies significant inequality in the bottom half of the productivity distribution and thus much larger distributional gains from taxing the moderately poor to increase lump-sum transfers benefiting the very poor (panel D).

However, those larger transfers mean smaller distributional gains from raising rates in the middle of the wage distribution, which translates to lower marginal tax rates there. Note

[37] 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: $\hat{\sigma}_\alpha^2 = \sigma_\alpha^2, \hat{\lambda}_\alpha^{-2} = ((1+\sigma)/\sigma)^2\sigma_\varepsilon^2$.

[38] If we eliminate the exponential *right* tail in the distribution for α, so that α is Normally distributed, then optimal marginal tax rates decline with productivity over most of the productivity distribution (see figure A6 in Appendix G.4). However, that finding is of more theoretical interest than practical relevance, since the heavy Pareto-like right tail in the empirical earnings distribution is a long-recognized feature of U.S. data, and clearly points to an exponentially distributed component in the right tail of productivity distribution.

[39] The percentile ratios of productivity are P5/P1=2.02, P10/P5=1.41, and P25/P10=1.64, compared to 1.46, 1.23, and 1.42 in the baseline economy that we reported in table 1.

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.[40]

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.[41] 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.[42] Thus, the required government tax take is 56 percent of GDP. In our baseline parameterization, the corresponding number is 41.8 percent (transfers are 21.5 percent of GDP, and purchases are 20.3 percent; see table 2).

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 4 (dashed blue lines), government purchases plus transfers are 59.9 percent (panel A) and 54.7 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.[43] 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.

[40] 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.

[41] 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.

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

[43] National Income and Product Accounts, Table 3.1.

5.3 Preferences without Income Effects à la Diamond (1998)

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

$$\log\left(c - \frac{h^{1+\sigma}}{1+\sigma}\right). \tag{19}$$

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).[44] 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 4. 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.

Given the utility function (19), the efficiency cost of taxation is given by

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

This expression makes clear that thinking about how the efficiency cost of taxation varies with productivity is of limited value in understanding the shape of the optimal tax schedule. First, besides the marginal tax rate itself, efficiency costs vary only because of exogenous variation in the inverse Mills ratio. We have shown that the level of government purchases G plays a key role in shaping the optimal tax schedule, but it does not show up in the efficiency cost expression (neither do Pareto weights $W(\alpha)$). Second, given our baseline specification with the EMG distribution for α, the inverse Mills ratio is increasing in α, suggesting a motive for declining marginal tax rates, while (as we will see) the optimal marginal tax schedule in our calibrated example is increasing in α.[45] Thus, the preference specification

[44] 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.

[45] Note that given an exponential distribution for α, the inverse Mills ratio would be constant and equal to λ_α. In that case, efficiency costs vary with productivity only because marginal tax rates do. So any slope to

without income effects clearly illustrates the importance of distributional gains in shaping the optimal tax schedule.

Given the utility function (19), optimal tax rates must satisfy

$$\underbrace{\int_{\alpha}^{\infty}\left\{1-\frac{u_c(s)}{\chi}\right\}dF_{\alpha}(s)}_{\tilde{D}(\alpha)\equiv[1-F_{\alpha}(\alpha)]D(\alpha)} \quad = \quad \underbrace{\frac{1}{1+\sigma}\frac{T'(\alpha)}{1-T'(\alpha)}f_{\alpha}(\alpha)}_{\tilde{E}(\alpha)\equiv[1-F_{\alpha}(\alpha)]E(\alpha)} \quad \text{for all } \alpha, \tag{20}$$

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 equation (9) in Diamond (1998) when his G function is logarithmic.

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.

Panel A of figure 5 plots the total distributional gain term $\tilde{D}(\alpha)$ under the optimal policy and the density $f_{\alpha}(\alpha)$ for a parameterization similar to the one described in Section 4.[46] Note that the distributional gain term peaks *after* $f_{\alpha}(\alpha)$ (i.e., $\alpha^* > \alpha_m$), so Diamond's condition

the optimal marginal tax schedule must be *entirely* driven by distributional concerns. The optimal marginal tax schedule in such a case is in fact strongly upward sloping (see figure A7 in Appendix G.4).

[46] 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.

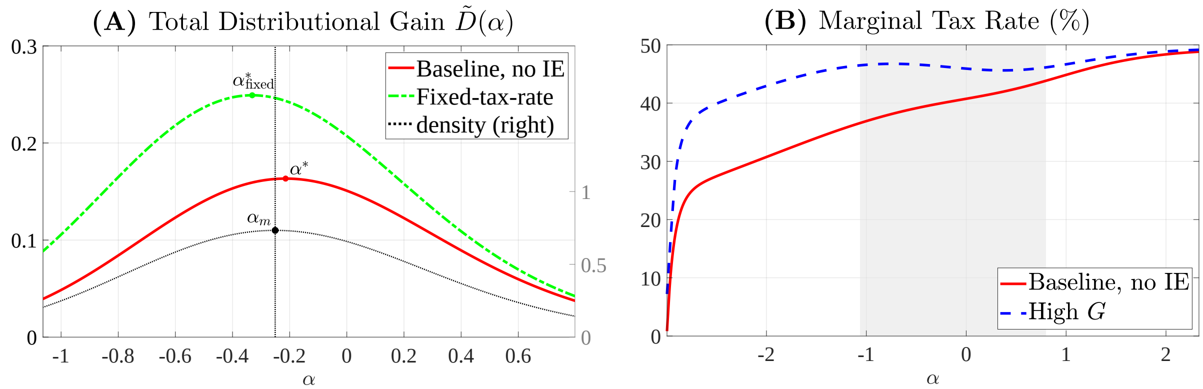

Figure 5: Preferences without Income Effects. Panel A plots distributional gains and the productivity density (between the 5th and 95th percentiles). The peak of each curve is indicated by a dot. The solid red line is the baseline case. The dotted green line corresponds to the interim case in which G 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.

for the optimal marginal tax schedule to have a downward-sloping portion is not satisfied. The optimal marginal tax schedule plotted in panel B is in fact everywhere increasing, as in our baseline calibration (see figure 2).[47]

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 (19), a reduction in lump-sum transfers (i) has no effect on efficiency costs, $\tilde{E}(\alpha)$, (ii) increases total 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 F.1. ■

Result (i) is trivial: lump-sum transfers do not affect labor supply given the preferences in eq. (19). Result (ii) is intuitive and reflects the fact that with lower lump-sum transfers, there 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,

[47] In discussing the condition $\alpha^* < \alpha_m$, 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). One might interpret Diamond as arguing here that the condition will be satisfied if the planner has a strong enough desire to redistribute. In Section 5.4 we show that when the planner has a strong taste for redistribution, the optimal marginal tax schedule does indeed have a downward-sloping portion.

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 5 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.[48] 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).[49]

5.4 Alternative Social Preferences

To this point, we have explored optimal policy assuming the planner is utilitarian ($\theta = 0$), the most common assumption in the literature. We now consider alternative Pareto weight functions. The one-parameter specification considered in eq. (18) nests several classic social preference specifications. The case $\theta = -1$ corresponds to a laissez-faire planner, with planner weights inversely proportional to equilibrium marginal utility absent redistributive

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

[49] 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 now satisfied, which accounts for the U-shape of the optimal profile in panel B.

taxation.[50] 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.[51]

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 (18) 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 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. (18) in which the taste for redistribution θ^{US} satisfies $\hat{\tau}(\theta^{\text{US}}) = \tau^{\text{US}}$.[52]

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.

[50] 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.

[51] 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.

[52] 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.

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:*

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

where g^{US} is the observed ratio of government purchases to output.[53]

Proof. See Appendix F.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 elasticity parameter σ, we can immediately infer θ^{US}.[54] Given our baseline parameter values and $g^{\mathrm{US}} = 0.188$, the implied empirically motivated taste for redistribution is $\theta^{\mathrm{US}} = -0.517$. 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

[53] 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.

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

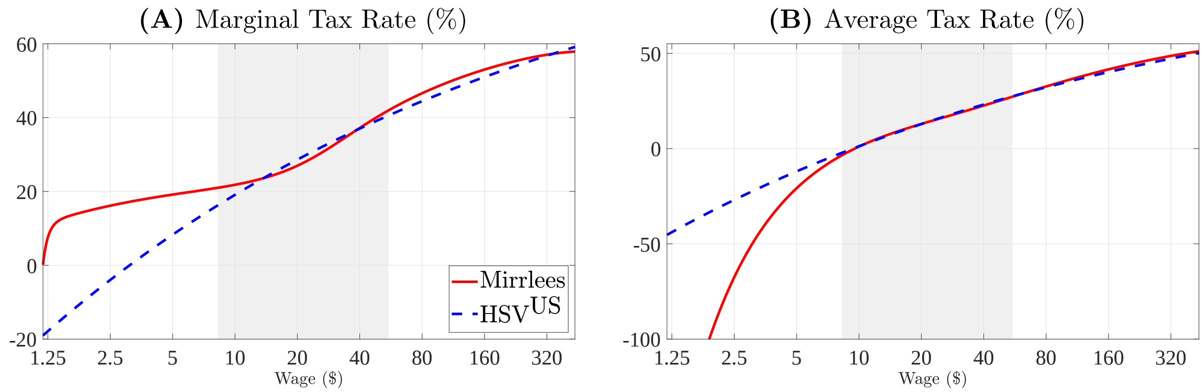

Figure 6: Optimal taxes and the HSV approximation with $\theta = \theta^{\text{US}}$. The figure contrasts tax rates under the current HSV tax system to those under the Mirrlees policy using our empirically motivated Pareto weight function.

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.[55]

Figure 6 plots the marginal and average tax schedules for the optimal Mirrleesian policy given $\theta = \theta^{\text{US}}$ against the schedules under our HSV approximation to the current tax system. The 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. The welfare gain of switching from the current HSV tax schedule to the optimal policy is tiny at 0.05 percent of consumption, indicating that the current tax system is close to efficient. The gain is so small because the current HSV tax schedule is very similar to that chosen by the Mirrleesian planner, especially in the shaded area where 90 percent of households are located.

Optimal Taxation under Alternative Social Preferences Panel A of figure 7 plots optimal Mirrleesian marginal tax rate profiles for $\theta \in \left\{-1, \theta^{\text{US}}, 0, 1, \infty\right\}$.[56] Panel B plots the

[55] 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.

[56] 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

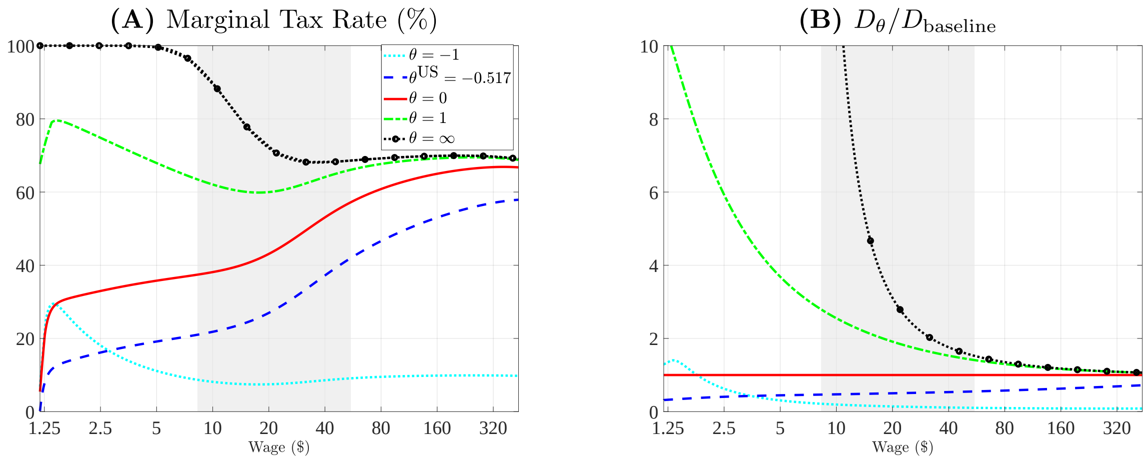

Figure 7: Alternative Social Preferences. Panel A plots optimal marginal tax schedules corresponding to Pareto weight functions with the following values for the taste for redistribution parameter: $\theta = -1$ (laissez-faire), $\theta = -0.517$ (empirically motivated), $\theta = 0$ (utilitarian), $\theta = 1$ (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 ($\theta = 0$).

corresponding distributional gain functions, in each case relative to the utilitarian baseline.

Considering Pareto weight functions with a stronger than utilitarian taste for redistribution, the optimal marginal tax schedule shifts upward. In addition, the shape of optimal schedule changes from upward sloping to U-shaped and eventually, under the Rawlsian objective, to downward sloping. 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.[57]

Note that in the laissez-faire case ($\theta = -1$), the optimal marginal tax schedule is quite different, with low and generally declining marginal rates. In this case, the planner does not

program.

[57] Welfare gains from tax reform here are very large, reaching 662 percent of consumption in the Rawlsian case (see Appendix G.5). They are so large because the least productive households rely almost entirely on transfers for consumption, and the Rawlsian planner therefore essentially maximizes transfers. Transfers in this case are $32,574.

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 (8.2 percent).

5.5 Summary of Findings

We take away several related messages from this analysis. First and foremost, for a range of plausible alternative model calibrations to the United States, the optimal tax and transfer system features marginal tax rates that are increasing in income.

Second, and related, the optimal policy is similar in spirit to a Universal Basic Income (UBI) system.[58] In particular, the optimal system features generous transfers, and these transfers are universal in that they are not quickly phased out or taxed away as household earned income rises. In contrast, in his exploration of optimal tax and transfer policy, Saez (2001) reports that "as in actual systems, the simulations suggest that the government should apply high rates at the bottom in order to target welfare only to low incomes" (p.223). Thus, one way to frame the distinction between our increasing optimal marginal rate schedule and Saez' U-shaped one is that transfers under our scheme have the flavor of UBI, while transfers in his have the flavor of means-tested benefits.

Third, the shape of the optimal schedule depends heavily on how much fiscal pressure the government faces. Reducing fiscal pressure on the government (e.g., by reducing G) both *increases* lump-sum transfers and *reduces* marginal tax rates on low incomes. Conversely, 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.[59]

[58] That in turn is a repackaging of the Negative Income Tax proposal in Friedman and Friedman (1962).

[59] 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.

Table 3: Mirrlees versus Ramsey Taxation

System	Parameters		Outcomes					
			$\overline{T'}$ (%)	Tr (\$)	$\frac{Tr}{Y}$ (%)	$\frac{Tr+G}{Y}$	ω (%)	ΔY (%)
HSV$^{\text{US}}$	$\lambda : 0.840$	$\tau : 0.181$	33.5	1,753	2.3	21.1	—	—
HSV	$\lambda : 0.817$	$\tau : 0.331$	46.6	4,632	6.4	26.5	1.65	−6.53
Affine	$\tau_0 : \$ - 20,747$	$\tau_1 : 49.2\%$	47.7	20,111	28.1	48.3	1.36	−7.31
Mirrlees			49.1	15,400	21.5	41.8	2.07	−7.32

Note: See the notes to table 2.

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 to the current tax system.

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 3 presents outcomes for each tax function. 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

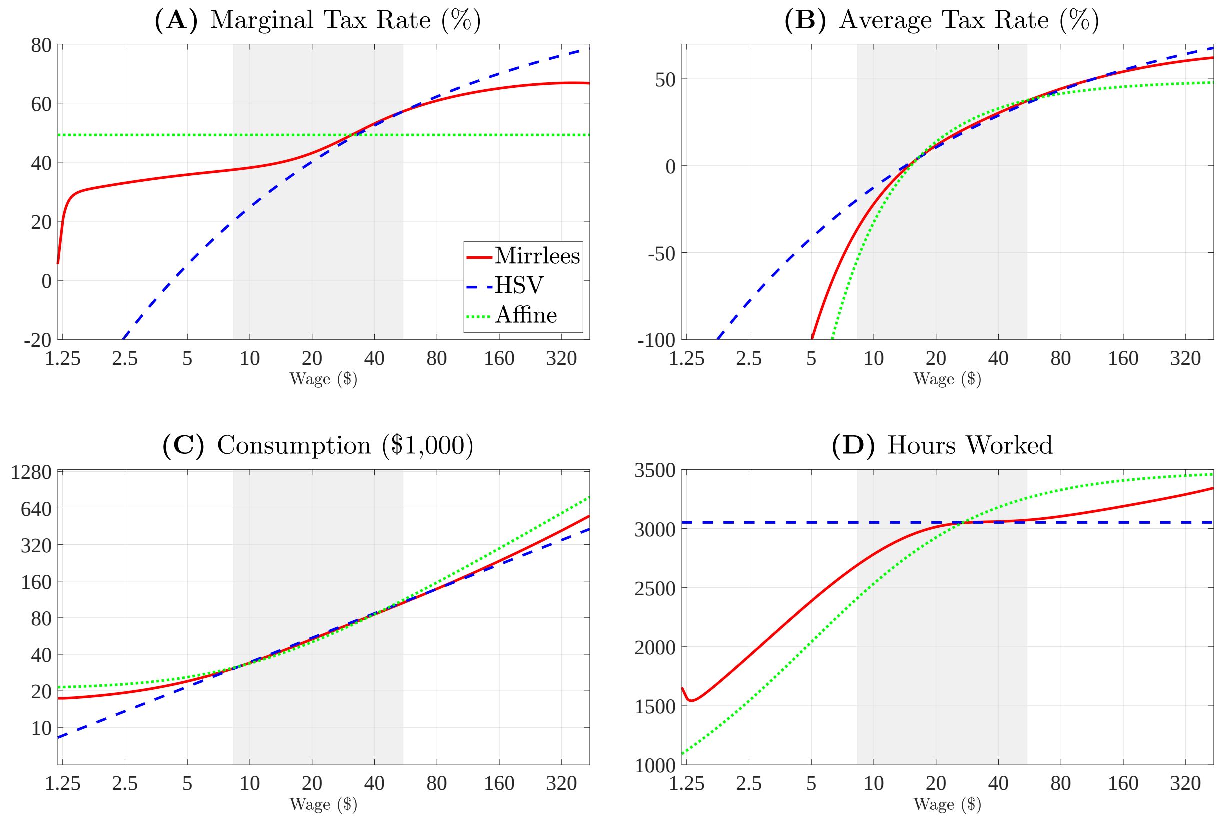

Figure 8: Mirrlees versus Ramsey Taxation. The figure contrasts allocations under the HSV tax system, the affine system, and the Mirrlees system. Panels A and B plot marginal and average tax schedules, and panels C and D plot decision rules for consumption and hours worked.

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. 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 8 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 households work too little. At the same time, because transfers are too small, very unproductive households work too much. However, the mass of households in these tails is small.

Panel A of figure 8 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 tax rates are too low at high income levels, high-productivity households consume too much.[60]

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 7 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 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 $U(\alpha,\alpha) \geq U^{\mathrm{US}}(\alpha)$ for all $\alpha \in \mathcal{A}$, where $U^{\mathrm{US}}(\alpha)$

[60]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.

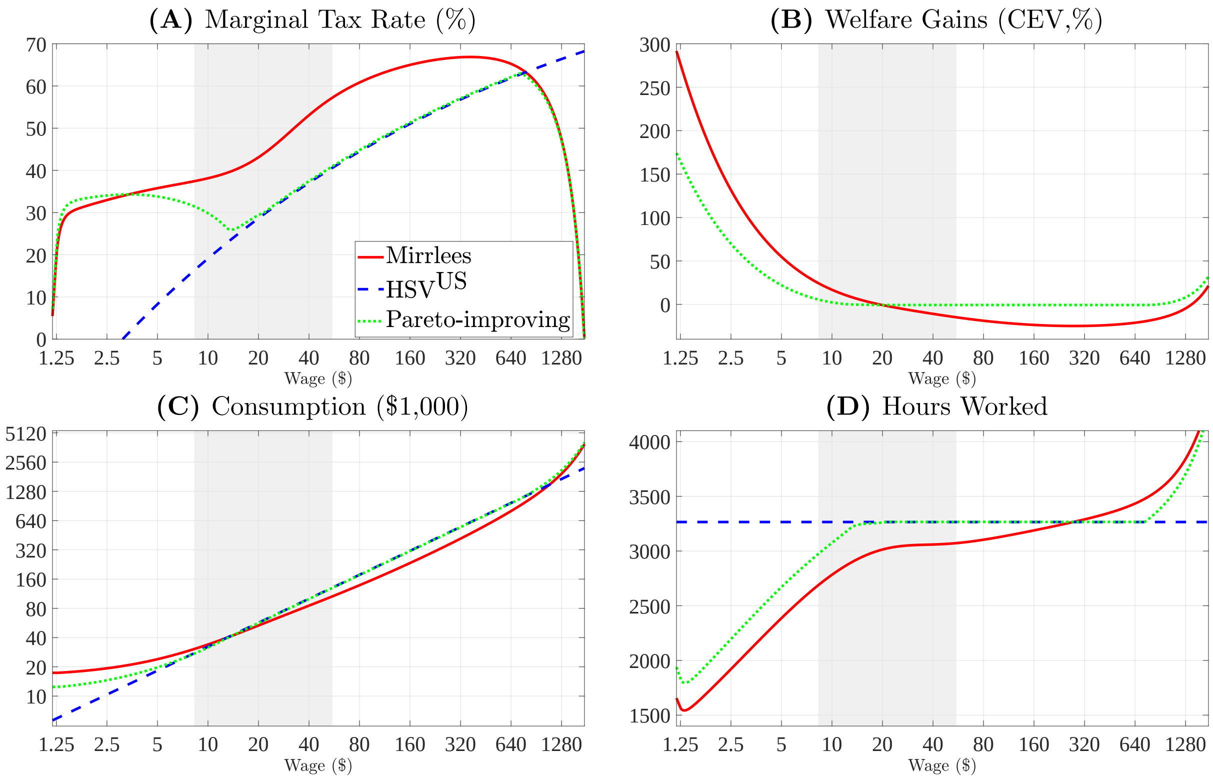

Figure 9: Pareto-improving Tax Reforms. The figure plots marginal tax schedule and welfare gains (CEV) (panels A and B) and decision rules for consumption and hours worked (panels C and D) under the Mirrlees system, the approximation to the current system, and the system that is optimal subject to being weakly Pareto improving.

denotes expected utility for a household of type α under our HSV approximation to the current U.S. tax and transfer system.[61]

Figure 9 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.

[61] 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 D.2.

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 F.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 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 5.4) 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 revisited the classic question of the optimal shape of the income tax schedule, in an economy calibrated to match the shape of the earnings distribution in the United States, and the extent of private insurance. We highlight five findings from our analysis.

First, a utilitarian planner would choose a system in which marginal tax rates increase in income and which delivers generous transfers. Low marginal rates at low income levels mean that these transfers are more akin to universal basic income than to means-tested benefits. The increasing optimal profile for marginal rates is robust to plausible alternative values for preference parameters.

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.

Choné, P. and G. Laroque (2018): "On the redistributive power of pensions," *Social Choice and Welfare*, 50, 519–546.

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.

Friedman, M. and R. D. Friedman (1962): *Capitalism and Freedom*, Chicago : University of Chicago Press.

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 (2021): "Practical Optimal Income Taxation," mimeo.

Hendren, N. (2020): "Measuring economic efficiency using inverse-optimum weights," *Journal of Public Economics*, 187, 104198.

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.

Ndiaye, A. (2020): "Flexible Retirement and Optimal Taxation," mimeo.

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.

Shourideh, A. and M. Troshkin (2017): "Incentives and Efficiency of Pension Systems," mimeo.

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 **1**
- A.1 Insurance via Family versus Insurance via Financial Markets 1
- A.2 Dynamic Model with Life-Cycle Shocks 3
- A.3 Individual- versus Family-Level Taxation 7

B Decomposition of Welfare Effects of A Tax Reform **8**
- B.1 Distributional Gain and Efficiency Cost 8
- B.2 Diamond-Saez Formula 11

C Theoretical Justification for Rescaling Variables **14**

D Computational Method **15**
- D.1 Mirrlees Optimal Taxation 15
- D.2 Pareto-Improving Tax Reforms 16

E Baseline Optimal Tax Policy **17**

F Proofs of Propositions **19**
- F.1 Proof of Proposition 1 19
- F.2 Proof of Proposition 2 21
- F.3 Proof of Proposition 3 24

G Sensitivity Analysis **25**
- G.1 Preferences 25
- G.2 Fiscal Pressure 27
- G.3 Extent of Private Insurance 27
- G.4 Shape of the Wage Distribution 28
- G.5 Alternative Social Preferences 32
- G.6 Type-Contingent Taxes 36

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

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

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 G.6).

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

$$\begin{aligned} & \int Q(\varepsilon_1)\lambda^{\frac{-1}{1-\tau}} c_1(\alpha,\varepsilon_1)^{\frac{1}{1-\tau}} d\varepsilon_1 + \iint Q(\varepsilon_1,\varepsilon_2)\lambda^{\frac{-1}{1-\tau}} c_2(\alpha,\varepsilon_1,\varepsilon_2)^{\frac{1}{1-\tau}} d\varepsilon_1 d\varepsilon_2 \\ \leq \; & \int Q(\varepsilon_1)\frac{2}{1+\rho}\exp(\alpha)\exp(\varepsilon_1)h_1(\alpha,\varepsilon_1)d\varepsilon + \iint Q(\varepsilon_1,\varepsilon_2)\frac{2\rho}{1+\rho}\exp(\alpha)\exp(\varepsilon_2)h_2(\alpha,\varepsilon_1,\varepsilon_2)d\varepsilon_1 d\varepsilon_2, \end{aligned}$$

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

$$\begin{aligned} \frac{1}{c_1(\alpha,\varepsilon_1)} &= \zeta\lambda^{\frac{-1}{1-\tau}}\frac{1}{1-\tau}c_1(\alpha,\varepsilon_1)^{\frac{1}{1-\tau}-1}, \\ \frac{1}{c_2(\alpha,\varepsilon_1,\varepsilon_2)} &= \zeta\lambda^{\frac{-1}{1-\tau}}\frac{1}{1-\tau}c_2(\alpha,\varepsilon_1,\varepsilon_2)^{\frac{1}{1-\tau}-1}. \end{aligned}$$

which immediately imply

$$\begin{aligned} c_1(\alpha,\varepsilon_1) &= c_2(\alpha,\varepsilon_1,\varepsilon_2) = c(\alpha), \\ \zeta &= (1-\tau)\lambda^{\frac{1}{1-\tau}}c(\alpha)^{\frac{-1}{1-\tau}}. \end{aligned}$$

[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

$$\begin{aligned}
h_1(\alpha,\varepsilon_1) &= \left[(1-\tau)\lambda^{\frac{1}{1-\tau}}c(\alpha)^{\frac{-1}{1-\tau}}\right]^{\frac{1}{\sigma}}\left(\frac{2}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\alpha\right)\exp\left(\frac{1}{\sigma}\varepsilon_1\right),\\
h_2(\alpha,\varepsilon_1,\varepsilon_2) &= \left[(1-\tau)\lambda^{\frac{1}{1-\tau}}c(\alpha)^{\frac{-1}{1-\tau}}\right]^{\frac{1}{\sigma}}\left(\frac{2\rho}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\alpha\right)\exp\left(\frac{1}{\sigma}\varepsilon_2\right).
\end{aligned}$$

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

$$\begin{aligned}
2\lambda^{\frac{-1}{1-\tau}}c(\alpha)^{\frac{1}{1-\tau}} &= \left[(1-\tau)\lambda^{\frac{1}{1-\tau}}c(\alpha)^{\frac{-1}{1-\tau}}\right]^{\frac{1}{\sigma}}\exp(\alpha)^{\frac{1+\sigma}{\sigma}}\left(\frac{2}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}\left\{\mathbb{E}\left[\exp(\varepsilon_1)^{\frac{1+\sigma}{\sigma}}\right]+\rho^{\frac{1+\sigma}{\sigma}}\mathbb{E}\left[\exp(\varepsilon_2)^{\frac{1+\sigma}{\sigma}}\right]\right\}\\
&\Rightarrow\ c(\alpha)=\lambda\,(1-\tau)^{\frac{1-\tau}{1+\sigma}}X^{\frac{\sigma(1-\tau)}{1+\sigma}}\exp(\alpha)^{1-\tau},\\
\text{where } X &= \mathbb{E}\left[\exp(\varepsilon)^{\frac{1+\sigma}{\sigma}}\right]\frac{1}{2}\sum_{t=1}^{2}\left(\frac{2\rho^{t-1}}{1+\rho}\right)^{\frac{1+\sigma}{\sigma}}.
\end{aligned}$$

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

$$\begin{aligned}
h_1(\alpha,\varepsilon_1) &= h_1(\varepsilon_1)\\
&= (1-\tau)^{\frac{1}{\sigma}}\lambda^{\frac{1}{\sigma(1-\tau)}}\left[\lambda\,(1-\tau)^{\frac{1-\tau}{1+\sigma}}X^{\frac{\sigma(1-\tau)}{1+\sigma}}\right]^{\frac{-1}{\sigma(1-\tau)}}\left(\frac{2}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\varepsilon_1\right)\\
&= (1-\tau)^{\frac{1}{1+\sigma}}X^{\frac{-1}{1+\sigma}}\left(\frac{2}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\varepsilon_1\right),\\
h_2(\alpha,\varepsilon_2) &= h_2(\varepsilon_2)\\
&= (1-\tau)^{\frac{1}{1+\sigma}}X^{\frac{-1}{1+\sigma}}\left(\frac{2\rho}{1+\rho}\right)^{\frac{1}{\sigma}}\exp\left(\frac{1}{\sigma}\varepsilon_2\right).
\end{aligned}$$

Earnings are given by

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

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

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

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

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

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

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,

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

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 due to two effects. First, it induces households at income level $\hat{y}$ to work less, due to a substitution effect, resulting in a loss of revenue, $S(\hat{y})<0$. Saez (2001) considers the virtual density $f_{y}^{*}(\hat{y})$ which is the density of incomes that would take place at $\hat{y}$ if the tax function $T(\cdot)$ were replaced by the linear tax schedule tangent to $T(\cdot)$ at level $\hat{y}$. Using this convenient notion, he then obtains the substitution effect as

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

where $e^{c}(\hat{y})>0$ is the compensated (Hicksian) labor supply elasticity. Using his equation (13), this can be written using the actual density as

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

Second, households earning more than $\hat{y}$ work more due to the income effect of paying one dollar more tax. Saez also obtains the income effect as

$$-I(\hat{y}) = -\int_{\hat{y}}^{\infty} \eta(y) \frac{T'(y)}{1-T'(y)} f_y^*(y) dy,$$

where $\eta(y) < 0$ is the elasticity of earnings with respect to a change in unearned income. Again using his equation (13), this can be written as

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

Note that 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.

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

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

Distributional Gains and Efficiency Costs as Functions of Productivity In Section 3.3, we defined 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 5.3, when the utility function is given by eq. (19) 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 5.2 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 Reproducing the Mirrlees planner's problem from eqs. (10-12), we have

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

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:

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

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:

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

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

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

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

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

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

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

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

$$\begin{aligned} \frac{T'\left(y(\alpha)\right)}{1 - T'\left(y(\alpha)\right)} &= (1+\sigma)\frac{1 - F_\alpha(\alpha)}{f_\alpha(\alpha)} \int_\alpha^\infty \left[1 - \frac{W(s)c(s)^{-\gamma}}{\zeta}\right] \frac{c(\alpha)^{-\gamma}}{c(s)^{-\gamma}} \frac{dF_\alpha(s)}{1 - F_\alpha(\alpha)}, \\ \text{where} \qquad \zeta &= \frac{1}{\int c(s)^\gamma dF_\alpha(s)}. \end{aligned}$$

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

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

where C denotes aggregate (and average) consumption.

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

$$\begin{aligned}
\frac{T'(y(\alpha))}{1-T'(y(\alpha))} &= A(\alpha) \times B(\alpha), \qquad\qquad\qquad\qquad \text{(A10)}\\
\text{where} \quad A(\alpha) &= (1+\sigma) \times \frac{1-F_\alpha(\alpha)}{f_\alpha(\alpha)},\\
B(\alpha) &= F_a(\alpha) \times \frac{\mathbb{E}\left[c(\tilde{\alpha})|_{\tilde{\alpha}\geq\alpha}\right] - \mathbb{E}\left[c(\tilde{\alpha})|_{\tilde{\alpha}<\alpha}\right]}{c(\alpha)}.
\end{aligned}$$

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 Section 5.2 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 4 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 Theoretical Justification for Rescaling Variables

As described in Section 4, we scale model earnings, consumption and taxes by a factor $\bar{Y}/Y$, and wages by $\bar{w} \equiv (\bar{Y}/Y)/(\bar{H}/H)$.

The theoretical justification for rescaling model variables in this fashion is that in the model one can scale wages by a factor ξ and introduce a constant φ in front of the disutility from hours worked term in the utility function, and essentially preserve the same equilibrium given appropriate rescaling of all equilibrium variables.

More precisely, consider a competitive equilibrium for our baseline economy in which $\xi = 1$ and $\varphi = 1$ given a differentiable tax schedule $T(y)$ with the associated marginal tax rates $T'(y)$. Let $w(\alpha, \varepsilon)$, $h(\alpha, \varepsilon)$, $y(\alpha, \varepsilon)$, and $c(\alpha, \varepsilon)$ denote wages, hours, earnings, and consumption in this equilibrium.

Now consider an alternative economy in which $\hat{\xi} \neq 1$ and $\hat{\varphi} \neq 1$. Let the tax schedule in this alternative economy be given by

$$\hat{T}(\hat{y}) = \hat{\xi}^{\frac{1+\sigma}{\sigma+\gamma}} \hat{\varphi}^{\frac{-1}{\gamma+\sigma}} T(\hat{\xi}^{-\frac{1+\sigma}{\sigma+\gamma}} \hat{\varphi}^{\frac{1}{\gamma+\sigma}} \hat{y}),$$

which implies

$$\hat{T}'(\hat{y}) = T' \left(\hat{\xi}^{-\frac{1+\sigma}{\sigma+\gamma}} \hat{\varphi}^{\frac{1}{\gamma+\sigma}} \hat{y} \right).$$

It is straightforward to verify that there is an equilibrium in the rescaled hatted economy in which

$$\begin{aligned}
\hat{w}(\alpha, \varepsilon) &= \hat{\xi} w(\alpha, \varepsilon), \\
\hat{h}(\alpha, \varepsilon) &= \hat{\xi}^{\frac{1-\gamma}{\sigma+\gamma}} \hat{\varphi}^{\frac{-1}{\gamma+\sigma}} h(\alpha, \varepsilon), \\
\hat{y}(\alpha, \varepsilon) &= \hat{\xi}^{\frac{1+\sigma}{\sigma+\gamma}} \hat{\varphi}^{\frac{-1}{\gamma+\sigma}} y(\alpha, \varepsilon), \\
\hat{c}(\alpha, \varepsilon) &= \hat{\xi}^{\frac{1+\sigma}{\sigma+\gamma}} \hat{\varphi}^{\frac{-1}{\gamma+\sigma}} c(\alpha, \varepsilon).
\end{aligned}$$

Note that $\hat{T}'(\hat{y}) = T'(y)$, so households of a given α type face the same marginal tax schedules in the two economies.

When rescaling model variables to match average earnings and hours in the data, we are effectively setting $\hat{\xi}$ and $\hat{\varphi}$ so that $\hat{Y} = \bar{Y}$ and $\hat{H} = \bar{H}$ which implies

$$\frac{\bar{Y}}{Y} = \hat{\xi}^{\frac{1+\sigma}{\sigma+\gamma}} \hat{\varphi}^{\frac{-1}{\gamma+\sigma}}$$

and

$$\frac{\bar{H}}{H} = \hat{\xi}^{\frac{1-\gamma}{\sigma+\gamma}} \hat{\varphi}^{\frac{-1}{\gamma+\sigma}}.$$

These two equations implicitly define the requisite adjustment factors $\hat{\xi}$ and $\hat{\varphi}$. Note that $\bar{w} = (\bar{Y}/Y)/(\bar{H}/H) = \hat{\xi}$.

D Computational Method

D.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 with the preference class we consider in Section 4, the local incentive compatibility constraints are necessary and sufficient for the global incentive compatibility constraints (12) to be satisfied (see Carroll 2012):

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

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: Deviation from Diamond-Saez formula

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

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.

D.2 Pareto-Improving Tax Reforms

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

$$\begin{aligned}
\max_{\{c(\alpha),y(\alpha)\}} \quad & \sum_i \pi_i W(\alpha_i)U(\alpha_i,\alpha_i), && \text{(A11)}\\
\text{subject to} \quad & \sum_i \pi_i c(\alpha_i) + G = \sum_i \pi_i y(\alpha_i), &&\\
& U(\alpha_i,\alpha_i) \geq U(\alpha_i,\alpha_j) \qquad \text{for all } i,j && \text{(A12)}\\
& U(\alpha_i,\alpha_i) \geq U^{\mathrm{US}}(\alpha_i) \qquad \text{for all } i. && \text{(A13)}
\end{aligned}$$

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

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

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

$$\begin{aligned}
\max_{\{c(\alpha),y(\alpha)\}} \quad & \sum_i \pi_i \left[W(\alpha_i)U(\alpha_i,\alpha_i) - \gamma \min\left\{U(\alpha_i,\alpha_i) - \bar{U}_i, 0\right\}^2\right], && \text{(A14)}\\
\text{subject to} \quad & \sum_i \pi_i c(\alpha_i) + G = \sum_i \pi_i y(\alpha_i), &&\\
& U(\alpha_i,\alpha_i) \geq U(\alpha_i,\alpha_j) \qquad \text{for all } i,j &&
\end{aligned}$$

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

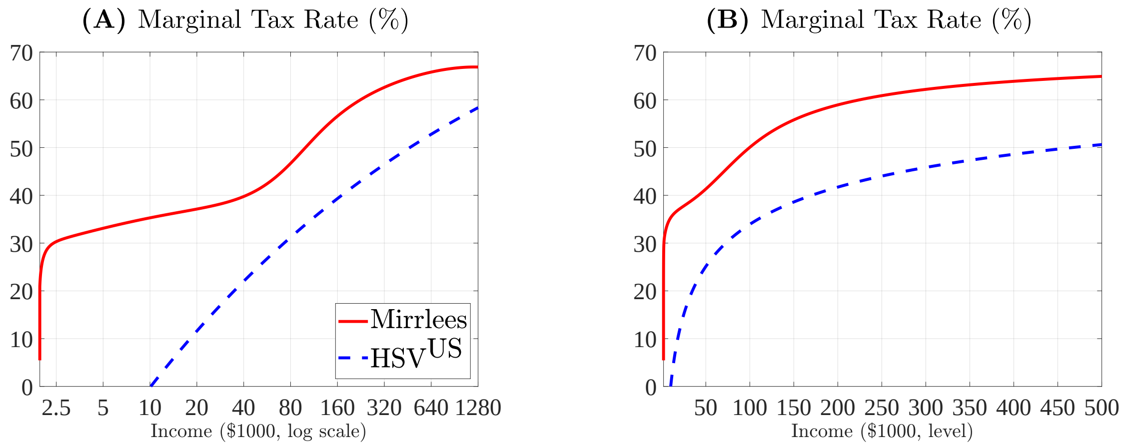

Figure A1: Optimal Tax Policy against Income. The figure plots the optimal Mirrleesian tax schedules against income (in log scale in panel A, in level in panel B) and the baseline HSV approximation to the U.S. tax and transfer system.

described in Appendix D.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 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).

E 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.)

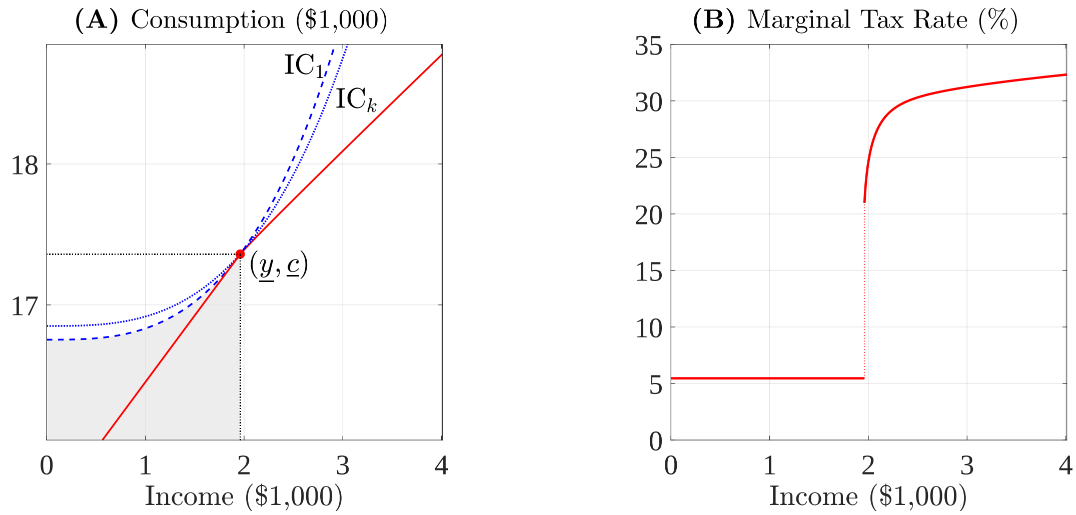

Figure A2: Allocations and Tax Rates for Low Income Households. Panel A plots household consumption against household income at the bottom of the income distribution (red solid line) and the indifference curves (blue dashed lines) for the least and most productive households in the bunched set, labelled 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.

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 $T(y) = y - c(y)$ 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 $\mathrm{IC}_1(y)$

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 $\mathrm{IC}_1(y)$ at $y = \underline{y}$. This translates into an upper bound on the marginal tax rate under any optimal tax scheme of 5.5 percent. 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 $\mathrm{IC}_k(y)$ at $y = \underline{y}$ in order to dissuade type k from delivering more income. This

[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.

translates into a lower bound on the marginal tax rate of 21.0 percent. 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.

F Proofs of Propositions

F.1 Proof of Proposition 1

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

Proof of Part (i) From eq. (20), $E(\alpha)$ is independent of Δ. ■

Proof of Part (ii) Given a utility function of the form (19), 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

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

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:

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

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 $\psi(x) > 0$. $\blacksquare$ $\mathcal{Q.E.D.}$

F.2 Proof of Proposition 2

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

$$c(\alpha;\lambda,\tau) = \lambda(1-\tau)^{\frac{1-\tau}{1+\sigma}}\exp\left[(1-\tau)\alpha\right]\exp\left(\frac{1-\tau}{\sigma}\frac{\sigma_\varepsilon^2}{2}\right), \tag{A16}$$
$$h(\varepsilon;\tau) = (1-\tau)^{\frac{1}{1+\sigma}}\exp\left(\frac{-1}{\sigma^2}\frac{\sigma_\varepsilon^2}{2}\right)\exp\left(\frac{\varepsilon}{\sigma}\right). \tag{A17}$$

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

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

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 (18), 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.

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

[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

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

$\hat{V}(\alpha) \geq |V(\alpha, \theta)|$ 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

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

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). $\mathcal{Q.E.D.}$

F.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

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

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

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

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 $y(\alpha; T_1) = y\left(\alpha; T_2\right)$. Therefore, if it is also the case that $U(\alpha; T_1) = U\left(\alpha; T_2\right)$, then $c(\alpha; T_1) = c\left(\alpha; T_2\right)$. 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 $T^{\mathrm{US}\prime}(y(\alpha; T^{\mathrm{US}})) = T^{\mathrm{PI}\prime}(y(\alpha; T^{\mathrm{PI}}))$. $\mathcal{Q.E.D.}$

G Sensitivity Analysis

We explore the sensitivity of the optimal Mirrleesian tax schedule to preferences, 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. We then consider 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.

G.1 Preferences

In Section 5.1, we discuss how optimal policies change when we consider different values for risk aversion γ and the labor elasticity parameter σ. Here we offer some more details on the sensitivity analysis with respect to these parameters. Figure A3 plots distributional gain/efficiency cost functions relative to the baseline, consumption and hours by productivity alongside optimal marginal tax schedules.

When $\gamma = 5$, the profile for hours by productivity under the optimal policy is rising for households with wages below around \$10 per hour, and decreasing thereafter. The upward-sloping portion reflects the fact that consumption (and the associated hours-reducing income effect) increases only slowly with wages at the bottom of the productivity distribution, given large lump-sum transfers. The declining pattern at the top reflects the existence of strong income effects coupled with a tax system that is effectively near proportional for high wage workers.

The welfare gains of switching from our HSV approximation to the current tax and transfer system to this much more redistributive one are huge, approaching 10 percent of consumption (see table A2), even though this reform depresses aggregate output by a similar amount.

Because changing the risk aversion parameter γ changes household decision rules, it also changes the models predictions for cross-sectional inequality, so that the model no longer replicates our empirical targets for earnings and consumption disperson. One way to change risk aversion without changing these predictions is to introduce preference heterogeneity in the disutility from work that is correlated with productivity. We have therefore explored introducing an idiosyncratic weight $\exp(\varpi\alpha)$ in front of the disutility of leisure term in utility. When the tax and transfer system is in the HSV case with progressivity parameter τ and when $\varpi = (1-\tau)(1-\gamma)$, the equilibrium decision rules for hours and consumption are identical (up to a constant of proportionality) for any value for γ. It follows that the

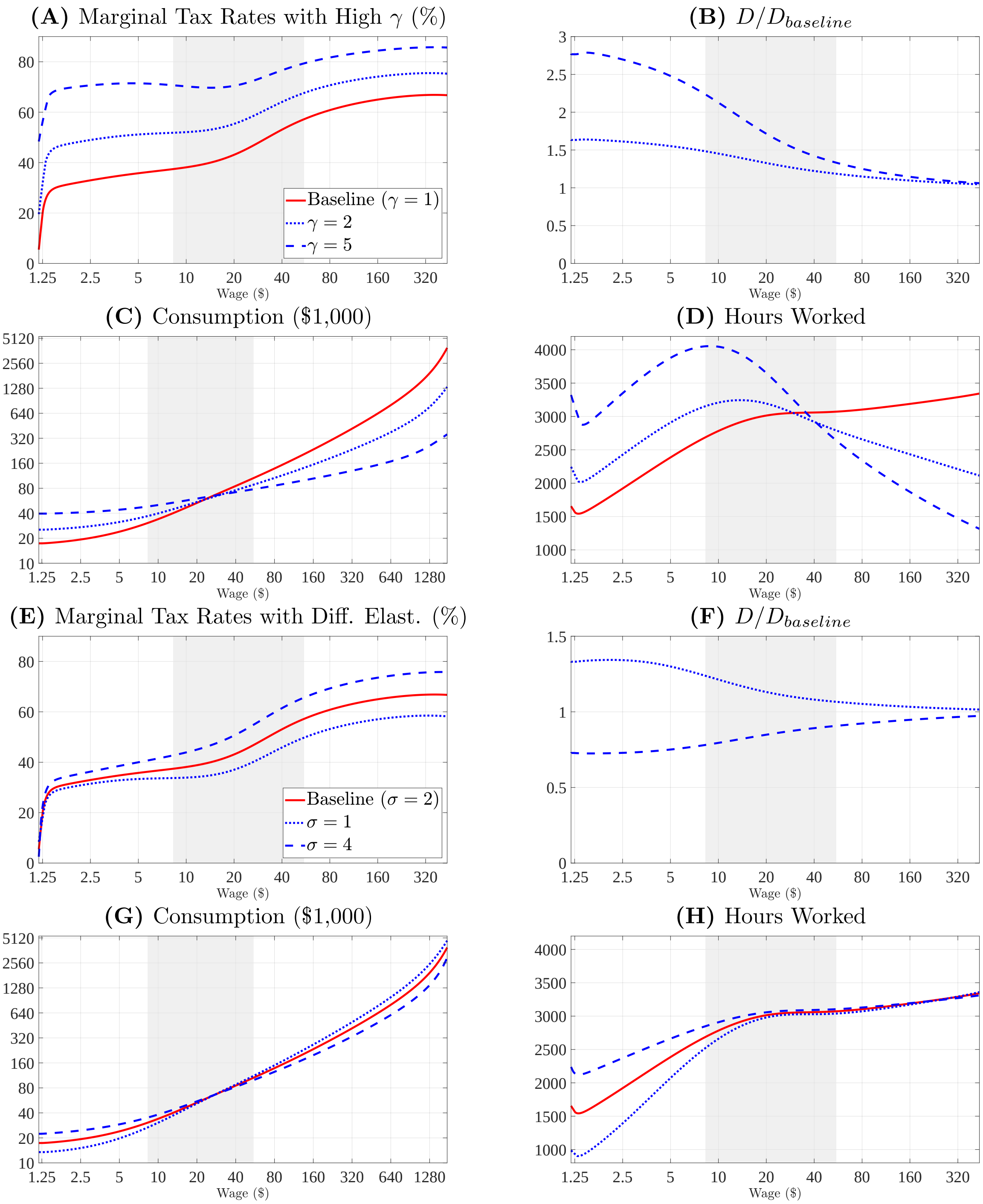

Figure A3: Optimal Policies with Different Preference Parameters. The figure plots the optimal Mirrleesian tax schedules, distributional gain/efficiency cost functions relative to the baseline, consumption and corresponding hours for higher risk aversion (panels A to D) and for different labor supply elasticities (panels E to H).

Table A2: Key Statistics for Additional Sensitivity

Model	Outcomes					
	$\overline{T'}$ (%)	Tr (\$)	$\frac{Tr}{Y}$ (%)	$\frac{Tr+G}{Y}$	ω (%)	ΔY (%)
HSV^{US}	33.5	1,753	2.3	21.1	—	—
Baseline	49.1	15,400	21.5	41.8	2.07	−7.32
High Risk Aversion: $\gamma = 5$	73.4	35,669	47.1	67.8	9.88	−9.59
Low Labor Elasticity: $\sigma = 4$	56.8	19,770	27.1	47.5	4.05	−7.31
High Govt. Expenditure: $g = 0.7$	59.4	−2,957	−3.5	59.9	8.35	−10.42
No Ins. Shocks: Exp. Left Tail	56.5	22,183	33.7	54.7	8.04	−10.59

Note: See the notes to table 2.

equilibrium values for all cross-sectional moments involving log wages, log hours and log consumption are then independent of the choice for γ. We have computed the optimal tax schedule in this model for different values for γ, and found that they are similar to the ones plotted above (see panel A of figure A4).

Finally, we have also computed a case in which we simultaneously increase risk aversion and eliminate insurable risk, thereby simultaneously increasing consumption inequality (by eliminating private insurance) and the welfare cost of consumption inequality (via greater curvature in utility). In this case, the optimal policy calls for very high but still increasing marginal tax rates across most of the income distribution (panel C).

G.2 Fiscal Pressure

In Section 5.2, we discuss how optimal policies change when fiscal pressure is higher. Figure A5 is supplemental to Figure 4. It plots optimal marginal tax schedules, distributional gain/efficiency cost functions relative to the baseline, consumption and hours by productivity for the economies with higher government expenditures or no insurable shocks.

G.3 Extent of Private Insurance

In Section 5.2, we showed that incorporating private insurance has an impact on the shape of the optimal tax and transfer schedule, and that our finding of an upward-sloping marginal tax profile is robust to any plausible variation in the extent of private insurance, as long as we retain the assumption that wages follow a Pareto lognormal distribution.

In this section, we provide more details about the optimal policies. Table A3 shows how allocations and tax schedules change when we rule out private insurance by setting $\sigma_\varepsilon^2 = 0$ and increase the variance of α_N, the normally distributed uninsurable component, so as to

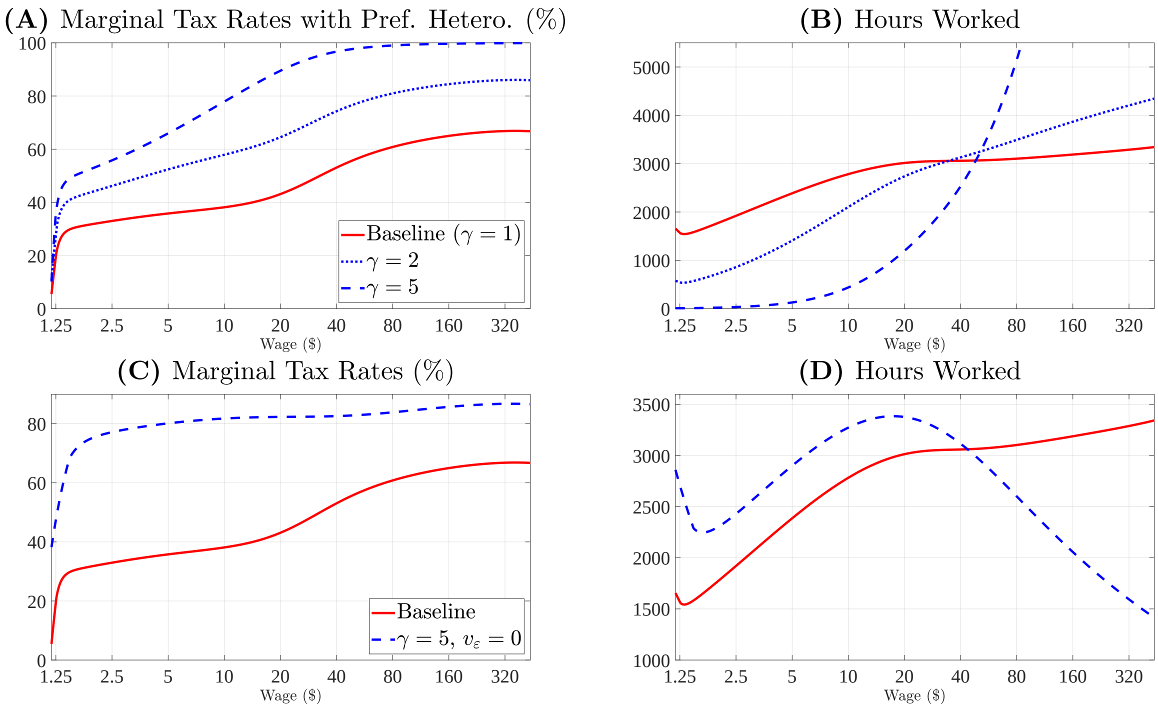

Figure A4: More Sensitivity with respect to Preference Parameters. Panels A and B plot the optimal Mirrleesian tax schedules and hours for the economy with preference heterogeneity. Panels C and D plot those for the economy with higher risk aversion and no insurable shocks.

leave the total variance of log wages unchanged. Since the dispersion of uninsurable shocks is now larger than in the baseline calibration, there would now be more poverty, absent public redistribution. Thus, Mirrlees 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 table also shows 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. 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.

G.4 Shape of the Wage Distribution

In Section 5.2, we showed that the shape of the wage distribution has an important quantitative impact on the shape of the optimal tax and transfer schedule. We now discuss this issue further.

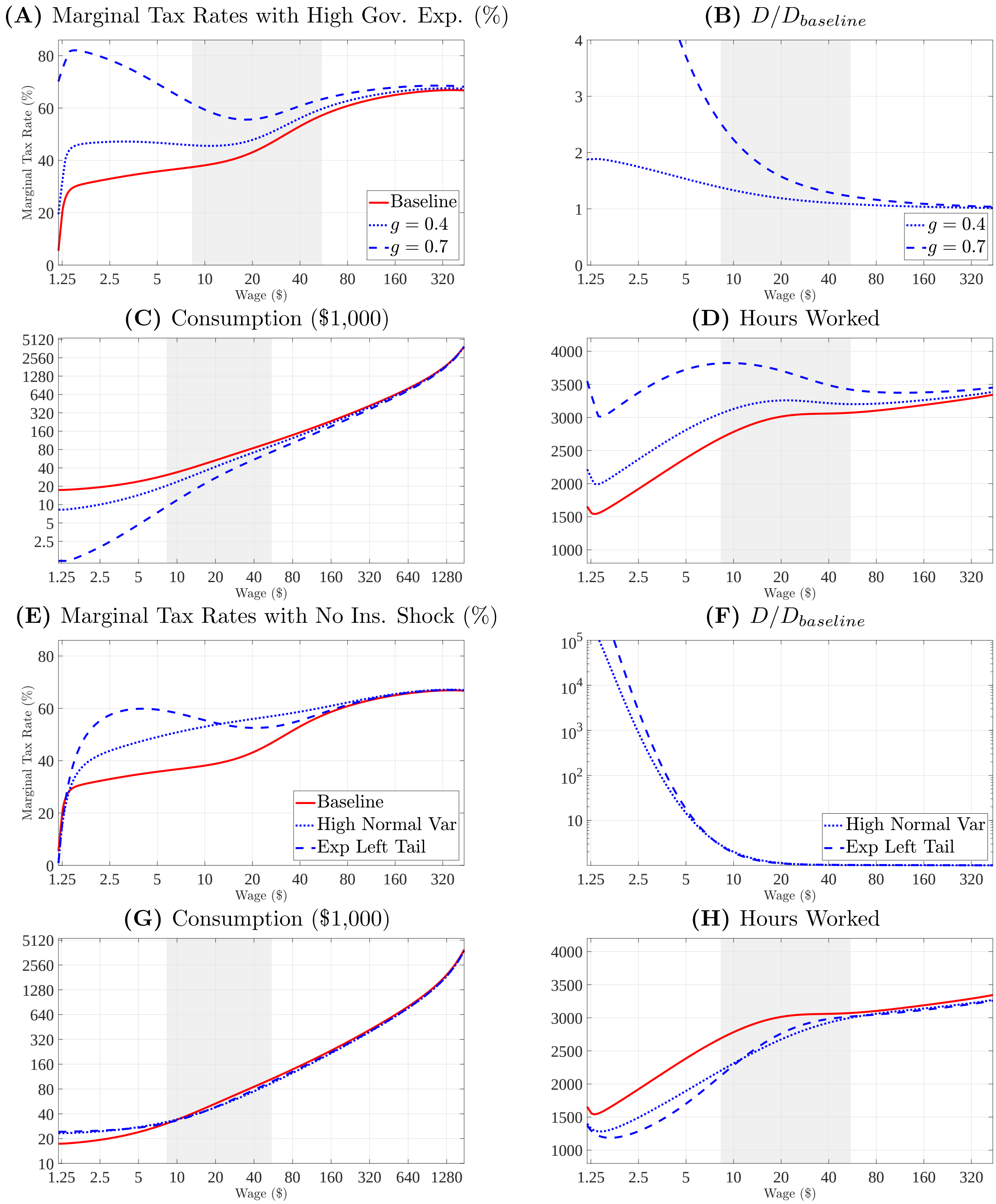

Figure A5: Optimal Policies with High Fiscal Presure. The figure plots the optimal Mirrleesian tax schedules, distributional gain/efficiency cost functions relative to the baseline, consumption and corresponding hours for higher government expenditures (panels A to D) and for no insurable shocks (panels E to H).

Table A3: Optimal Tax and Transfer System with No Insurable Shocks

System	**Parameters**		**Outcomes**					
			$\overline{T'}$ (%)	Tr ($)	$\frac{Tr}{Y}$ (%)	$\frac{Tr+G}{Y}$	ω (%)	ΔY (%)
HSV$^{\text{US}}$	$\lambda : 0.851$	$\tau : 0.181$	33.5	1,809	2.4	21.2	—	—
HSV*	$\lambda : 0.802$	$\tau : 0.422$	54.4	6,906	10.3	31.5	5.62	−10.97
Affine	$\tau_0 : \$ - 25,303$	$\tau_1 : 58.1\%$	58.1	23,951	36.0	57.2	8.35	−11.36
Mirrlees			58.6	21,586	32.5	53.8	8.63	−11.57

Note: See the notes to table 2.

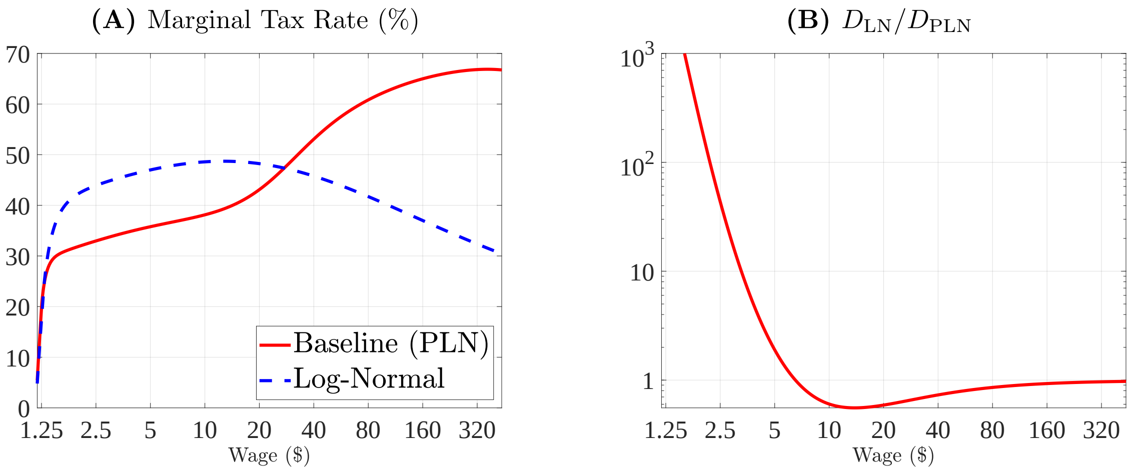

Figure A6: Lognormal versus Pareto Lognormal Wage Distribution. Panel A plots the profiles of optimal marginal tax rates. Panel B plots the distributional gain function of lognormal case, relative to that of the baseline Pareto lognormal case.

Lognormal Productivity 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 the taste for redistribution parameter to $\theta = 0$. Figure A6 plots the optimal tax schedule in this lognormal case (red dashed line) relative to the baseline Pareto lognormal economy.

We find that the optimal 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 lognormal productivity 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

Table A4: Optimal Tax and Transfer System with Lognormal Wage Distribution

System	Parameters		Outcomes					
			$\overline{T'}$ (%)	Tr (\$)	$\frac{Tr}{Y}$ (%)	$\frac{Tr+G}{Y}$	ω (%)	ΔY (%)
HSV$^{\text{US}}$	$\lambda : 0.828$	$\tau : 0.181$	33.5	1,674	2.2	21.0	—	—
HSV*	$\lambda : 0.813$	$\tau : 0.287$	42.7	3,561	4.8	24.5	0.64	−4.50
Affine	$\tau_0 : \$ - 18,508$	$\tau_1 : 45.1\%$	45.1	17,869	24.4	44.3	1.96	−5.32
Mirrlees			44.4	18,735	25.5	45.3	2.06	−5.05

Note: See the notes to table 2.

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 A4 compares the optimal Mirrlees policy to the best-in-class HSV and affine tax schemes given a lognormal productivity distribution. A key result is that the best affine tax and transfer system dominates the best system in the HSV class and very closely approximates the second-best allocation. Thus, assuming a lognormal 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.

Economy with No Income Effects and Pareto Distribution In Section 5.3, we considered the case of preferences that have no income effects. Given the utility function (19), the efficiency cost of taxation then simplifies to

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

Note that, besides the impact of the marginal tax rate itself, efficiency costs vary with productivity only because of exogenous variation in the inverse Mills ratio.

We now further assume a Pareto distribution for productivity, assuming that α is exponentially distributed. In this case, the inverse Mills ratio becomes constant and equal to λ_α. Efficiency costs thus vary with productivity only because marginal tax rates do. So any slope to the optimal marginal tax schedule must be entirely driven by distributional concerns.

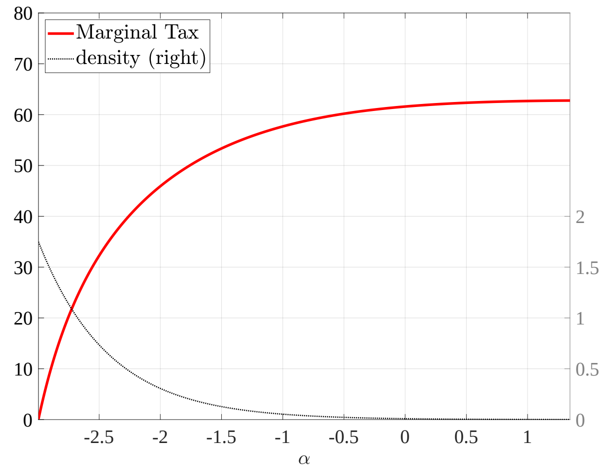

Figure A7: Model with No Income Effects and Pareto Wage Distribution. The figure plots the profile of optimal marginal tax rates for the economy with no income effects and a Pareto wage distribution.

Table A5: Alternative Social Preferences

Social Preferences		**Mirrlees Allocations**			**Welfare Gain** ω **(%)**			
	θ	$\overline{T'}$ (%)	Tr ($)	ΔY (%)	Mirrlees	HSV*	Affine	$\frac{\omega(\text{HSV}^{\text{US}},\text{HSV}^*)}{\omega(\text{HSV}^{\text{US}},\text{Mirrlees})}$
Laissez-Faire	-1	8.2	$-6,925$	10.57	3.71	3.54	3.70	95%
Emp. Motivated	-0.517	33.4	$5,235$	0.03	0.05	—	-0.53	0%
Utilitarian	0	49.1	$15,400$	-7.32	2.07	1.65	1.36	80%
Rawlsian	∞	71.1	$32,574$	-21.98	661.6	329.3	606.2	50%

Note: See the notes to table 2.

Figure A7 shows that the optimal marginal tax schedule in this case is upward-sloping.

G.5 Alternative Social Preferences

Table A5 shows summary statistics for the optimal Mirrlees policies under the different Pareto weight functions discussed in Section 5.4.[7] 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

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

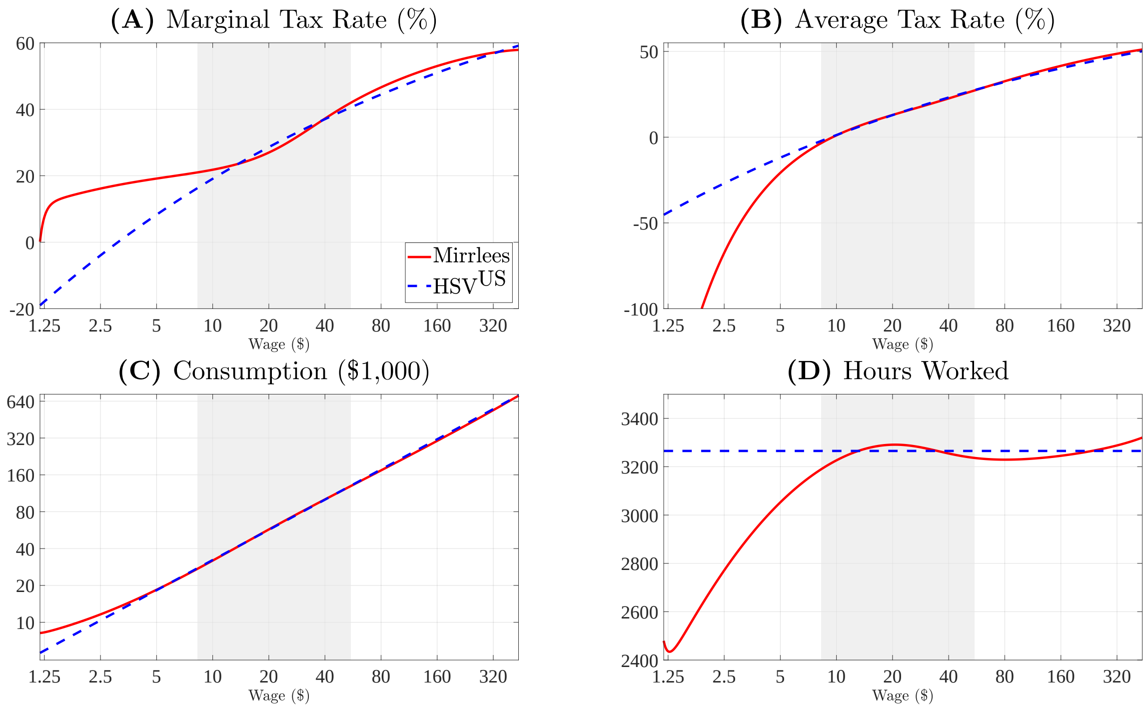

Figure A8: HSV versus Mirrlees tax functions with $\theta = \theta^{\mathrm{US}}$. The figure contrasts tax rates and allocations under the current HSV tax system to those under the Mirrlees policy using our empirically motivated Pareto weight function.

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 A8). Allocations are more different at the extremes of the distribution, but the population density in those ranges is very small.[8]

Table A5 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 A9 offers another perspective on the properties of optimal allocations at the

[8] 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 the productivity distribution.

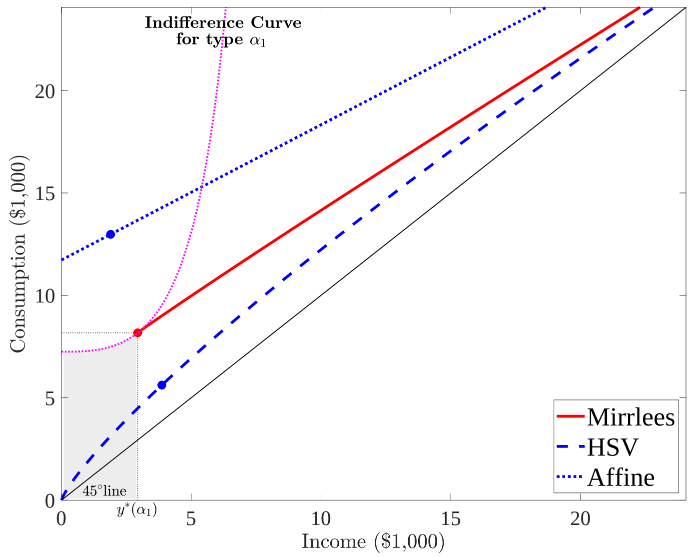

Figure A9: Allocations for low income households. The figure plots household consumption against household income at the bottom of the income distribution under the Mirrlees (red solid), HSV (blue dashed), and best-in-class affine (blue dotted) tax systems. Each tax scheme is best-in-class given the empirically motivated Pareto weight function.

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 in table A5, 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 A9 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

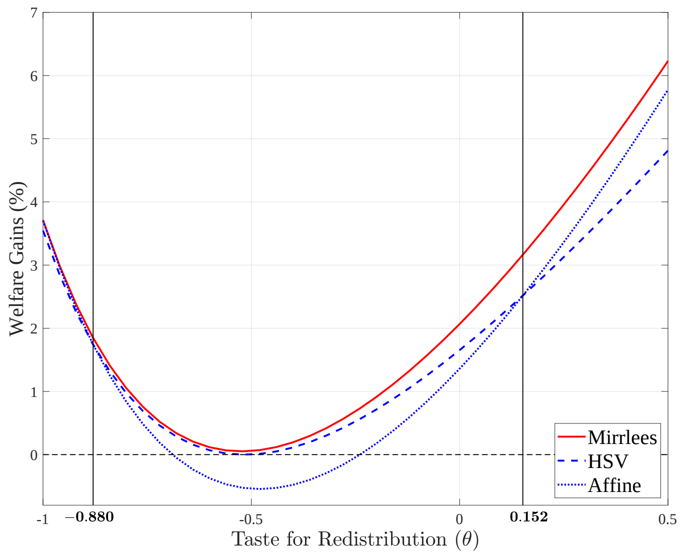

Figure A10: Maximum welfare gains from tax reform. The figure plots the maximum possible gains from tax reform for a range of values for the taste for redistribution parameter θ. Three lines are plotted, corresponding to the best policies in the unrestricted Mirrlees class (red solid), the HSV class (blue dashed), and the affine class (blue dotted).

in part because the fact that a portion of wage dispersion is privately insurable reduces the need for public insurance.

Figure A10 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.

The first message from figure A10 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.

G.6 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).[9]

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

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

Table A6: Type-Contingent Taxes

Tax System			**Outcomes**				
			$\overline{T'}$ (%)	Tr (\$)	$\frac{Tr}{Y}$ (%)	ω (%)	ΔY (%)
HSV$^{\text{US}}$	$\lambda: 0.834$	$\tau: 0.181$	9.6	1,443	1.9	—	—
			23.9	1,901	2.5		
HSV*	$\lambda^L: 1.067$	$\tau^L: 0.481$	9.0	11,181	14.8	5.85	−2.09
	$\lambda^H: 0.596$	$\tau^H: 0.075$	32.7	−1,424	−1.9		
Affine	$\tau_0^L: -0.402$	$\tau_1^L: 0.346$	10.2	31,986	42.1	5.79	−1.84
	$\tau_0^H: -0.034$	$\tau_1^H: 0.453$	32.0	858	1.1		
Mirrlees			10.0	28,068	37.0	6.18	−1.84
			31.9	681	0.9		

Note: See the notes to table 2.

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.

References

Carroll, G. (2012): "When Are Local Incentive Constraints Sufficient?" *Econometrica*, 80, 661–686.

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.