Optimal Income Taxation: Mirrlees Meets Ramsey

Jonathan Heathcote

Federal Reserve Bank of Minneapolis and Center for Economic and Policy Research

Hitoshi Tsujiyama

Goethe University Frankfurt

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 United States, 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.

I. 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, Weinzierl, and Yagan (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 as well as 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% of gross domestic product (GDP) under the optimal policy featuring increasing marginal rates. The corresponding figure for the United States in 2015 was 33.5%. Saez's (2001) calibrations are associated with much higher spending levels of around 56% 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 US 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, Storesletten, and Violante (2017), where taxes net of transfers are given by the following function of income: $T(y) = y - \lambda y^{1-\tau}$. In this scheme (henceforth, 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}, which 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, Troshkin, and Tsyvinski (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, Storesletten, and Violante (2017), who explore the HSV form developed by Feldstein (1969), Persson (1983), and Benabou (2000). Relative to those papers, the advantage of our nonparametric 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 nonparametric profile for social welfare weights that precisely rationalizes a particular observed tax system (see Bourguignon and Spadaro 2012; Brendon 2013). The approach in this paper restricts the Pareto weight function to a one-parameter functional form that allows for only 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 nonparametric 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 US 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 nonmarginal 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: (1) we consider optimal Mirrleesian tax policy in addition to affine tax systems, and (2) our analysis is more quantitative in nature.

II.   Environment

A.   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.

---

[1] In app. sec. 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.

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]

B. 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).

C. 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

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

D. 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.

---

[2] We discuss this interpretation further in app. sec. 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, Storesletten, and Violante (2014) estimate that life cycle uninsurable shocks account for only 17% 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; Ndiaye 2020).

E. 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

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

F. 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 are

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

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

The first condition 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 1 minus the marginal tax rate on family income. If the tax function satisfies

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

---

[3] In app. sec. 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.

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.

G. Equilibrium

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

1. the decision rules $\{c, h\}$ solve the family's maximization problem (4);
2. the resource feasibility constraint (1) is satisfied; and
3. the government budget constraint is satisfied: $\int T(y(\alpha)) dF_\alpha(\alpha) = G$.

III. Planner's Problems

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

A. 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.

1. 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

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

$$\text{subject to} \int c(\alpha, \tilde{\alpha}, \varepsilon) \, dF_{\varepsilon}(\varepsilon) = c(\tilde{\alpha}), \tag{8}$$

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

Solving this problem gives the following indirect utility function:

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

$$\text{where } \Omega = \left(\int \exp (\varepsilon)^{(1+\sigma)/\sigma} dF_{\varepsilon}(\varepsilon) \right)^{-\sigma}. \tag{9}$$

2. First-Stage Planner's Problem

The planner maximizes social welfare, evaluated according to $W(\alpha)$, subject to the resource constraint and to incentive constraints:

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

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

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

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.

---

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

3.   Decentralization with Income Taxes

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

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

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

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

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

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

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

---

[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

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

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

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

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

C. 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 nonoptimal.[9]

Consider the effect of increasing the marginal tax rate at some income level $\hat{y}$ so as to collect \$1 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 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% 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.

IV.   Calibration

A.   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, Storesletten, and Violante (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, Storesletten, and Violante (2014). The compensated (Hicks) elasticity of hours with respect to the marginal net-of-tax wage is approximately equal to $1/(1 + \sigma)$ (see Keane 2011, eq. [11]), which, given $\sigma = 2$, is equal to $1/3$. Again, this value is consistent with empirical estimates: Keane reports an average estimate across 22 studies of 0.31. Given our model for taxation, the elasticity of average income with respect to 1 minus the average income-weighted marginal tax rate is also

----

[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 app. sec. B.1 for the derivation.

equal to $1/(1 + \sigma)$.[11] According to Saez, Slemrod, and Giertz (2012), the best available estimates for the long-run version of this elasticity range from 0.12 to 0.40.

B. 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, Storesletten, and Violante (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 2000–2006, Heathcote, Storesletten, and Violante (2017) estimate $\tau = 0.181$.

The remaining fiscal policy parameter λ is set such that government purchases G is equal to 18.8% 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.

C. Wage Distribution and Insurance

Our strategy for calibrating the model distribution of wages and the relative importance of uninsurable versus insurable shocks is as follows. First, we assume that log wages are drawn from an exponentially modified Gaussian (EMG) distribution. Second, we parameterize the overall wage variance and the fraction of this variance that is privately uninsurable to replicate the observed cross-sectional variances of log earnings and log consumption, exploiting the standard result that uninsurable wage risk will show up in consumption, whereas insurable shocks will not.[13]

---

[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, Storesletten, and Violante 2017, eq. [4]).

[12] National Income and Product Accounts (table 1.1.10). Heathcote, Storesletten, and Violante (2017) use the same value.

[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, app. sec. 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.

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 $\alpha \sim \text{EMG}(\mu_\alpha, \sigma_\alpha^2, \lambda_\alpha)$. It follows that the log wage, $\log w = \alpha + \varepsilon$, is also 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 distributions with[14]

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

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

Our calibration strategy is to first use an empirical distribution for log earnings to estimate the normal variance $\sigma_y^2 = [(1 + \sigma)/\sigma]^2 \sigma_\varepsilon^2 + \sigma_\alpha^2$ 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 equation (17). Finally, given σ_α^2 and λ_α, the variance of log earnings (16) residually exactly identifies σ_ε^2.

As Mankiw, Weinzierl, and Yagan (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

---

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

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

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

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

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 $1 + 1/\sigma$, respectively. Consumption does not respond to insurable shocks, and the elasticity of consumption to uninsurable shocks is $1 - \tau$.

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, Perri, and Violante (2010, fig. 13) report a variance of 0.332 in 2006.[17] However, Heathcote, Storesletten, and Violante (2014, table 3) estimate that 29.6% 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 equation (16) to residually infer σ_ε^2 gives $\sigma_\varepsilon^2 = 0.120$. In section V.B, we will explore 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, Perri, and Violante (2010, fig. 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.

---

[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% 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, figs. 6, 7) report 90th/50th and 50th/10th percentile ratios in the mid-2000s that are both close to 2. The same ratios are also close to 2 in Heathcote, Perri, and Violante (2010, fig. 13). Attanasio and Pistaferri (2014, fig. 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.

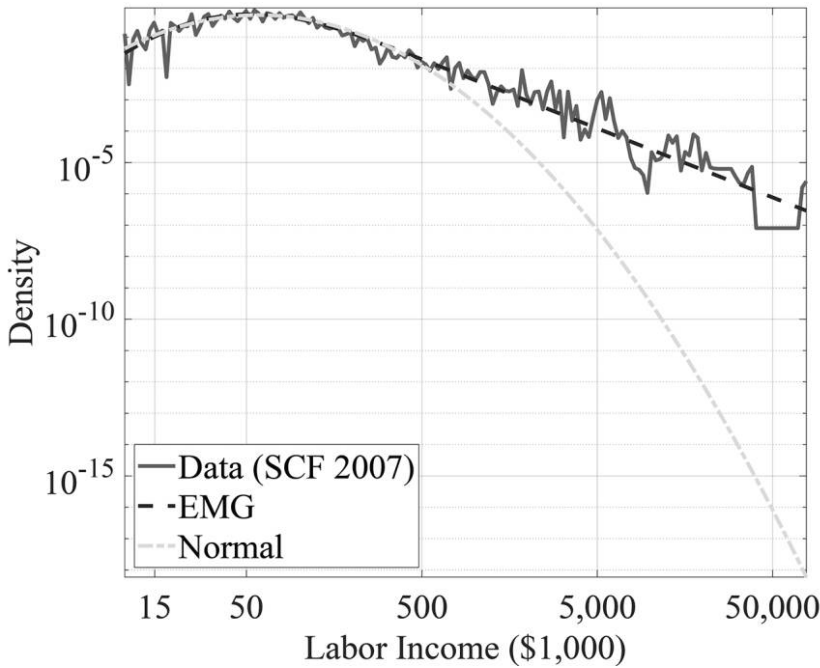[19] We thank Low and Pistaferri for sharing their estimates.

Fɪɢ. 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. A color version of this figure is available online.

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

TABLE 1
Mᴏᴅᴇʟ Pʀᴏᴅᴜᴄᴛɪᴠɪᴛʏ Dɪsᴛʀɪʙᴜᴛɪᴏɴ ᴀɴᴅ Oꜰꜰᴇʀᴇᴅ Wᴀɢᴇ Dɪsᴛʀɪʙᴜᴛɪᴏɴ
ɪɴ Lᴏᴡ ᴀɴᴅ Pɪsᴛᴀꜰᴇʀʀɪ (2015)

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

Nᴏᴛᴇ.—Px denotes the xth percentile.

is similar at 2.69, providing empirical support for our assumption that the exponential component of log wages is uninsurable.

D. 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 the 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 years) 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 (apps. A–G are available online), we provide the theoretical justification for rescaling model variables in this fashion.

E. Discretization

In solving the Mirrlees problem to characterize efficient allocations, the incentive constraints apply only to the uninsurable component of the wage α, and the distribution for ε appears only in the constant Ω. Thus, there is no need to approximate the distribution for ε, and we therefore assume that 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 $\{\alpha_1, ..., \alpha_I\}$ with corresponding probabilities $\{\pi_1, ..., \pi_I\}$ as follows. We make the end points 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 $\exp(\alpha_1)/\Sigma_i(\pi_i \exp(\alpha_i)) = 0.05$ and set α_I such that $\exp(\alpha_I)/\Sigma_i(\pi_i \exp(\alpha_i)) = 74$, which corresponds to household labor income at the 99.99th percentile of the SCF labor income distribution (\$6.17 million).[21] We read corresponding

---

[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 underrepresentation of high-income households in the CPS.

[21] Given $Y/H = \exp(\sigma_\varepsilon^2/\sigma) = 1.06$, the average hourly wage is $\bar{w} = (\bar{Y}/Y)/(\bar{H}/H) = \$23.68$, so 5% 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.

probabilities π_i directly from the continuous EMG distribution, rescaling to ensure that (1) $\Sigma_i \pi_i = 1$, (2) $\Sigma_i \pi_i \exp(\alpha_i) = 1$, and (3) 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.
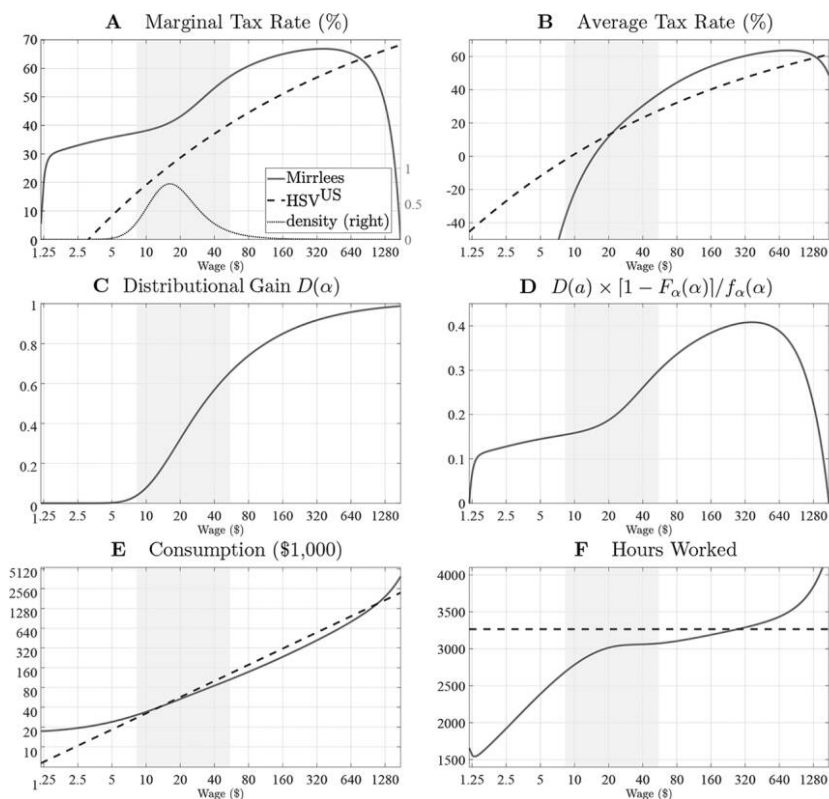

FIG. 2.—Optimal tax policy. Panels A and B plot the optimal Mirrleesian tax schedules against the baseline HSV approximation to the US tax and transfer system and the productivity density. Panels C and D plot the distributional gain (equivalently, the efficiency cost) under the optimal policy. Panels 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. A color version of this figure is available online.

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

A.   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, e.g., 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 (see also Diamond and Saez 2011; Golosov, Troshkin, and Tsyvinski 2016).

  Figure 2 plots the marginal and average tax schedules (panels A, B) that decentralize the constrained efficient allocation against the baseline HSV approximation to the current US tax and transfer system (HSV[US]).[23] In contrast to Diamond and Saez, optimal marginal tax rates are always increasing in income (except at the very top).[24] The marginal rate starts at 5.5% for the least productive households, is fairly flat (between 30% and 40%) for households making up to around $15 per hour, and rises rapidly to peak at 66.9% for those making around $350 per hour.[25] Panel A indicates that the optimal schedule imposes much higher marginal tax rates

---

[22] In app. sec. D.1, we explain how we numerically solve the Mirrlees optimal tax problem.
[23] The profiles for marginal and average tax rates look very similar plotted against log household income rather than against log household productivity. Figure A1 (figs. A1–A10 are available online) plots the marginal tax rate against the level of income.
[24] 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.
[25] 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.

than our approximation to the current US system: the average income-weighted marginal tax rate is 49.1% compared with 33.5% 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 sec. III.C for the definitions of these measures).[26] 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 sec. V.C). This resemblance indicates that if we can understand what drives distributional gains, we can better understand optimal taxation.[27]

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% of every hypothetical marginal dollar in tax revenue leaks away via behavioral responses. As we increase α further, the efficiency cost rises further toward 1. 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.[28] 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% 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)

---

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

[27] In sec. V.C, 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.

[28] 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 $\bar{T}' = [1 + \bar{\zeta}^u + \bar{\zeta}^c(\lambda_y^* - 1)]^{-1} = (\sigma + \gamma)/(\sigma + \lambda_y^*)$, 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, $\bar{\zeta}^u = (1 - \gamma)/(\sigma + \gamma)$, $\bar{\zeta}^c = (\sigma + \gamma)^{-1}$, and $\lambda_y^* = \lambda_\alpha[(\sigma + \gamma)/(1 + \sigma)]$. In the baseline case where $\gamma = 1$, we obtain $\bar{T}' = (1 + \sigma)/(\sigma + \lambda_\alpha)$. 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% at the top.

TABLE 2
KEY STATISTICS

	OUTCOMES					
MODEL	$\overline{T'}$	Tr	Tr/Y	$(Tr + G)/Y$	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: $\gamma = 2$	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	.87	−5.85
High government expenditure: $g = .4$	52.9	5,633	7.3	47.8	.19	−1.07
No insurable shocks: high normal variance	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. Tr/Y is transfers as a percentage of average income. $(Tr + G)/Y$ 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.

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% of each marginal tax dollar would leak away. Because 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.[29]

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

[29] Plotted against income, marginal tax rates jump upward at the income bunching value. We discuss this bunching in more detail in app. E.

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.[30]

We start by considering $\gamma = 2$ and $\gamma = 5$, and we compare the optimal tax schedule in these cases with the one under our baseline logarithmic specification ($\gamma = 1$). Panel A of figure 3 plots optimal marginal tax rates.[31] 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 γ is, the stronger income effects in labor supply choices are. 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% when $\gamma = 1$ to 59.8% 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 (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.

B. 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

---

[30] 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 app. G, we discuss these exercises in more detail and conduct extensive additional sensitivity analyses.

[31] To avoid the visual distraction of the zero-top-tax-rate property, we have truncated the visible range for wages at the 99.95th percentile of the baseline model distribution for α in this figure and in subsequent similar ones.
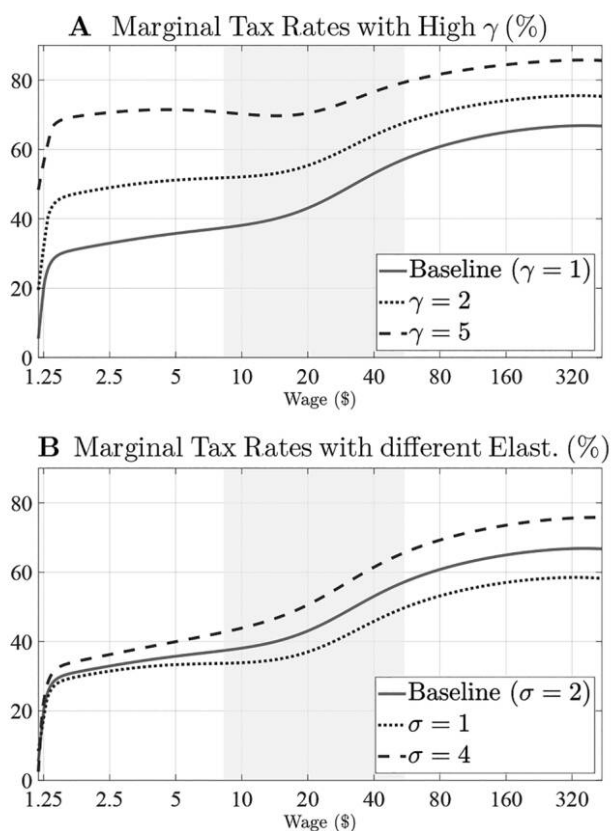
Fig. 3.—Sensitivity. The figure plots the optimal Mirrleesian tax schedules with higher risk aversion (panel A) and with a higher/lower labor supply elasticity (panel B). A color version of this figure is available online.

there are large distributional gains from imposing high marginal tax rates at low income levels. Such gains can arise when (1) the government must deliver high government consumption, which crowds out lump-sum transfers, or (2) when there are many households with very low productivity. 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 V.D, we show that a U-shaped optimal profile also emerges when the planner puts very high welfare weight on low-productivity households.

1.   Increasing Government Purchases

Panel A of figure 4 plots the optimal marginal tax schedules when we increase G from the baseline value (18.8% of output under the HSV[US]
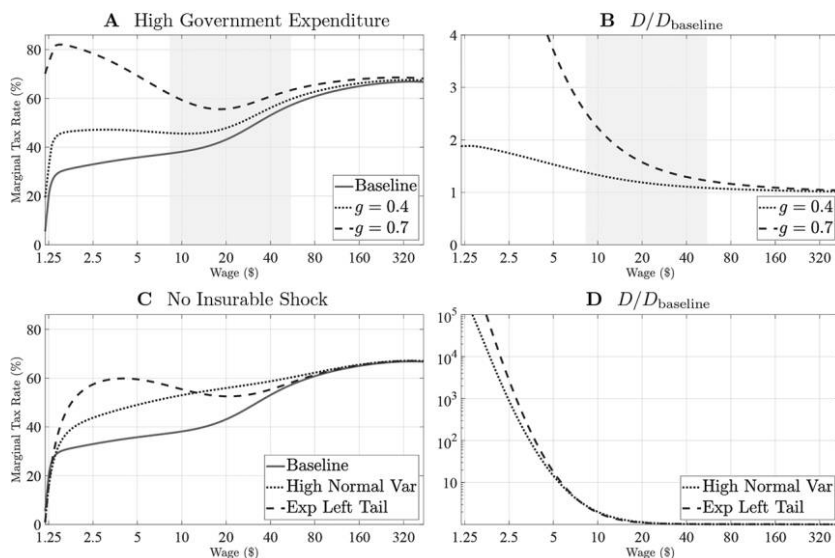
F_IG. 4.—Increasing fiscal pressure. Panels A and B plot the optimal marginal tax sched-ules and the corresponding distributional gains for different values of government pur-chases. For example, the line labeled $g = 0.4$ corresponds to a value for G equal to 40% of output under the HSV[US] policy. Likewise, panels C and D are the economies with no in-surable shocks and higher uninsurable risk. The dotted lines indicate when the uninsur-able risk has a high normal variance, and the dashed lines indicate when the uninsurable risk has a thick left exponential tail. A color version of this figure is available online.

policy) to higher levels (40% and 70%). Panel B plots the corresponding distributional gain/efficiency cost functions, in each case relative to the baseline.

Raising required expenditure leads the government to raise marginal tax rates across the productivity distribution and by much more at low pro-ductivity levels. The result is that the schedule eventually becomes generally U-shaped.[32] 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 mar-ginal tax rates at relatively low productivity levels, thereby sacrificing re-distribution to the middle class in order to focus on the very poorest.[33]

---

[32] 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.

[33] Slemrod et al. (1994) explored the sensitivity of optimal policy with respect to the gov-ernment's revenue requirement in a two-tax-bracket economy. They found that the opti-mal 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.

Distributional gains near the top of the income distribution are always close to 1 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.[34]

The Diamond-Saez implicit formula for optimal marginal tax rates offers 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.[35] 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.

2.   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 that 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.

---

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

[35] See app. sec. B.2 for the derivation and a longer discussion.

We run two versions of this experiment.[36] 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.[37]

Second, we introduce a heavy left tail in the wage distribution so that more individuals have very low (and privately uninsurable) productivity.[38] 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 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.[39]

3. 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,

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

[37] 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 fig. A6). 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 US data and clearly points to an exponentially distributed component in the right tail of productivity distribution.

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

[39] 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.

and optimal transfers are smaller in our model, in part because we allow for private insurance.[40] In Saez's calibration reported in column 3 of his table 2, optimal transfers are 31% of GDP, and government purchases are 25% of GDP.[41] Thus, the required government tax take is 56% of GDP. In our baseline parameterization, the corresponding number is 41.8% (transfers are 21.5% of GDP, and purchases are 20.3%; 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 lines), government purchases plus transfers are 59.9% (panel A) and 54.7% (panel C) of output.

In 2015, total US government spending including public consumption, gross investment, transfer payments, and interest on debt was 33.5% of GDP.[42] 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.

C.   Preferences without Income Effects à La Diamond (1998)

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

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

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).[43] 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

---

[40] Mankiw, Weinzierl, and Yagan (2009) and Golosov, Troshkin, and Tsyvinski (2016) also find U-shaped marginal rates. Both papers abstract from private insurance. The Golosov, Troshkin, and Tsyvinski (2016) calibration implies that most households have very low productivity, while Mankiw, Weinzierl, and Yagan (2009) assume that 5% 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.

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

[42] National Income and Product Accounts (table 3.1).

[43] 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 $G(\cdot)$ function in his eq. (1) is logarithmic.

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.[44] Thus, the preference specification 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

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

where χ is the average marginal utility of consumption in the population (see sec. III.C) 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 that $D(\alpha)$ 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

---

[44] 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 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 fig. A7).

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
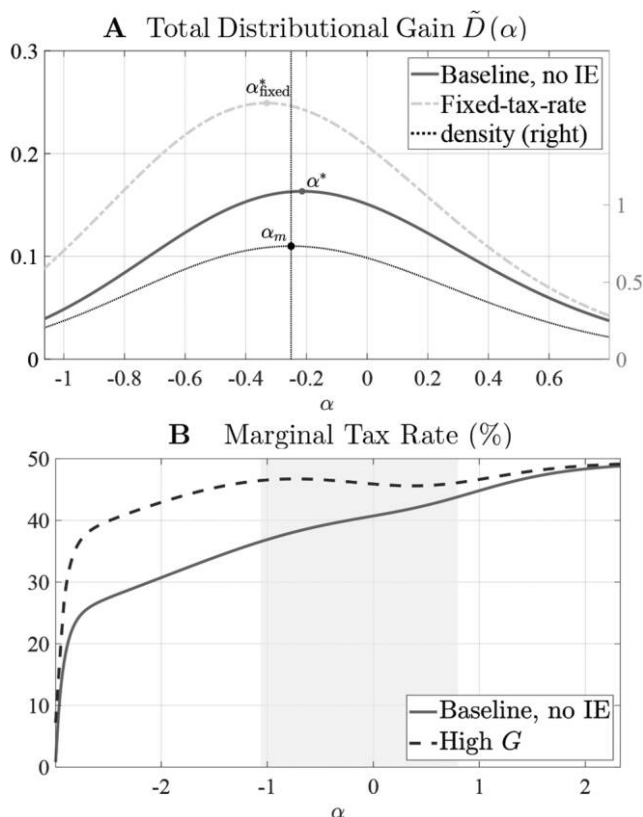

Fig. 5.—Preferences without income effects. Panel A plots distributional gains and the productivity density (between the 5th and 95th percentiles). The peak of each curve is indicated by a circle. The solid line is the baseline case. The dashed line corresponds to the interim case in which G is higher but marginal tax rates are unchanged. Panel B plots the optimal marginal tax schedules for the baseline (solid line) and high G (dashed line) cases. A color version of this figure is available online.

the one described in section IV.[45] Note that the distributional gain term peaks after $f_\alpha(\alpha)$ (i.e., $\alpha^* > \alpha_m$), so Diamond's condition 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 fig. 2).[46]

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 (1) has no effect on efficiency costs, $\tilde{E}(\alpha)$, (2) increases total distributional gains $\tilde{D}(\alpha)$ for all finite α, and (3) reduces the value α^* at which $\tilde{D}(\alpha)$ is maximized.

Proof.—See appendix section F.1. QED

Result 1 is trivial: lump-sum transfers do not affect labor supply, given the preferences in equation (19). Result 2 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 3 is that reducing lump-sum transfers hurts the poor disproportionately, in the sense that marginal utility becomes a more convex function of productivity. Thus, distributional gains increase relatively more at low income levels.

Panel A of figure 5 illustrates proposition 1 with a numerical example. The dashed line plots distributional gains when G is increased but the marginal tax schedule is unchanged relative to the (initially optimal) baseline.[47] 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 1 and 2, it cannot be optimal to finance an increase in G solely by reducing lump-sum transfers: distributional gains (dashed line) would then exceed efficiency costs (solid line) at all productivity levels. This explains why optimal marginal tax rates increase across the distribution. Second, thanks to result 3, increasing G shifts the argmax of the $\tilde{D}(\alpha)$

[45] 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% of output under the tax system estimated in sec. IV. Given these choices and under that tax system, the distribution for labor earnings would be identical to the EMG distribution estimated in sec. IV.

[46] In discussing the condition $\alpha^* < \alpha_m$, Diamond (1998, 87) 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." One might interpret Diamond as arguing here that the condition will be satisfied if the planner has a strong enough desire to redistribute. In sec. V.D, we show that when the planner has a strong taste for redistribution, the optimal marginal tax schedule does indeed have a downward-sloping portion.

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

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).[48]

D.   *Alternative Social Preferences*

To this point, we have explored optimal policy assuming that 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 equation (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 taxation.[49] 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.[50]

1.   Empirically Motivated Pareto Weight Function

We are especially interested in the value for θ that rationalizes the extent of redistribution embedded in the actual US 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 US tax and transfer system. We define an *empirically motivated Pareto weight function* $W(\alpha; \theta^{\text{US}})$ as the

---

[48] Let $\alpha^*_{G^+}$ 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.

[49] 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.

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

special case of the function defined in equation (18) in which the taste for redistribution θ^{US} satisfies $\hat{\tau}(\theta^{US}) = \tau^{US}$.[51]

The Pareto weight function $W(\alpha; \theta^{US})$ 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 $\theta = \theta^{US}$, any tax system that delivers higher welfare than the HSV function with $\tau = \tau^{US}$ 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 $\theta = \theta^{US}$ isolates the efficiency gains from replacing the HSV parametric function with the optimal nonparametric schedule.

2. 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:

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

where g^{US} is the observed ratio of government purchases to output.[52]

Proof.—See appendix section F.2. QED

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}.[53] Given our baseline parameter values and $g^{US} = 0.188$, the implied empirically motivated taste for redistribution is $\theta^{US} = -0.517$. Thus, the fact that the current US tax and transfer system is only modestly redistributive points to a weaker than utilitarian taste for redistribution.

---

[51] 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.

[52] 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.

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

Our finding of a negative θ may be interpreted in two ways. One is that the US political system appropriately aggregates Americans' preferences, so we should use these weights to evaluate social welfare. Consistent with this idea, Weinzierl (2017) reports survey support for the idea that there should be a link between taxes paid and government benefits received and that respondents who emphasize that principle are not enthusiastic about using the tax system to reduce inequality. An alternative interpretation is that the political system has been captured by the elites and that a utilitarian (or Rawlsian) objective would better reflect the preferences of average Americans. Gilens and Page (2014) find that the preferences of affluent citizens have a much greater impact on policy outcomes than the preferences of those in the middle of the income distribution. The probabilistic voting model (see Persson and Tabellini 2000) is one model that can account for this pattern.[54]

Figure 6 plots the marginal and average tax schedules for the optimal Mirrleesian policy, given $\theta = \theta^{\mathrm{US}}$ 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% 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% of households are located.

3. Optimal Taxation under Alternative Social Preferences

Panel A of figure 7 plots optimal Mirrleesian marginal tax rate profiles for $\theta \in \{-1, \theta^{\mathrm{US}}, 0, 1, \infty\}$.[55] Panel B plots the corresponding distributional gain functions, in each case relative to the utilitarian baseline.

Considering Pareto weight functions with a stronger than utilitarian taste for redistribution, we find that the optimal marginal tax schedule shifts upward. In addition, the shape of optimal schedule changes from

---

[54] 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.

[55] When we compute the Rawlsian case, we simply maximize welfare for the lowest α type in the economy, subject to the usual feasibility and incentive constraints. A numerical value for θ is not required for this program.
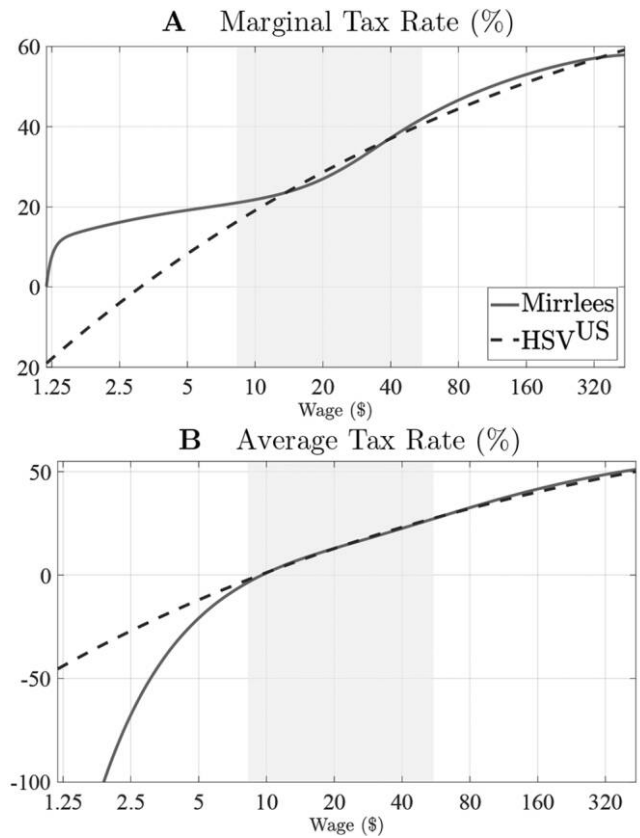
Fig. 6.—Optimal taxes and HSV approximation with $\theta = \theta^{US}$. The figure contrasts tax rates under the current HSV tax system to those under the Mirrlees policy using our empirically motivated Pareto weight function. A color version of this figure is available online.

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.[56]

---

[56] Welfare gains from tax reform here are very large, reaching 662% of consumption in the Rawlsian case (see app. sec. 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.
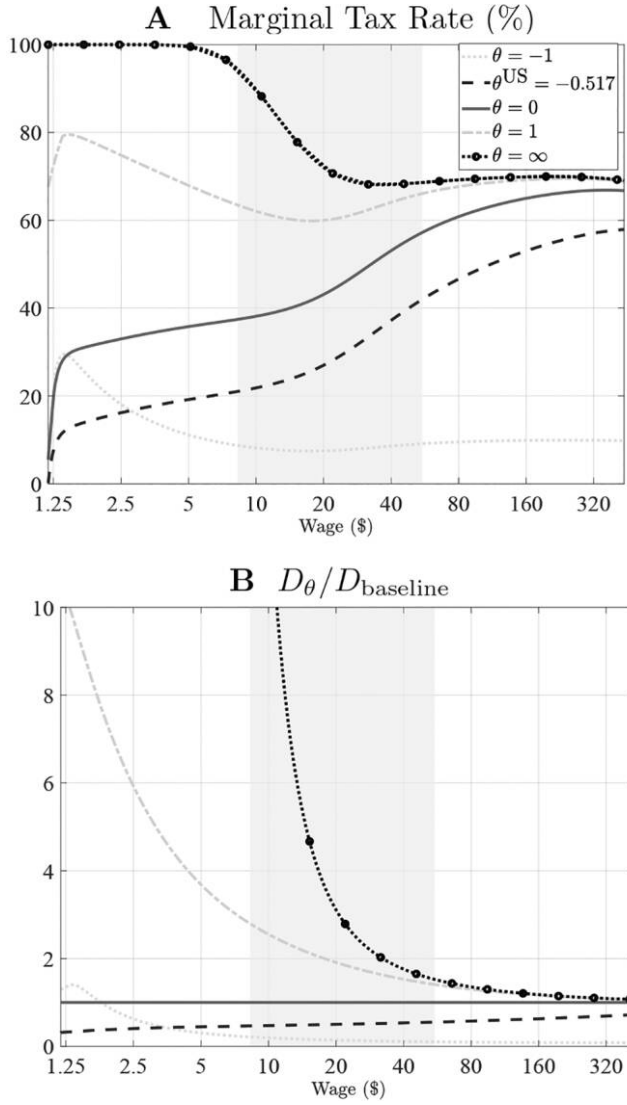
FIG. 7.—Alternative social preferences. Panel A plots optimal marginal tax schedules corresponding to Pareto weight functions with the following values for the taste for redistribution parameter: $\theta = -1$ (laissez-faire), $\theta = -0.517$ (empirically motivated), $\theta = 0$ (utilitarian), $\theta = 1$ (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 ($\theta = 0$). A color version of this figure is available online.

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 perceive large distributional gains from downward redistribution and focuses 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%).

E. 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 system.[57] 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, 223) 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." Thus, one way to frame the distinction between our increasing optimal marginal rate schedule and Saez's U-shaped one is that transfers under our scheme have the flavor of universal basic income, 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.[58]

VI. Further Explorations

We extend our exploration in two different directions. First, we compare optimal nonparametric 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.

[57] That in turn is a repackaging of the negative income tax proposal in Friedman and Friedman (1962).

[58] 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.

A. Mirrlees versus Ramsey Taxation

We compute the best tax and transfer systems in two parametric classes: (1) the HSV class and (2) 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 US tax and transfer system.

Table 3 presents outcomes for each tax function. Moving from the baseline policy HSV^{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% decline in output relative to the baseline. However, the additional redistribution translates into an overall welfare gain of 2.07%.

When we restrict the new fiscal policy to the parametric HSV class, we find an increase from 0.181 to 0.331 in the progressivity parameter τ. This reform generates a welfare gain equivalent to giving all households 1.65% more consumption, which is 80% of the gain under the best possible Mirrleesian policy. The best policy in the affine class does less well, delivering only 66% 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, B) and decision rules for consumption and hours (panels C, D) for each best-in-class tax and transfer scheme. Over most of the shaded area, which covers 90% 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

TABLE 3
MIRRLEES VERSUS RAMSEY TAXATION

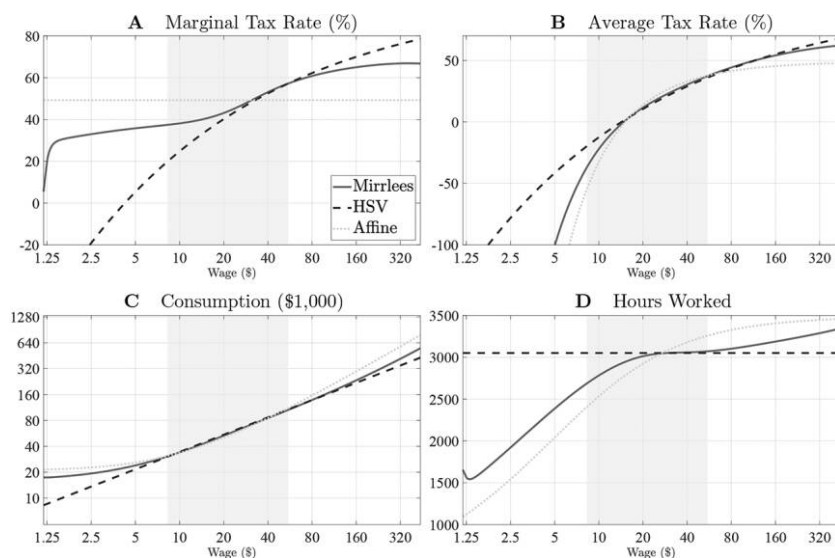| SYSTEM	PARAMETERS	OUTCOMES					
		$\overline{T'}$ (%)	Tr ($)	Tr/Y (%)	$(Tr + G)/Y$	ω (%)	ΔY (%)
HSV^{US}	$\lambda = .840$, $\tau = .181$	33.5	1,753	2.3	21.1	. . .	. . .
HSV	$\lambda = .817$, $\tau = .331$	46.6	4,632	6.4	26.5	1.65	−6.53
Affine	$\tau_0 = -\$20,747$, $\tau_1 = 49.2\%$	47.7	20,111	28.1	48.3	1.36	−7.31
Mirrlees		49.1	15,400	21.5	41.8	2.07	−7.32

NOTE.—See table 2.

FIG. 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. A color version of this figure is available online.

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.[59]

B. 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

---

[59] 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.

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 $U(\alpha, \alpha) \geq U^{US}(\alpha)$ for all $\alpha \in \mathcal{A}$, where $U^{US}(\alpha)$ denotes expected utility for a household of type α under our HSV approximation to the current US tax and transfer system.[60]

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 that 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% of households worse off.

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

PROPOSITION 3.   Let T^{US} be the current tax system, and let T^{PI} be the optimal Pareto-improving system. If the Pareto-improving constraints bind in an open interval $\Gamma \subset \mathcal{A}$—that is, $U(\alpha; T^{US}) = U(\alpha; T^{PI})$ 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 section F.3. QED

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.

---

[60] 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 app. sec. D.2.
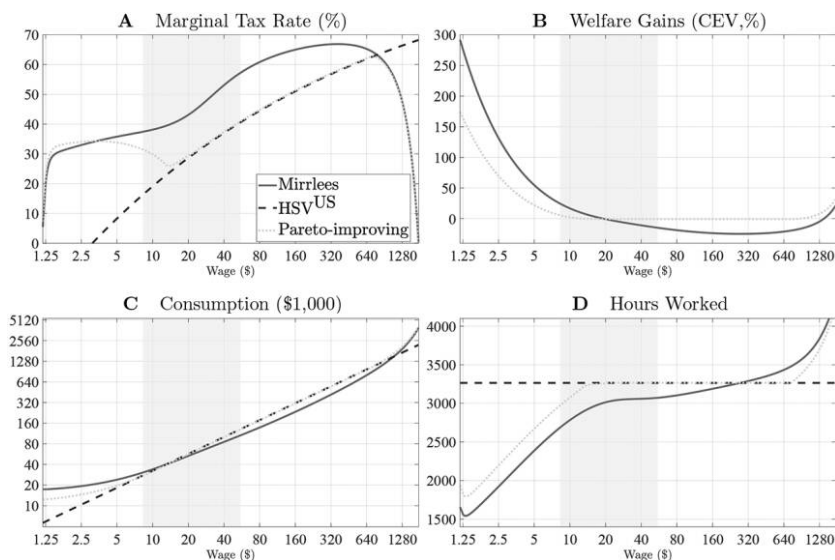
FIG. 9.—Pareto-improving tax reforms. The figure plots marginal tax schedule and welfare gains (CEV; panels A, B) and decision rules for consumption and hours worked (panels C, D) under the Mirrlees system, the approximation to the current system, and the system that is optimal subject to being weakly Pareto improving. A color version of this figure is available online.

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% more consumption compared with a 2.07% 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 (sec. V.D) 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.

VII.   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 as well as 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, Storesletten, and Violante (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." *A.E.R.* 104:122–26.

Benabou, R. 2000. "Unequal Societies: Income Distribution and the Social Contract." *A.E.R.* 90 (1): 96–129.

Bourguignon, F., and A. Spadaro. 2012. "Tax-Benefit Revealed Social Preferences." *J. Econ. Inequality* 10:75–108.

Brendon, C. 2013. "Efficiency, Equity, and Optimal Income Taxation." EUI Working Paper no. 2013/22, Max Weber Programme, European Univ. Inst.

Chetty, R., and E. Saez. 2010. "Optimal Taxation and Social Insurance with Endogenous Private Insurance." *American Econ. J. Econ. Policy* 2:85–114.

Choné, P., and G. Laroque. 2018. "On the Redistributive Power of Pensions." *Soc. Choice and Welfare* 50:519–46.

Conesa, J. C., and D. Krueger. 2006. "On the Optimal Progressivity of the Income Tax Code." *J. Monetary Econ.* 53:1425–50.

Diamond, P., and E. Saez. 2011. "The Case for a Progressive Tax: From Basic Research to Policy Recommendations." *J. Econ. Perspectives* 25:165–90.

Diamond, P. A. 1998. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *A.E.R.* 88:83–95.

Farhi, E., and I. Werning. 2013. "Insurance and Taxation over the Life Cycle." *Rev. Econ. Studies* 80:596–635.

Feldstein, M. S. 1969. "The Effects on Taxation on Risk Taking." *J.P.E.* 77:755–64.

Friedman, M., and R. D. Friedman. 1962. *Capitalism and Freedom.* Chicago: Univ. Chicago Press.

Gilens, M., and B. I. Page. 2014. "Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens." *Perspectives Politics* 12:564–81.

Golosov, M., M. Troshkin, and A. Tsyvinski. 2016. "Redistribution and Social Insurance." *A.E.R.* 106:359–86.

Golosov, M., and A. Tsyvinski. 2015. "Policy Implications of Dynamic Public Finance." *Ann. Rev. Econ.* 7:147–71.

Gouveia, M., and R. P. Strauss. 1994. "Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis." *Nat. Tax J.* 47:317–39.

Heathcote, J., F. Perri, and G. L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006." *Rev. Econ. Dynamics* 13:15–51.

Heathcote, J., K. Storesletten, and G. L. Violante. 2014. "Consumption and Labor Supply with Partial Insurance: An Analytical Framework." *A.E.R.* 104:2075–126.

———. 2017. "Optimal Tax Progressivity: An Analytical Framework." *Q. J.E.* 132:1693–754.

Heathcote, J., and H. Tsujiyama. 2021. "Practical Optimal Income Taxation." Working paper.

Hendren, N. 2020. "Measuring Economic Efficiency Using Inverse-Optimum Weights." *J. Public Econ.* 187:104198.

Keane, M. P. 2011. "Labor Supply and Taxes: A Survey." *J. Econ. 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." *A.E.R.* 105:1478–508.

Low, H., and L. Pistaferri. 2015. "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off." *A.E.R.* 105:2986–3029.

Mankiw, N. G., M. Weinzierl, and D. Yagan. 2009. "Optimal Taxation in Theory and Practice." *J. Econ. Perspectives* 23:147–74.

Meyer, B. D., and J. X. Sullivan. 2017. "Consumption and Income Inequality in the U.S. Since the 1960s." Working Paper no. 23655, NBER, Cambridge, MA.

Mirrlees, J. A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Rev. Econ. Studies* 38:175–208.

Ndiaye, A. 2020. "Flexible Retirement and Optimal Taxation." Working paper.

Persson, M. 1983. "The Distribution of Abilities and the Progressive Income Tax." *J. Public Econ.* 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." *Econ. J.* 37:47–61.

Saez, E. 2001. "Using Elasticities to Derive Optimal Income Tax Rates." *Rev. Econ. Studies* 68:205–29.

Saez, E., J. Slemrod, and S. H. Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *J. Econ. Literature* 50:3–50.

Saez, E., and S. Stantcheva. 2016. "Generalized Social Welfare Weights for Optimal Tax Theory." *A.E.R.* 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 Univ. Press.

Shourideh, A., and M. Troshkin. 2017. "Incentives and Efficiency of Pension Systems." Working paper.

Slemrod, J., S. Yitzhaki, J. Mayshar, and M. Lundholm. 1994. "The Optimal Two-Bracket Linear Income Tax." *J. Public Econ.* 53:269–90.

Toda, A. A., and K. Walsh. 2015. "The Double Power Law in Consumption and Implications or Testing Euler Equations." *J.P.E.* 123:1177–200.

Tuomala, M. 1990. *Optimal Income Tax and Redistribution.* Oxford: Clarendon.

———. 2010. "On Optimal Non-Linear Income Taxation: Numerical Results Revisited." *International Tax and Public Finance* 17:259–70.

Weinzierl, M. 2014. "The Promise of Positive Optimal Taxation: Normative Diversity and a Role for Equal Sacrifice." *J. Public Econ.* 118:128–42.

———. 2017. "Popular Acceptance of Inequality due to Innate Brute Luck and Support for Classical Benefit-Based Taxation." *J. Public Econ.* 155:54–63.

Werning, I. 2007. "Pareto Efficient Income Taxation." Working paper.